International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 25th Jun 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 05/28

Title **DEEP LEARNING MODEL FOR BREAKING CAPTCHAs**

Volume 11, SPL ISSUE 05, Pages: 183-191

Paper Authors

Mr. K. Gopala Reddy, Saride Srisivaphani Srisai Sudharam, Uppalapati Tejaswi, Rudru Renuka keerthi

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DEEP LEARNING MODEL FOR BREAKING CAPTCHAs

Mr. K. Gopala Reddy[1], Saride Srisivaphani Srisai Sudharam[2], Uppalapati Tejaswi[3] Rudru Renuka keerthi[4]
[1]Associate professor, Dept. of CSE, [2]18ME1A0599, [3]18ME1A05B5, [4]18ME1A0593
Ramachandra College of Engineering, A.P., India
emailme.kallam@rcee.ac.in, saisudharam123@gmail.com, Uppalapati.teju@gmail.com, Rudrurenu@gmail.com

Abstract

CAPTCHAs - Completely Automated Public Turing test to tell Computers and Humans Apart are distortions of phrases or images so that they may assess whether a user providing response is human or not. Additionally, many CAPTCHAs also inject noise or other features such as blobs or lines into the image to make it difficult to recognise. They are primarily used as security measures on websites to prevent bots from accessing or performing transactions on the site.

Deep learning neural networks have been taught with remarkable success on related challenges such as handwritten digit recognition.This motivates us to build an ML-based CAPTCHA breaker that maps CAPTCHAs to their solutions. Deep learning neural networks have been taught with remarkable success on related challenges such as handwritten digit recognition.This motivates us to build an ML-based CAPTCHA breaker that maps CAPTCHAs to their solutions. used CTC loss as loss function.

Here we have used convolutional neural networks to extract features from the image these features are passed through the recurrent neural network to produce the predictions. Convolutional neural networks are outperforming other neural networks in case image, audio data using this we can assess the security of captcha based system and also we can say that captcha based systems are vulnerable.

Introduction

CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart. In other words, CAPTCHA assesses whether the user is real or a spam robot. CAPTCHAs stretch or alter letters and numbers and rely on human abilities to determine which symbols they are.[7]

CAPTCHAs were designed to block spamming software from making comments on pages or purchasing surplus products at once. The most frequent version of CAPTCHA is an image with many distorted letters.

Uses of captcha:

CAPTCHA is a type of verification tool that is used on a number of websites to ensure that the operator is not a robot.

CAPTCHA is first and principally used to verify internet polls. In 1999, Slashdot hosted a poll in which visitors were invited to vote for the finest graduate school for computer science. Students from Carnegie Mellon and MIT constructed bots, or automated programmes, to vote for their institutions

frequently. Thousands of votes were cast for several schools, whereas only a few hundred were cast for others. CAPTCHA was included to prevent people from abusing the polling mechanism.

These steps are difficult because segmentation is difficult because the characters might overlap with one another, deformation in characters also make this task difficult, unknown scale of character and orientation of the characters will make segmentation difficult modules mentioned above are optimized independently, we can instead train a neural network We just give images to the network, allowing it to learn characteristics from the images and use these features for recognition to ensure that all modules are in alignment. Here we have used convolutional neural networks to extract features from the image these features are passed through the recurrent neural network to produce the predictions. Convolutional neural networks are outperforming other neural networks in case image, audio data using this we can assess the security of captcha based system and also we can say that captcha based systems are vulnerable.

A research paper was published by Google We applied the same concept to train datasets for convolutional neural networks, which they utilised to recognise arbitrary multi-digit numerals from Street View imagery [5]. our idea is mainly baser on paper [6] by Xinjie Feng,Hongxun Yao, and Shengping Zhang.

CAPTCHA is also used in registration forms on websites where consumers can create free accounts, such as Yahoo! Mail or Gmail. Spammers can't use bots to establish a slew of spam email accounts because of CAPTCHAs.

CAPTCHA is also used by ticket services like Ticketmaster to prevent ticket scalpers from buying too many tickets for big events. This helps honest people to buy tickets in a fair manner while preventing scalpers from putting hundreds of orders.

Finally, CAPTCHA is used to prevent spamming messages or comments on online pages or blogs that have message boards or contact forms. It does not prevent cyberbullying, but it does stop bots from automatically sending messages. Here we have used convolutional neural networks to extract features from the image these features are passed through the recurrent neural network to produce the predictions. Convolutional neural networks are outperforming other neural networks in case image, audio data using this we can assess the security of captcha based system and also we can say that captcha based systems are vulnerable.

Related Work:

Greg Mori and Jitendra Malik [1] have broken EZ-Gimpy with 92% success and Gimpy with 33% success Using complex object recognition techniques on these captchas. In comparison to earlier works we are trying to train an end to end neural network system that will extract sequence of features required for categorization. Neural networks have shown great achievements in new applications including natural language processing [2], voice [3] and picture processing. One does not require any domain knowledge to alter inputs in certain manner in order to supply the features that a model could learn from, useful features are extracted by hidden layer in the neural network.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

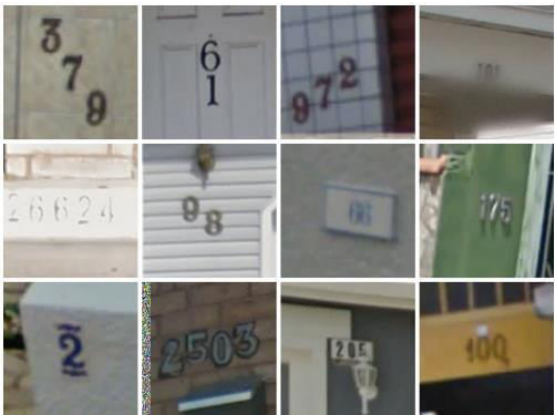

Figure 1. Internal street number dataset [5]

Few papers [4] use the following sequence of steps for character recognition in captcha

- Pre-processing
- Dividing the image into segments
- Training the model for individual characters
- Predict character for each segment and generate the sequence with highest probability

These steps are difficult because segmentation is difficult because the characters might overlap with one another, deformation in characters also make this task difficult, unknown scale of character and orientation of the characters will make segmentation difficult modules mentioned above are optimized independently, we can instead train a neural network We just give images to the network, allowing it to learn characteristics from the images and use these features for recognition to ensure that all modules are in alignment.

A research paper was published by Google We applied the same concept to train datasets for convolutional neural networks, which they utilised to recognise arbitrary multi-digit numerals from Street View imagery [5]. our idea is mainly baser on paper [6] by Xinjie Feng,Hongxun Yao, and Shengping Zhang.

Problem Statement:

Existing system and its disadvantages:

Some papers [4] use the following steps for character recognition in captcha Pre-processing the image, Dividing the image into segments, Training the model for individual characters, Predict character for each segment and generate the sequence with highest probability. These are very difficult steps to do because of the following reasons:

- Segmentation is difficult because some digits could overlap
- Deformity of digits is also a major problem
- Scale of the characters is not known
- Character orientation makes recognition difficult

Proposed System

We use two CNN layers to get information from the images, and then use an RNN to sequentially parse each letter available. As for the loss function, we use the CTC Loss to find the error sequentially. Artificial neural networks (ANNs) in the convolutional neural network (CNN, or ConvNet) class are most frequently used to assess visual imagery.

A Connectionist Temporal Classification Loss, also known as a CTC Loss, is made for applications where alignment between sequences is required but is challenging, such as aligning each character to its position in an audio recording.

IMPLEMENTATION

Pre-processing images:

Converting the image into RGB format

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

RGB is simply three 8-bit values for red, green and blue

Resize the images to specific size 300x75

The procedure that modifies the range of pixel values is known as normalisation. To bring the image to a range that is normal to senses is the goal of normalisation.

Linear Normalization is used where data is linear.

$$I_N = (I - \text{Min})\frac{\text{newMax} - \text{newMin}}{\text{Max} - \text{Min}} + \text{newMin}$$

Figure – 2 [8]

$$I_N = (\text{newMax} - \text{newMin})\frac{1}{1 + e^{-\frac{I-\beta}{\alpha}}} + \text{newMin}$$

Figure – 3 [8]

Non-Linear Normalization used when is no linear relationship between old image and new image. Example Normalization follows a sigmoid function then Normalized formula will be: Now convert this image into NumPy array

Pre-processing targets:

In our case the names of the files are the target labels we must extract the labels from the filenames. Now after extracting the names find individual characters in these names and convert these labels into integers for further processing**.** A Connectionist Temporal Classification Loss, also known as a CTC Loss, is made for applications where alignment between sequences is required but is challenging, such as aligning each character to its position in an audio recording.

Working:

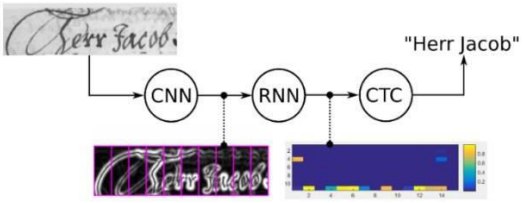

Figure 4. A general CRNN architecture for handwriting recognition [9]

If you need a computer to identify text, neural networks remain the best option because they currently outperform alternative techniques. In such circumstances, the neural network has a Convolutional layer that extracts a series of features from the picture, and the recurrent layer receives input from this sequence. For each sequence element, it generates character scores. we can instead train a neural network We just give images to the network, allowing it to learn characteristics from the images and use these features for recognition to ensure that all modules are in alignment.

Our model has this similar architecture

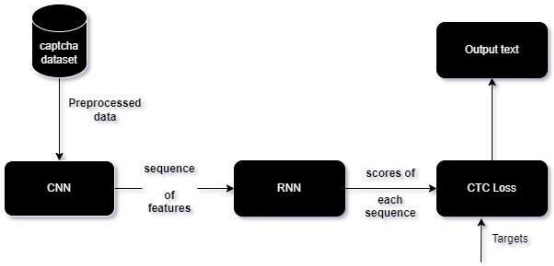

Figure 5. architecture of the model

Convolutional network Layer:
Convolutional neural networks differ from other neural networks in that they function better with picture, voice, or audio signal inputs. They have three different sorts of layers:

1. Convolutional layer
2. Pooling layer
3. Fully-connected (FC) layer

Convolutional Layer

Convolutional layer is the initial layer of a convolutional network. Many more convolutional layers or pooling layers follow these convolutional layers. Finally, the Fully connected layer is added. With each layer, the CNN becomes more complicated, allowing it to detect more sections of the picture. Earlier layers concentrate on basic elements like colours and borders. As the visual data passes through the CNN's layers, it begins to detect bigger components or forms of the object, eventually identifying the desired item[10].

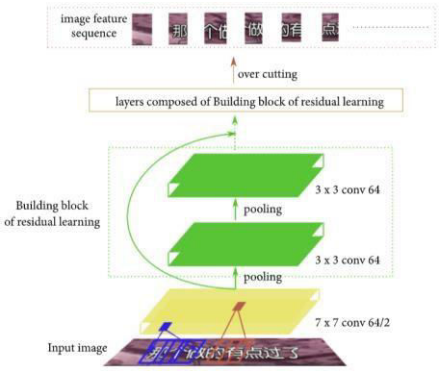

Figure 6. Feature extraction using convolutional networks [6]

The parameter sharing method ensures that the weights of the feature detector remain constant as it moves across the image. During training, various parameters, like as weight values, are changed using gradient descent and backpropagation. But before the neural network can be trained, three hyperparameters that control the output volume size must be supplied. They include: amount of filters, stride, padding[10]. A Connectionist Temporal Classification Loss, also known as a CTC Loss, is made for applications where alignment between sequences is required but is challenging, such as aligning each character to its position in an audio recording.

A CNN's convolutional layer serves as its computational core and is where most of the computing happens.Along with other things, it needs input data, a filter, and a feature map. Assume the input is a colour image made up of a 3D pixel matrix. As RGB in an image, this suggests the input will have three dimensions.The receptive fields of the image will be scanned by a feature detector, also known as a kernel or filter, to see if the feature is there. This process is known as convolution.[10]

A two-dimensional (2-D) array of weights represents part of the picture in the feature detector. The filter size, which can vary in size, is typically a 3x3 matrix, which also influences the receptive field's size. The dot product of the input pixels and the filter is then calculated after the filter has been applied to a selected area of the image. The dot product is then loaded into an output array after that. Once the kernel has swept through the entire image, the filter moves by a stride and the process is repeated. The end result of a series of dot products from the input and the filter is a feature map, activation map, or convolved feature.[10]

Pooling Layer

A dimensionality reduction approach called downsampling, commonly referred to as pooling layers, lowers the number of input components. Like the convolutional layer, the pooling method sweeps a filter across the entire input, but this filter lacks weights. Instead, the kernel populates the output array from the values in the

receptive field using an aggregation function. There are two main types of pooling:

Max pooling : As it moves over the input, the filter selects the pixel with the greatest value to transmit to the output array. This approach is applied more frequently than average pooling. Even though the pooling layer loses a lot of information, the CNN does benefit in some ways from it. They help to lessen complexity, boost effectiveness, and lessen the risk of overfitting.

Average pooling : As the filter passes over the input, the average value within the receptive field is computed and transferred to the output array. [10]

Fully-Connected Layer

The name of the full-connected layer is self-explanatory. The pixel values of the input image are not, as was previously indicated, immediately related to the output layer in partially linked layers. On the other hand, each node in the output layer links directly to a node in the layer before it in the fully linked layer. Based on the attributes that the previous layers and their respective filters have obtained, this layer performs categorization tasks. While FC layers commonly utilise a softmax activation function to produce a probability from 0 to 1, convolutional and pooling layers frequently use ReLu functions to classify inputs[10]. A Connectionist Temporal Classification Loss, also known as a CTC Loss, is made for applications where alignment between sequences is required but is challenging, such as aligning each character to its position in an audio recording.

Recurrent neural network:

A type of artificial neural network called a recurrent neural network (RNN) analyses time series or sequential data.Examples of problems where deep learning techniques are used include language translation, natural language processing (nlp), speech recognition, and image captioning. Their "memory," which enables them to influence current input and output by drawing on knowledge from prior inputs, is what makes them unique. Recurrent neural networks' output is dependent on the previous components of the sequence, whereas normal deep neural networks assume that inputs and outputs are independent of one another. Unidirectional recurrent neural networks cannot take into account future occurrences in their forecasts, despite the fact that they may be helpful in defining a sequence's output [11]. we can instead train a neural network We just give images to the network, allowing it to learn characteristics from the images and use these features for recognition to ensure that all modules are in alignment.

The Gated Recurrent Unit (GRU) is a step forward from the regular RNN (recurrent neural network). Kyunghyun Cho et al first proposed it in 2014.[12]

Another intriguing feature of GRU is that, unlike LSTM, it lacks a distinct cell state (Ct). It only has one state: hidden (Ht). GRUs are easier to train because of their simpler design.[12]

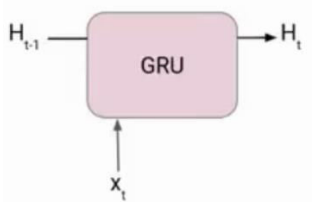

Fig 7. GRU [12]

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

The architecture of Gated Recurrent Unit:

It accepts an input Xt and the hidden state Ht-1 from the preceding timestamp t-1 at each timestamp t. It then generates a new hidden state Ht, which is then sent to the following timestamp. As the filter passes over the input, the average value within the receptive field is computed and transferred to the output array. [10]

In contrast to an LSTM cell, which has three gates, a GRU now has just two gates. The Reset gate is the first, while the Update gate is the second.[12]

Reset Gate:

The network's short-term memory, or hidden state, is controlled by the Reset Gate (Ht). The Reset gate has the following equation.

$$r_t = \sigma(x_t * U_r + H_{t-1} * W_r)$$

This is extremely close to the LSTM gate equation. Due to the sigmoid function, the value of rt will vary between 0 and 1. For the reset gate, Ur and Wr are weight matrices.[12]

Update gate :

We also have an Update gate for long-term memory, which has the equation below. The only variation is the weight metrics, which are Uu and Wu.[12]

How CTC works :

The CTC loss function will lead the NN-training. The CTC loss function only receives the NN's output matrix and the accompanying ground-truth (GT) text.

How does it know where each character appears, though? Well, it has no idea. Rather, it attempts all potential alignments of the GT text in the picture and averages the results. In this case, a GT text's score is high if the total of the alignment-scores is high.

Why we want to use CTC:

1. Annotating a data set on a character level takes a long time (and is tedious).
2. We just get character-scores, so we'll have to do some more processing to retrieve the final text. Because the "o" is a broad letter, we may get "ttooo" because it spans numerous horizontal locations, as illustrated in Fig. 2. All duplicate "t"s and "o"s must be removed. But what if the text that was identified was "too"? Then we obtain the erroneous result by eliminating all duplicate "o"s. What should I do?

CTC takes care of both issues for us, All we have to do now is tell the CTC loss function what text appears in the picture. As a result, both the position and width of the letters in the picture are ignored.

The identified text does not require any additional processing.

Encoding the text:

There was also the problem of encoding

$$u_t = \sigma(x_t * U_u + H_{t-1} * W_u)$$

redundant characters (remember what we mentioned about the word "too"?) The problem is overcome by using a pseudo-character named blank (not to be confused with a "real" blank, which is a white-space character). In the next text, this special character will be marked as "-." To tackle the duplicate-character problem, we utilise a creative coding scheme: while encoding a sentence, we may introduce an unlimited number of blanks at any location, which will be erased when decoding it. However, between repeated characters, such as "hi," we must put a blank. We may also repeat each character as many times as we like. Academic success is critical for the a success improvement of younger human beings in society. We additionally evaluate the KNN, SVM,

Logistic regression which one is extra correct in locating the output. Based on attributes we will perceive the scholar dropouts and the scholar who want the unique interest from instructor to offer counselling which improves the overall performance in students. By the usage of gadget mastering we examine the scholar overall performance.

some examples:

- “to” → “---ttttttooo”, or “-t-o-”, or “to”
- “too” → “---ttttto-o”, or “-t-o-o-”,or “to-o”, but **not** “too”

As you can see, this schema also makes it simple to construct alternative alignments of the same text, for example, "t-o," "too," and "-to" all express the same word ("to"), but with different picture alignments. The NN is programmed to generate encoded text (encoded in the NN output matrix).

We additionally evaluate the KNN, SVM, Logistic regression which one is extra correct in locating the output. Based on attributes we will perceive the scholar dropouts and the scholar who want the unique interest from instructor to offer counselling which improves the overall performance in students. By the usage of gadget mastering we examine the scholar overall performance. We can accomplish our objective assuming we follow those means. The calculation's key advances are recorded underneath. Each dataset ought to be standardized. Partition the first dataset into testing and preparing datasets. Create IDS models utilizing Logistic Regression, Decision Tree, Random Forest, and MLP. Evaluate the presentation of each model. convolutional neural networks to extract features from the image these features are passed through the recurrent neural network to produce the predictions. Convolutional neural networks are outperforming other neural networks in case image, audio data using this we can assess the security of captcha based system and also we can say that captcha based systems are vulnerable.

Conclusion:

In this paper, we attempted to use deep neural networks to decode an image-based CAPTCHA. Instead of employing the traditional method of first cleaning a CAPTCHA image, segmenting the image, and identifying the individual characters, we have used convolutional neural networks and recurrent neural networks. We have taken advantage of RNNs' ability to work with sequences and CNNs' ability to work on images.

Finally, we are able to offer a complete neural network system. We will be able to crack the captcha given an image.

References:

[1] Mori, G., Malik, J.: Recognizing objects in adversarial clutter: breaking a visualcaptcha

[2] Tomas Mikolov Et al. Distributed Representations of Words and Phrases and their Compositionality

[3] Awni Hannun, Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng. Deep Speech: caling up end-to-end speech recognition

[4] Kumar Chellapilla, Patrice Y. Simard Using Machine Learning to Break VisualHuman Interaction Proofs (HIPs) Microsoft Research, one microsoft way, WA.

[5] Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, Vinay Shet.

Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks
[6] Xinjie Feng, Hongxun Yao , and Shengping Zhang.Focal CTC Loss for Chinese Optical Character Recog-nition on Unbalanced Datasets. Hindawi ComplexityVolume
[7] https://www.pandasecurity.com/en/mediacenter/panda-security/what-is-captcha/
[8] https://medium.com/@shoaibrashid/what-is-image-normalization-d8305bf328c0
[9] https://towardsdatascience.com/intuitively-understanding-connectionist-temporal-classification-3797e43a86c
[10] https://www.ibm.com/cloud/learn/convolutional-neural-networks
[11] https://www.ibm.com/cloud/learn/recurrent-neural-networks
[12] https://www.analyticsvidhya.com/blog/2021/03/introduction-to-gated-recurrent-unit-gru/