Letter Recognition Using Machine Learning Algorithms

Merima Ćeranić, Samed Jukić
International Burch University,
Sarajevo, Bosnia and Herzegovina
merima.ceranic@stu.ibu.edu.ba

Original research

Abstract: *Optical character recognition represents the mechanical or electronic conversion of handwritten, typed or printed images into coded text. Optical character recognition is widely used as a form of data entry from records that have been printed, and it can include invoices, bank statements, passports and many more. In the research, Optical character recognition reads data from the Re-Captcha dataset of images, converts them into strings, and these strings are used for testing, training and calculating prediction accuracy. The methodologies used are Convolutional neural network and Recurrent neural network. The convolutional neural network consist of neurons that receive data and group them according to similarity. A recurrent neural network cycle can be created between the connections of nodes, allowing the output from nodes to influence the subsequent input to other nodes. For data were used Re-Captcha images, and for the prediction of characters from images was used TensorFlow with Keras. The best results that are produced can be compared between first and last result, where the loss for first result was 20.63 and value loss was 16.45, while last result has loss of 0.56 and value loss of 2.96.*

Keywords: Keras, OCR, Re-Captcha, Tensorflow.

1. Introduction

Institutions that focus on finance, such as banks, are involved in the process of creating the latest records for their clients or may even include the conclusion of new deals, which creates many records that are in paper form. Thus, documents become numerous, and everything is in the form of paper, and they can become a big concern for banks. These documents are crucial, and the ideal solution is to digitize these data, and using a data entry service is the ideal solution for this problem. Today's security is in digital documents, which means simple storage and very fast retrieval of documents. Banking institutions must simplify organizational processes in order to provide their users with the best possible services. In this way, the search time for all documents used by banks, for example, would be reduced. The technology that is unique to this is Optical character recognition (OCR), which is used by banks as part of the extraction of huge amounts of information. With OCR, banks can process, evaluate and monitor payments including huge amounts of data about their customers. These data are most often personal data or security data.

OCR is a technology that can extract all text from images, documents or scanned files. It enables banks to minimize human error, and to save time and effort while simultaneously improving the user experience. Banks should properly authenticate customers for routine or banking transactions, account opening and numerous other functions. For example, with the help of OCR or machine learning, banks can extract data in real time from passports or other documents. In this way, they can quickly identify clients before transferring money or opening a new bank account. OCR provides a Software development kit (SDK) that includes personal document understanding, data identification, and data validation. It can check whether the signature on the personal document matches the signature of a real person. Some of the tasks of Optical character recognition are mentioned below, where are listed some of the most important usings. Task of Optical character recognition are:

- The density of the text on the written station represents dense text. So, for example, every day we can notice a STOP sign on the streets, where the text is scarce.
- Text structure: Generally, the text on the station is structured in strong lines, while the example works in external conditions - the wilderness, the text can be scattered anywhere.
- Fonts: Written letters are darker, while printed fonts are lighter because they are more structured than written letters.

- Type of sign: We can distinguish many languages, so the text can also differ. One example can be numbers, where we immediately have a difference in house numbers.
- Artifacts
- Location: some tasks include a centralized text, while sometimes the text can be scattered in random places [1].

Also, in addition to the previously mentioned ways of using OCR, in everyday life we encounter applications that require additional authentication whether the end user is a robot or not. In those cases, the most common display of authentication is expressed in the form of Re-Captcha images. The paper shows how to use OCR for Re-Captcha images. Furthermore, OCR can help to read the data from the image, and further this data can be converted into audio recordings. If Re-Captcha images are converted into audio recordings, these recordings can be used by blind people.

2. Literature Review

Character recognition has become an integral part of computer analysis and vision. Several corporations are working on improving the technique and the current situation. Algorithms that can recognize notes and handwritten numbers are being actively developed by several industries. In the era of digitization, editing, indexing, finding and storing information in digital documents is much easier than spending a lot of time flipping through printed documents. Searching for data in a non-digital document is not only time-consuming, but there is a very likely possibility that some information will be missed while manually searching a document. Every day computers are getting better and better at doing tasks that people thought only they could do.

Related work

In the research by Karishma Tyagi and Vedant Rastogi, it can be observed that OCR recognizes any multimedia content like videos and images. We can use character positioning, image processing, neural network to solve the problem of image recognition in text. According to the research of these two authors, there are approaches for identifying links:

- *HMM approach*
  The Hidden Mark model is a stochastic two-step process. An established stochastic process that can be observed as another stochastic process that produces a series of

observations, but is not visible.

- *A neural network approach*
  Recognition of registrar characters on license plates plays a significant role in the optical recognition system. This can be directly related to the success or failure of system recognition.

- *Normalization of character*
  The necessary step is to frame the letters, characters or numbers to some standard size. Character normalization to one fixed size can be performed to simplify the task of optical character recognition.

- *Correlation method for recognizing the lower sign*

- *Pre-processing*
  This step involves converting images to grayscale. Then the image is converted into a binary image. This process is also known as the process of image digitization. It is important to note here that the picture may have some defects or difficulties. The result may be some unnecessary details that are present in the image.

- *Segmentation*
  In this step, the position of the object is learned. The size of the image can be expressed according to the size of the template [2].

In the article by Konica Minolta, How optical character recognition works, OCR allows users to convert scanned images into text and to convert paper documents into digitized documents. Digitizing documents helps reduce the amount of physical space required to store documents. Also, digitizing documents can reduce the risk of lost or incorrectly archived documents and, in many cases, eliminate the need for manual processing of documents. Manual processing of documents in most cases leads to errors. According to Konica Minolta, OCR analyzes the patterns of light and darkness that makeup letters and numbers to turn a scanned image into text. OCR recognizes characters in a variety of fonts. Early OCR systems were designed to work with one specific font, which was created specifically for this purpose. Today, modern OCR systems can even recognize people's handwriting. The technology that recognizes human handwriting is called intelligent character recognition (ICR). OCR programs work on the principle of recognizing text character by character. It can also check errors during the process at the end of the process.

OCR as a technology has existed since the late 1920s. Today, OCR can convert large documents, where only a few errors can occur. It is important to mention that there are six key ways OCR helps businesses [3]:

- *Automate workflows*
  Businesses that work with large amounts of paperwork can save time, and thus can increase productivity through scanning.

- *Turn read-only files into editable text*
  OCR allows users to read PDF documents that can be later edited, and used in other documents, but can also be searched.

- *Create audible files*
  It saves time spent reading long and complex documents. It enables the conversion of documents into a document that the user can listen to (while going to work), and it is considered that in these situations the user becomes more productive.

- *Translation of foreign documents*
  Some OCR solutions can convert documents into more than 180 foreign languages.

- *Manage forms and questionnaires*

- *Achieve faster, more accurate data entry*

In an article by Nitin Ramesh, Aksha Srivastava and K. Deeba, Improving Optical Character Recognition Techniques, it is stated that there are two types of character recognition, printed and handwritten. In the printed character recognition type, OCR searches for written text and reviews it one by one. On the other hand, Intelligent character recognition (ICR) can also work with text that is handwritten [4]. For the handwritten type of character recognition, the offline way to recognize is static document processing, while the online version is much more advanced and uses handwriting motion analysis. The most commonly used algorithm for learning patterns, the online mode allows us to record the movement, that is, the order in which the segments are drawn and what is their direction of movement [5].

3. Methods

In this part of the paper, the methods used will be described, as well as a database. The method will be presented as a flowchart, where each process will be described step by step.

Machine Learning Methods

The research includes Recurrent neural network (RNN) and Convolutional neural network (CNN) as machine learning methods. A convolutional neural network (CNN) is a machine learning algorithm used for image processing, recognition and classification for face detection or object identification. It consists of neurons that receive data that later assign importance to them and group them according to similarity. It is also called "ConvNet", in order to make accurate predictions it can look at the surroundings of the object. They will look at smaller parts or letters, rather than looking at the whole picture to determine features [6]. The recurrent neural network (RNN) represents artificial neural networks, where a cycle can be created between the connections of nodes, allowing the output from some nodes to influence the subsequent input to other nodes. RNN can use internal state to process variable-length nodes. The term "recursive neural network" will be used to denote the class of networks with infinite impulse response, while the previously mentioned CNN refers to the class of finite impulse responses. Both classes represent temporal dynamic behavior. A recursive network with a finite impulse is a directed acyclic graph that can be replaced by a neural network with strict forward connection. A recursive network with infinite momentum is a directed cyclic graph that cannot be wrapped [7].

Importing Libraries

The first step was to import appropriate libraries that will be required for the needs of the project. Some of the libraries that were using are numpy, matplotlib, tensorflow. Numpy is Python library that provides multidimensional array object. The elements in NumPy are all required to be the same data type. Matplotlib is data visualization and graphical plotting library for Python ant its numerical extensions NumPy. Tensorflow is open source library for numerical computation that makes machine learning and development neural network faster and easier.

Collecting Data

For this purpose is decided to use Re-Captcha images, which are collected from the internet. The dataset that is used in the project contains 1040 Re-Captcha PNG files. The label for each sample is a string.

Processing

Data processing is the method of transforming data into a graspable format. Collecting data is usually incomplete, noisy, inconsistent, and redundant. Processing data is an important step to reinforce data effectiveness. This step is done in the Jupyter notebook and includes mapping characters to integers, and then again, mapping integers back to original characters.

Create Dataset Objects

In this stage were created training and validation datasets. This was done by using TensorFlow functions.

Visualize Data

Visualization is important step because through this step can be seen with what data we are working, and what is real output of it. In this part of the work was used function from matplotlib.

Build a Model

In this stage computing the training time loss value and adding it to layer using self.add_loss() function was done. In this part of the research was used functions from tensorflow.keras – keras, and was used keras from tensorflow.

Train a Model

Training is a step that follows the steps of Building a model. In this step, we could see what is value for loss and what is value for value loss. In this part of the research was used function from keras. Import step of keras is mentioned in *G* step.

Printing the Interface – Printing the Results

The last step that we were using in our research was to print the results. Through this step we can see how is model trained. *Figure 1* presents stages that were used for research process.

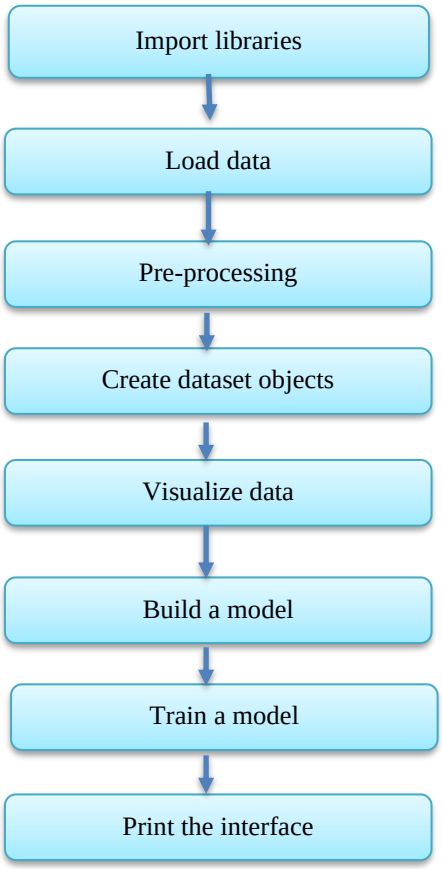

Figure 1: Diagram – Stages (B – I) needed for research

4. Results

To calculate the loss of value and calculate the loss, the model.fit was used in the work. Model.fit and everything that works smoothly can be used to calculate loss of value and loss. Model.fit is a customized training algorithm. Progressive complexity discovery is the basic tenet of Keras. How well a machine learning model generalizes data similar to what it was trained with is the basic story of a model measure. A good model fit refers to a model that accurately approximates the output when it has invisible inputs. Setting up readers to adjust parameters in the model to improve accuracy involves running the algorithm on data for which the target variable is known to produce the machine learning model. Model results are compared to actual values of the target variable to determine accuracy. The next step involves adjusting the standard parameters of the algorithm in order to reduce the error rate and make the model as precise as possible in determining the relationship between the target variable and the features. Until the model finds the optimal prediction parameter with significant accuracy, this process is repeated. *Table 1* shows the first five and the last five epochs of the training model where the time of execution, loss of value and loss can be seen. In the results can be seen the progress of the model. At the first epoch, loss was

20.6370 and value loss was 16.6513, while on the last epoch loss was 0.5616 and value loss was 2.9697. How well model is trained can be seen with comparison of loss 16.6513 and 0.5616, and value loss 16.6513 and 2.9697.

Table 1. Epochs – time – loss and values loss for first five epochs

Epoch n/100	Time	Loss	Value loss
1/100	43s 402ms/step	20.6370	16.4513
2/100	20s 339ms/step	16.3671	16.4555
3/100	20s 332ms/step	16.3534	16.4488
4/100	19s 325ms/step	16.3467	16.4365
5/100	22s 375ms/step	16.3346	16.4167
96/100	2s 29ms/step	0.6674	3.1233
97/100	2s 29ms/step	0.6018	2.8405
98/100	2s 28ms/step	0.6322	2.832
99/100	2s 29ms/step	0.5889	2.8786
100/100	2s 28ms/step	0.5616	2.9697

Figure 2 presents a prediction of a model. There can be seen sixteen examples and what is prediction for each of them. Sixteen examples present visual presentations of results, so it can be easily seen that for last example in forth column is 6p2ge.

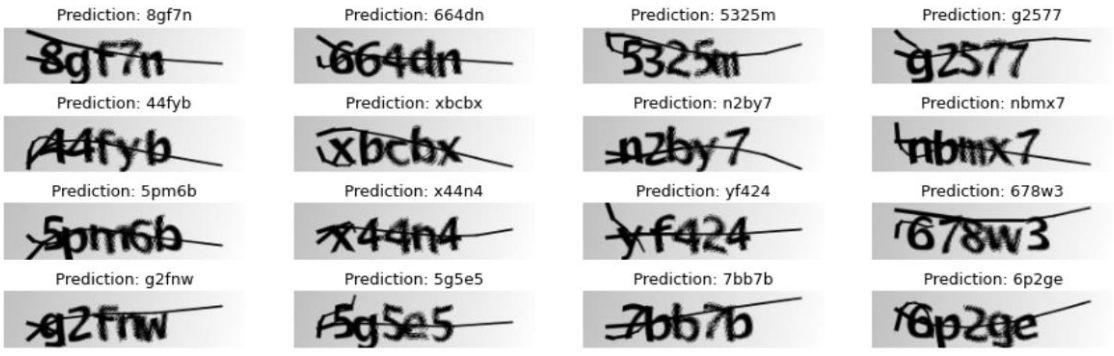

Figure 1. Prediction of a model – Visualization for 12 examples

5. Discussion

By following TensorFlow using Keras, has been seen how can be calculated accuracy of reading letters from Re-Captcha images. Analysis of data in the research is done by using machine learning (ML) tools. By looking at accuracy, that is presented above images/examples (*Figure 2*), it can be seen that accuracy is correct. Most letters from images are correct, which is pretty amazing. It can be concluded that model is trained well when the comparison between first and last epoch is made. For example, the results for first epoch were 20.6370 for loss and 16.4513 for value loss, while for the last epoch the results were better 0.5616 for loss and 2.9697 for value loss. Optical character recognition is useful for cases of scanning documents and transforming them into digital form. Besides, optical character recognition is useful for scanning documents, such as passports, transforming documents into audio format, etc. For example, if we take a look at one of the images (*Figure 3*) that was used in the paper, it can be seen that the prediction was pretty correct.

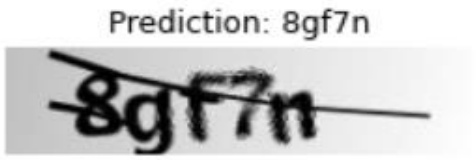

Figure 2. Prediction for single PNG file

In the research by Ondrej Bostik and Jan Klecka, Recognition of Captcha Characters by Supervised Machine Learning Algorithms [8], they reveal that all used algorithms can classify objects into the first class with a 99% success rate. Computational costs are the main difference in algorithms. According to the research, Pettr's neutron network proved to be the best algorithm. It has good precision and good computational costs. Feed-forward is the name for an inappropriately used neural network that was not optimized for pattern recognition. It had great difficulties during the learning phase when achieving correct performance.

In the research conducted by Ashish Renjan, Varun Nagesh Jolly Behra and Mothahar Reza, OCR using Computer Vision and Machine Learning [9], OCR basics were studied. Preprocessing is an important step in OCR systems. The authors state that if the input image contains tabular data, the processing becomes more complex if the tabular structure is to be

preserved. In conclusion, contours can help in extracting tabular data. The extraction succeeds if the borders are visible, but fails if the borders are not visible.

In the research by Jangid and Srivastava's from 2019, they used deep learning techniques to improve existing OCR approaches for recognizing Chinese capital letters [10]. The deeper the number of neural network layers and the more parameters, the more accurate the results. Accurate results mean that more computer resources are used. Removing data that is not essential for research is of enormous importance. This can be achieved by identifying the most connected neurons using the Average Percentage of Zeros algorithm, and removing some unnecessary network neurons and keeping the weight parameters that are key to reducing the network parameters in order to reduce the reasonable complexity of the model. The accuracy was reduced by 1.26%, but a 96.5% net weight reduction was achieved.

In the research of Khaled S. Younis and Abdullah A. Alkhateeb, A New Implementation of Deep Neural Networks for Optical Character Recognition and Face Recognition [11], using TensorFlow to classify the ubiquitous MNIST dataset, they designed a multilayer neural network. The accuracy was 98.48%.

7. Conclusion

Optical character recognition helps users to recognize numbers, letters and written characters. It can convert images and scanned documents into electronic data. This technology is at the center of a growing trend when it comes to workflow modernization. Artificial intelligence (AI) opens the door to many possibilities, and OCR can be used for many purposes. If we pay attention to the work of Konica Minolta and make a comparison with this work, we can conclude that they have the same goal. In working with Konica Minolta, the main goal was to document data and how to preserve it from loss. During my research, the goal was to see how accurately we can read from images, how accurate the prediction is for each image.

According to research by Himini Kohli, Jyoti Agarwal and Manoj Kumar [12], printed characters are easy to recognize because they have a defined size and shape. OCR faces the difficulties of handwriting, as each individual has a different handwriting. To solve the problem, the OpenCV technique is used in the research, which has a focus on testing and training the model. From the research, 99.5% training accuracy and 99% testing accuracy were achieved.

Reference

1. G. Shperber, “A gentle introduction to OCR”, 2018
2. K.Tyagi, V. Rastogi, “Survey on Character Recognition using OCR Techniques”, 2014
3. K. Minolta, “How optical character recognition works”, 2018
4. N. Ramesh, A. Srivastava, K. Deeba, “Improving Optical Character Recognition Techniques”, 2018
5. D. G.Pelli, C W.Burns, B. Farell and D. C. Moore-Page, “Feature detection and letter identification”, 2006
6. S. Saha, “A Comprehensive Guide to Convolutional Neural Networks - the ELI5 way”, 2018
7. N. Laskowski, “Recurrent neural networks”, 2021
8. O. Boštík, J Klečka, “Recognition of CAPTCHA Characters by Supervised Machine Learning Algorithms”, 2018
9. A. Ranjan, V. Behera, M. Reza, “OCR Using Computer Vision and Machine Learning”, 2021
10. Yin Y, Zhang W, Hong S, Yang J, Xiong J, Gui G., “Deep Learning-Aided OCR Techniques for Chinese Uppercase Characters in the Application of Internet of Things”, 2019
11. K. Younis and A. A. Alkhateeb, “A New Implementation of Deep Neural Networks for Optical Character Recognition and Face Recognition”, 2017
12. H. Kohli, J. Agarwal and M. Kumar, “An improved method for text detection using Adam optimization algorithm”, 2022