

A Deep Learning Approach to Recognize Cursive Handwriting

Vijay Shelake¹, Heena Patil²

¹ Associate Professor, Dept. of Computer Engineering, Yadavrao Tasgaonkar college of Engineering and management, Mumbai, India.

² Master of Engineering Student, Dept. of Computer Engineering, Alamuri Ratnamala Institute of Engineering and Technology, Mumbai, India.

Abstract – Now a days cursive handwriting recognition for cursive is a very difficult due to the unique styles of writing from person to person. Various research has been conducted in this field since around four decades. Character recognition of cursive handwritten script is a very challenging task. In English cursive writing, the characters in a word are connected to each other. So, the feature extraction and segmentation of cursive script is more difficult.

This work aims to represent the work related to recognition of cursive English handwriting. An offline cursive writing character recognition system is implemented using an

Artificial Neural Network. The features of each character written in the input are analyze, extracted and then passed to the neural network. Data sets, containing texts written in cursive by different people are used to train the system. The proposed cursive handwriting recognition system gives high levels of accuracy as compared to the conventional approaches because it uses online Data set to train the system. This system can efficiently recognise cursive texts and convert them into structural form like; notepad. It used for converting scanned handwritten notes into digital text available for search and storage in any digital platform.

In the proposed work, deep learning CNN algorithm used for classification recognition and pattern detection.

Key Words: Recognition, Deep Learning, CNN, Neural Network.

1. INTRODUCTION

Handwritten character recognition is an ongoing field of research for artificial intelligence, computer vision, and pattern recognition. An algorithm that performs handwriting recognition can acquire and detect characteristics from pictures or uploaded files and convert them to a machine-readable form. There are two basic types of handwriting recognition systems first is online recognition and another one is offline recognition. Both types can be implemented in applications to learn while performing offline learning on data. Several implementations have been used for online and offline handwriting recognition fields, such as statistical and structural methods, neural networks and syntactic methods. Some recognition system identifies strokes and apply

recognition on a single character or entire word or complete file using deep learning.

1.1 Objective

The main objective is to represent the work related to recognition of cursive English handwriting. An offline cursive writing character recognition system is implemented using an Artificial Neural Network. The features of each character written are extracted like; resize and split the words and then passed to the neural network. Online data sets, containing texts written by different people are used to train the system. The proposed recognition system gives high levels of accuracy as compared to the other handwritten recognise systems. This system can efficiently recognise cursive texts and convert them into structural form. It used for converting scanned handwritten notes into digital text available for search and storage in any digital platform.

1.2 Problem Definition

Cursive writing recognition is one of the most challenging topics in the field of image processing and pattern recognition. Each person has a different style of writing the same alphabet and word. There are also variations in the text written by the same person with different time period. The development of cursive writing recognition systems makes an improved interaction between human and machine. Various research works have been conducted focusing on new techniques that aim at reducing the processing time while providing higher accuracy.

Deep learning has been widely used to recognise handwriting. In offline handwriting recognition, text is analysed after written and converted in image or file format. The only information that can be analysed is the binary output of a character against a background and then shifts towards digital stylus for writing gives more information, such as pen stroke, pressure and speed of writing, there is still a necessity for offline methods, when online is inaccessible. It is necessary for stored historical documents, archives, or mass digitisation of hand-filled cursive documents

2. LITERATURE SURVEY

It's clear, then, that for computers to recognise and digitize handwritten documents and messages, there's a lot to learn. There are the different letters, characters and digits. But there's also the importance of being able to identify them despite differences due to different handwriting styles. This is where deep learning and neural networks are coming into handwriting recognition. Deep learning allows machines to learn over time, and adapt their output using weights.

In other words, the machine can learn how to identify letters despite different handwriting. More weight can sit on the factors that stay largely the same across handwriting. This means that deep learning is more adaptable to handwriting changes.

2.1 Different Approaches

Handwriting character recognition is one of the research fields in computer vision, artificial intelligence, and pattern recognition. A computer application that performs handwriting recognition have the ability to detect characters in pictures, paper documents, and other sources and convert them into electronic format. The system may obtain Handwriting sources from a piece of paper through scanning and converted them in to image or file format.

Following are some different methods of handwriting recognition system with its advantages and disadvantages.

Table -1: Methods used for Handwritten Recognition

Method	Advantage	Disadvantage
Convolutional Neural Network Method	1) When CNN is trained then the image recognition will be accurate 2) It has been proven successful used for many handwritings and computer recognition	1) In order for high accuracy in the training process should use many Samples. 2) Long training time
Semi Incremental Method	1) Waiting time will not be too visible 2) Considers the latest strokes and previous segments	1) Should be accompanied by other methods can not only semi-incremental only. 2) The way it works is more complicated than the pure incremental method
Line and Word Segmentation Method	1) Line and Words Segmentation Approach for Printed Documents to project a powerful process	1) Can't detect pattern
Slope and Slant Correction Method	1) Segmentation process simpler and the accuracy of the writing recognition	1) Not maximal for handwriting recognition

Zoning Method	1) Accuracy results are quite high.	1) The number of zones in an image must be many, because if a little then the accuracy will be smaller.
---------------	-------------------------------------	---

2.2 Motivation

The research on cursive handwritten word recognition is a large area motivated by many application areas, such as doctors cursive handwriting. students handwritten notes, historical documentation, text-voice conversion, security, etc.

3. PRAPOSED SYSTEM

There are two basic types of handwriting recognition systems first is Online recognition and another one is offline recognition. Both types can be implemented in applications to learn while performing offline learning on data. Several implementations have been used for both type of handwriting recognition fields, such as statistical and structural methods, neural networks and syntactic methods. Some recognition system identifies strokes and apply recognition on a single character or entire word or complete file using deep learning.

3.1 System Design

The system proposed to recognize the character present in the image of PDF files. This character can be digital or human handwriting characters. The architecture of proposed scheme is shown in Fig.1 consisting of processing the provided input and generating the result based on the trained Keras-OCR model. To improve result, training of the data sets will be required

Image pre-processing:

Image pre-processing is most important part in the recognition for correct character prediction. These cursive recognitions typically include noise removal, image segmentation, cropping, scaling, and more. This system first take input as a scanned image or pdf of cursive writing. These images can be in JPEG or PNG format.

Digital capture and conversion of an image contains noise, shadow, contrast, sharpness, which makes it hard to identify what is actually a part of the object of interest. Considering the problem of character recognition, we want to reduce as much noise and all the image constraint which effect on conversion as possible for correct classification

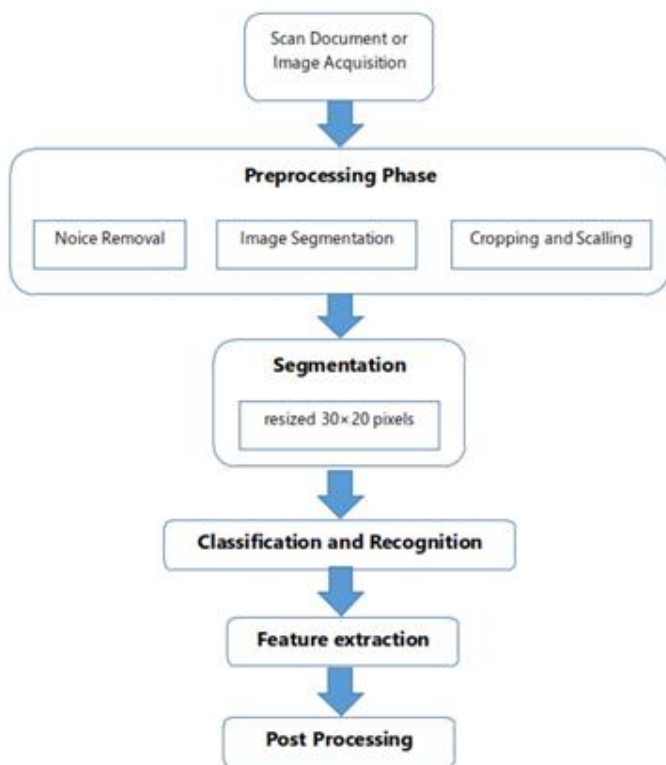


Fig. Architecture of Cursive Handwriting Recognition

Segmentation:

In the segmentation stage, a sequence of characters is segmented into a sub-image of an individual character and resized into pixels.

Classification and Recognition:

This stage is the decision-making stage of the recognition system. These contains two layers, which are used for activation function to train the algorithm.

Feature extraction:

The features of properties of input data under observations, which is used to analyses or classify these instances of data. The task of feature extraction is to identify the features that see the difference between these two instances that are independent of each other.

Character extraction:

Individual characters are recognized by ease with OCR. Cursive handwriting, which having connected letters added more issues with evaluation. It is difficult to interpret cursive handwriting because cursive writing changed by person to person with no distinct separation between characters.

3.2 Algorithm

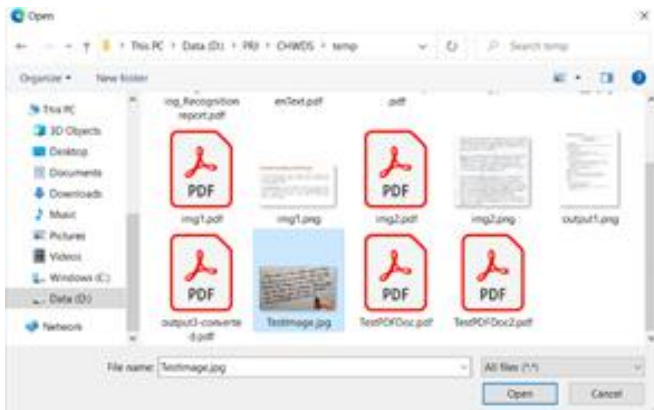
In this section, we introduce algorithms for Cursive Handwriting detection. This algorithm converts the input file from pdf to jpeg in case pdf format is provided. This image has cursive handwritten format, which needs to slant to straighten an image. Post on the straightened image, segmentation is performed This includes word and character segmentation. In Centre align process, it takes in the segmented character images, adds padding, resizes and Center aligns the images, which is ready to be fed to the model for prediction. In Model build module, it builds the CNN model using Keres, in this training of the Neural network is done. In recognition, it uses the model built to recognize the characters in the images.

It consists of following steps.

- 1] Start
- 2] Server is hosted on local host, visit the localhost URL
- 3] Take input from user in PDF or Image format and upload the file
- 4] If it is PDF, it converts it into image/s
- 5] Post on the converted image, slating and grey scale operations is performed.
- 6] Press Start Detection
- 7] Post Slanting, Segmentation of the image is performed
- 8] Adds padding, resizes and center aligns the image
- 9] Train CNN model using Keras
- 10] Recognise the characters in the images using Keras Model.
- 11] File will get stored at file location on server
- 12] Output file will be available for download
- 13] Stop

4. EXPERIMENTAL RESULTS

The Cursive Handwriting Detection system is based on the Deep Learning and NLP based system designed and implemented using the Python and Keras-OCR framework



In this screen, user is allowed to upload file with format as PDF or image file formats. If other file format is uploaded, then it will return an error. Validation for the file format is handled in backend and response is provided to web page

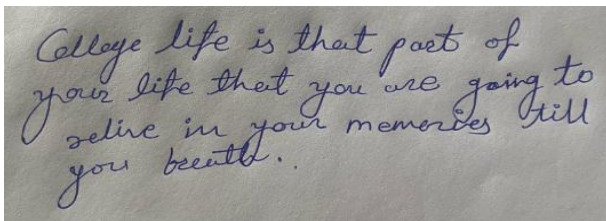


Fig. Handwritten Cursive Writing

An experiment is conducted to show the existing Handwriting detection methods with this system.

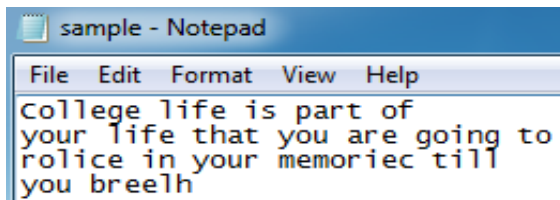


Fig. Converted Text data

The system analyse the provided input with high accuracy depends upon the inputs files quality. The process time to analyse the result is very low and output is store in the plain file.

CONCLUSION

The problem of the cursive handwriting is made complex, the letters in a word are generally linked together, poorly written, and may even be missing. As a result, cursive script recognition requires lots of experienced techniques, which uses a large amount of shape information and which compensates for the ambiguity by the use of contextual information

In this system, we provide a smart way to converting scanned handwritten notes into digital text. By using the Conventional

Neural Network (CNN) to recognize handwritten text and return readable digital text. This makes the ordinary persons to understand what written in cursive also make readable in digital format to search and storage in any digital platform using deep machine learning. This system will also help to convert the old books to the textual format, which will help In researches and analytics of the handwritten data.

REFERENCES

- [1] Savitha Attigeri; "Neural Network based Handwritten Character Recognition system", International Journal of Engineering and Computer Science ISSN:2319-7242, Volume 7 Issue 3, March 2018.
- [2] Megha Agarwal, Shalika, Vinam Tomar, Priyanka Gupta; "Handwritten Character Recognition using Neural Network and Tensor Flow", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue- 6S4, April 2019.
- [3] Sara Aqab, Muhammad Usman Tariq; "Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, 2020
- [4] Sudharshan Duth P, Amulya B ; "Recognition of Handwritten and Printed Text of Cursive Writing Utilizing Optical Character Recognition", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore Part Number:CFP20K74-ART; ISBN: 978-1-7281-4876-2 Paris, France 22-23, June 2018.
- [5] Nikita Mehta, Jyotika Doshi; "A Review of Handwritten Character Recognition", International Journal of Computer Applications (0975 - 8887) Volume 165 - No.4, May 2017.
- [6] Salma Shofia Rosyda, Tito Waluyo Purboyo; "A Review of Various Handwriting Recognition Methods", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 2, 2018.