

Handwritten Character Recognition

NAVYA A¹, MUDALIYAR M JANVI J MOHAN², NISHCHITHA R HATHWAR³, PAVAN KUMAR K N⁴, RUDRAMURTHY V C⁵

^{1, 2, 3, 4} Student, Global Academy of Technology

⁵ Assistant Professor, Global Academy of Technology

Abstract— Scientists have made numerous improvements to Deep Learning for Computer Vision, which have aided in the development of millions of smart gadgets. Even though scientists have revolutionized in the field of image processing, one of the most difficult aspects in it is the identification of handwritten text that are available in the documents. It's a method for verifying the veracity of these kinds. Character Recognition is the term for this. The idea of this research is to read a handwritten word and characterize it into a digitally accessible format. Through the use of neural network, models are developed that can identify the handwritten characters in the image.

Indexed Terms-- Convolutional neural networks (CNN), Deep learning, Handwriting Recognition, Image classification.

I. INTRODUCTION

- A computer having the capability of recognizing handwritten characters is said to be able to acquire the knowledge of detection of characters and text in documents, images, smart devices, and other sources, and transform them into machine-encoded form.
- It's usage is found in optical character recognition and more complex text recognition systems. These systems generally implement machine learning mechanisms such as neural networks.
- Most of these systems nowadays implement machine and deep learning mechanisms such as neural networks. Machine learning is a sub unit of artificial intelligence that defines the process of studying from a set of data which can be used to solve a wide range of problems. This is inspired by human psychology and biology.
- Neural networks are learning models used in machine learning whose purpose is to imitate the

working process that occurs in a human brain. These are one of the most efficient models, which are applicable in automation of activities when the time taken by a human being to do the same task is very long and not accurate.

- Using scanning technology and hard-wired logic, Jacob Rabinowitz's early postal readers were able to distinguish monospaced fonts. A more advanced scanner by Allum et al. provided for much more variances over how the text was generated and encrypted the intelligence into a barcode printed directly on the letter.

DEFINITIONS

- HCR: Handwritten Character Recognition is the process of recognizing the handwritten characters present in the image.
- CNN: Convolutional Neural Network is a type of neural network which is widely used for recognizing the figures in the image.
- LeNet Architecture: LeNet is a type of architecture that possesses a number of layers such as convolutional layer, pooling layer and fully connected layer.
- GUI: Graphical User Interface is the user interface used by the user to write the characters and letters.
- SRS: System Requirements Specification is the basic requirements of the system to perform HCR.
- DFD: Data Flow Diagram depicts the flow of data in a process or a system.
- MNIST: Modified National Institute of Standards and Technology.

II. REVIEW OF LITERATURE

This chapter evaluates the current work with previous work. It depicts the current implementation that overcomes the previous challenge.

A. *System Study*

1. Handwritten Character Recognition using Convolutional Neural Networks in Python with Keras

- This project classifies an individual handwritten character so that handwritten text can be translated to a digital format.
- The project uses neural networks for developing a system that can recognize handwritten English alphabets. In this project, the type of neural network used is Convolutional Neural Network (CNN) which can be implemented with various architectures to train a model that can accurately classify characters.
- In this experiment the number of epochs were kept low which reduced the training time of the model but failed to recognize similar looking alphabets distinctly.

2. A Review of Various Handwriting Recognition Methods

- This paper discusses the various methods that can be used to effectively recognize a character.
- The methods discussed are Convolutional Neural Network (CNN), Semi-Incremental Segmentation, Incremental Recognition, Lines and Words Segmentation, Part Based Method, Slope and Slant Correction Method, Ensemble Method and Zoning Methods.
- Out of these, the CNN method to recognize characters is discussed in detail as it produces results with the highest accuracy among all the methods, the lowest being Slope and Slant Correction Method.
- In conclusion, the recognition of handwritten characters is more challenging as each and every person will have different handwritings which will be more difficult to detect compared to writings from computers that already have a definite standard form.

3. Handwritten Character Recognition Using HOG, COM by OpenCV & Python

- Experimental results using KNN and SVM classification techniques are presented in this paper.

- This paper applies the HOG features for character recognition followed by center of mass of image with SVM algorithm.
- HOG features are extracted by deskewing the image converted to center of the image. This experiment uses the HASY dataset which contains 168233 instances of 369 classes.
- The result of the experiment is that in KNN, the errors occur due to improper segmentation due to which features extracted being unambiguous.
- Therefore, the paper concludes that SVM approach leads to better results with higher accuracy.

4. A Literature Survey on Handwritten Character Recognition

- This paper discusses the working of a handwritten character recognition system and also has a detailed literature survey related to the same method.
- The paper describes the six phases of character recognition that are image acquisition, pre-processing, segmentation, feature extraction, classification and post processing.
- The paper reviews all advances in the area of handwritten character recognition in depth.
- A number of techniques have been described for character recognition here.
- The paper concludes that the selection of the classification algorithm and the character-feature extraction methods play a major role in attaining maximum efficiency in recognizing the character.
- The paper also reveals that there is scope of improving the algorithms and feature extraction methods as well as the rate of recognition of characters.

5. Prospective Classification over Various Handwritten Character Recognition Algorithms – A Survey

- This paper is a survey about the various algorithms that are used to perform Handwritten Character Recognition.
- The various methods discussed are Support Vector Machine (SVM), Multilayer Perception (MLP), Convolutional Neural Network (CNN).
- This paper differentiates the methodologies used and the execution time taken and gives a clear idea on the flow of action.

- SVM has maximum precision when it comes to preparing information, while CNN has maximum precision when it comes to testing datasets.
- This paper incorporates the correctness given by the SVM.

B. Proposed Work

We propose a Handwritten Character Recognition system to recognize the characters and digitize it. This system reduces the error rate in processing of handwritten work and converts them in lesser time. It takes into consideration, every type of handwriting and converts it thus avoiding confusion and makes the handwritten content easily understandable. Results illustrate that the proposed system provides high accuracy than the existing systems.

C. Scope of the Project

Although the world has evolved technologically today, still many people tend to use the traditional method of writing using pen and paper. However, there are drawbacks with respect to this. Physically handwritten data takes up more physical storage space and accessing it also becomes difficult. It is difficult to search through them effectively. Chances of data loss is very high in these cases as the information does not get transferred to digital format at all.

Handwritten Character Recognition is also used in signature identification in several government organizations and banks where the characters and the patterns are recognized by the system to identify its originality.

This is also used in Research related activities where the old documents and manuscripts found are recognized using Handwritten Character Recognition and the information is transferred to digital format.

Handwritten Character Recognition can also be used in converting a handwritten application form to a digital form. Here, the filled form is recognized using the model and the characters are converted into their respective digital characters which helps in easier understanding and faster processing of the forms.

III. SYSTEM REQUIREMENT SPECIFICATION

The System Requirement Specification (SRS) is a key document that establishes the outline of the product development process. It records the necessities of a framework as well as has a depiction of its significant highlight. It's a two-way approach that makes sure that both the customer and the association comprehend each other's necessities from the same viewpoint at any given point in time.

A. Functional Requirements

This section defines the functional needs of the system for those requirements which are expressed in the natural language style.

1. The system should convert the character written by the user to an image format.
2. System should be able to detect the characters present in the image.
3. System should be able to retrieve the characters detected and display them to the user.

B. Non-Functional Requirements

These are requirements that are not functional in nature, that is, these are constraints within which the system must work.

1. Performance: Handwritten characters in the supplied image will be identified with a 90% or higher accuracy rate.
2. Functionality: This software will be able to deliver all the functional needs mentioned in the document.
3. Availability: The system will fetch those regions which contain handwritten text in it.
4. Learnability: The software is simple to use and saves the amount of time spent learning.

C. Hardware Components

- System : Intel I3.
- Hard Disk : 5 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 2 GB

D. Software Requirements

- Operating System: Windows 8 or above
- Programming language: Python 3.8

- Development environment: Visual Studio code

IV. SYSTEM DESIGN

This chapter gives an in-depth insight about the project. This chapter describes the basic theoretical information regarding each component and the related aspects of the project, such as system design, system architecture, flow diagrams, sequence diagrams.

System design is the process of defining the architecture, components, modules, interface, and data for a system to satisfy specified requirements. System design can be perceived as the application of system theory to product development. There is some overlap with respect to the fields of system analysis, system architecture and system engineering.

A. Design Overview

The process of implementing multiple requirements and creating the physical reality of the system is defined as "design". Various design features are followed to develop the system based on the design specification which describe the features of the system, elements of the system and their appearance to the end-users.

B. System Architecture

Fig 1 describes the System Architecture of Handwritten Character Recognition project. In this diagram, a character is taken from the user and it is converted to an image. Then the image is pre-processed. The pre-processed image will be resized to 28 x 28 pixel image and given to Convolution Neural Network (CNN) model. The CNN model would have been trained from MNIST dataset. The trained CNN model recognizes the character present in the image and displays it in the output terminal.

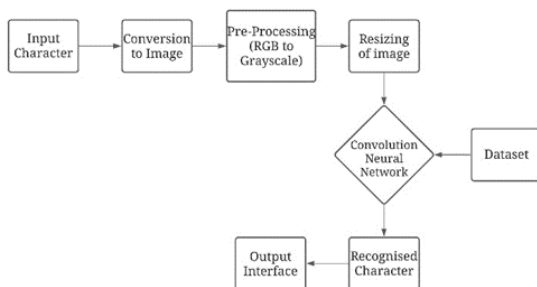


Fig 1: System Architecture

C. Data Flow Diagram

A data-flow diagram is a way of representing the flow of data through a system or a process. The DFD also contains information on each entity's input and output, as well as the process itself.

Fig 2 describes the DFD Level 0 diagram. In this diagram the input character image is given to CNN and the CNN will recognize the character present in the image and displays it in the output terminal.

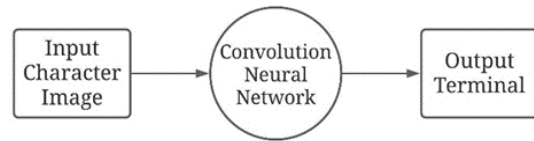


Fig 2: DFD 0

Fig 3 describes the DFD Level 1 diagram which gives a more detailed idea of Level 0 diagram. The CNN is split into 3 layers, the layer consists of Convolution, Pooling and Fully Connected layer which will extract the features from the image and classify the character present in the image.

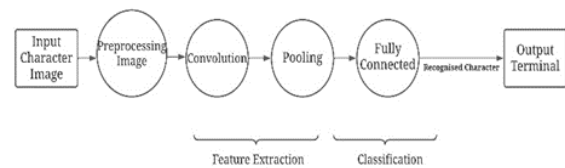


Fig 3: DFD 1

V. MODULES

- Training the model: To train the CNN model with a given dataset.
- Image Preprocessing: To pre-process the image into required format.
- Classify Handwriting: To recognize the handwriting character and display it and also store it in a file.

VI. CONCLUSION

Handwritten Character Recognition is used for conversion of written text to typed text format dynamically. A character will be taken from the user

and will be preprocessed and given to CNN to recognize the character and it will display the recognized character. User no longer needs keyboard to write the text. User can write text on the screen which will be converted to typed text and will be displayed on the screen and also it will be stored in a file. The GUI will be user friendly for writing text on the screen.

REFERENCES

- [1] Hanu Priya Indiran, “Handwritten Character Recognition using Convolutional Neural Networks in Python with Keras”, Asian Journal of Convergence In Technology, Volume V, Issue III, ISSN No: 2350-1145 I.F-5.11.
- [2] Salma Shofia Rosyda1 and Tito Waluyo Purboyo2, “A Review of Various Handwriting Recognition Methods”, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 13, Number 2 (2018) pp. 1155-1164.
- [3] Rushikesh Laxmikant Kulkarni, “Handwritten Character Recognition Using HOG, COM by OpenCV & Python”, International Journal of Advance Research in Computer Science and Management Studies, Volume 5, Issue 4, April 2017.
- [4] Ayush Purohit, Shardul Singh Chauhan, Centre for Information Technology, University of Petroleum and Energy Studies, Dehradun, India, (IJCSIT), “A Literature Survey on Handwritten Character Recognition,” International Journal of Computer Science and Information Technologies, Vol. 7 (1), 2016, 1-5.
- [5] Pradheeba U,1, Bhavani M, Yuvaraj B R and Krithika V, Department of CSE, R.M.K. College of Engineering and Technology, India Department of CSE, Rajalakshmi Engineering College, India, Department of IST, Anna University, India, Department of R&D and RDAF, National Institute of Wind Energy, India, Advances in Parallel Computing Technologies and Applications, D.J. Hemanth et al. (Eds.)© 2021, “Prospective Classification over Various Handwritten Character Recognition Algorithms – A Survey” The authors and IOS Press, doi:10.3233/APC210166.