International Journal of Innovative Research in Engineering
Volume 3, Issue 6 (November-December 2022), PP: 15-16.
www.theijire.com
ISSN No: 2582-8746

OCR Based Mark sheet Digitization

Jeyaseelan R[1], Krishnaraj R[2], Rishikeshan J[3]
[1,2,3] *Computer Science and Engineering, Bannari Amman Institute of Technology, TN, India.*

How to cite this paper:
Jeyaseelan R[1], Krishnaraj R[2], Rishikeshan J[3], "OCR Based Marksheet Digitization", IJIRE-V3I06-15-16.

Copyright © 2022 by author(s) and5th Dimension Research Publication.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

Abstract: *Every Educational institute need some kind of formatted mark sheet. Here in our project, we make work simpler for these institutes. Transforming printed or handwritten documents directly to the database. In this software, theuser just needs to scan the copy of the mark sheet and the rest of the thing is doneby software. The scanned file is stored as an Image file so this Image file undergoes processing and the user data is extracted. This data is stored in a database, thus reducing the manual burden. Our project works for any kind of format, so this makes our project Dynamic. Our project has to undergo various Image Processing tasks.*

Key Word: *OCR, tesser act..*

I.INTRODUCTION

Traditionally, document storage was done with paper and traditional file systems, however, this form of storage is immune to degradation due to time and natural dilapidation. Another alternative is the digitization of all information manually. This is why Digital Image processing of documents is proving to be vital function for any organization.

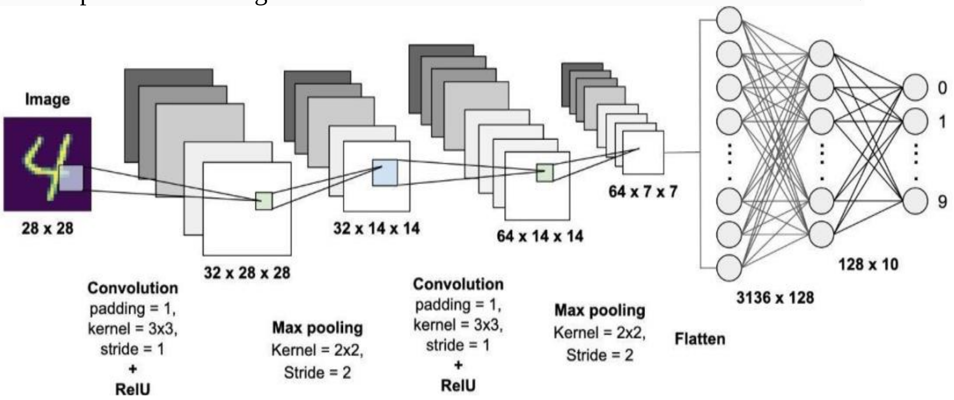
Growing demand for the need to digitize data an automated tool for converting hard-coded data into digital form format.Data cannot be obtained directly from images; this has to be done manually. Only OCR can detect text, however, image preprocessing is very essential before using OCR because the image is raw the form cannot be processed by OCR. Also after OCR detects text from images, the data obtained should be stored in a database where they can be handled and easyto process. Currently, there are mark sheet details manually entered into the database. That requires a lot of human effort and is also time-consuming. in addition, there is a risk of human error as it is tedious job. That's why we try to do this using image processing methods to automate the entire process of creating a student database from brands.

II. LITERATURE SURVEY.

Character recognition is not a new problem, but its roots can be traced back to systems before the invention of computers. It has ability to reliably read text is still far below that of a human. Therefore, current OCR research is being conducted to improve the accuracy and speed of OCR for documents of various styles printed/written in unconstrained environments. There was no open-source or commercial software available for complex languages

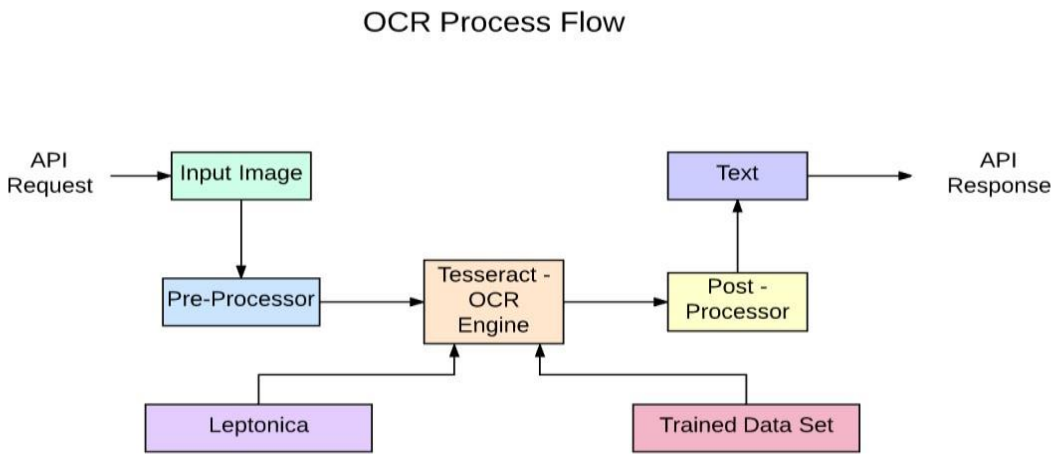
III.PROPOSED METHODOLOGY

OCR work on character recognition system that supports character recognition of multiple languages. And also eliminates the problem of different character recognition and supports multiple functions that need to be performed on a document. Grid infrastructure in this context means infrastructure that supports a group of specific sets of languages. Optical character recognition approaches and also described graphical representation of OCR algorithmic variations with their handled steps for processing various levels of text and the flow of methodology. This descriptive graphical representation will be helpful to all upcoming researchers in the innovative OCR field. Graphical representation flow is easy to understand and simple to gain the basic important knowledge to focus in their OCR research for further innovations.



OCR is built on convolutional neural network (CNN), a popular deep neural network architecture. Traditional CNN classifiers are able to learn the important features present in the images and classify them, the classification is done usinga soft-max layer. We presented OCR by combining CNN and ECOC (Error Correcting Output Code) classifiers. In this CNN used for feature extraction and for classification of image ECOC is used.

To find a suitable feature extraction CNN. which can be used in combination with the classifier to accurate handwritten character recognition, several popular CNN classifiers were investigated. CNN-ECOCs are trained and validated using theNIST handwritten character dataset. The simulation result shows that CNN-ECOC provides higher accuracy.



IV.OCR (OPTICAL CHARACTER RECOGNITION)

OCR is optical character reader, the conversion of images, printed text into editable text, whether from a scanned document, photographic document, photograph of a scene, or caption text overlaid on an image. Instead, OCR extracts the relevant information and automatically enters it. The result is accurate and efficient processing of information without time.

V.CONCLUSION

OCR treats typical compound characters (combinations of half letters) as separate classes to improve segmentation accuracy.

References

1. *Savita Ahlawat, Amit Choudhary, Anand Nayyar, Saurabh Singh and Byungun Yoon, "Improved Handwritten Digit Recognition Using Convolutional Neural Networks(CNN): Sensors", MDPI, 2020*
2. *Chaudhuri Arindam, Mandaviya Krupa, Badelia Pratixa, Ghosh Soumya K, etal.“Optical Character Recognition System Optical CharacterRecognition Systems for Different Languages with Soft Computing, Springer (2017)*