JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Text Extraction from an Image using CNN

**[1] P. Rajeshwari [2] D. Vinay Sekhar Reddy ,[3] G. Pranay,[4] Mohd Arbaz Mazharuddin**
[1] Assistant Professor,Department of Computer Science and Engineering,
Anurag Group Of Institutions,Telangana,India,
[2,3,4] Student,Department of Computer Science and Engineering,
Anurag Group Of Institutions,Telangana,India,

## ABSTRACT

The primary goal of this research is to automatically extract text from images and summarise visual information using a deep learning-based technology called Convolutional Neural Network (CNN) for text extraction from images. Target text region identification and classification are two separate phases utilised in text extraction. In the first step, rectangular text-containing zones were discovered, which included neural processing and geometrical limitations. The resulting target text regions were then sent through many layers of CNN, which initially created a convolutional feature map before classifying the candidate regions as text or not-text. As a result, we can extract the text from the given image.
**Keywords**: Text extraction,Text comments,Image extraction using CNN.

## INTRODUCTION

Keeping, modifying, indexing, and finding information in a digital document is considerably easier than scrolling through printed/handwritten/typed documents in this era of digitization. Furthermore, searching for something in a huge non-digital document is not only time intensive, but it is also likely that we will overlook information while manually scrolling the text. As a result, text extraction from photos has been introduced. Passport recognition, automatic number plate recognition, translating handwritten texts to digital text, converting typed text to digital text, and so on are some of the applications. Many Computer Vision jobs need the detection of text in a natural scene. This project's main goal is to extract text from an image, which can be either documented text or scene text.

## RELATED WORKS

This section gives a high-level summary of the current state of text recognition research. Text recognition has been around for quite some time. The performance of the Tesseract Optical Character Recognition (OCR) engine is improved by considering images with coloured backgrounds and describing a preprocessing procedure. The text is separated from the colourful background via text segmentation, which is accomplished by separating the original image into k images. The image with text is then recognised by a classifier. When compared to the Tesseract OCR performance, preprocessing resulted in a 20 percent improvement. One of the earlier methods relies on a text extraction pipeline, which extracts text from photos of varying quality collected from social media. Their major task is to divide the input photos into distinct classes, after which preprocessing is carried out according to the classes. Following that, an OCR engine is used to recognise text. This research makes use of a dataset gathered from social media. The text component of photographs is identified via OCR. The authors present a method for extracting text from a scanned page in their paper. The Otsus method was employed for segmentation, and the Hough transform was used for skew detection in this study. Characters have also been identified using the OCR technology. They conducted tests and validated the suggested algorithm on a variety of photos obtained from diverse sources. The average accuracy was determined to be 93%. The approach uses a combination of powerful optical character recognition (OCR) and text-based searching technologies to discover and recognise text in pictures and video frames. Text characters in photos, unfortunately, can be any gray-scale value (not usually white), variable size, low-resolution, and imbedded in complicated backgrounds. Text segmentation frequently uses texture as a characteristic. Many academics working on text detection and thresholding algorithms used a variety of ways to get good results, which were limited by certain constraints. To bridge the gap between image documents and the input of a normal OCR system, effective identification and segmentation of text characters from the backdrop is required. Previously, three methods were offered, separated into bottom-up and top-down ways. Bottom-up approaches divide images into regions, which are then grouped into words as "character" regions. As a result, the recognition algorithm's performance is determined by the image's complexity and segmentation algorithm. Top-down algorithms identify text regions in photos and then divide them into text and background. Although they can handle more complicated images than bottom-up techniques, they still have problems at the detection and segmentation identification stages. Unless the text is anticipated to appear at pre-determined page places, such as in forms processing, the text must be positioned in some way. Text detection in document processing is frequently seen as a simple procedure. Typically, this entails looking for text lines in a binarized image. Processing and classifying related components are two further ways. Nagy provides a summary of the area of document image analysis as represented by publications in Transactions on Pattern Analysis and Machine Intelligence. Detecting text in scenes or low-resolution video is more difficult, however challenges sponsored by the International Conference on Document Analysis and Recognition in 2003 and 2005 brought these issues to the forefront. There have been a variety of approaches for detecting and recognising text in photographs. In, you'll find extensive surveys. Text detection methods can be divided into three types: texture-based methods, region-based methods, and hybrid methods. Furthermore, these algorithms are mostly focused on detecting horizontal texts. Color, grey scale, and alignment attributes of a text region, as well as differences in background properties, are used in region-based approaches. Using segmentation or clustering, extract candidate text regions first, then delete non-text parts. Different document or online, e-mail images, where text characters are normalised and correct resolutions are used, natural scene photos, where text can be embedded in a variety of sizes, shapes, and orientations into a complex background Because of the complex background, it is impossible to discern text in photos using OCR software. This is all about various text detection approaches that accurately detect text and aid in better recognition.

## PROPOSED SYSTEM

Our method is built on deep learning, and the to-go libraries for extracting text from photos are OpenCV, EasyOCR, and PyTesseract. Optical character recognition (OCR) is performed using traditional computer vision algorithms, and from the fourth edition forward, neural network components such as LSTM are included. In Python, the OpenCV package is used to perform computer vision operations. We'll utilise it to overlay photos with their corresponding identified sentences later. To display photos, we'll require the Matplotlib module. Text detection algorithms are needed to recognise text in a picture and construct a bounding box around the text-containing area. Standard strategies for detecting objections will also work here. For text region detection, geometric parameters such as eccentricity, bounding box, solidity, and euler number are also taken into account. For detecting the region of interest, connected components within a region are also taken into account. The stroke-width is taken into account while removing non-text parts. When it comes to stroke widths of lines and curves, text characters have a lot of variation, whereas non-text areas have a lot of variation. As a result, regions with a lot of variance in stroke width are deleted because they're more likely to be non-text regions. OCR (Optical Character Recognition) is used to digitise the discovered text sections and to detect and extract the text from the image.

## ADVANTAGES

1. It helps you save time and effort.
2. Increases precision
3. Requires less manpower

## CLASSIFIER USED

Convolutional Neural Networks are complicated feed forward neural networks used in machine learning. Because of its great accuracy, CNNs are employed for picture categorization and recognition. It aids the computer in detecting details that might otherwise go unnoticed if an image were merely flattened into its pixel values. In comparison to a typical neural network, CNNs are constructed differently. CNN has three-dimensional layers in terms of width, height, and depth. Decoding Facial Recognition, Analyzing Documents, Historic and Environmental Collections, Understanding Climate, Grey Areas, Advertising, and Other Fields are some of the CNN applications.

## RESULT

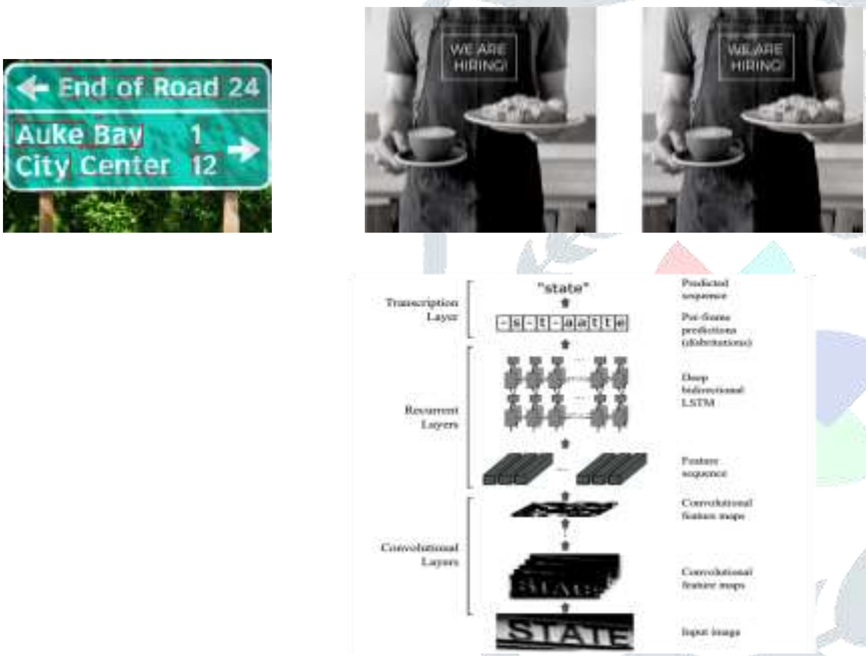


Figure 1: The Network Architecture

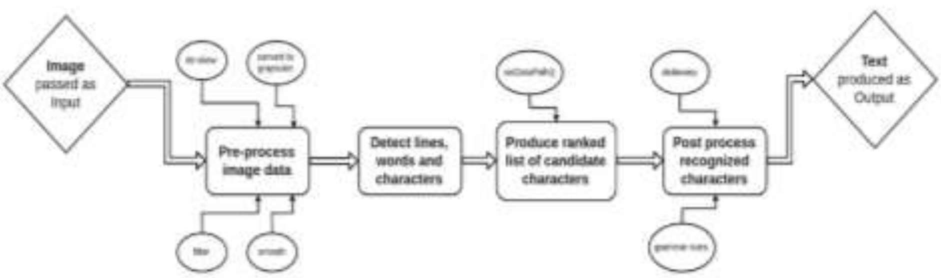


Figure 2: Working Architecture of EasyOCR


Figure 3: Realtime Output

## CONCLUSION

We can effectively develop a deep learning model that can extract data from a given image in both structured and unstructured environments with the highest possible accuracy using this project and its implementations. We hope that by completing this research, we will be able to eliminate the need for manual text reading in images. Certain characters may not be recognised correctly in a few circumstances, i.e. a character may be misidentified as another character. This could be due to a mismatch between the processed input matrix's pixel intensity values and the multiple stored character matrices. The experiments on this model show that the majority of the text is effectively recognised, and that the recommended methodology is effective.

## FUTURE ENHANCEMENTS

A web-based application can be created in which the user can upload an image or a video and extract all of the text material included in them, which can then be stored in a database. We can also remove any text that is present in the image or video.

## REFERENCES

[1] Text detection and recognition in images and video frames Datong Chen, Jean-Marc Odobez, Herve Bourlard Received 30 December 2002; accepted 20 June 2003.

[2] X. Chen and A. Yuille, “Detecting and reading text in natural scenes,” in Computer Vision and Pattern Recognition, vol. 2, 2004.

[3] Chucai Yi,YingLi Yian “Text string detection from natural scene by structure based partition and grouping “in IEEE transaction on image processing sep 2011.

[4] K. C. Jung, K. I. Kim, and A. K. Jain, “Text information extraction in images and video: A survey," Pattern Recognit., vol. 5, pp. 977, May 2004.

[5] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, S.Hwang, “Automatic detection and recognition of Korean text in outdoor signboard images”, Pattern Recognition Letters,2010.

[6] Datong Chen, Jean-Marc Odobez, Herve Bourlard "Text detection and recognition in images and video frames" Pattern Recognition, vol.37 ,2004,pp 595 .

[7] Charles Jacobs, Patrice Y. Simard"Text Recognition of Lowresolution Document Images",IEEE conference of document analysis, 2004.

[8] Asif Shahab, Faisal Shafait, Andreas Dengel,"ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images",Int. Jour. on Document Analysis and Recognition,2011.

[9] Y. Pan, X. Hou, and C. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. on Image Processing, vol. 20, no. 3, pp. 800, Mar. 2011.

[10] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Text detection in indoor/outdoor scene images," in 1st Int. Workshop on Camerabased Document Analysis and Recognition, Seoul, Korea, Aug. 2005, pp. 127-132