# An Enhanced Approach towards Object Detection Using Deep Learning Techniques

[1]S. Priyadarsini, [2]G.VijiPriya, [3]Dr.E.S. Shamila, [4]A.Praveena, [5]A.Rathika

[1]Student Scholar, [2,4,5]Assistant professor, [3]Associate Professor
Department of Computer Science and Engineering,
Jansons institute of Technology, Coimbatore, India.

***Abstract:*** **Text recognition in images and videos have advanced as an active research area over last few years. Text detection from the scene image is a process by which text regions are segmented from non-textual ones and they are organized in accordance with their correct direction of reading. Different text patterns and variant background interferences are the challenges that affect the consistency of text character extraction. Digitization of text documents is frequently united with the progress of optical character recognition (OCR). OCR tool gives good outcomes obtained to read the text from an image. The objective study of this paper is to propose a text recognition model that is designed using rich supervision to accelerate training and achieve state-of-the-art performance on several benchmark datasets. A deep neural network based system is used to read scene text and show that scene text reading can be effectively applied for the purpose of retrieving objects. In this paper, we examine a modest but powerful approach to make robust use of HOG and LBP features for text recognition. Histograms of Oriented Gradients (HOGs) and Local Binary Patterns (LBPs) have proven to be an effective descriptor for object recognition in general and text recognition in specific. Experimental results over the ICDAR 2013 and SVT 2010 dataset demonstrate the efficiency of the proposed approach.**

***Keywords*****: Deep learning, Image processing, OCR, HOG, LBP.**

## I. INTRODUCTION

In recent years, text recognition from images and videos gained popular interest with progression in pattern recognition and computer vision technology. The import of semantic or high-level text data present in an image is that it can easily describe an image with good clarity and can be extracted using low-level features like color, texture etc., which in turn varies with language, font, style and background, thus making the task of text extraction a challenging one [13]. Recognition of text is yet another challenge for the researchers as low resolution text with small fonts may be present in an image or video with complex or textured background [14]. The aim of the proposed method is to detect and recognize the text- candidates from a scene-text image. The first stage of this method is text detection and the second stage deals with recognition of the detected text characters. The system presented in this paper is an approach to detect, localize and recognize the text in natural scene images which consists of four stages. In the initial, pre-processing stage, to detect text regions in each layer of the image, a text region detector is designed by using a widely used feature descriptor: in addition enhancement technique like Simultaneous, Noise Removal, Filtering and Contrast Enhancement are done. Second stage covers segmentation process in which morphological operations are done. Third stage is feature extraction where certain deep learning algorithms like histogram of oriented gradients (HOG) and local binary pattern (LBP). Fourth stage is classification which uses Artificial Neural Network for abstracting image from the data set. Clustering of the extracted text candidates from region filtered image is done using morphological operations. An OCR is then used for recognizing the characters. The rest of this paper is organized as follows. The related works done in previous papers based on the text extraction algorithm is discussed in section II, the proposed methodology is detailed in section III, the performance evaluation is given in section IV and finally, the inference and the future scope is concluded in section V.

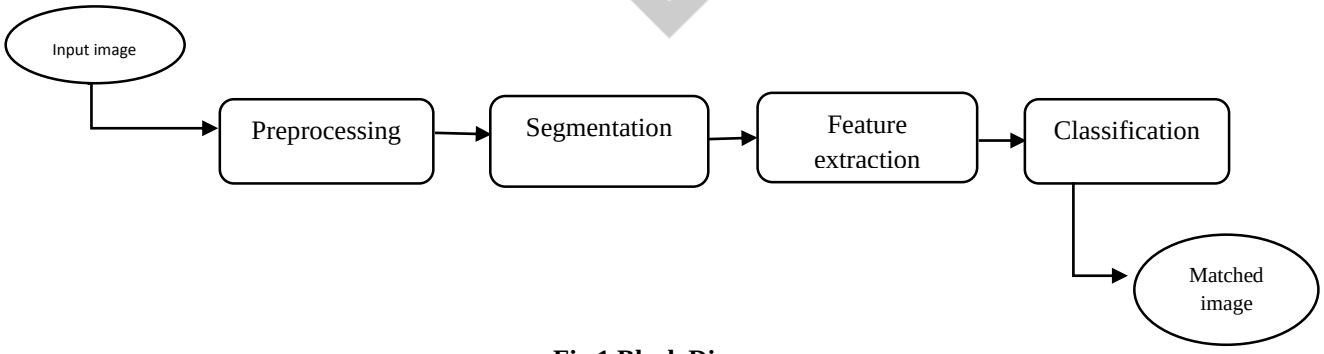


**Fig.1 Block Diagram**

## II. RELATED WORKS

Over the last few decades, detection and recognition of text in images and videos have been developed as an active research area. The unique features of text help us to distinguish it from non-textual regions. Most commonly used text extraction methods are summarized in this section.

[1]Jiji Mol1 et al (2017) proposed a novel system for text detection and recognition in images. The proposed method uses Fractional Poisson enhancement for removing Laplacian noise of the input image. Then Edge-enhanced Maximally Stable Extremal Regions

(MSERs) is obtained from the pre-processed image. Region filtering is used to filter non-text regions and is then recognized by an Optical Character Recognition (OCR) system. Then Edge-enhanced Maximally Stable Extremal Regions (MSERs) is obtained from the pre-processed image. Region filtering is used to filter non-text regions and is then recognized by an Optical Character Recognition (OCR) system.

[2] Peng et al (2017) introduce hashing layer to CNNS, and learn the image features and hash functions simultaneously, and construct hash functions with the independence and quantization error minimized. The hashing layer contains two parts. The first part includes slice layer, fully-connected layer, activation layer and concat layer. The first part is used to map the features to continuous codes and generate independent hash functions.

[3] A deep recurrent model, building on long short-term memory (LSTM), is developed to robustly recognize the generated CNN sequences, departing from most existing approaches recognizing each character independently. Pan He et al (2016) has proposed a number of appealing properties in comparison to existing scene text recognition methods: (i) It can recognise highly ambiguous words by leveraging meaningful context information, allowing it to work reliably without either pre- or post-processing; (ii) the deep CNN feature is robust to various image distortions; (iii) it retains the explicit order information in word image, which is essential to discriminate word strings; (iv) the model does not depend on pre-defined dictionary, and it can process unknown words and arbitrary strings.

[4] A Character recognition mechanisms is needed to perform Document Image Analysis (DIA) which transforms documents in paper format to electronic format. In this paper they have discuss method for text recognition from images. Manwatkar et al (2015) recognition of text from image for better understanding of the reader by using particular sequence of different processing module. He try to extract basic constituent of the script, which are certainly characters. In this project, he perform the segmentation of character from image by applying Line detection and Character detection algorithm.

## III.PROPOSED METHODOLOGY

The proposed method consists of four steps are briefly described below.

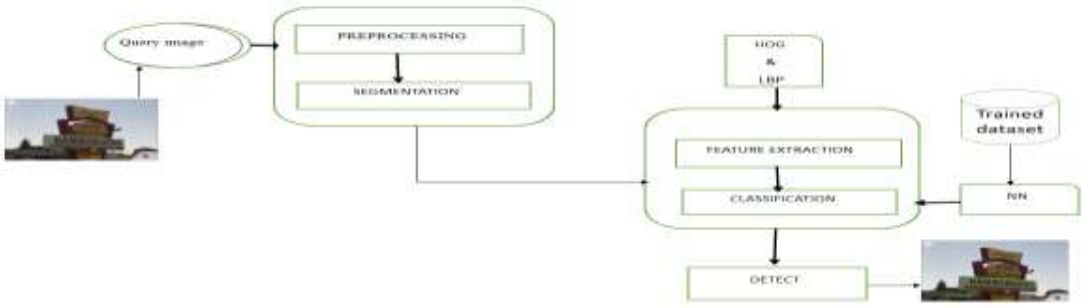


**Fig.2 Schematic overview of the proposed framework**

In this architectural diagram it is explained about the actual process that is done in the entire work. Initially a query image is given to the preprocessing phase where various enhancement techniques are applied and the result is passed to the segmentation, feature enhancement and classification phases. Here with some techniques of deep learning and image processing we recognize and match the images from the data set.

### A. ENHANCEMENT

Image enhancement techniques have been widely used in many applications of image processing where the subjective quality of images is important for human interpretation. OCR reads the image and it identify the characters in the image. Contrast is an important factor in any subjective evaluation of image quality. Contrast is created by the difference in luminance reflected from two adjacent surfaces. Our visual system is more sensitive to contrast than absolute luminance; therefore, we can perceive the world similarly regardless of the considerable changes in illumination conditions. Noise removal and Filtering algorithms for accomplishing contrast enhancement have been developed and applied to problems in image processing.

### B. SEGMENTATION

Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image. Morphological operations is applied to greyscale images such that their light transfer functions are unknown. Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element is positioned at all possible locations in the image and it is compared with the corresponding neighborhood of pixels.

### C. FEATURE EXTRACTION

#### 1. HOG

The histogram of oriented gradients (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is

similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy.

**2. LPB**

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings.

## D. CLASSIFICATION

An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. In machine learning, a convolutional neural network (CNN or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analyzing visual imagery.

## IV. RESULT

In this project, we have presented a convolutional neural network-based text detection system that learns to automatically extract its own feature set instead of using a hand-crafted one. By leveraging large, multi-layer CNNs, we train powerful and robust text detection and recognition modules. The main purpose for building such a system is to fulfill the needs of growing content-based multimedia indexing, retrieval and management. The result achieved is analyzed and compared with certain other techniques like Ostu, AdaBoost, SVM, OCR in terms of accuracy, error rate and precision recall value and tabulated as below.

**Table. I Result of various classifier**

| METHOD | ACCURACY | ERROR RATE | PRECISION CALL |
|---|---|---|---|
| Ostu | 64.40% | 39 | 30 |
| AdaBoost | 75.04% | 41 | 34 |
| SVM | 78.80% | 44 | 31 |
| OCR | 82.35% | 36 | 35 |

## V. CONCLUSION

In this paper the text is retrieved from the input image and it is matched with other images in the trained dataset. Similar images are matched based on the feature extracted and displayed in the output. Accuracy of 82.3% is achieved during the image retrieval which is higher than the existing techniques used. This is done efficiently with several image processing and deep learning techniques. Matching of images in database can be used in many applications like book management, object matching using text, etc. This is able to process both hand written text and digitalized text. Background color of the picture effects the retrieval accuracy that can be taken into account for future enhancement. In addition only certain font style are accepted by this method this also can be concentrated for future enhancement.

## REFERENCES

[1]. Jiji Mol1, Anisha Mohammed2 and Mahesh B S2 "Text Recognition Using Poisson Filtering and Edge Enhanced Maximally Stable Extremal Regions" 2017 International Conference on Intelligent Computing,Instrumentation and Control Technologies (ICICICT).

[2]. Peng Tian-qiang, Li Fang "IMAGE RETRIEVAL BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS AND BINARY HASHING LEARNING" 2017 IEEE.

[3] Pan He,∗1, 2 Weilin Huang,∗1, 2 Yu Qiao,1 Chen Change Loy,2, 1 and Xiaoou Tang2, 1 "Reading Scene Text in Deep Convolutional Sequences" Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16).

[4] Mr. Pratik Madhukar, Manwatkar Mr. Shashank H. Yadav "Text Recognition from Images" IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS)2015.

[5] Cong Yao, Xiang Bai, Wenyu Liu, "A Unified Framework for Multioriented Text Detection and Recognition" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 23, NO. 11, NOVEMBER 2014.

[6] Jerod J. Weinman "Typographical Features for Scene Text Recognition" 2010 International Conference on Pattern Recognition

[7] Mr. Pratik Madhukar, Manwatkar Mr. Shashank H. Yadav "Text Recognition from Images" IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded and Communication systems (ICIIECS)2015.

[8] Amruta Pise, S.D.Ruikar "Text Detection and Recognition in Natural Scene Images" International Conference on Communication and Signal Processing, Apri[ 3-5, 20 14, India.

[9] Rhead, Mke, "Accuracy of automatic number plate recognition (ANPR) and real world UK number plate problems." IEEE International Carnahan Conference on Security Technology (ICCST), 2012.

[10] Naveen Sankaran and C.V Jawahar, "Recognition of Printed Devanagari Text Using BLSTM Neural Network," IEEE, 2012.

[11] Wang, K.; Babenko, B.; and Belongie, S. 2011. End-to-end scene text recognition. IEEE International Conference on Computer Vision (ICCV).
[12] Wang, T.; Wu, D.; Coates, A.; and Ng, A. Y. 2012. End-to-end text recognition with convolutional neural networks. IEEE International Conference on Pattern Recognition (ICPR).

[13] Keechul Jung, Kwang In Kim and Anil K. Jain, "Text information extraction in images and video: a survey," *The journal of the Pattern Recognition society*, 2004

[14] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, "TextFinder: An Automatic System to Detect and Recognize Text in Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 11, November 1999.