International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

# Text Detection in Natural Scene Images: A Survey

**Manish Narayan B S[1], Chintan S A[2], Kaushik S[3], Krupashankari S S[4]**

*[1,2,3]Student, Dept. of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.*
*[4]Assistant Professor, Dept. of Information Science and Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.*

---

**Abstract –** *As extracting text from different places using machine learning is being developed, in this survey, we aim to present the different methods employed in creating a text detection model for natural scene images and the possible implementations by discussing about a method called Progressive Scale Expansion Network (PSENet) , and how this is being developed and used to overcome the challenges faced.*

***Key Words*: Text detection and recognition, Image Processing, Natural Scene Images, Convolutional Neural Networks.**

## 1. INTRODUCTION

The growing necessity to find methods to expand the use of technology in the visual medium has paved way for various advancements in the field of computer vision and machine learning. Optical character recognition and scene text detection are the fields which have seen a rapid increase in development. Our focus lies mainly on text detection in scene images. Certain images have text which is present in the view captured, this is known as scene text.

The conventional methods for developing a text detection model majorly use the bounding box technique to detect text which adds certain challenges like in-accuracy for locating texts present in arbitrary shapes and also texts present in close vicinities [3].

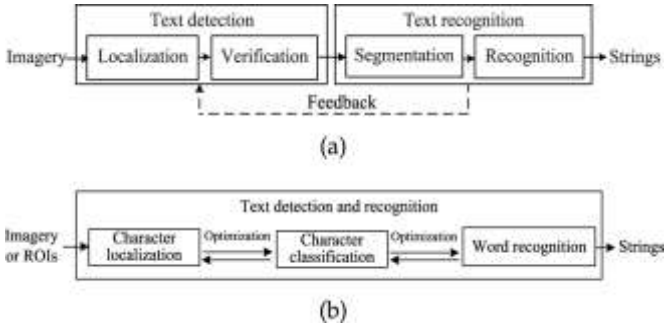


**Fig-1**: Frameworks of two commonly used text detection and recognition methodologies a) Stepwise Methodology, b) Integrated methodology.

Here we took upon the task of exploring ways to find a better and efficient text detection models which use various machine learning based approaches to detect text that exists in different orientations in natural scene images. This paper takes into account of the different techniques and methods that were presented for detection of text and describes the contributions made to develop an efficient model.

## 2. LITERATURE SURVEY

Fagui Lui, Cheng Chen et al. [1] proposed a framework that is a combination of Feature Pyramid Network (FPN) and Bidirectional Long Short-Term Memory (Bi - LSTM) Networks. Then a text connector is used to connect the detected text into lines. The results were based on several public datasets like ICDAR2013, ICDAR 2015. The target of the paper is to have a multi-scale and multi oriented detection in natural scene images.

Zhida Huang et al. [2] used a Mask - RCNN based text detection approach which requires the challenging task of Instance segmentation. They propose to use Mask -RCNN incorporated with a Pyramid Attention Network (PAN) to strengthen the feature representation ability instead of using the usual Feature Pyramid Network (FPN). They have used a Region Proposal Network (RPN) that generates rectangular text proposals from which corresponding quadrilateral bounding boxes can be obtained as outputs.

Traditional bounding box give rectangular bounding boxes which is inaccurate for curved and multi oriented texts in natural scenes. So, Wenhai Wang et al. [3] proposed a Progressive Scale Expansion Network (PSENet) which can precisely detect text in different shapes and orientations. Firstly, it is a Segmentation based method where PSENet performs pixel level-based segmentation which locates the text instance precisely even if it is an arbitrary shape. Then, a progressive scale expansion algorithm is used which can successfully identify different adjacent text instances.

Densely Convolutional Networks (DenseNet) were proposed in 2016 and since have been very successfully used in object detection and recognition. Here, Mitra Behzadi and Reza [4] proposed a Fully Convolutional DenseNet approach to text detection. They perform semantic segmentation with 3 classes on images which allows the model to learn to separate close words. They use minor post processing on the output in the testing phase to get better results. Their method was tested on the ICDAR 2013 dataset.

Asghar Ali and Mark Pickering [5] propose a network that can accurately detect Arabic and Urdu text in natural scene images. The network proposed is a Fast RCNN network which is based on a pretrained VGG16 convolutional network on the ImageNet dataset. They used the pre-trained VGG16 model for the initial layers and the later convolutional layers are trained on the multi lingual image text dataset.

Lionel Prevost et al. [6] proposed a detection technique which is based on a cascade of boosted ensemble and a localizer using standard image processing techniques. In this approach various overlapping text segments are extracted from images containing text lines. They used a set of 39 features that are capable of detecting various type of text in grey level natural scene images. Then, the coordinates of the rectangles around the detected text are obtained through a localizer. This scheme is tested on the ICDAR 2003 robust reading and text locating database.

Shangxuan Tian et al. looked to address some of the issues that were present in the prevalent scene text detection approach [7]. They proposed Text Flow a unified scene text detection system which has the usual first step - Character Candidate detection but, it combines the next three sequential steps into a single process. A fast cascade boosting technique is used for character candidate detection. Then a min-cost flow network handles the second unified step that is to take the character candidates as inputs and output the text lines. This model outperforms the current techniques on the ICDAR 2011 and 2013 dataset.

Jinsu Kim et al. looked towards deep networks [8]. Deep networks generally perform better for classification problems than localization problems. They proposed a method that aims to localize and recognize text with four steps which use Maximally Stable Extremal regions (MSERs) for path extraction an ensemble of ResNets for patch classification. Then text regions are identified by filtering out non character patches. Since localization problems are formulated to classification problems and Residual Networks are used the error rate of the proposed model is reduced.

**TABLE-1**: Comparison of results from different papers.

| Author | Network Architecture | Dataset | Accuracy |
|---|---|---|---|
| Cheng Chen et al. 2019 | CNN + FPN along with RNN(Bi-LSTM) | ICDAR 2015 | 72.8 % |
| Lei Sun et al. 2019 | Mask - RCNN with PAN (Pyramid Attention Network) | ICDAR 2017 MLT | 73.3 % |
| Mitra Behzadi et al. 2018 | Fully Convolutional DenseNets. | ICDAR 2013 | 70 % |
| Mark Pickering et al. 2019 | Fast - RCNN with RNN | ICDAR 2017 | 46.15 % on Arabic 33.27 % on Urdu. |
| Lionel Prevost 2008 | Image Processing | ICDAR 2003 | 50.7 % |
| Shangxuan Tian et al. 2015 | Cascade Boosting with min cost flow network | ICDAR 2013 | 80.25 % |
| Jinsu Kim et al. 2017 | An Ensemble of ResNets | ICDAR 2013 | 85.7 % |
| Wenhai Wang et al. 2019 | Progressive Scale Expansion network | ICDAR 2015 | 74.3 % |

## 3. CONCLUSION

In this paper we discussed the various methods, techniques and the network architectures used for implementation of different text detection and recognition model for images present in natural scenes. The results from various papers were analyzed, compared and tabulated.

## REFERENCES

[1]FTPN: Scene Text Detection with Feature Pyramid Based Text Proposal Network: FAGUI LIU, CHENG CHEN, DIAN GU, AND JINGZHONG ZHENG: IEEE.

[2] Mask R-CNN with Pyramid Attention Network for Scene Text Detection: Zhida Huang, Zhuoyao Zhong, Lei Sun, Qiang Huo 978-1-7281-1975-5/19/ ©2019 IEEE.

[3] Wenhai Wang, Enze Xie, Xiang Li, Wenbo Hou, Tong Lu, Gang Yu, Shuai Shao. Shape Robust Text Detection with Progressive Scale Expansion Network. arXiv:1903.12473v2 [cs.CV] 29 Jul 2019.

[4] Text Detection in Natural Scenes using Fully Convolutional DenseNets: Mitra Behzadi, Reza Safabakhsh, 978-1-7281-1194-0/18/ ©2018 IEEE.

[5] Convolutional Feature Fusion for Multi-Language Text Detection in Natural Scene Images: Asghar Ali Chandio, Mark Pickering 978-1-5386-9509-8/19/ ©2019 IEEE.

[6] A Cascade Detector for Text Detection in Natural Scene Images: Shehzad Muhammad Hanif, Lionel Prevost, Pablo Augusto Negri 978-1-4244-2175-6/08/ ©2008 IEEE.

[7] Text Flow: A Unified Text Detection System in Natural Scene Images: Shangxuan Tian, Yifeng Pan, Chang Huang, Shijian Lu, Kai Yu, and Chew Lim Tan 1550-5499/15 © 2015 IEEE.

[8] Jinsu Kim, Yoonhyung Kim, and Changick Kim. 2017. A Robust, Ensemble of ResNets for Character Level End-to-end Text Detection in Natural Scene Images. In Proceedings of CBMI, Florence, Italy, June 19-21, 2017.