**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p.2802-2811
https://publishoa.com
ISSN: 1309-3452

# Contemporary Methods on Text Detection and Localization from Natural Scene Images and Applications

**[1]Awatef Balobaid,**
Assistant Professor, Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia.

**[2]Dr. S Murali Mohan,**
Professor, Department of ECE, Mother Theresa Institute of Engineering and Technology, Palamaner, India.

**[3]V. Mahalakshmi,**
Assistant Professor, Department of Computer Science, College of Computer Science and Information Technology, Jazan University, Jazan, Saudi Arabia.

**[4,*] Sateesh Kumar Reddy C,**
Research Associate, SENSE, VIT Chennai Campus, Chennai, India.

**[5]P. Anjaiah,**
Assistant Professor, Department of C.S.E, Institute of Aeronautical Engineering, Dundigal, India.

**[6,*]M. Sudhakara,**
Assistant Professor, School of Computing and Information Technology, REVA University, Bangalore.

[1]asbalobaid@jazanu.edu.sa
[2]muralimohanieee@gmail.com
[3]mahabecs@gmail.com
[4]satishreddic@gmail.com
[5]p.anjaiah@iare.ac.in
[6]malla.sudhakara@reva.edu.in
**Received** 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

**Abstract**

A significant research problem being investigated for a very long time is Text detection and localization from natural scene images. It is a necessary technique for providing solutions to various applications such as intelligent transport systems, content-based retrieval systems, blind assistants and tour guide systems etc. Text found in scene images contains beneficial information for the needs and safety of people. Hence, it is essential to detect text and extract it for further application processing by computer vision systems. Text detection from scene images poses several challenges: varying illumination, low resolution, font size and style variations, complex backgrounds, presence of noise and other degradations. Various attempts are made to localize text from outdoor natural scene images, and this paper provides an exhaustive literature review of all existing methods and provides comparatives analysis. The way forward for future researchers is also reported.

**Keywords**: Text detection, Text localization, Machine Learning, Feature Extraction.

## 1. Introduction

Nowadays, everyone needs to give a unique name with a remarkable text style to their shops, foundations, hostels, and restaurants to draw in individuals with their unique names. The display boards and signboards in natural scenes also have

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p.2802-2811
https://publishoa.com
ISSN: 1309-3452
varying font styles and sizes. It is genuinely challenging to see such text information influenced by low lights, high lights, shadow, climate conditions, the text language, and direction. Extracting such text data will benefit various computer vision systems such as ordering colossal picture and video data sets by their printed content [1], content-based image retrieval, intelligent transportation systems, portable vision systems, industrial automation and other areas [2-6].

"Text detection" was initially presented during the 1990s. With the quick progression of Internet innovations and versatile cell phones, an ever-increasing number of situations requiring message extraction from picture information is being created. Text detection is now a vital part of vision and recognition algorithms as a focal point for research work in scene text understanding. Some of the most prestigious international conferences in the relevant area are ICDAR and ICCV. Some applications in different sectors are Mobile transliteration/translation technologies, Scene understanding, and Address identification. The street name or address is generally written to familiarize people with the location. Everyone needs to be able to identify the location. The street name is usually written on a plain background with distinct font size. However, the camera's noise makes it difficult to distinguish the text when it is captured. Figure 1 shows some examples of such images.


Figure 1: Example of representing the address

The solution to text detection and localization problems comprises various processing steps. Figure 2 describes the steps involved to achieve the task.

The methods invented for text detection will pre-process the image for removal of noise, degradations, suppression of background and edge detection. Further, feature engineering is done on the pre-processed image to extract suitable characteristics for categorizing discriminating text from non-text regions. Later, the features are analyzed by classifiers to extract text regions.

In this research work, the technical advancements were examined using scene text detection in-depth to motivate and advise young researchers to practice. We also first review the history and development of text recognition before classifying the conventional approaches and highlighting the significant challenges and strategies that go along with them. Then, by comparing state-of-the-art algorithms, we present frequently used benchmark datasets and assessment methods. Lastly, we review our findings and make predictions about future research areas. Parameters, i.e., Accuracy, Recall, Precision and Error rate, are also briefed to evaluate the comparison's performance.

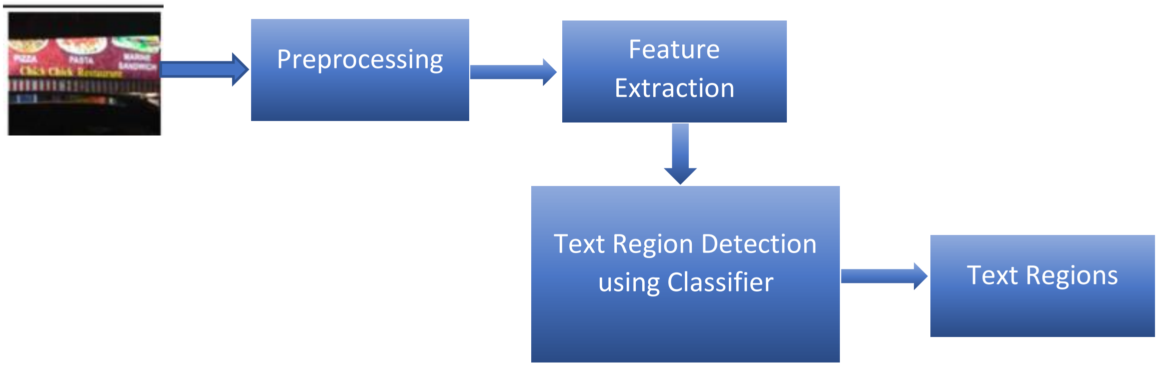



Figure 2: Block Diagram of Text Detection from Image

## 2. Technologies for Text Detection and Localization

Text detection and localization are based on some methods like (i) Connected Component Analysis, (ii) Edge Analysis and (iii) Image Blocks Analysis. Xiaoqing Liu et al. [7] presented Text detection using an edge analysis approach comprising various steps: character extraction, candidate text detection discovery and text extraction. The immensity of the second derivative of intensity is employed as a value of edge strength in the first stage. It allows peaks for better recognition that generally describes text. The average edge strength inside a window is used to compute the edge density. Four orientations ($0^0$, $45^0$, $90^0$, $135^0$) are used to find the variance of orientations in terms of effectiveness and efficiency, where $0^0$ means horizontal direction, $90^0$ denotes vertical direction, $45^0$ and $135^0$ are the diagonal directions. Implementation of the Multi-scale method is used for Edge detection in which Gaussian pyramids are used to generate multi-scale images after applying a smoothing filter and down sampling the raw image in vertical and horizontal directions.

Segmentation [8] from low resolution and texture-based text detection in images is introduced. In D.C.T. (Discrete Cosine Transform) domain, a high pass filtering is used to suppress the background. Textural features were computed on image blocks to identify, detect and segment the text regions in the picture. Based on newly developed discriminant functions, each case is categorized as either text or non-text. A merging algorithm is identified for combining various text blocks to generate text regions. Later, those areas are fine-tuned and utilized for post-production. Texture features, Block categorization, D.C.T. domain, and Text block merging and refining text regions are the five-step approach were determined for removing constant background in the future matrix D generation on every 50x50 frames.

A Comparative analysis [9] of various approaches based on texture and morphology multi-scale analysis. MATLAB 7.0 is utilized to complete the examination. The analysis shows that the combined Morphological-Region method outperforms the Multi-scale and Texture techniques in terms of noise tolerance and the ability to recognize all types of text regions even against various backdrops.

A novel Connected Component (CC) analysis technique [10] is applied for text extraction, text contours and stroke inside regions individually and combining from images. This approach efficiently creates character candidates. Compared to [10] MSERs-based and SWT-based techniques, test outcomes convey that the suggested approach creates a few candidates yet recalls the majority of text C.C.s. The C.C.s is subjected to filtering by CC, line categorization and line grouping to get the final result.

A technique for detecting text areas [11] was invented, and the region suggestions were quickly created by the classification based on connected components; then, the subsequent classification of those regions effectively removed all false detections using in-depth features. Here, a dataset for manga text areas was established for noise-free learning and appropriate assessment of the detection algorithms. The techniques achieve a recall value of 0.851 and an F-measures value of 0.466 in the manga dataset, which is higher than the primary methods shown in Figure 3.

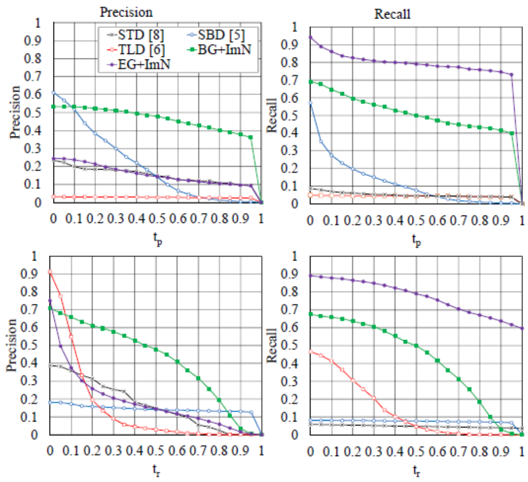



Figure 3: Precision and recall achieve higher performance

A method [12] is proposed for reducing Laplacian noise in the input image, and Edge-Enhanced Maximally Stable Extremal (EEMSE) regions are generated using the pre-processed image (MSERs). Non-text portions are filtered using region filtering, and an Optical Character Recognition (O.C.R.) technique is used to identify them. This approach's output exceeds Peak Signal to Noise Ratio (PSNR) and Structural Similarity Measures (SSIM). Also, this approach was tested on the standard ICDAR dataset, which consists primarily of real-time photographs and achieves 90.30% of better character confidence and 96.42% of high character accuracy and less error rate of 3.58%.

In [13], A novel technique is explained for text identification depending on Deep Learning Region Classification (DLRC) and Super pixel-based Stroke Feature Transform (SSFT). The ICDAR2011, ICDAR2013, and street view text datasets are used in this technique. The SSFT was intended for up-and-comer character area (C.C.R.) extraction, which involves separating an information picture into numerous areas utilizing super pixel-based clustering, deleting the majority of the regions that do not meet the characters' established criteria, and then the remaining regions can be post-processing for finding out the C.C.R.s using a stroke width map. Then, based on high-level features, colour, geometrical, and deep convolution neural networks were used to identify character areas. Finally, the suggested approach yields F-Measure values of 0.876, 0.885, and 0.631, respectively.

Weighted median filters and techniques called Maximal Stable Extremal Region are used for identifying and extracting text areas in natural pictures [14]. Towards hold and smooth the edges, the picture is passed using a weighted middle channel, trailed by MSER up-and-comer locale extraction. Non-text components are filtered using heuristics criteria. Finally, classifiers are used to categorize potential text areas and non-text regions based on three text-specific characteristics, followed by clustering to group text components in text lines [15]. This technique is designed to extract text areas from low-contrast pictures reliably. The method's efficiency in terms of accuracy, recall, and F-measure is demonstrated using the ICDAR 2011 testing dataset.

A texture-based method in two phases; Text Detection and Text Localization. [16] This technique comprises two steps, namely Pre-processing to increase the quality of the natural image by removing the unwanted noise. In the localization stage, texture analysis is used based on texture features by considering the parameters like variance and contrast. Then the blocks are divided into text and non-text blocks, where text blocks are identified as spatially blended, and text areas are eventually contained within a bounding box. Variance is determined using first-order statistics, where r and c are several columns and columns, A is the input of the image matrix, and contrast is measured from p (i, j), which is the grey level co-occurrence matrix.

Author [17] understood the challenges of text localization and proposed a hybrid methodology that extracts text from Natural images. This method consists of four stages; firstly, superimposed text regions are extracted from natural images based on features such as Horizontal crossings, Bounding Box, Perimeter, Area, and Euler number. Next, extracted text regions are tested on the character descriptor and the Support Vector Machine classifier. In the Third step, recognition of line segmentation and multiple lines in localized text regions is executed using horizontal profiles. In the last step, each portioned line is extracted using vertical profiles. Font style, font size, scale, illumination, and Orientation may vary throughout the text. The suggested method uses a canny edge operator with a median filter to retain the sharp edges of text in the picture, resulting in more accurate text region localization. Extracting characters from the individually segmented line is successful when the localized text areas, including several lines of text in various scripts, are segmented.

A methodology that uses a median filter by considering the canny edge operator conserves a sharp edge of the text in the image and gives more accuracy to text regions [18].An algorithm for new text regions extraction from an image with complex background. This proposed method is highly dependent on the Recurrent Convolutional Neural Network (R-CNN) for the faster mechanism by taking the features like the area to identify the text from the image [19].

Text detection on sign boards using ANNs from the natural scene images is divided into different levels for text recognition [20]. First, the image is acquired from the outdoor scene, and then the edges of the signboard are detected whereas, in the second stage is for detecting and recognizing the text in two different languages: English and Urdu. The last step uses ANNs to classify and recognize text extracted from natural scenes or an outside environment. As part of this research, a developed image database was used to generate experimental results. The work is multilingual and gives output in English as well as Urdu. In terms of image recognition, the system compares well to other systems, providing an overall 85% level of accuracy.

A system for scene text detection for Accuracy Analysis and Complexity of CNNs is established, and the investigations made on the detection accuracy and indicators of these three strategies R-CNN's network architecture is reliable compared to RRPN and CTPN [21]. The F-measure of RRPN concerning R-CNN and CTPN is 0.18 and 0.15, respectively. In the process of detecting scenes texts, however, each of these methods puts a different emphasis on text detection. The CTPN is superior for detecting laterally distributed scene texts; however, the RRPN is superior for detecting tilt-angle scene texts.

A method for a Unified Video Text Detection (UVTD) with Network Flow, In this research, scene detection using an FCNN is presented. The method analyses frames and detects text information from the images. The experimental results were discussed regarding the accuracy of the UVTD [22].

Kannada text detection using Combined Edge and Connected Component in natural scene images is developed [23]. Three aspects make the system sturdy, and the initial approach was to detect edges morphologically. An edge image is created by determining a threshold level. Second, a text feature filtering technique is applied that extracts as many non-letter components as possible. Furthermore, this approach produces a wide range of letter components in the mean height of the residual bounding boxes. The text region is then binarized using a threshold value from the residual bounding boxes. Own collection of 200 Kannada images was used to test the proposed method, and the results are satisfactory.

Text Detection with Text Line Construction aims to develop a new text detection algorithm capable of accurately localizing text in natural images. Mainly R-CNN is used as a faster mechanism for converting small text areas into object detection for small text areas. This method allows a neural network to detect text data and categorize the text/non-text information within the image evaluated on the ICDAR dataset [24].

This paper presents an attention-guided multi-scale regression method for detecting scene text, which shows promising results [25]. Two different scales are used to create the pipeline, enhancing its perception of large and long texts. Additionally, the addition of the attention module improved the prediction performance by more accurately separating text from the background. Compared to ICDAR 2015 and MSRA-TD500, the proposed text detector has achieved comparable results in terms of accuracy and precision.

Recently, academics and industry have become interested in detecting traffic panels. The reason for this is that traffic panels provide several difficulties. To acquire accurate text detection in traffic panels with text data, there is a need to detect the traffic panel. A new dataset is created that includes Persian text on traffic panels in the streets because there are no available text-based traffic panel datasets. In the dataset, there are two sets of figures, i.e., the first batch has 9294 images, while the second set contains 3305 images [26]. The homogeneity of the first dataset is lower than that of the second. As a result, the first dataset is a supplement to the second dataset. Therefore, a prior trained model using the different datasets is trained using the primary dataset. To train and evaluate the model, the tiny YOLOv3 algorithm is

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p.2802-2811
https://publishoa.com
ISSN: 1309-3452
used with low complexity and quick execution. Finally, the K-fold cross-validation technique is used to access the algorithm's efficiency and shows that precision is 0.973, recall is 0.945, and F-$_{measure}$ is 0.955.

A challenge is that Context-aware applications have drawn more attention to automatic text reading in scenes in recent years. The author developed a method for automatic text extraction from scenes, which takes advantage of the object proposal in generic object identification [27]. The max-pooling-based method is designed to find scene text ideas within a feature map obtained from picture edges. The results of the search are then sorted using a scoring algorithm depending on the histogram of the oriented gradient. The dataset used for this research is taken from publicly available scene datasets: the ICDAR2015 t and the Street View Text dataset. And the proposed technique exceeds the state-of-the-art performance, mainly when a small number of proposals are examined, and when combined with a scene text recognition model, it also achieves better efficiency in scene text recognition.

This research proposes a strategy for increasing text localization in Scene Images using a New Defect Detection method using post-processing approaches. The suggested method extracts properties such as phase congruency, entropy, and compactness for text detection to improve feature extraction's discriminatory power [28]. The characteristics of the text component are determined using a Support vector machine classifier and the Gaussian distribution of text components. The weights are multiplied by the characteristics using clustering to detect defect components. The bounding boxes have been redrawn, resulting in a precisely aligned bounding box and producing good results. The trials on MSRA-TD-500 and SVT benchmark datasets for detection and recognition before and after defect identification to evaluate the effectiveness of defect detection will improve performance dramatically after fault detection. Table 1 summarizes the recent work in text detection and localization.

Table 1: Recent Research works on Text Detection and Localization.

| References | Dataset | Approach | Metrics |
|---|---|---|---|
| [1] | ICDAR 2013<br>ICDAR 2015<br>Street View Text | Clustering for external regions and OCR Classifier | Recall -71.3% precision- 82.1%<br>F-measure- 76.3% |
| [3] | Traffic signs | MSERs and HSV thresholding | F-measure-0.93<br>0.89 -recognition, 0.87 -entire system |
| [7] | Four types of 75 test images, including book covers<br>480 × 640 resolution | Multi-scale Edge-Based Text Extraction Algorithm | NA |
| [8] | 100 varieties of low-resolution natural scene<br>images each of size 240x320 | D.C.T. based high pass filter and texture feature | 96.6% Detection rate |
| [10] | Born-digital dataset of ICDAR2013 | CC based filtering | Recall -90.51% precision- 48.33%<br>F-measure- 63.44% |
| [11] | Manga text regions | Connected Components | Recall - 0.851<br>F-measure -0.466, |
| [13] | ICDAR 2013<br>ICDAR 2015<br>Street View Text | SSFT and DLRC<br>Algorithm | F-measures o-0.876, 0.885, and 0.631 |
| [14] | ICDAR 2011 | AdabooostM1<br>and K-NN Classifier | Recall-0.826<br>Precision-0.62<br>F-measure-0.71 |
| [16] | ICDAR 2011 | GLCM features like variance and contrast | Accuracy 90 |
| [29]<br>(2022) | 14100 Urdu Cropped image | CNN, RNN, CTC | C.R.R. of 95.75%, a W.R.R. of 87.13% and a WRR1F of 94.21%, |

**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p.2802-2811
https://publishoa.com
ISSN: 1309-3452

| | | | |
|---|---|---|---|
| [30] (2022) | 12,120 Images | Text stroke mask using TSDNet | NA |
| [31] (2022) | Self-acquired | Key point Detection<br>Key point Link Prediction | Recall-74.7 Precision 77.3 and F-Score76.0 |
| [32] (2022) | Self-acquired | CRAFT, Four Stage Network | NA |
| [34] (2022) | MSRA-TD500, CTW1500, Total-Text, ICDAR2015 | CM-Net | F-Measure-82.8 |
| [35] (2021) | Self-acquired | Canny edge detection algorithm _ k-means | Accuracy-90.5% |
| [36] (2022) | ICDAR 2013, ICDAR 2015, MSRA-TD500,<br>and SCUT-CTW1500 | progressive region prediction network | F-measure of 86.0% on ICDAR 2015 |

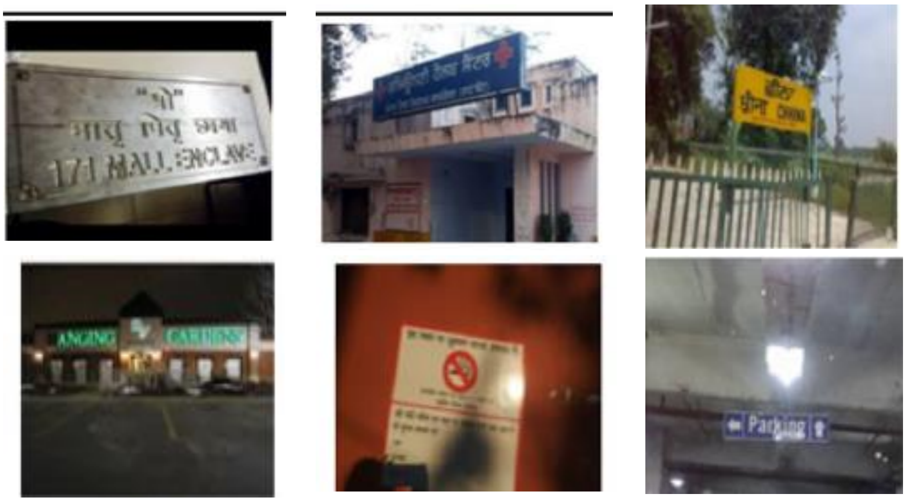
## 3. Gaps Involved in Text Analysis

Some of the challenges found in image analysis for text detection and localization are as follows.

- Environment: - The images captured are subjected to distortions due to environmental conditions. Some distortions include Uneven Lightening, Image Complexity, perspective distortion and noise.
- Image Acquisition: The acquisition of images may also introduce distortion and other degradation due to lens and device configuration issues.
- Text Content: The challenges related to text content are multi-oriented/curved text, different font styles, different sizes, unique colour, unique font alignment, multilingual environments and variation of aspect ratio and fonts are some of the challenges that affect the performance of detection and localization of texts.
- Multi Orientation It is to attract the viewers. Text may be vertical, horizontal, curved or multi-oriented. Methods used to find horizontal text cannot be used detection of text in Orientation. Henceforth multi, Orientation is a challenging task for researchers.
- Complexity in backgrounds- Background includes rocks, grasses, bricks, and signboards, which leads to more complexity for identifying text.

**Image Capturing Angle**

**Noise Image**



**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p.2802-2811
https://publishoa.com
ISSN: 1309-3452


Figure 4: Challenges in Text detection and localization

## 4. Applications and Future Avenues

- Text recognition is used in various fields to automate tasks, including automatic license plate reading at toll booths, automatic signature reading on check leaves, and image tagging and analysis of image data.
- In the health care industry, text recognition allows storage and retrieval of enormous documentation. It can also create a searchable database where the text can be found easily. The created databases can be easily updated and accessed, which saves time and a lot of paperwork. Automated text detection with voice assistance is also helpful to the visually impaired. Text recognition in videos can also be achieved using automated text recognition.
- Automatic text identification can also be used to make transportation systems more intelligent. It can also be used in airports to check passports and extract information. Text detection is also used for automated data entry for business documents. Text detection also allows for the publication of an array of books online that can be shared and preserved. Automation in the business world is also made possible by text recognition, which supports automated reading of labels and numbers.
- Further advancements in technology have led to text recognition finding applications in many fields.

### Future Research Avenues

Some of the future research problems than can be taken up by the researchers are below.

- Text extraction from drone camera/traffic camera-based images.
- Text extraction from items kept in big malls.
- Text extraction from display boards in an unknown environment

## 5. Conclusion

Several methods for text detection have been developed previously. Text detection was a natural extension of document analysis, from scanned page images to camera recorded photos, pre-processing, detection, and O.C.R. technologies. Although scene text was thought to be a more challenging problem, relatively little work had been done with it. Recently, academics have looked into methods for capturing text in various combinations that have proven useful. Some options include unsupervised feature learning, Convolutional neural networks, and other techniques. This paper presents a comprehensive survey of various methods researched so far. Machine learning and deep learning methods are more accurate and provide better results. The challenges involved in the problem are also discussed. Several approaches have been introduced to overcome the existing challenges: AdaBoost, CNN, TSD-Net, and CM-Net. The researchers can continue exploring the solutions' usage for various relevant applications.

**References**

[1] Lukas Neumann, Jirı Matas, "Real-Time Lexicon-free Scene Text Localization and Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38 pp. 1872 – 1885, 2015.

[2] Yin, X.-C.; Zuo, Z.-Y.; Tian, S.; Liu, C.-L. "Text Detection, Tracking and Recognition in Video: A Comprehensive Survey". IEEE Trans. Image Process. 2016, 25, 2752–2773.

[3] Greenhalgh, J.; Mirmehdi, M. "Recognizing Text-Based Traffic Signs". IEEE Trans. Intell. Transp. 2015, 16,1360–1369.

[4] Shilkrot, R.; Huber, J.; Liu, C.; Maes, P.; Nanayakkara, S.C. Finger Reader: A wearable device to support text reading on the go. In CHI'14 Extended Abstracts on Human Factors in Computing Systems; A.C.M.: New York, NY, U.S.A., 2014; pp. 2359–2364.

[5] Hedgpeth, T.; Black, J.A., Jr.; Panchanathan, S. A demonstration of the iCARE portable reader. In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY, U.S.A., 23 October 2006; Volume 279.

**[6]** Ham, Y.K.; Kang, M.S.; Chung, H.K.; Park, R.-H.; Park, G.T. Recognition of raised characters for automatic classification of rubber tires. Opt. Eng. 1995, 34, 102.

[7] Xiaoqing Liu, Jagath Samarabandu, "Multi-scale Edge-Based Text Extraction from Complex Images",
1-4244-03677/06 ©2006 *IEEE*

[8] S. A. Angadi and M. M. Kodabagi, "Text region extraction from low resolution natural scene images using texture features," 2010 IEEE 2nd International Advance Computing Conference (IACC), 2010, pp. 121-128, doi: 10.1109/IADCC.2010.5423026.

[9] C. P. Sumathi and S. N. Priya, "Text extraction from assorted images using morphological-region, texture and multi-scale techniques - A comparative study," *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 2013, pp. 294-299, doi: 10.1109/ICICES.2013.6508383.

[10] K. Chen, F. Yin and C. Liu, "Effective Candidate Component Extraction for Text Localization in Born-Digital Images by Combining Text Contours and Stroke Interior Regions," *2016 12th IAPR Workshop on Document Analysis Systems (D.A.S.)*, 2016, pp. 352-357, doi: 10.1109/D.A.S.2016.30

[11] Y. Aramaki, Y. Matsui, T. Yamasaki and K. Aizawa, "Text detection in manga by combining connected-component-based and region-based classifications," *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 2901-2905, doi: 10.1109/ICIP.2016.7532890.

[12] J. Mol, A. Mohammed and B. S. Mahesh, "Text recognition using poisson filtering and edge enhanced maximally stable extremal regions," *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2017, pp. 302-306, doi: 10.1109/ICICICT1.2017.8342578.

[13] Y. Tang and X. Wu, "Scene Text Detection Using Superpixel-Based Stroke Feature Transform and Deep Learning Based Region Classification," in *IEEE Transactions on Multimedia*, vol. 20, no. 9, pp. 2276-2288, Sept. 2018, doi: 10.1109/TMM.2018.2802644.

[14] R. Soni, B. Kumar and S. Chand, "Extracting Text Regions from Scene Images using Weighted Median Filter and MSER," *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2018, pp. 915-920, doi: 10.1109/ICACCCN.2018.8748492.

[15] Neumann, L.; Matas, J. A" Method for Text Localization and Recognition in Real-World Images". In Proceedings of the Asian Conference on Computer Vision, Queenstown, New Zealand, 8–12 November 2011; Springer:Berlin/Heidelberg, Germany, 2011; Volume 770–783.

[16] T. Kumuda and L. Basavaraj, "Detection and localization of text from natural scene images using texture features," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2015, pp. 1-4, doi: 10.1109/ICCIC.2015.7435688.

[17] K. S. Satwashil and V. R. Pawar, "Integrated natural scene text localization and recognition," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 371-374, doi: 10.1109/ICECA.2017.8203708.

[18] K. S. Satwashil and V. R. Pawar, "English text localization and recognition from natural scene image," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 555-559, doi: 10.1109/ICCONS.2017.8250523.

[19] Z. Liu, Q. Shen and C. Wang, "Text Detection in Natural Scene Images with Text Line Construction," 2018 IEEE International Conference on Information Communication and Signal Processing (ICICSP), 2018, pp. 59-63, doi: 10.1109/ICICSP.2018.8549799.

[20] M. A. Panhwar, K. A. Memon, A. Abro, D. Zhongliang, S. A. Khuhro and S. Memon, "Signboard Detection and Text Recognition Using Artificial Neural Networks," *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Beijing, China, 2019, pp. 16-19.

[21] H. Li and D. Chen, "Complexity and Accuracy Analysis of Convolutional Neural Networks in Scene Text Detection," *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Xiamen, China, 2019, pp. 1416-1419.

[22] X. Yang, W. He, F. Yin and C. Liu, "A Unified Video Text Detection Method with Network Flow," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 331-336.

[23] S. Siddiqua, C. Naveena and S. K. Manvi, "A Combined Edge and Connected Component Based Approach for Kannada Text Detection in Images," *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, Bangalore, India, 2017, pp. 121-125.

[24] Z. Liu, Q. Shen and C. Wang, "Text Detection in Natural Scene Images with Text Line Construction," *2018 IEEE International Conference on Information Communication and Signal Processing (ICICSP)*, Singapore, 2018, pp. 59-63.doi: 10.1109/ICICSP.2018.8549799

[25] G. Huang, "Attention Guided Multi-Scale Regression for Scene Text Detection," *2021 2nd International Conference on Computing and Data Science (C.D.S.)*, Stanford, CA, U.S.A., 2021, pp. 498-502.

[26] S. Kheirinejad, N. Riaihi and R. Azmi, "Persian Text Based Traffic sign Detection with Convolutional Neural Network: A New Dataset," *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, 2020, pp. 060-064.

[27] D. N. Van, S. Lu, X. Bai, N. Ouarti and M. Mokhtari, "Max-Pooling Based Scene Text Proposal for Scene Text Detection," *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, 2017, pp. 1295-1300.

[28] H. Mokayed, P. Shivakumara, M. Liwicki and U. Pal, "A New Defect Detection Method for Improving Text Detection and Recognition Performances in Natural Scene Images," *2020 Swedish Workshop on Data Science (SweDS)*, Lulea, Sweden, 2020, pp. 1-7.

[29] A. A. Chandio, M. Asikuzzaman, M. R. Pickering and M. Leghari, "Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network," in IEEE Access, vol. 10, pp. 10062-10078, 2022, doi: 10.1109/ACCESS.2022.3144844.

[30] Detection of Adversarial Examples in Text Classification: Benchmark and Baseline via Robust Density Estimation (Yoo et al., Findings 2022).

[31] Bian, X., Wang, C., Quan, W. *et al.* Scene text removal via cascaded text stroke detection and erasing. *Comp. Visual Media* **8,** 273–287 (2022). https://doi.org/10.1007/s41095-021-0242-8.

[32] Chuhui Xue, Shijian Lu, Steven Hoi,Detection and rectification of arbitrary shaped scene texts by using text keypoints and links,Pattern Recognition,Volume 124,2022,108494,ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2021.108494.

[33] https://towardsdatascience.com/pytorch-scene-text-detection-andrecognition-by-craft-and-a-four-stage-network-ec814d39db05

[34] Yang C, Chen M, Xiong Z, Yuan Y, Wang Q. CM-Net: Concentric Mask Based Arbitrary-Shaped Text Detection. IEEE Trans Image Process. 2022;31:2864-2877. doi: 10.1109/TIP.2022.3141844. Epub 2022 Apr 8. PMID: 35349439.

[35] Wu, F., Zhu, C., Xu, J. et al. Research on image text recognition based on canny edge detection algorithm and k-means algorithm. Int J Syst Assur Eng Manag 13, 72–80 (2022). https://doi.org/10.1007/s13198-021-01262-0

[36] Yuanhong Zhong, Xinyu Cheng, Tao Chen, Jing Zhang, Zhaokun Zhou, Guan Huang,PRPN: Progressive region prediction network for natural scene text detection,Knowledge-Based Systems,Volume 236,2022,107767, ISSN 0950-7051,https://doi.org/10.1016/j.knosys.2021.107767.

[37] Sudhakara, M., et al. "A Statistical Analysis of Fruit and Vegetables Quality Detection and Disease Classification for Smart Farming." *JOURNAL OF ALGEBRAIC STATISTICS* 13.2 (2022): 1426-1438.

[38] Maddela, Vijayasanthi, et al. "Security Assaults, Its Counter Measures and Difficulties over a Sensor Cloud." *JOURNAL OF ALGEBRAIC STATISTICS* 13.2 (2022): 1633-1643.

[39] LNC Prakash, et al. "Real-Time Conceptual Video Interpretation for Surveillance Systems using Euclidean Norms", *JOURNAL OF ALGEBRAIC STATISTICS* 13.2 (2022): 753-759.