Sri Lankan Journal of Technology
Research

SLJoT
2025 Sp.Issue 85-96
ISSN-2773-6970

Extracting English Text From Complex Natural Scene Imagery

T.Kiniththa[1], U.A.P Ishanka[2]

[1] Department of Information and Communication Technology, South Eastern University, Sri Lanka
[2] Sabaragamuwa University, Sri Lanka

*Corresponding Author: tkiniththa05@gmail.com || ORCID: 0009-0002-3487-7380

Received: 26-11-2024. * Accepted: 11-04-2025 * Published Online: 05-05-2025

Abstract- Text detection in natural scene images has become an essential area of research in image processing, driven by the increasing demand for automated text extraction in various applications. Unlike traditional Optical Character Recognition (OCR) tasks, detecting text from natural scenes presents unique challenges, including complex backgrounds, varying text orientations, and diverse font types. This paper investigates the use of EasyOCR, an OCR tool built with deep learning techniques, combined with the Character Region Awareness for Text detection (CRAFT) algorithm, to detect English text in natural scene images. The study utilizes a dataset of 450 images captured under diverse conditions, emphasizing the impact of noise reduction techniques on detection accuracy. To address the challenges posed by noise and background complexity, various filters, including Gaussian, Median, Average, and Bilateral filters, are applied during the preprocessing stage to enhance text detection accuracy. The experimental results demonstrate that Bilateral Filtering, followed by Gaussian Filtering, significantly improves the text detection accuracy, achieving an overall accuracy of 95.56% and word accuracy of 82.43%. The study highlights the importance of preprocessing in improving OCR performance in natural scene images and offers insights into the effectiveness of different filters in enhancing detection accuracy. Despite the progress, the paper identifies areas for future improvement, such as addressing text occlusion, curved text, and variations in illumination. This research contributes to the ongoing efforts to improve automated text detection in complex environments, providing a foundation for further advancements in the field.

Keywords: Deep Learning, Image Preprocessing, Natural Scene Images, Optical Character Recognition (OCR), Text Detection

Kiniththa, T., & Ishanka, U. A. P. (2025). Extracting English text from complex natural scene imagery. *Sri Lankan Journal of Technology.* 85-96.

1. Introduction

In 2021, image processing has emerged as one of the most popular research areas across various industries and fields, driven by its wide range of applications. A key focus within this domain is text detection, a crucial task that involves extracting textual information from images. This capability has broad applications in data digitization, document analysis, and scene understanding, making it an essential component in advancing technologies like augmented reality, autonomous systems, and artificial intelligence.

Optical Character Recognition (OCR) is a technology used to extract text from images and convert it into machine-readable formats, enabling further data processing tasks such as editing, searching, and indexing. Efficient OCR systems have played a significant role in computer program design for many years, making them essential tools in various industries.

Two commonly used tools for character recognition today are Tesseract-OCR and EasyOCR (Madre & Gundre, 2018). EasyOCR, built with Python and the PyTorch deep learning library, leverages modern machine learning techniques to enhance the accuracy and efficiency of text detection. By utilizing a GPU, the process of detection can be further accelerated. EasyOCR employs the CRAFT (Character Region Awareness for Text detection) algorithm for the detection phase and uses the CRNN (Convolutional Recurrent Neural Network) model for text recognition, making it highly effective for detecting and recognizing text in various image types.

Although Optical Character Recognition (OCR) technologies are highly effective in extracting text from well-structured documents, they struggle when applied to natural scene images, where text is often embedded in complex backgrounds, appears in varying orientations, and is subject to fluctuating lighting conditions. These challenges make accurate text detection in natural scene images a significantly more complex task. (Sidhwa *et al.,* 2018).

Image-processing technologies can be used in combination with optical character recognition to improve recognition accuracy and to improve the efficiency of extracting text from images. The core problem addressed in this research is the accurate detection of text in natural scene images, where traditional OCR methods falter due to the unstructured nature of the input. Enhancing text detection in such images is critical for various applications, from improving accessibility tools to advancing autonomous systems that rely on visual scene understanding.

Text detection from camera captured natural scene images is complex due to various challenges. The main challenges are background complexity, different directions of the text, complexity in backgrounds, and diversity of scene text and interference factors. The challenges present in the text detection attract many researchers to contribute in this area. Text detection and text recognition are two main steps involved in text processing applications.

Text detection is the process of locating and extracting the text present in the image. By optimizing the preprocessing and detection pipeline, this study is expected to yield higher detection accuracy under challenging conditions. Additionally, the comparative analysis of different filtering methods will contribute to a deeper understanding of their impact on text detection performance, providing insights for future applications in fields such as augmented reality, autonomous driving, and data digitization.

This work proposes a method that combines deep learning-based OCR models, such as EasyOCR, with advanced filtering techniques to reduce noise and improve text detection accuracy in natural scene images. By applying these methods, this research aims to overcome the limitations of existing approaches and provide more reliable results under challenging conditions. Addressing the challenges of text detection in natural scenes will not only advance the field of image processing but also enable more robust applications in fields such as artificial intelligence, augmented reality, and real-time information extraction.

2. Literature review

A Number of approaches for text detection in images has been proposed into the past. Automatic detection and translation of text in images done using different techniques proposed. Text detection and recognition in images and video frames, is a process of combination of advanced optical character recognition (OCR) and text-based searching technologies. Unfortunately, text characters contained in images can be any gray-scale value, variable size, low-resolution and embedded in complex background.

In 2019 Matteo Brisinello, Ratko Grbi, Mario Vranjes and Denis Vranjes they are conduct a research regarding review about text detection in natural scene images. There are so many text detection methods. Traditional method divided into two categories sliding window methods which sliding a window over the whole image in different scales and connected component-based method which mainly detect single characters and then group them into words or text-line regions. This method offer unique advantages but also present notable limitations. Sliding window method adaptable and flexible but suffer from high computational costs and potential inaccuracies in complex images. Connected component-based methods provide effective character-level detection but struggle with noise, occlusion, and complex backgrounds. Understanding these strengths and weaknesses is crucial for selecting the appropriate method for specific text detection tasks and highlights the need for further research to address these limitations. (Brisinello *et al*., 2019).

In 2010 Boris Epshtein, Eyal Ofek and Yonatan Wexler, used Stroke Width Transform (SWT) for their research. Stroke Width Transform is a Text detection algorithm; it is local image operator which computes per pixel the width of the mostly likely stroke containing the pixel. The Output of the SWT is an image of size equal to the size of the input image. While SWT provides a robust framework for detecting text in varying conditions, its limitations in handling complex backgrounds and noise, along with its computational requirements, highlight areas for further research and improvement (Epshtein *et al*., 2010).

In order to more efficiently handle scene text with cluttered background information, several hybrid methods are proposed by various researchers, which make use of the advantages of different methods and combine with specific schemes. Fabrizio et al presented a hybrid text detector in 2016, which adopts connected component (CC) method to generate text candidates and also applies texture analysis to compose text string or discard false positives (Fabrizio et al., 2016). There are some proposed algorithms for text detection that algorithms consists of several steps. That steps are finding regions with maximally stable extremal regions (MSER) features, geometric elimination, finding stroke widths with SWT, and connecting characters to obtain text groups (Ozgen *et al*., 2018).

Layman's terms extraction from scene images is divided into two sub-processes namely Scene text detection and, Scene text recognition. In the text detection process, the locations where the text is present in the image are identified. This process is also known as text localization. There can be two types of approaches for the task of text detection; conventional approach and a deep neural network based approach. However, they come with higher computational costs and complexity, and their effectiveness is heavily dependent on the quality of training data (Goel *et al*., 2019).

In 2018 Pooja Patil, Kajol Patil and Dr. Anagha Kulkarni used tesseract OCR method to extract text from image. The input to the system will be an image, particularly a colored one. The image will be processed to detect the area containing the text, the crucial features that uniquely identify the text characters will be detected and extracted. And finally the text is extracted into a text file. The text detection process begins with the denoising of the image followed by converting it to a grayscale image followed by a binarization of that gray scaled image. Once gray scaled, the extracted text is written into a text file. The proposed model is robust to different font sizes, font colors, background colors. While Tesseract OCR offers a robust solution for text extraction from colored images, it is crucial to address its preprocessing dependency and limitations in handling complex backgrounds and unconventional text scenarios. These factors should be considered when evaluating its applicability to specific use cases and real-time requirements. (Patil *et al.,* 2018).

In 2019 Vaibhav Goel, Vaibhav Kumar, Amandeep Singh Jaggi and Preeti Nagrath presents a technique that uses a combination of the Open Source Computer Vision Library (OpenCV) and the Convolutional Neural Networks (CNN), to extract English text from images efficiently. The CNN model is based on a two-stage pipeline that uses a single neural network to directly detect the characters in the scene images. It eliminates the unnecessary intermediate steps that are present in the previous approaches to this task making them slower and inaccurate, thereby improving the time complexity and the performance of the algorithm (Goel *et al*., 2019).

In 2020 Ebin Zacharias, Martin Teuchler and Benedicte Bernier used tesseract OCR method to extract text from image. They used specific application with constrained images as input. The overall accuracy of 83% was achieved which can be slightly improved with the quality of the input images. Text recognition is not flawless when subjected to complex environmental situations. Tesseract OCR is most suitable for the printed texts. The deep learning-based approach can be helpful to evaluate the dependability of such patterns of text in an image and also to analyze the influence of font type in text recognition (Ebin *et al.,* 2020).

In 2020 Shuping Liu, Yantuan Xian, Huafeng Li, and Zhengtao Yu present a novel method to detect text from scene images. Firstly, they decompose scene images into background and text components using morphological component analysis (MCA), which will reduce the adverse effects of complex backgrounds on the detection results. After that, the text in the query image can be detected by applying certain heuristic rules. The results of experiments show the effectiveness of the proposed method. It is essential to address its limitations, including dependence on heuristic rules, computational complexity, scalability, and potential Overfitting. These factors should be considered when evaluating the method's practical applicability and effectiveness in various real-world scenarios. (Liu *et al*., 2020).

Previous research on text detection has identified several persistent challenges, including difficulties with detecting closely spaced text instances, occluded words, long text strings, and curved text. Additionally, issues such as false positives and image noise have been prevalent. Earlier studies often relied on text with simple backgrounds and uniform fonts, which limited their applicability to more complex scenarios. In contrast, this research employs the EasyOCR method and the CRAFT (Character Region Awareness for Text detection) algorithm to enhance text detection. Furthermore, it incorporates a range of filters namely Bilateral Filter, Gaussian Filter, Average Filter, and Median Filter to mitigate noise and improve detection accuracy. By addressing the limitations of previous approaches and handling varied fonts and complex backgrounds, this study aims to achieve more reliable text detection in challenging conditions.

3. Methodology

A. Research Design

This section proposes a pipeline for detecting text in the natural scene images. “Figure 01” shows a flowchart depicting the major steps involved in text detection. A detailed description of the steps involved is discussed in the following sections.

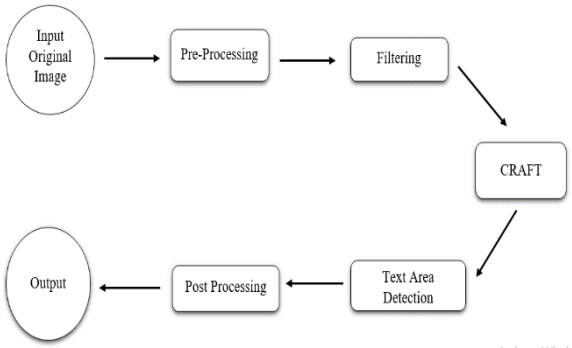

Figure 1. *Flowchart for the Text Detection Model*

B. Data Collection

The natural scene images used in this research were captured using standard cameras, and the dataset comprises nearly 450 images containing English text to be detected. These images include noise and require pre-processing. Additionally, because the images were captured around the clock, there is significant variation in their brightness.

C. Data Preprocessing

Preprocessing is a critical step that significantly influences the effectiveness of subsequent stages. There are significant amounts of noise available in the collected images, and the camera was set at a slightly inclined angle relative to the text to be detected. The preprocessing phase, essential for accurate text extraction, involves several key strategies: image cropping, brightness adjustment, and grayscale conversion. Due to the low illumination of night photographs and direct flash light reflecting off the text, achieving optimal results was challenging. Properly adjusting the image's lighting is crucial for the EasyOCR engine to function effectively. “Figure 02” illustrates some sample images used in this research.

1) Image Cropping:

The images acquired are limited in that the text of interest is consistently positioned at the lower part of the image. These constraints and the unique characteristics of the input images facilitated the establishment of a cropping region to eliminate unnecessary portions of the images. Consequently, the images are cropped based on a predefined crop region.

Figure 2. *Sample Images in the Dataset*

2) Brightness Check

After cropping the images, the next challenge was addressing the varying brightness levels. The night photos were further complicated by the presence of an additional light source, which significantly affected the illumination. To address this, the brightness of the images was carefully assessed and adjusted to ensure consistent differentiation among the photos.

3) Gray Scaling:

Grayscale is a black-and-white monotone that uses just varying shades of gray instead of black and white. Gray scale is just a method of lowering complexity, and it is used to reduce image noise before the detection process begins.

D. Filtering

To reduce noise in the images, various filters are applied. These filters work by blurring the image through the use of low-pass filter kernels, which effectively smooth out high-frequency components and reduce noise, though they can also blur edges. Gaussian filter, Average Filter, Median Filter & Bilateral Filtering these are the filters used to remove the noises in the image. In Gaussian Filtering, instead of using a box filter with identical filter coefficients, a Gaussian kernel is employed in this approach. Median Filtering computes the median of all pixels within the kernel window, and this value is used to replace the center pixel. This is a great way to get rid of salt-and pepper noise. Bilateral Filtering, this is not the case with the bilateral filter,

which was designed to remove noise while keeping edges and is quite good at it. Averaging it simply averages all of the pixels in the kernel area and uses that average to replace the core element.

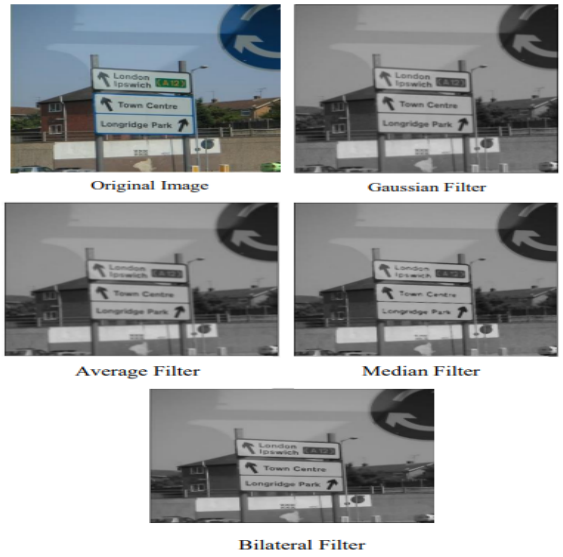

Figure 3. *Filtered Images*

E. Text Detection

As the next step CRAFT text detector is used to detect the text in the filtered images. Text detector that effectively detects text area by exploring each character region and affinity between characters. Text detection is performed using Easy OCR. Easy OCR is a machine learning-based text detection model which not only possesses high accuracy levels but also supports a wide variety of Languages.

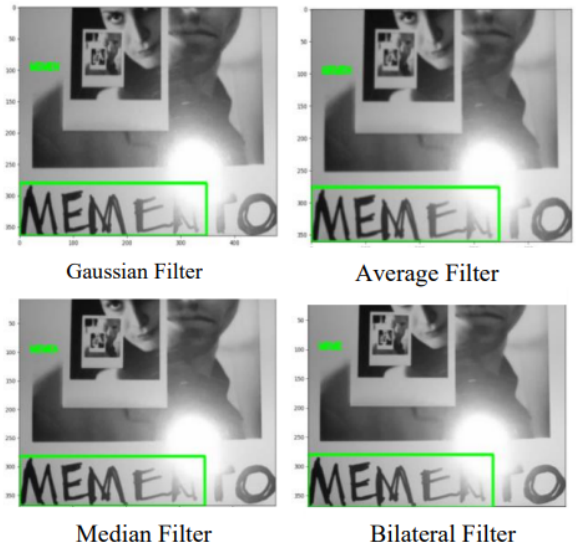

Figure 4. *Text Detection after Filtering*

F. Post Processing

The detected text is then evaluated using text post-processing. Draw the bounding box and text over the image on which we have to perform our detection. After doing the processes

mentioned above, can get the output as the detected text with a bounding box and find the accuracy for that using the parameter.

Figure 5. *Bounding Box around Text*

G. Performance Evaluation

In the text detection part. To evaluate the validity of the text detection of the images, this paper uses accuracy to evaluate the performance of the model. A confusion matrix is used as a parameter to find the accuracy of the text detection. The calculation formula is as shown in formula (A) & formula (B).

$$\text{Text Detection Accuracy} = (TP + TN) / (FP+FN+TP+TN)$$

Formula-(A)

- TP means that positive samples judged as positive samples
- FN means that positive samples judged as negative samples
- FP means that negative samples determined as positive samples
- TN means that negative samples determined to be negative samples

$$\text{Word Detection Accuracy} = (n - \#error) / n$$

Formula – (B)

- Here n = Total Words

4. Result & discussion

The results of this study were represented using the accuracy of the text detection. It was observed that the system or tools detected the target words successfully. Initially, a dataset of 450 images was used for testing. After text post-processing and verifying the checksum based on known calculation, the overall Text Detection accuracy was 93.28%. If one of the text in the image was detected wrong, for find and minimizing this problem word accuracy of the detected text is calculated. The word Detection accuracy of the image showed as 74.06%. These calculations are done before using the filers for the images.

After using a variety of filters to the images again text detection accuracy and word detection accuracy are calculated. The "Table 01" shows the accuracy of the dataset after using different filters. Using filters increases the accuracy of the image. Filters are used for remove the noise in the image and background of the image.

Table 1
Accuracy of Total Images

Methods	Word Detection Accuracy (%)	Text Detection Accuracy (%)
Without Filter	74.06	93.28
Gaussian Filter	86.07	95.59
Average Filter	79.19	95.07
Median Filter	77.55	95.27
Bilateral Filter	82.43	95.56

Word Detection Accuracy refers to how many words are correctly identified or detected from an image that contains text. It measures the ability of a system to locate and detect the presence of words in an image, regardless of whether the words themselves are recognized correctly. For example, if there are 10 words in an image and the system correctly detects 8 of them, the word detection accuracy is 80%.

Text Detection Accuracy refers to how accurately the detected words are recognized or interpreted. This means how correctly the characters or words detected from the image are understood by the system. For example, if the system detects 10 words and correctly recognizes 7 of them (spelling, characters) the text detection accuracy would be 70%.

In order to compare the accuracy while using each filters. Bilateral Filtering gives the best result next the Gaussian filter then the Median Filter and finally the average filter. So using the filter can reduce the noise of the image and increase the accuracy of text detection and word accuracy.

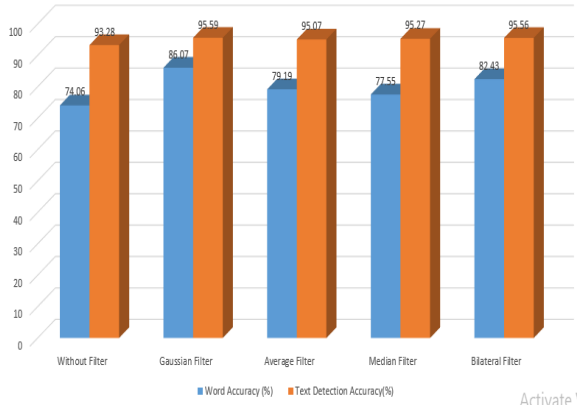

Figure 6. *Accuracy Chart*

The effectiveness of filters in this study echoes similar observations in past research. For example, Epshtein *et al*. (2010) highlighted the limitations of text detection algorithms in noisy environments and suggested the use of preprocessing techniques to mitigate these issues. The improvements achieved through filtering in this study validate the importance of preprocessing as outlined in these earlier studies. Moreover, the use of filters in this research complements the approaches seen in works by Brisinello *et al*. (2019) and Patil et al. (2018), where preprocessing steps were crucial for enhancing text detection accuracy. The findings also align with the approaches used by Liu *et al*. (2020), who employed morphological component analysis to reduce background noise, similarly demonstrating the need for effective noise reduction strategies.

5. Conclusion

Optical Character Recognition (OCR) continues to be an extensively researched field, with text detection in natural scene images presenting significant challenges. This study focused on a specific application involving constrained images, resulting in an impressive overall text detection accuracy of 95.56%. This high level of accuracy was primarily achieved through the application of various filters to reduce image noise, with the Bilateral Filter demonstrating the most substantial improvement in accuracy. The process involved calculating word accuracy to further assess and minimize false positives, which was initially measured at 74.06% before filtering. After applying filters, the word accuracy improved, with the Bilateral Filter achieving 82.43%, showcasing the filter's effectiveness in enhancing detection precision. Despite these advancements, text detection remains imperfect, particularly under complex environmental conditions. Small variations in illumination and text angles led to notable errors in detection. Although EasyOCR performed well with printed text, its accuracy varied significantly with scene text images, highlighting the need for further refinement. To address these challenges, training and implementing a deep learning-based model specifically designed for scene text extraction could offer promising improvements in both detection accuracy and robustness, even in more complex scenarios.

References

Brisinello, M., Grbic, R., Vranjes, M., & Vranjes, D. (2019, September 1). Review on Text Detection Methods on Scene Images. IEEE Conference Publication | IEEE Xplore.

Cho, H., Sung, M., & Jun, B. (2016). Canny Text Detector: Fast and Robust Scene Text Localization Algorithm. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) | IEEE Xplore.

Epshtein, B., Ofek, E., & Wexler, Y. (2010, June 1). Detecting text in natural scenes with stroke width transform. IEEE Conference Publication | IEEE Xplore.

Fabrizio, J., Robert-Seidowsky, M., Dubuisson, S., Calarasanu, S., & Boissel, R. (2016). TextCatcher: A method to detect curved and challenging text in natural scenes. International Journal on Document Analysis and Recognition, pp. 99–117.

Goel, V., Kumar, V., Jaggi, A. S., & Nagrath, P. (2019). Text Extraction from Natural Scene Images using OpenCV and CNN. International Journal of Information Technology and Computer Science, pp. 48–54.

Hanif, S. M., Prevost, L., & Negri, P. A. (2008). A cascade detector for text detection in natural scene images. Proceedings - International Conference on Pattern Recognition.

Jain, A., Peng, X., Zhuang, X., & Natarajan, P. (2014, May 1). Text detection and recognition in natural scenes and consumer videos. IEEE Conference Publication | IEEE Xplore.

Kumar, A. (2014). An Efficient Approach for Text Extraction in Images and Video Frames Using Gabor Filter. International Journal of Computer and Electrical Engineering, pp. 316–320.

Lin, H., Yang, P., & Zhang, F. (2019). Review of Scene Text Detection and Recognition. Archives of Computational Methods in Engineering, pp. 433–454.

Liu, X., Meng, G., & Pan, C. (2019). Scene text detection and recognition with advances in deep learning: a survey. International Journal on Document Analysis and Recognition, pp. 143–162.

Liu, S., Xian, Y., Li, H., & Yu, Z. (2020, January 1). Text detection in natural scene images using morphological component analysis and Laplacian dictionary. IEEE/CAA Journal of Automatica Sinica| IEEE Xplore.

Liu, Z., Shen, Q., & Wang, C. (2018). Text Detection in Natural Scene Images with Text Line Construction. IEEE International Conference on Information Communication and Signal Processing (ICICSP)| ResearchGate.

Madre, S.C., & Gundre, S. B. (2018). OCR Based Image Text to Speech Conversion Using MATLAB. 2018 Second International Conference on Intelligent Computing and Control Systems | IEEE Xplore.

Neeta Devi, C., Mamata Devi, H., & Das, D. (2015, November 1). Text detection from natural scene images for Manipuri Meetei Mayek script. 2015 IEEE International Conference on Computer Graphics, Vision and Information Security | IEEE Xplore.

Nimasha, Ranathunge, Jayawickrama, Mahaliyanaarachchi, & Subhagya. (2018, September 1). A Robust Algorithm for Text Extraction from Signage Images. 18th International Conference on Advances in ICT for Emerging Regions | IEEE Xplore.

Ozgen, A. C., Fasounaki, M., & Ekenel, H. K. (2018). Text detection in natural and computer-generated images. 2018 26th Signal Processing and Communications Applications Conference (SIU) | IEEE Xplore.

Patil, P., Patil, K., Kulkarni, A., Iyer, S., & Natu, I. (2018, August 1). Detection of Hindi Textual Characters from an Image. Fourth International Conference on Computing Communication Control and Automation | IEEE Xplore.

Sun, Y., Dawut, A., & Hamdulla, A. (2018, December 1). Retracted: A Review: Text Detection in Natural Scene Image. 3rd International Conference on Smart City and Systems Engineering | IEEE Xplore.

Su, Y. M., Peng, H. W., Huang, K. W., & Yang, C. S. (2019). Image processing technology for text recognition. 24th International Conference on Technologies and Applications of Artificial Intelligence.

Sidhwa, H., Kulshrestha, S., Malhotra, S., & Virmani, S. (2018, October 1). Text Extraction from Bills and Invoices. International Conference on Advances in Computing, Communication Control and Networking | IEEE Xplore.

Up, E., & Iqubal, S. (2018). Text Detection in Natural Scene Images. International Journal of Latest Trends in Engineering and Technology| ResearchGate.

Xue, M., Shivakumara, P., Zhang, C., Xiao, Y., Lu, T., Pal, U., Lopresti, D., & Yang, Z. (2021). Arbitrarily-Oriented Text Detection in Low Light Natural Scene Images. IEEE Transactions on Multimedia, pp. 2706–2720.

Zacharias, E., Teuchler, M. and Bernier, B. (2020) Image processing based Scene-Text detection and recognition with Tesseract. ResearchGate.

