Asia-Pacific Journal of Information Technology and Multimedia 13(02) 2024: 221 – 231
https://doi.org/10.17576/apjitm-2024-1302-05

Corner Pixel-Based Method for Selecting Binary Text in Scene

Kaedah Berasaskan Piksel Sudut untuk Memilih Teks Binari Dalam Pemandangan

Ednawati Rainarli[1], Suprapto[2], Wahyono[2]*

[1]Faculty of Engineering and Computer Science, Universitas Komputer Indonesia, Bandung, 40132, West Java, Indonesia

[2]Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, 55281, D.I. Yogyakarta, Indonesia

**Corresponding author: ednawati.rainarli@email.unikom.ac.id*

Received 31 May 2024
Accepted 11 September 2024, Available online 9 October 2024

ABSTRACT

Text detection in natural images is a process to indicate the location and presence of text appearing in images. The complexity of the background images, the similarity of text shapes to non-text objects, and the variability in text shapes and colours make automatic text detection in natural images challenging to achieve using traditional image processing techniques alone. The machine learning methods are one way to perform filtering to eliminate non-text candidates. We used secondary data as additional training data, such as in the ICDAR 2011, ICDAR 2013, and ICDAR 2015. The diversity of text colours in these datasets makes binary image processing not uniformly applicable to each image. Therefore, in this study, we proposed a method to process text images and automatically select between binary or negative binary images by checking pixels at the four corners of the binary and negative binary images. If the number of white pixels is greater than or equal to two, select the negative binary image; otherwise, select the binary image. This way, we automatically selected suitable images for feature extraction before using them to build text and non-text classification models. For low-resolution text images and digitally created text images, in ICDAR 2011, the accuracy of selecting binary text images reached 85.00%. For focused text taken with specific purposes and horizontal text appearances, like in ICDAR 2013, the accuracy of selected binary text images reached up to 92.10%. The accuracy of binary text image selection reached 66.67% for incidental text with multi-oriented text positions. Based on the research results, the proposed strategy can work optimally, especially for focused text with various colours, including white or black coloured text, with diverse sizes and types of text.

Keywords**:** binary text; scene text detection; text candidates; text segmentation

ABSTRAK

Pengesanan teks dalam imej semula jadi adalah proses untuk menunjukkan lokasi dan kewujudan teks yang muncul dalam imej. Kerumitan imej latar belakang, persamaan bentuk teks dengan objek bukan teks, serta kepelbagaian bentuk dan warna teks menjadikan pengesanan teks secara automatik dalam imej semula menjadi sukar dicapai dengan menggunakan teknik pemprosesan imej tradisional sahaja. Kaedah pembelajaran mesin adalah salah satu cara untuk melakukan penapisan bagi menghapuskan calon bukan teks. Kami menggunakan data sekunder sebagai data latihan tambahan, seperti dalam ICDAR 2011, ICDAR 2013, dan ICDAR 2015. Kepelbagaian warna teks dalam set data ini menjadikan pemprosesan imej binari tidak dapat diaplikasikan secara seragam pada setiap imej. Oleh itu, dalam kajian ini, kami mencadangkan satu kaedah untuk memproses imej teks dan secara automatik memilih antara imej binari atau binari negatif dengan memeriksa piksel pada empat sudut imej binari dan binari negatif. Jika bilangan piksel putih adalah lebih besar daripada atau sama dengan dua, pilih imej binari negatif; jika tidak, pilih imej binari. Dengan cara ini, kami secara automatik memilih imej yang sesuai untuk pengekstrakan ciri sebelum menggunakannya untuk membina model klasifikasi teks dan bukan teks. Untuk imej teks resolusi rendah dan imej teks yang dicipta secara digital, dalam ICDAR 2011, ketepatan pemilihan imej teks binari mencapai 85.00%. Untuk teks fokus yang diambil dengan tujuan khusus dan teks yang muncul secara mendatar, seperti dalam ICDAR 2013, ketepatan pemilihan imej teks binari mencapai sehingga 92.10%. Ketepatan pemilihan imej teks binari mencapai 66.67% untuk teks sampingan dengan kedudukan teks berbilang orientasi. Berdasarkan hasil kajian, strategi yang dicadangkan boleh berfungsi secara optimum, terutamanya untuk teks fokus dengan pelbagai warna, termasuk teks berwarna putih atau hitam, dengan saiz dan jenis teks yang pelbagai.

Kata kunci: teks binari; pengesanan teks dalam pemandangan; calon teks; segmentasi teks

INTRODUCTION

Text detection in natural images is a dynamic and intricate field of computer vision research. It is the initial step in developing systems that can autonomously identify text within natural images. The output of text detection is represented by text localization using bounding boxes. There are many applications of text recognition in natural images, including automatic image captioning, information extraction from natural images, content-based image retrieval, and object recognition based on text in images. However, text detection in natural images poses unique challenges (Mahajan and Rani, 2021). These challenges arise from the complexity of background scenes in natural images, variations in appearing text, and diverse image acquisition conditions from the camera. As a result, researchers are constantly innovating to develop models for text detection in natural images that can effectively detect text under various conditions.

According to Rainarli et al., (2021), one approach to scene text detection utilizes a bottom-up pipeline. Candidate letter extraction, non-letter candidate filtering, letter merging into words, candidate text feature extraction, and text classification are stages in natural image text detection with a bottom-up approach. Text classification employs methods in machine learning. Methods such as SVM (Francis and Sreenath, 2017; Zheng et al., 2017), AdaBoost (Naiemi et al., 2020; Qiu et al., 2018; Soni et al., 2018), or CNN (Tian et al., 2017; Turki et al., 2018; Yasmeen et al., 2020; Zhang et al., 2018) are some classification methods for filtering text classes. Training these models requires both text and non-text training data. Adding training data, in addition to using data augmentation, can be done to address the imbalance in the number of text and non-text candidates from text candidate extraction results. This thorough approach ensures the reliability and robustness of our findings.

In training, we can use binary images of text and non-text candidates. The color variation of text appearing in scene data makes direct binarization challenging. For example, in white text images and black text images, grayscale images are used for binarization in white text images. In contrast, negative grayscale images convert black text images to binary images. This method also applies to text of other colors. We propose a method for selecting binary or negative binary images by first checking the color intensity of the four corner pixels in an image to address this issue.

Moreover, uneven lighting conditions, background scene complexity, and text variation necessitate image processing to improve image quality before binarization. Some studies Lashkov et al., (2023); Manju et al., (2019); Paul and Aslan, (2021); Yuan et al., (2023) use Contrast Limited Adaptive Histogram Equalization (CLAHE) to address uneven lighting conditions in images and images with low contrast. CLAHE is not only used in medical image processing (Hayati et al., 2022) but can also be used for vehicle detection at night (Lashkov et al., 2023) and as a preprocessing step for real-time face recognition (Paul and Aslan, 2021). Therefore, in this study, CLAHE will be used in preprocessing before conversion into grayscale and binary images.

Our research brings two contributions. The first is a unique framework that leverages CLAHE to process text images into binary images. Second, we propose an algorithm for automatically selecting binary images, a approach allows us to add text augmentation data for training text and non-text classification automatically. The structure of this manuscript starts with a comprehensive background of the problem, highlighting the necessity for binarization in text images and the challenges encountered during text image binarization. The subsequent section details the description and characteristics of the scene text data, along with our proposed binarization methods. We presented our testing and discussion in the third section. Finally, we conclude with the results obtained and our plans for further research development, promising exciting avenues for future exploration in this field.

METHODOLOGY

In this study, we use three public datasets (Karatzas et al., 2015, 2013; Shahab et al., 2011). These datasets are from challenges in the Robust Reading Competition over three years. Each dataset has different characteristics and challenges. Table 1 describes those datasets. In Table 1, the number of words comes from the word count in the training data for each dataset. ICDAR 2015 is the dataset with the highest image complexity among these datasets. The ICDAR 2011 dataset has relatively homogeneous background images. Meanwhile, the ICDAR 2013 dataset has some images with uneven lighting. We include sample images from these three datasets in Figure 1 for comparison.

Figure 1(a) shows examples of words from the ICDAR 2011 dataset. This dataset includes word art, text with low contrast, and variations in the background image color that are not uniform. Figure 1(b) provides examples from the ICDAR 2013 dataset. Text appearing with occlusion, blurry text, and uneven lighting are additional challenges in the ICDAR 2013 dataset. For the ICDAR 2015 dataset in Figure 1(c), text with multi-orientation poses another challenge for text detection in natural images. Some text appearing among clusters of trees will pose a challenge during word segmentation.

TABLE 1. Description of the datasets

Dataset	Number of words	Description
ICDAR 2011	3567	Data originates from text images in electronic documents such as web and email. There are variations in text size, font type, and color of the scene text.
ICDAR 2013	848	Text is the focus of natural images. Horizontal text images vary in color, size, and font type. The image can have uneven lighting, blur, and occlusion.
ICDAR 2015	2073	Text incidentally appears during image capture. Variations in text orientation, low-resolution text, and background image complexity resembling text pose additional challenges in detecting scene text.

(a) (b) (c)

FIGURE 1. Examples of images from the ICDAR 2011 (a), ICDAR 2013 (b), and ICDAR 2015 (c) datasets

To obtain binary images, we follow several steps. Figure 2 illustrates the stages in selecting binary images from scene text images. Our detection process starts by enhancing the image quality using CLAHE. The images are separated into three channels: R, G, and B. Then, the system applies CLAHE to each channel to improve contrast. Afterwards, the system recombines the three channels into an enhanced RGB image. The purpose of using CLAHE is to enhance objects in images with low contrast. This enhancement not only sharpens text but also non-text objects. Therefore, the next step is to remove noise. The noise removal process is done using a Gaussian filter. Once the noise is removed, the image is resized to 32×16 pixels. The system then converts the resized image to grayscale. Next, the grayscale image is transformed into two binary images using binarization methods Otsu thresholding. The results are in two binary images, including a negative binary image. The final step is selecting the binary image to be used in the feature extraction process.

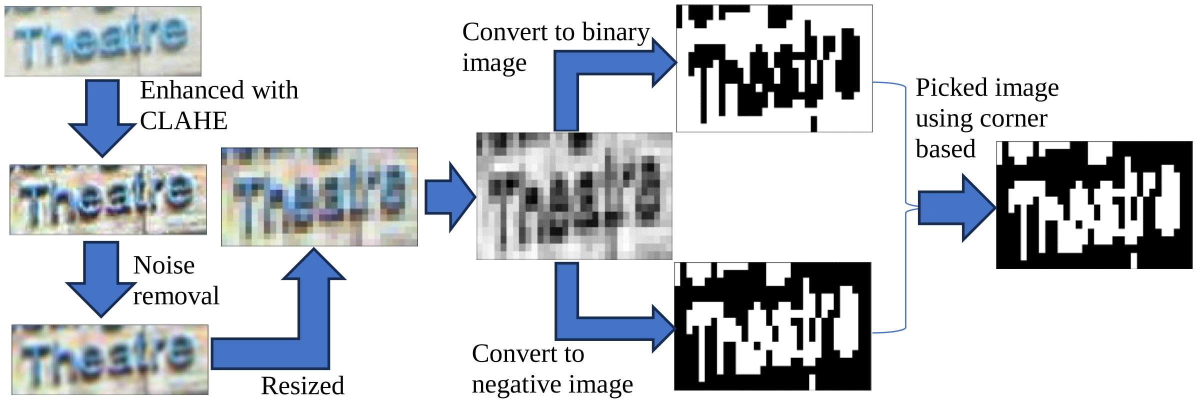


FIGURE 2. Illustration of image processing to obtain binary images

The text that appears in natural images is not always white or black. Therefore, it is necessary to check beforehand to determine which binary image should be used for feature extraction. Algorithm 1 presents the pseudocode of the process we propose to automatically select the appropriate binary image. We check the intensity values of the four corner pixels of the binary image. If the intensity value of a corner pixel is 255, the corresponding pixel_value is assigned a value of 1; otherwise, it is assigned a value of 0. If all four corner pixels are white, the total number of elements in pixel_value will be 4. This indicates that the background of the image is white, and the text is black. In this case, the negative binary image is used; otherwise, we use the binary image. This method allows us to segment the text into a white-on-black image. Algorithm 1 shows the complete rule in lines 10-13.

ALGORITHM 1. Corner-Pixel Based
Input: *grayscale*
Output: *keep_image*
1 $binary \leftarrow threshold(grayscale)$
2 $negative\_binary = 255 - binary$
3 $h, w \leftarrow size\ of\ binary$ # size of binary image
4
5 $pixel\_value[0] \leftarrow integer(binary[0,0] == 255)$
6 $pixel\_value[1] \leftarrow integer(binary[w,0] == 255)$
7 $pixel\_value[2] \leftarrow integer(binary[0,h] == 255)$
8 $pixel\_value[3] \leftarrow integer(binary[w,h] == 255)$
9
10 if $sum\ (pixel\_value) < 2$ then
11 $\quad keep\_image \leftarrow binary$
12 else
13 $\quad keep\_image \leftarrow negative\_binary$
14
15 Return idx_keep

RESULT AND DISCUSSION

We observed the effect of using CLAHE on scene text images. The second row of Figure 3 shows several results after the image is enhanced with CLAHE, while the first row of Figure 3 represents the original images. In Figures 3(a) and 3(d), CLAHE successfully enhances the contrast of the text objects to be recognized. Using CLAHE under uneven lighting conditions with low intensity, as shown in Figures 3(b) and 3(c), produces better images. However, under certain conditions, such as in Figures 3(e) and 3(f), CLAHE introduces noise into the text. Therefore, a noise removal process is necessary in the subsequent steps.

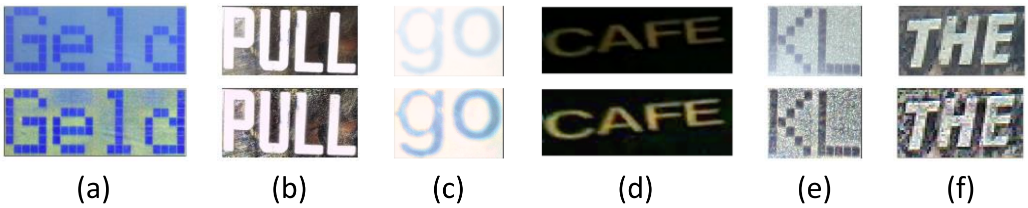


FIGURE 3. Sample of image results after CLAHE enhancement

We also conducted quantitative observations to measure the extent of the impact of using CLAHE. We utilized the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) Score to assess the success of the image enhancement process (Mittal et al., 2011). This method is beneficial in image processing and computer vision to measure the naturalness and quality of images without requiring corresponding reference images as benchmarks. The smaller the BRISQUE score, the better the perceptual quality and naturalness of the image (Mittal et al., 2011).

Our findings, as detailed in Table 2, revealed some intriguing results. The most successful use of CLAHE was for the ICDAR 2013 dataset, with a score reduction of 8.04. However, for the ICDAR 2015 dataset, the score was insignificant, only 0.26. This slight difference may be due to the complexity of the background images in the ICDAR 2015 dataset. The most unexpected result came from the ICDAR 2011 dataset, where the use of CLAHE increased the BRISQUE score, reaching 8.68. This result is possible because the images in the ICDAR 2011 dataset have more uniform background images than the ICDAR 2013 dataset. These unexpected findings open up new avenues for further research and exploration.

Based on qualitative observations, many images after the CLAHE process showed more noise, as seen in Figures 3(e) and 3(f). We minimized the influence of noise by using a Gaussian Filter before converting the images to grayscale.

TABLE 2. The comparison of BRISQUE scores before and after the use of CLAHE

Dataset	Average of BRISQUE Score	
	Original image	After enhancement
ICDAR 2011	113.74	122.42
ICDAR 2013	85.26	77.22
ICDAR 2015	73.33	73.07

We also tested all three public datasets to test the success of our binary text image selection. We manually evaluated the binary images. If the text is segmented with white color in the image, then the binary image selection process was successful. The accuracy calculation utilizes Equation (1), which is:

$$accuracy = \frac{number\ of\ binary\ images\ selected\ accurately}{number\ of\ image} \times 100\% \tag{1}$$

Table 3 shows the accuracy measurement results for the ICDAR 2011, ICDAR 2013, and ICDAR 2015 datasets. The highest average accuracy value in Table 3 is for the ICDAR 2013 dataset, which is consistent with the BRISQUE score measurement results. The lowest average accuracy is in the ICDAR 2015 dataset. The complexity of the images and suboptimal use of CLAHE are suspected causes of poor binary text segmentation.

TABLE 3. The accuracy of the three datasets

Dataset	**The average of accuracy (in percent)**
ICDAR 2011	85.00
ICDAR 2013	92.10
ICDAR 2015	66.67

Furthermore, to test the impact of using binary data as augmentation in text and non-text classification, we compared the classification model results of Extreme Learning Boosting (XGB), Random Forest (RF), and Support Vector Machine (SVM) with and without the addition of binary data. The training text data was sourced from the ICDAR 2013 dataset. The additional text data, converted to binary, was also sourced from the ICDAR 2013 dataset. The ground truth, which is the bounding box from the ICDAR 2013 dataset, was used to label non-text data. If the Intersection over Union (IoU) of the candidate component is zero, then the component is classified as non-text. Table 4 shows the composition of the training data, additional binary data, and test data used in the text and non-text classification. A total of 806 binary text data were added as part of the training data.

TABLE 4. The number of data from each training data, augmentation data, and testing data

	Training	**Biner**	**Testing**
Non-text	4668	-	642
Text	845	806	214

Feature extraction from each text and non-text component was conducted using the Histogram of Oriented Gradients (HOG) method. We used the features and class targets from the training data to fit the hyperparameters of each classification method: XGB, RF, and SVM. Once the models were obtained, we conducted testing using the test data. As a reference, we also trained and tested the text classification models without the addition of binary data.

TABLE 5. Comparison of text classification performance using the XGB, RF, and SVM

Method	Accuracy		Precision		Recall		F1	
	Initial	Proposed	Initial	Proposed	Initial	Proposed	Initial	Proposed
XGB	0.8470	0.8715	0.8822	0.8885	0.7048	0.7601	0.7426	0.7981
RF	0.8540	0.8738	0.8772	0.9022	0.7235	0.7586	0.7615	0.7991
SVM	0.8855	0.8890	0.8947	0.8998	0.7913	0.7967	0.8262	0.8319

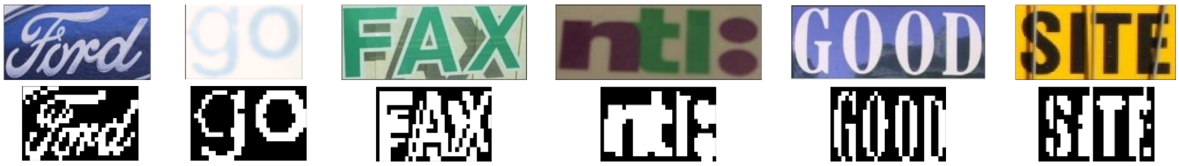
Table 5 shows that the addition of text data during training increased the recall value the most for XGB, by almost 6%. Among the three classification methods, SVM showed a smaller improvement compared to XGB and RF. However, there was still an improvement, which may be attributed to the addition of class weighting in SVM, providing a robust mechanism for handling data imbalance. This similar result was also reported by research (Zhang and Zhong, 2020). This indicates that binary data augmentation still provides benefits, albeit on a smaller scale. Generally, data augmentation improves the performance of all tested methods, with the greatest improvement seen in XGB and RF.

Figure 4 shows several results of binary image selection from samples of the ICDAR 2011 dataset that were successfully segmented. The second row of Figure 4 depicts several successful results of binary text segmentation. Differences in text color, as seen in Figure 4(d), and contrasting colors between the text and background image, as seen in Figure 4(b) and (c), are reasons for the successful binarization process we conducted. Particularly for Figure 4(a), the text appears on a textured background. However, the significant color difference between the background image and the text resulted in successful segmentation.

(a) (b) (c) (d)

FIGURE 4. Sample results of binary image segmentation from the ICDAR 2011 dataset

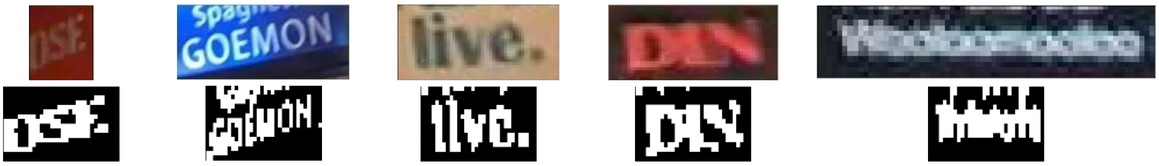
In the ICDAR 2013 dataset, for connected text, as shown in Figure 5(a), text with shadow effects, as in Figure 5(a), was successfully segmented well. Text appearing less contrasted with the background image, as in Figure 5(b), can be segmented effectively. As in Figure 5(f), text appearing with occlusion is well segmented, but occlusion with a fence causes the segmented text to be separated into several parts, such as in the letters S and E. This fraction will affect the letter recognition process. Therefore, additional image processing, such as dilation or closing, can handle this condition. Text with more complex background images, as in Figure 5(e), or text appearing in different colors, as in Figure 5(d), can still be segmented effectively.



(a) (b) (c) (d) (e) (f)

FIGURE 5. Sample results of binary image segmentation from the ICDAR 2013 dataset

We present the text segmentation results from the ICDAR 2015 dataset in Figure 6. Images with low contrast, such as Figure 6(a), and images with uneven lighting conditions, as shown in Figure 6(b), were successfully segmented well. Blurry text images like Figures 6(c) and 6(d) also produced good binary text images. Figures 6(d) and 6(e) show the well-segmented texts. However, the quality of the text shapes could be better because the images in Figures 6(d) and 6(e) are blurry and out of focus, making the text unreadable even to the human eye. We can apply techniques for image sharpening or super-resolution techniques to enhance low-resolution images can be used before converting the images into grayscale.



(a) (b) (c) (d) (e)

FIGURE 6. Sample of binary image segmentation from the ICDAR 2015 dataset

We present the text segmentation results from the ICDAR 2015 dataset in Figure 6. Images with low contrast, such as Figure 6(a), and images with uneven lighting conditions, as shown in Figure 6(b), were successfully segmented. However, not all blurry text images were successfully segmented. Figure 7 depicts several images that failed to be segmented. The rule of checking pixels at the four corners of the image may fail if the text to be segmented has a color similar to some corners of the image, as seen in Figure 7(e) and Figure 7(a). Failures in the binarization process, as observed in Figure 7(c), Figure 7(d), and Figure 7(f), indicate that the Corner Pixel-Based algorithm was unable to accurately select the binary text image. Additionally, the influence of lighting and text positions too close to the edges of the image, as in Figure 7(b), resulted in the letters 'P' and 'L' in the word 'Pull' being identified as white in all three-pixel corners. This condition caused the Corner Pixel-Based algorithm to fail in selecting the binary text from the image.

(a) (b) (c) (d) (e) (f)

FIGURE 7. Sample of binary text images that failed to be segmented

CONCLUSION

Processing text images into binary text images is necessary as additional data for text classification in the text detection process on natural images. We have proposed a method to obtain binary text images by utilizing binary image checking at the four corner pixels of the binary image. We also used CLAHE to enhance the image quality before transforming the image into a grayscale image. Test results show that mainly for the ICDAR 2013 dataset, the strategy of selecting images using the four-corner pixel checking successfully automatically selected binary text images with an accuracy of 92.10%. The results of text classification testing indicate that augmentation with binary data generally enhances performance for all methods tested, with the most significant improvements observed in XGBoost and Random Forest. Using CLAHE also improved the image quality in the ICDAR 2013 dataset. The primary constraint is the appearance of text under various conditions, such as uneven lighting and shallow contrast images, which hinder the binarization process after using CLAHE from working optimally. The modification of the CLAHE in determining the clip limit value and adaptive grid size could be further researched. The text classification model formed using the SVM method will be used as part of the text detection process.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Universitas Komputer Indonesia for providing financial support for this publication.

REFERENCES

Francis, L.M. & Sreenath, N. 2017. TEDLESS - Text detection using least-square SVM from natural scene. *J. King Saud Univ. - Comput. Inf. Sci.*, 32:287–299.

Hayati, M., Muchtar, K., Roslidar, Maulina, N., Syamsuddin, I., Elwirehardja, G.N., & Pardamean, B. 2022. Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Comput. Sci.*, 216:57–66.

Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L. Chandrasekhar, V.R., Lu, S., & Shafait, F. 2015. ICDAR 2015 competition on robust reading. *13th International Conference on Document Analysis and Recognition (ICDAR)*. France: IEEE, 1156–1160.

Karatzas, D., Shafait, F., Uchida, S., Iwamura, M., Bigorda, L.G.I., Mestre, S.R., Mas, J., Mota, D.F., Almazan, J.A., & De Las Heras, L.P. 2013. ICDAR 2013 robust reading competition. *Proceedings of the International Conference on Document Analysis and Recognition* (ICDAR). Washington, DC: IEEE, 1484–1493.

Lashkov, I., Yuan, R., & Zhang, G. 2023. Edge-Computing-Facilitated Nighttime Vehicle Detection Investigations With CLAHE-Enhanced Images. *IEEE Trans. Intell. Transp. Syst*., 24:13370–13383.

Mahajan, S. & Rani, R. 2021. Text detection and localization in scene images: a broad review, *Artificial Intelligence Review*, 54(6):4317–4377.

Manju, R.A., Koshy, G., & Simon, P. 2019. Improved Method for Enhancing Dark Images based on CLAHE and Morphological Reconstruction. *Procedia Comput. Sci.*, 165:391–398.

Mittal, A., Moorthy, A.K. & Bovik, A.C. 2011. Blind/referenceless image spatial quality evaluator. *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.* California: IEEE, 723–727.

Naiemi, F., Ghods, V. & Khalesi, H. 2020. Scene text detection using enhanced extremal region and convolutional neural network. *Multimed. Tools Appl.* 79:27137–27159.

Paul, K.C. & Aslan, S. 2021. An Improved Real-Time Face Recognition System at Low Resolution Based on Local Binary Pattern Histogram Algorithm and CLAHE. *Opt. Photonics J.*, 11:63–78.

Qiu, Q., Feng, Y., Yin, F., & Liu, C.L. 2018. A flattened maximally stable extremal region method for scene text detection. *Commun. Comput. Inf. Sci*., 757:252–262.

Rainarli, E., Suprapto, & Wahyono. 2021. A decade: Review of scene text detection methods. *Comput. Sci. Rev.*, 42:100434.

Shahab, A., Shafait, F., & Dengel, A. 2011. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. China: IEEE, 1491–1496.

Soni, R., Kumar, B., & Chand, S. 2018. Extracting text regions from scene images using weighted median filter and MSER. *Proceedings International Conference on Advances in Computing, Communication Control and Networking, ICACCCN*. India: IEEE, 915–920.

Tian, C., Xia, Y., Zhang, X., & Gao, X. 2017. Natural scene text detection with MC–MR candidate extraction and coarse-to-fine filtering. *Neurocomputing*, 260:112–122.

Turki, H., Ben Halima, M., & Alimi, A.M. 2018. Text Detection Based on MSER and CNN Features. *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*. Japan: IEEE, 949–954.

Yasmeen, U., Shah, J.H., Khan, M.A., Ansari, G.J., Rehman, S.U., Sharif, M., Kadry, S., & Nam, Y. 2020. Text detection and classification from low quality natural images. *Intell. Autom. Soft Comput*., 26:1251–1266.

Yuan, Z., Zeng, J., Wei, Z., Jin, L., Zhao, S., Liu, X., Zhang, Y., & Zhou, G. 2023. CLAHE-Based Low-Light Image Enhancement for Robust Object Detection in Overhead Power Transmission System. *IEEE Trans. Power Deliv*., 38:2240–2243.

Zhang, J.J. & Zhong, P. 2020. Learning Biased SVM with Weighted Within-Class Scatter for Imbalanced Classification. *Neural Process. Lett*., 51:797–817.

Zhang, X., Gao, X., & Tian, C. 2018. Text detection in natural scene images based on color prior guided MSER. *Neurocomputing*, 307:61–71.

Zheng, Y., Li, Q., Liu, J., Liu, H., Li, G., & Zhang, S. 2017. A cascaded method for text detection in natural scene images. *Neurocomputing*, 238:307–315.