IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 18, 2021, accepted July 15, 2021, date of publication August 9, 2021, date of current version September 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103279

Anomaly Detection in Natural Scene Images Based on Enhanced Fine-Grained Saliency and Fuzzy Logic

HAMAM MOKAYED[1], **PALAIAHNAKOTE SHIVAKUMARA**[2], **RAJKUMAR SAINI**[1], **MARCUS LIWICKI**[1], **(Member, IEEE), LOO CHEE HIN**[3], **AND UMAPADA PAL**[4], **(Senior Member, IEEE)**

[1]Department of Computer Science, Electrical and Space Engineering, Lulea University of Technology, 971 87 Luleå, Sweden
[2]Department of System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia
[3]Department of Computer Science, Asia Pacific University, Kuala Lumpur 57000, Malaysia
[4]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata 700108, India

Corresponding author: Hamam Mokayed (hamam.mokayed@ltu.se)

ABSTRACT This paper proposes a simple yet effective method for anomaly detection in natural scene images improving natural scene text detection and recognition. In the last decade, there has been significant progress towards text detection and recognition in natural scene images. However, in cases where there are logos, company symbols, or other decorative elements for text, existing methods do not perform well. This work considers such misclassified components, which are part of the text as anomalies, and presents a new idea for detecting such anomalies in the text for improving text detection and recognition in natural scene images. The proposed method considers the result of the existing text detection method as input for segmenting characters or components based on saliency map and rough set theory. For each segmented component, the proposed method extracts feature from the saliency map based on density, pixel distribution, and phase congruency to classify text and non-text components by exploring a fuzzy-based classifier. To verify the effectiveness of the method, we have performed experiments on several benchmark datasets of natural scene text detection, namely, MSRATD-500 and SVT. Experimental results show the efficacy of the proposed method over the existing ones for text detection and recognition in these datasets.

INDEX TERMS Natural scene detection, natural scene text recognition, rough set, saliency, fuzzy logic, anomaly text classification.

I. INTRODUCTION

Researchers have focused on scene text detection and recognition using deep learning [1], [2], beating the previous state-of-the-art methods and becoming practically useful. However, when it comes to scene text associated with symbols, logos, or non-text components that share text properties, the performance of such methods degrade [1], [2]. Since these situations are pretty common in real natural scenes, this work focuses on the improvement in these areas. We consider such components as anomalies because these components are unexpected and are not related to the normal text. As such components are introduced to the bounding boxes along with text components, text detection performance degrades, resulting in poor recognition performance.

Furthermore, text recognition degrades if the extracted features clash with those non-text components due to shared attributes. This makes sense because the non-text components' shapes may be similar to the alphabets' shapes. Therefore, separating anomalies from text is challenging and is essential for detecting and recognizing text in images. As a result, non-text components must be distinguished from text components by assessing local information and determining precise bounding boxes. This results in a good performance in text detection and recognition as well.

The main impact of the contribution of this paper is illustrated in Fig. 1. The information of the characters and the affinity among characters are used in Character Region Awareness for Text Detection (CRAFT) [3] to detect the text.

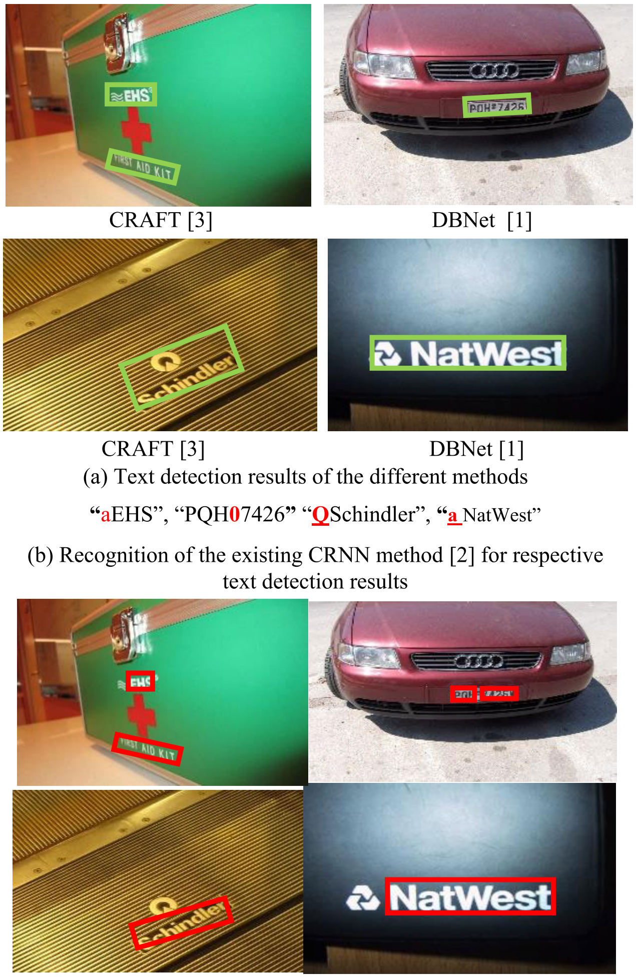
The associate editor coordinating the review of this manuscript and approving it for publication was Senthil Kumar.

Differentiable Binarization Network (DBNet) [1] does binarization (during segmentation) for detecting the text. Both do not find precise bounding boxes in the two different scenes, as depicted in Fig. 1(a). It can be noticed that decorative non-text components and the symbols(both are called anomalies in this work) cause poor text detection. CRNN [2] does not recognize the text correctly from such detected bounding boxes. In the example of Fig. 1(b), characters in red color indicate improper recognition. With this scenario, it is logical to argue that current text detection and identification systems have a fault in detecting and recognizing accurate text. As shown in Fig. 1(c), the bounding boxes can be precisely detected and get good recognition performance. This is the motivation for proposing a new method for anomaly detection from the result of text detection. It is based on text and non-text component classification, as demonstrated in example findings in Fig. 1(c). The anomalies were successfully removed in all four cases, and the recognition results were corrected using the same procedure. As a result, this research aims to provide a new method for detecting anomalies in text detection results.

II. RELATED WORK

Deep learning models are used in most of the methods to achieve the desired results. A text detection in natural scene images using character region awareness was proposed by Baek *et al.* [3]. To detect the text in the images, the method extracts information at the character level as well as the relationships between the characters. Similarly, Wang *et al.* [4] proposed a text detection method based on progressive scale expansion network. A segmentation-based detector with multiple predictions for each text instance in the images is used in the text detection process. Liao *et al.* [1] proposed a differentiable binarization network-based method for text detection in natural scene images. Several thresholding techniques for binarization are used in the methods, which result in the segmentation of text from the input images. Wang *et al.* [5] proposed a method based on two-stage network architectures for scene text in the wild. Quadrilateral regression algorithms are proposed for generating quadrilaterals in this method. The method's performance is improved by pooling weighted ROI data.

Zhu and Du [6] proposed an instance segmentation-based method for scene text detection. The method focuses on separating text from non-text regions by determining the text center direction. It entails determining the relationship between the text border and the text center in order to detect text in images. A multi-scale context-aware features aggregation-based method for text detection in natural scene photos was proposed by Dai *et al.* [7]. Based on text-related features, the method creates an enhancement module. An arbitrarily shaped scene text detection using a mask tightness text detector was proposed by Liu *et al.* [8]. It predicts pixel-wise masks based on polygonal boundary information. It then achieves mutual promotion by incorporating a branch for each text region's polygonal boundary.



(a) Text detection results of the different methods

“aEHS”, “PQH07426” “QSchindler”, “a NatWest”

(b) Recognition of the existing CRNN method [2] for respective text detection results

(c) Removing anomalies added to text region from different position (top, left, right, and center).

“EHS”, “PQH7426”, “Schindler“, “NatWest”

(d) Recognition of the existing CRNN method [2] for corrected text detection results

FIGURE 1. Illustrating the need for defect detection to improve text recognition and text detection performances.

In summary, the methods described above recognize the natural scene text by employing several deep learning models. However, isolated characters, symbols, and non-text components with the same appearance as text characters add difficulties in detection and recognition. Context information (high-level features) are necessary in most deep learning models in order to achieve better results. However, when the scene text contains single characters and non-text components, the algorithms lose context information, resulting in poor performance.

For enhancing text detection performance, some approaches combine feature extraction and deep learning. Roy *et al.* [9] suggested an approach for recognizing text from multiview natural sceneries based on Delaunay triangulation. Nag *et al.* [10] proposed detecting text in sports photographs by combining features and deep learning. The method's primary focus is to find text in the marathon and sports video images. The approach can recognize isolated characters, but it is limited to images with humans in them.

Similarly, an arbitrarily oriented text detection in low light natural scene photos was proposed by Xue *et al.* [11]. The approach detects text candidates using MSER and the cloud of line distribution principle and then employs CNN for text detection. According to the review of available approaches, detecting and recognizing isolated character and non-text components that look as text is neglected.

There are various powerful approaches for recognition that explore deep learning models, like text detection approaches. Shi *et al.* [2] suggested a scene text recognition with an end-to-end deep network. To achieve a high recognition rate, the method includes feature extraction and sequence modeling. For online handwriting recognition, Carbune *et al.* [12] proposed using an LSTM deep network. The approach is limited to handwriting recognition rather than scene text recognition. The method's model, on the other hand, can detect scene text. A multi-branch guided attention network to recognize scene text was proposed by Wang and Liu [13]. The method is based on the mutual guidance mechanism notion. Zhang *et al.* [14] suggested a scene text recognition with a scale-aware hierarchical attention network. The method is based on the pyramidal structure of deep convolutional neural networks, which aids in extracting flexible receptive fields and abundant spatial semantic features.

Lee *et al.* [15] proposed a 2D self-attention network-based technique for scene text recognition. The method extracts the dependencies between word tokens in a sentence. This helps to extract 2D spatial dependencies between two characters in a scene text image. Long *et al.* [16] presented a character anchor pooling approach for scene text recognition. With this step, the method gathers more vital information for recognizing text in the images. Shang *et al.* [17] suggested a character awareness network-based technique for scene text recognition. A 2D character attention is used in the model, which boosts foreground text instances based on character awareness.

However, while the methods handle many of the issues of scene text identification, they produce poor results when text loses context information, particularly when text comprises language that looks like symbols. As a result, we may conclude that text detection and recognition algorithms are still not optimal for achieving improved results in many situations.

As a result, this work provides a new way to overcome the limitations of existing methods, resulting in improved text detection and identification performance. Motivated by local pixel information rather than the shape of text components, we explore pixel density and distribution-based features for detecting anomaly components in the text. Since the shapes of the anomaly component share with the text components, sometimes the extracted features may lose discriminative power and results in uncertainty. To remove uncertainty and strengthen the feature extraction, a fuzzy-based classifier for the classification of anomalies in the text has been considered here. Before extracting features, the proposed method uses a saliency map to enhance the fine details in the image and explore rough set theory for the classification of foreground and background information of the images. This results in component segmentation from text detection results. The way the proposed method combines saliency, rough set theory, and fuzzy concept for successful classification of anomaly components is novel and this is the main contribution of this work.

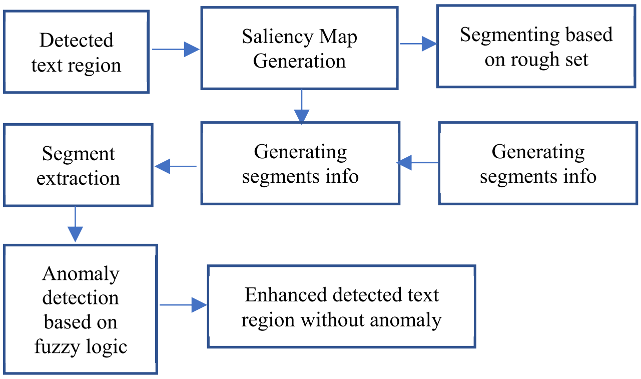

Detected text region
Saliency Map Generation
Segmenting based on rough set
Segment extraction
Generating segments info
Generating segments info
Anomaly detection based on fuzzy logic
Enhanced detected text region without anomaly

FIGURE 2. **Proposed Work for anomaly detection.**

III. PROPOSED METHOD

This research aims to enhance the detected text by removing all the noise attached to it, this step will play major role in enhancing the accuracy of the text recognition. The proposed method consists of two stages, namely (i) segmenting components, which include anomaly component also from the text, and (ii) classifying anomaly component from the text components. The anomaly components may confuse the text due to the similar shape. The local information makes difference there. Motivated by this observation, we propose to obtain saliency for the text, which enhances the fine details in the images [18]–[20]. When the fine details of edges are enhanced, the proposed method uses a rough set approach as the knowledge discovery using granular information [18] for separating foreground (character edges) and background. This results in a binary image of the input image. The use of a rough set is to deal with vague, imprecise, inconsistent, and uncertain knowledge introduced by the saliency map. The binary results are used to segment text components.

It is observed that the intensity distribution in the generated salient map and angular information of pixels make difference between text and anomaly component. With this notion, we propose to use phase congruency [21] for extracting angular information, pixel density, and distribution for classifying anomaly components from each segmented component. The extracted features are fed fuzzy rule-based classifier [22] for anomaly component detection. The block diagram of the proposed method can be seen in Fig. 2, where it considers the result of existing text detection method [1]–[3] as input for anomaly text detection in this work.

A. SEGMENTATION STAGE

a pyramid of different 8 grey-levels images is calculated with ratio of one-quarter of the previous one. As a later stage, the differences among centers are calculated as an across-scale difference between coarse and fine scales. The value of s from 2-4 are used to define fine scales, while the values of s from 5-8 are used to define coarse scales. An across-scale difference is calculated by scaling the coarse-scale into the fine-scale and then executing pixel by pixel subtraction. The system calculates on-center and off-center differences in the three images that shows the scales from 2-4. Centers are represented as a pixel and two surrounding values (σ) are used: 3 and 7, based on the work of [19]. Therefore, 12 intensity submaps are generated. The process of calculating these submaps is as follows: at first center and surround are defined and then every pixel of each intensity submap is calculated as defined in Equation (1).

$$int_{s,\sigma}(x, y) = max\{center(x, y, s) - surround(x, y, s, \sigma), 0\} \tag{1}$$

where $s \in \{2, 3, 4\}$ represents the image scale, $\sigma \in \{3, 4\}$ the surround.

Next, an on-center intensity map is calculated. This is obtained by scaling the six on-center intensity submaps into the largest scale, and then summing pixel by pixel as defined in Equation (2). The above process outputs saliency for the input images as shown in Fig. 3, where we can see text lines with anomaly components and respective saliency maps. It is observed from Fig. 3(b) that the fine details like edges are enhanced compared to its background.

$$Intensity = sum(int_{s,\sigma}) \tag{2}$$

The saliency maps are subjected to a rough set approach for approximation of sets using granular information [22]. This approach provides structures for the overlapping boundary in given domain knowledge. If the boundary region of the set Y is empty then it is a crisp set, otherwise, if the boundary region is non-empty then it is rough set. Given a saliency image, the process for granularization (g) is dividing the full resolution of the image window into $g = 4$, several 4×4 resolution sub-windows. The sub-window is the granules of knowledge where the pixel value classifies as foreground or background. The uncertainty of a rough set is measured by the roughness. Let $p(B)$ and $p(O)$ is the representation of two properties, for gray level intervals $0, 1, \ldots, T$ and $T + 1, T + 2, \ldots, L - 1$ that characterize background and object regions respectively. As there is object and background in the provided salient map, the rough set representation will be applied on the two sets as defined in Equation (3) to Equation (8).

Let inner approximation of the object be ($\underline{O}_T$)

$$\begin{aligned} \underline{O}_T &= Ui\, G_i | P_j > T, \\ &\qquad \forall j = 1, \ldots, mn \; and \; P_j \; is \; a \; pixel \; belonging \; to \; G_i \end{aligned} \tag{3}$$

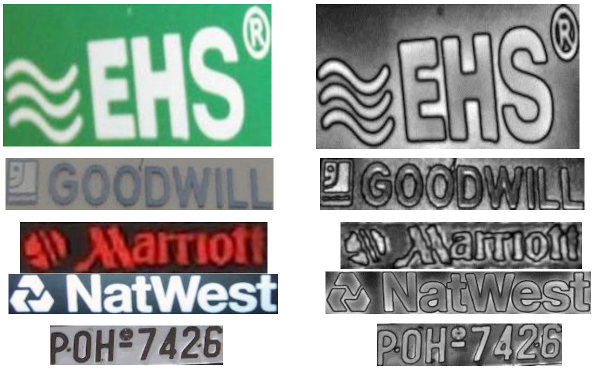


FIGURE 3. Saliency map of the detected text region.

Outer approximation of the object (O_T):

$$\begin{aligned} O_T &= Ui\, G_i, \exists j \\ &\qquad j = 1, \ldots, mns.t.P_j > T \; where \; P_j \; is \; a \; pixel \; in \; G_i \end{aligned} \tag{4}$$

Inner approximation of the background ($\underline{B}_T$):

$$\begin{aligned} \underline{B}_T &= Ui\, G_i | P_j \leq T \\ &\qquad \forall j = 1, \ldots, mn \; and \; P_j \; is \; a \; pixel \; belonging \; to \; G_i \end{aligned} \tag{5}$$

Outer approximation of the object (B_T):

$$\begin{aligned} B_T &= Ui\, G_i, \exists j \\ &\qquad j = 1, \ldots, mns.t.P_j \leq T \; where \; P_j \; is \; a \; pixel \; in \; G_i \end{aligned} \tag{6}$$

Therefore, the rough set representation of the image, for instance, object O_T and background B_T for a given $I_{m\times n}$ and it is depending on the value of T. Let the roughness of object O_T and background B_T as the definition $R_{OT} = 1 - |\underline{O}_T| / |O_T|$ as object roughness and $R_{BT} = 1 - |\underline{B}_T| / |B_T|$ as background roughness, where $|\underline{S}_T|$ and $|S_T|$ are cardinality of lower and upper approximation of set S as refer to the object and background, respectively. The rough entropy will be calculated to find the best threshold for the two-classification problem

$$RE_T = -e/2[R_{OT} log_e(R_{OT}) + R_{BT} log_e(R_{BT})] \tag{7}$$

$$T^* = arg_{max_T} RE_T \tag{8}$$

The output of rough set theory can be seen in Fig. 4, where one can see for the saliency maps in Fig. 4(a), binary results given by the rough set as shown in Fig. 4(b). It is observed from Fig. 4(b) that there is a clear spacing between the components including anomaly components. Therefore, the proposed approach uses two-pass connected component labeling algorithm for segmenting each component from the binary results as defined in Equation (9) to Equation (11). Let Ω be the spatial space of the image. The connectivity of two points $P, Q \in \Omega$ is represented in the following terms:

$$con(P, Q) = \begin{cases} 1, & if \; P, Q \; connected \\ 0, & otherwise \end{cases} \tag{9}$$

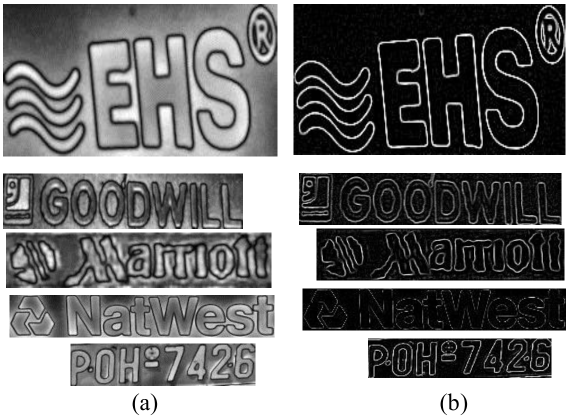


FIGURE 4. Binary image generated by applying rough set theory on the Saliency map. (a) Saliency Map. (b) Generated binary image based on classifying the pixels using rough set theory.

The scanning-masks for 4/8 connected labeling are calculated based on the following:

$$M4(P) = \{P, P_{upper}, P_{left}\},$$
$$M8(P) = \{P, P_{upper}, P_{left}, P_{upper} - P_{left}, P_{upper} - P_{right}\} \quad (10)$$

Note that if P is at the edge of an image, some points P in the upper definition do not exist. In this case $con\,(P, P)$ is set to 0 by definition. Let M be $M4$ or $M8$, then,

$$con_m\,(P) = \{Q \in M(P) | con\,(P, Q) = 1\} \subset \Omega \qquad (11)$$

B. ANOMALY DETECTION

As discussed earlier, the pixel values in saliency map, distribution, and angular information are extracted from each segmented component. It can be seen from Fig. 5 that there is a clear difference in pixels distribution (probability) for text and anomaly (non-text) components. This shows that though the shape of the anomaly component appears as a text component, the pixel distribution make difference for classification.

To classify anomaly components from text components, the proposed method uses fuzzy logic as a classifier by feeding extracted features as input. The feature values of the image undergo fuzzification into fuzzy sets. The prediction of the class label is done based on the features of each segmented component. The prerequisite of the classifier is a training data set which will be used to train the fuzzy classifier to predict class labels as defined in Equation (12).

$$g_k\,(x) = \left(\sum\nolimits_i \beta_{k,i}\tau_i(x)\right) / \left(\sum\nolimits_i \tau_i(x)\right) \qquad (12)$$

The increasing membership function utilized to distinguish the intensity of saliency in the image in two classes foreground as text, while logo and background as non-text as shown in Fig. 6, where the membership function defined for text and noise parts.

IV. EXPERIMENTAL RESULTS

There is no standard dataset available in the literature to assess the proposed anomaly detection. As a result, number of natural scene images extracted from MSRA-TD-500 [9] and SVT [11] datasets are used for evaluation. We choose the MSRA dataset since it was designed for arbitrary text detection, this dataset has images consist of text with anomaly attached to it. The reason of having this kind of images is that finding a text that is arbitrarily inclined is more complicated than determining the textual region for horizontal text. Thus, more opportunities to improve improper bounding boxes are existed, which include noises and many other factors. on the same page, the second chosen dataset contains street view photos with information such as building names, street names, trees, and so on. As a result, text identification in such photos is difficult. Aside from the complicated background, the photographs are acquired from an oblique perspective, which affects the quality of the images. Therefore, there is an excellent opportunity to correct the text lines' erroneous bounding boxes. Figure 7 shows sample photos from the MSRA and SVT datasets. The text lines are coupled with special symbols and logos. In this study, forty-four photos out of five hundred from the first dataset and thirty-nine photos out of three hundred fifty from the second dataset were chosen for experimentation. Despite the limited dataset size, the photos are sophisticated enough to test the proposed anomaly finding technique.

We implemented the following text finding and recognition methods to show that the proposed anomaly component detection is effective. Character Region Awareness for Text Detection was proposed by Baek *et al.* [3]. (CRAFT).

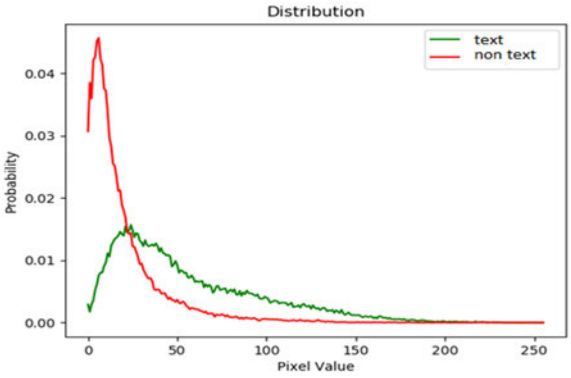


FIGURE 5. Density distribution information.

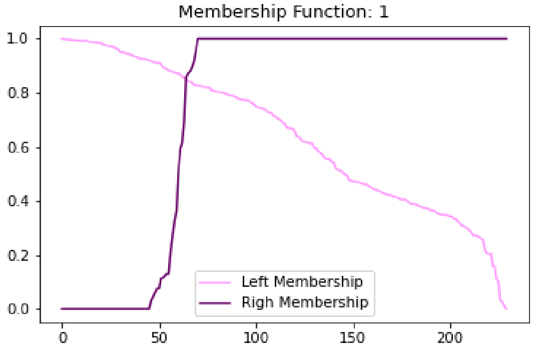


FIGURE 6. Membership function.

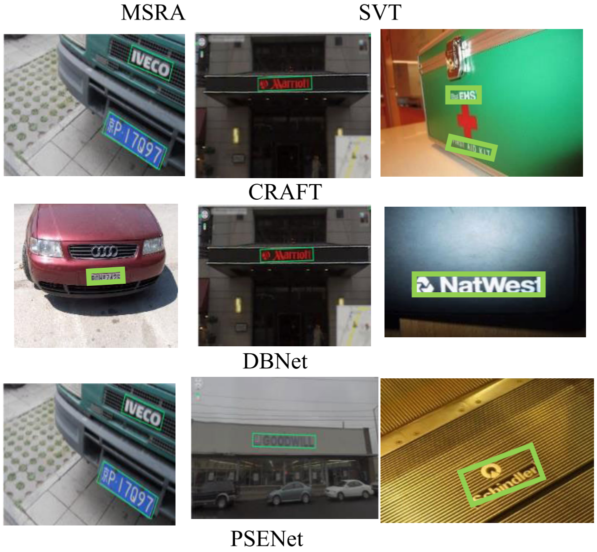


FIGURE 7. Text detection results of the different methods on MSRA and SVT datasets before anomaly detection.

Different techniques proposed for text detection will be used in the evaluation and testing cycle, Differential Binarization Network (DBNet) is one of them [1]. Progressive Scale Expansion Network (PSENet) method that used for word detection will be also tested [4]. End-to-end trainable neural network for image-based sequence recognition and applied it to scene text for recognition is nominated as one of the selected works [2]. It recognizes text in natural scene images using a Convolutional Recurrent Neural Network (CRNN). Carbune *et al.* [12] propose a handwriting recognition system based on LSTM. We utilize LSTM [12] to recognize scene text in the images in this study since it is good at managing complex scenarios and can adapt to varied datasets and applications. The above approaches were chosen to demonstrate the efficiency of the suggested anomaly detection since they are state-of-the-art and can handle complex scenarios. Furthermore, while deep learning-based methods effectively solve complex problems, they fall short in some typical cases. We use the standard measures of recall (R), precision (P), and F-measure ($F1$) to evaluate the performance of the proposed anomaly component detection. Text detection experiments employ the same measures. To demonstrate the efficiency of the proposed, accuracy of text detection measured before and after anomaly component finding. Various text detection methods are used to show that the proposed anomaly component detection is adequate. In experiments, the text detection result without anomaly components is fed into the same text detection methods as the text detection result with anomaly components. After anomaly component detection, it is expected that text detection performance will improve significantly. We use the word-level recognition rate in our recognition experiments. Detected texts before and after anomaly removal are sent to various recognition methods to verify the increase in the text recognition rate. In this paper, we use the instructions from [9], [11] to measure the performance of both recognition and detection.

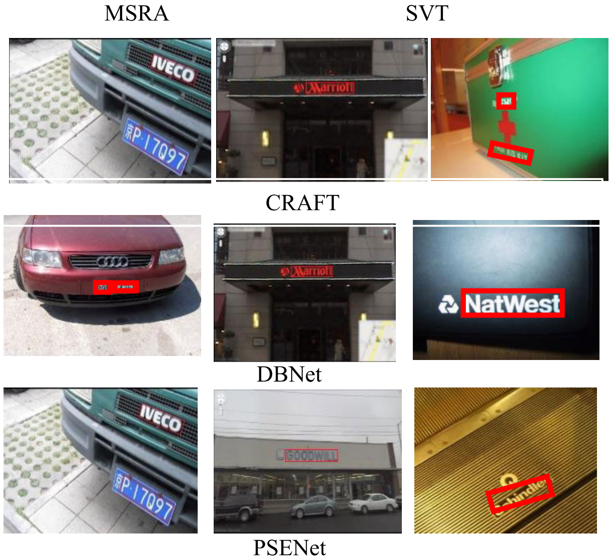


FIGURE 8. Text detection results of the different methods (CRAFT, DBNet and PSENet) on MSRA-TD-500 and SVT datasets after anomaly detection.

A. EVALUATING ANOMALY DETECTION

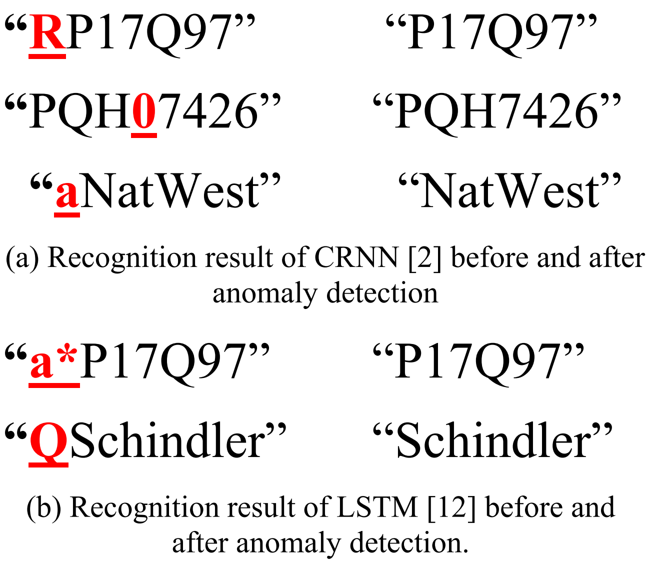
Table 1 shows the quantitative results of the proposed anomaly component detection for the MSRA-TD-500 and SVT datasets. We calculate measures for evaluating the anomaly component detection step based on the output of each text detection method. The proposed method performs reasonably well for the MSRA-TD-500 and SVT datasets, as shown in Table 1.

TABLE 1. Enhancement of text detection using different known techniques by adding the anomaly removal stage.

Techniques	SVT			MSRA		
Measures	Recall	Precision	$F1$	Recall	Precision	$F1$
CRAFT [3]	83.2	90.8	86,8	82.4	87.6	84,9
DBNet [1]	88.1	91.8	89,9	89.1	92.6	90,8
PSENet [4]	84.6	89.5	87,0	87.6	89.8	88,7

B. VALIDATING THE EFFECTIVNESS OF ANOMALY DETECTION

Figures 8 and 9 show the qualitative results of used text detection and recognition methods before and after anomaly detection. Text detection methods, for example, fix proper bounding boxes after removing anomaly components from the text detection results, as shown in Fig. 8. Similarly, the methods correctly recognize text after anomaly detection, as shown in Fig. 9. Fig. 9(a) and (b) show the recognition results before and after anomaly detection, respectively. This demonstrates that the proposed anomaly detection aids in text detection and recognition. Tables 2 and 3 show the quantitative results of text detection and recognition methods before and after anomaly detection, respectively. The performance of the text detection and recognition methods improves significantly after anomaly detection for both datasets showing the effectiveness of the proposed method.



(a) Recognition result of CRNN [2] before and after anomaly detection

(b) Recognition result of LSTM [12] before and after anomaly detection.

FIGURE 9. The effect of anomaly detection on recognition performance.

TABLE 2. Accuracy calculation for text detection with/without proposed anomalies detection stage.

Techniques	MSRA						SVT					
	without			with			without			with		
	P	R	F1	P	R	F1	P	R	F	P	R	F1
CRAFT [3]	88.2	78.2	82.9	89.8	81.5	85.4	73.1	87.2	79.5	75.8	89.6	82.1
DBNet [1]	85.9	52.0	64.5	89.0	58.2	70.4	69.8	54.0	60.8	73.7	63.2	68.0
PSENet [4]	91.5	79.2	84.9	92.5	81.6	86.7	72.5	62.2	67.0	75.8	68.2	71,8

TABLE 3. Accuracy calculation for text recognition with/without proposed anomalies detection stage.

Techniques	MSRA				SVT			
	without		with		without		with	
	LSTM	CRNN	LSTM	CRNN	LSTM	CRNN	LSTM	CRNN
CRAFT [3]	66.9	73.2	69.8	75.4	76.2	82.7	78.8	84.9
DBNet [1]	43.1	46.5	47.7	52.4	44.5	47.2	50.2	55.5
PSENet [4]	69.5	74.6	72.1	77.9	52.4	55.6	57.9	63.0

V. CONCLUSION AND FUTURE WORK

This paper proposes a new method for detecting anomalies in text detection results generated by text detection methods. The removal of these anomalies aids text detection and recognition methods in improving their performance. The proposed method uses a saliency map and rough set theory to segment characters and anomaly components. The proposed method extracts feature from the saliency map of segmented components based on pixel distribution and phase congruency for each segmented component. The features are then fed into a fuzzy logic classifier for anomaly and text component classification. The proposed anomaly detection performs well on the MSRA-TD-500 and SVT datasets, according to experimental results. Furthermore, investigations on two benchmark datasets before and after anomaly detection reveal that text detection and recognition algorithms perform much better after anomaly detection showing the effectiveness of the proposed method. The proposed method's performance may be affect-ed by excessive noise, blur, and other aberrations. The next goal is to develop a reliable system for detecting anomalies in noisy and blurred environments.

REFERENCES

[1] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differen-tiable binarization," in *Proc. AAAI*, 2020, pp. 1–8.

[2] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[3] Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9365–9374.

[4] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, and S. Shao, "Shape robust text detection with progressive scale expansion network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9336–9345.

[5] S. Wang, Y. Liu, Z. He, Y. Wang, and Z. Tang, "A quadrilateral scene text detector with two-stage network architecture," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107230.

[6] Y. Zhu and J. Du, "TextMountain: Accurate scene text detection via instance segmentation," *Pattern Recognit.*, vol. 110, Feb. 2021, Art. no. 107336.

[7] P. Dai, H. Zhang, and X. Cao, "Deep multi-scale context aware feature aggregation for curved scene text detection," *IEEE Trans. Multimedia*, vol. 22, no. 8, pp. 1969–1984, Aug. 2020.

[8] Y. Liu, L. Jin, and C. Fang, "Arbitrarily shaped scene text detection with a mask tightness text detector," *IEEE Trans. Image Process.*, vol. 29, pp. 2918–2930, 2020.

[9] S. Roy, P. Shivakumara, U. Pal, T. Lu, and G. H. Kumar, "Delaunay triangulation based text detection from multi-view images of natural scene," *Pattern Recognit. Lett.*, vol. 129, pp. 92–100, Jan. 2020.

[10] S. Nag, P. Shivakumara, U. Pal, T. Lu, and M. Blumenstein, "A new unified method for detecting text from Marathon runners and sports players in video (PR-D-19-01078R2)," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107476.

[11] M. Xue, P. Shivakumara, C. Zhang, Y. Xiao, T. Lu, U. Pal, D. Lopresti, and Z. Yang, "Arbitrarily-oriented text detection in low light natural scene images," *IEEE Trans. Multimedia*, early access, Aug. 7, 2020, doi: 10.1109/TMM.2020.3015037.

[12] V. Carbune, P. Gonnet, T. Deselaers, H. A. Roweley, A. Daryin, M. Calvo, L. L. Wang, D. Keysers, S. Feuz, and P. Gervais, "Fast multi-language LSTM-based online handwriting recognition," *Int. J. Document Anal. Recognit.*, vol. 23, pp. 83–102, Jun. 2020.

[13] C. Wang and C.-L. Liu, "Multi-branch guided attention network for irregular text recognition," *Neurocomputing*, vol. 425, pp. 278–289, Feb. 2021.

[14] J. Zhang, C. Luo, L. Jin, T. Wang, Z. Li, and W. Zhou, "SaHAN: Scale-aware hierarchical attention network for scene text recognition," *Pattern Recognit. Lett.*, vol. 136, pp. 205–211, Aug. 2020.

[15] J. Lee, S. Park, J. Baek, S. J. Oh, S. Kim, and H. Lee, "On recognizing texts of arbitrary shapes with 2D self-attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 2326–2335.

[16] S. Long, Y. Guan, K. Bian, and C. Yao, "A new perspective for flexible feature gathering in scene text recognition via character anchor pooling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2458–2462.

[17] M. Shang, J. Gao, and J. Sun, "Character region awareness network for scene text recognition," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2020, pp. 1–6.

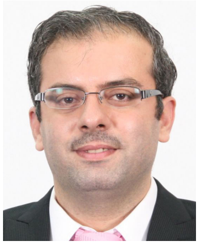
[18] C. F. Flores, A. Gonzalez-Garcia, J. van de Weijer, and B. Raducanu, "Saliency for fine-grained object recognition in domains with scarce training data," *Pattern Recognit.*, vol. 94, pp. 62–73, Oct. 2019.

[19] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[20] S. Montabone and A. Soto, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," *Image Vis. Comput.*, vol. 28, no. 3, pp. 391–402, Mar. 2010.

[21] A. Verikas, A. Gelzinis, M. Bacauskiene, I. Olenina, S. Olenin, and E. Vaiciukynas, "Phase congruency-based detection of circular objects applied to analysis of phytoplankton images," *Pattern Recognit.*, vol. 45, no. 4, pp. 1659–1670, Apr. 2012.

[22] X. Liu, W. Pedrycz, T. Chai, and M. Song, "The development of fuzzy rough sets with the use of structures and algebras of axiomatic fuzzy sets," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 3, pp. 443–462, Mar. 2009.



HAMAM MOKAYED received the Ph.D. degree from UITM, Malaysia, in 2015. He is currently working as a Postdoctoral Researcher with Prof. Marcus Liwicki at the EISLAB Machine Learning, Luleå tekniska universitet (Luleå University of Technology), Sweden. He was previously working as a Senior Staff Researcher at MIMOS, Kuala Lumpur, Malaysia. His research interests include machine learning, pattern recognition, and artificial intelligence with expertise in document analysis, vehicle intelligent systems, and natural scene images. His research interests include machine learning, image processing, and system design in real time environment.

PALAIAHNAKOTE SHIVAKUMARA received the B.Sc., M.Sc., and M.Sc. degrees in technology and the Ph.D. degree in computer science from the University of Mysore, Karnataka, India, in 1995, 1999, 2001, and 2005, respectively. He is currently an Associate Professor with the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. Previously, he was with the Department of Computer Science, School of Computing, National University of Singapore, from 2008 to 2013, as a Research Fellow on video text extraction and recognition project. He has published more than 200 papers in conferences and journals. His research interests include image processing and video text analysis. He had received the prestigious award ''Dynamic Indian of the Millennium'' from KG Foundation, India, for his research contribution to computer science field. He has been serving as an Associate Editor for *ACM Transactions Asian and Low-Resource Language Information Processing* (TALLIP), *CAAI-Transactions on Intelligence in Technology*, *Springer Nature Computer Science* (SNCS), and *Pattern Recognition* journal.

RAJKUMAR SAINI received the Ph.D. degree from the Department of Computer Science and Engineering, IIT Roorkee, Roorkee, India. He is currently a Postdoctoral Researcher with the EISLAB Machine Learning, Luleå University of Technology, Sweden. His research interests include computer vision, machine learning, pattern recognition, human–computer interface, brain signal analysis, and digital image processing.

MARCUS LIWICKI (Member, IEEE) received the M.S. degree in computer science from the Free University of Berlin, Germany, in 2004, and the Ph.D. degree from the University of Bern, Switzerland, in 2007. He worked as a Senior Researcher and a Lecturer with the German Research Center for Artificial Intelligence (DFKI). He is currently a Professor and the Head of machine learning subject with LTU University. His research interests include knowledge management, semantic desktop, electronic pen-input devices, online and offline handwriting recognition, and document analysis. He is a coauthor of the book *Recognition of Whiteboard Notes Online, Offline, and Combination* (World Scientific, 2008). He has 20 publications, including five journal articles. He is a member of the International Association for Pattern Recognition (IAPR). He is a Program Committee Member of the IEEE ISM Workshop on Multimedia Technologies for E-learning. He is a Frequent Reviewer of international journals, including the IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Audio, Speech, and Language Processing, and *Pattern Recognition* and a reviewer of several IAPR conferences and the IGS conference. (Based on document published on 30 May 2008)

LOO CHEE HIN received the B.S. degree from Asia Pacific University, Malaysia, in 2021, where he is currently pursuing the master's degree in artificial intelligence. His research interests include machine learning, image processing, image classification, vehicle intelligent systems, and natural scene images.

UMAPADA PAL (Senior Member, IEEE) received the Ph.D. degree from Indian Statistical Institute. He did his Postdoctoral at INRIA (Institut National de Recherché en Informatiqueeten Automatique), France. Since January 1997, he has been a Faculty Member of the Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, where he is currently a Professor and the Head. His fields of research interests include different pattern recognition problems, like digital document processing, optical character recognition, camera/video text processing, biometrics, document retrieval, and keyword spotting. He has published more than 390 research articles in various international journals, conference proceedings, and edited volumes. Because of his significant impact in the document analysis research domain of Indian language, TC-10, and TC-11 committees of International Association for Pattern Recognition (IAPR), where he received the ICDAR Outstanding Young Researcher Award in 2003. He is the Co-Editor-in-Chief of the *Springer Nature Computer Science Journal*. He is also the Editorial Board Member of several journals, like *PR*, *IJDAR*, *PRL*, *ACM Transactions on Asian Language Information Processing*, and *IET-Biometrics*. He is a fellow of IAPR.