Grenze International Journal of Engineering and Technology, Jan Issue

Curved Text Detection Methods for Scenic Images : A Review

Prachi chhabra[1] and Dr Ali Imam Abidi[2]
[1]Jss Academy of Technical Education, Noida /Information Technology, Noida, India
Email: Prachichhabra@jssaten.ac.in
[2-3]Sharda University, Greater Noida /CSE Greater Noida, India
Email: ali.abidi@sharda.ac.in

***Abstract*—The objective of this paper is to present a comprehensive exploration of deep learning based curved text detection methods. The main goal of text detection is to determine the position of text from input natural scenic image The position of the text area in the image is obtained using text detection and localization techniques. In this current era of deep learning, computer vision applications have evolved and taken new routes, such as smartphone translation technologies, supportive ways for the visually impaired, and so on. This paper is aimed at summarizing and analyzing the major changes and significant progresses of curved scene text detection and recognition in the deep learning era. This paper tends to highlight the Challenges arises in capturing and recognizing curved text from natural images. To identify publicly released scene image data sets as well as appropriate standard evaluation methods for scenic images.**

***Index Terms*— Deep learning, Scenic Images, Challenges, Curved text Detection methods, public curved text Dataset.**

I. INTRODUCTION

Text in natural settings can be found in almost every stage of life. From the facade of the buildings in our neighbourhood to the cover of a library book. Writing is undoubtedly one of mankind's most creative and powerful creations. In view of a variety of patterns, cluttered backgrounds, and text orientations, text detection in images collected in an unconstrained environment is significant but challenging computer vision task. This is a problem that both machine learning and deep learning techniques face. It has been found that Language difficulties, bounding box sizes, and text recognition in multidirectional and curved text images are major challenges which need to be addressed. Based on different research strategies and different existing deep learning approaches in specific ways, its performance when identifying curved characters (e.g., horizontal bounding boxes, rotated rectangles, or squares) decreases drastically. Therefore, the detection of curved text, which is quite common in natural situations, is of enormous importance. Deep learning based methods are sorted into two categories region proposal based methods and segmentation based methods [1] .

A. Region Proposal Based Methods

A region proposal for scene text identification is a convolution-based text detection approach that works from end to end. The proposal-based technique is primarily motivated by object detection frameworks in general. This

Grenze ID: 01.GIJET.9.1.715

approach targets the full-text region first, then applies horizontal/quadrilateral boxes, and then performs bounding box regression to provide correct coordinate offsets of anticipated text boxes. It can enable us to focus the component and then use CNN to detect and label items.

Advantages

1. For scene text detection in clutters, region proposal techniques are particularly efficient.
2. Multi-scale rectangular windows are concatenated over text appearances for more accurate text detection, and all of the windows are regressive with a high overlap fraction with text areas.
3 This approach is most commonly used in the wild with horizontal and multi-oriented aligned texts.
4. Some of the examples are RCNN, Faster -RCNN, SSD and YOLO.

Disadvantages

1. The proposal-based methods are mostly limited to simple texts with linear shapes.
2. It fails to detect curved texts accurately in most cases.

B. Segmentation Based Methods

Segment-based regression focuses on a portion of the texts, and for final identification, a merging approach is non-suppression. This approach targets the full-text region first, then applies horizontal/quadrilateral boxes, and then performs bounding box regression to provide correct coordinate offsets of anticipated text boxes.
subsequently, a non-suppression merging technique is applied for final detection

Advantages

1.In comparison to proposal-based strategies for irregular text identification, these methods are more flexible.
2. They can recognize text sections in a complicated environment with ease.
3. This approach aims to locate a tiny portion of text occurrences, which necessitates a limited receptive field of CNN baselines.
4. Some of the examples are YOLO, EAST, and Single shot multi box detector.

These methods are more flexible compared to proposal-based techniques for irregular text detection. They can effectively detect text regions in complex environment. These methods attempt to localize small part of text instances that requires small receptive field of CNN baselines.
These methods are more flexible compared to proposal-based techniques for irregular text detection. They can effectively detect text regions in complex environment. These methods attempt to localize small part of text instances that requires small receptive field of CNN baselines.
These methods are more flexible compared to proposal-based techniques for irregular text detection. They can effectively detect text regions in complex environment. These methods attempt to localize small part of text instances that requires small receptive field of CNN baselines.

Disadvantages

1. In the domain of text form, the linkage technique leads to more complicated algorithms and is more adaptable.
2. This strategy focuses on curved and multi-oriented texts in natural images.

II. Various Curved Text Detection Methods

In this section various models by different authors are covered for curved text detection methods Most of the scene text detectors are applicable for horizontal or multi-oriented text detection. It has been observed that performances of conventional approaches drop significantly in detecting curved texts in the wild Different models have different methodologies, bounding boxes with their performances. Some of the models and work done by the various researchers in this area are discussed from 2018-2022. Few of them are based on proposal-based methods and latter is based on segmentation-based methods

A. Proposal-Based Methods

Liao et al. (2018a) [2] implemented an end-to-end trainable fast scene text detector, named TextBoxes++, which detects arbitrary-oriented scene text with both high accuracy and efficiency in a single network forward pass. No post-processing other than efficient non-maximum suppression is involved. To overcomes the shortcoming of TextBoxes by introducing quadrilateral boxes instead of horizontally aligned boxes to detect linearly aligned scene.

Zhu et al. (2019) [3] have implemented a novel text detection algorithm for curved texts using bounding box regression method. In which candidate text proposals are generated using RPN and bounding box regression of text proposals are performed gradually in a two-stage manner, which produces more accurate text bounding boxes. The shortcoming of this method is that it can't achieve real-time detection compared with one-stage detector.

Liu et al. (2019 a) [4] designed a novel polygon-based curved text detector (CTD) for curve text detection. It represented a one-of-a-kind curved text detector based on polygons (CTD). It comprises of the regression module, the backbone, and the region proposal network (RPN) are the three parts. This approach is designed as a universal method, meaning it can be trained using rectangular angular or quadrilateral bounding boxes, requiring no extra effort. The curved text in this dataset is tightly labeled using polygons, which does not require significant manpower. Additionally, we proposed a novel CTD approach that may be the first attempt to directly detect curved text. An shortcoming of CTD is that it is slightly slower than a rigid rectangle-based detector.

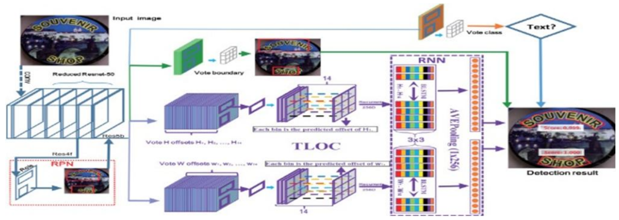

Figure 1: Overall structure of our Curve Text Detector (CTD) [1]

Xue et al. (2019)[5] proposed A multi-scale shape regression-based framework is designed by where central point of each candidate text proposal is identified using a triangulation method within polygon region. Then, distances from nearest boundary pixel to central pixel in both x and y direction are computed to produce a dense boundary of arbitrary shaped text instances. This method is claimed to be superior to general quadrilateral box regression methods that generally include additional background region. This method predicts dense text boundary points instead of sparse quadrilateral vertices that are prone to produce large regression errors while dealing with long text lines.

Chuang Yang et al.**2021[6]** proposed a Bidirectional Perspective strategy-based Network (BiP-Net) a new text representation strategy is proposed to represent text contours from a top-down perspective, which can fit highly curved text contours effectively. Moreover, a contour connecting (CC) algorithm is proposed to avoid the information loss of text contours by rebuilding interval contours from a bottom-up perspective. BiP-Net network mainly depends on the image scales, which means this method can run faster for smaller size images only.

Zobeir Raisi, et al.2022[7] end-to-end trainable architecture based on Detection using Transformers (DETR), this proposed method leverages a bounding box loss function that accurately measures the arbitrary detected text regions' changes in scale and aspect ratio. This is possible due to a hybrid shape representation made from Bezier curves, that are further split into piece-wise polygons. The proposed loss function is then a combination of a generalized-split-intersection-over union loss defined over the piece-wise polygons, and regularized by a Smooth-ln regression over the Bezier curve's control points.

B. Segmentation Based Methods

Xiqi Wang et.al. (2020) [8] proposed a novel method called Rotational You Only Look Once (R-(YOLO), a robust real-time convolutional neural network (CNN) model to detect arbitrarily-oriented texts in natural image scenes. It is a rotated anchor box with angle information is used as the text bounding box over various orientations. Then features of various scales are extracted from the input image to determine the probability, confidence, and inclined bounding boxes of the text. Finally, Rotational Distance Intersection over Union Non-Maximum Suppression is used to eliminate redundancy and acquire detection results with the highest accuracy.

Richardson et al. (2019)[9] have designed a segmentation-based model that approximately locate the text instances without processing all the pixels in the image, resulting in a robust and tighter text detection.

Xu et al. (2019a)[11] have proposed a novel text detection method where a VGG-16 based network learns a directional feld map consisting of 2-D vector (magnitude and direction) generated from an input image and finally using this directional information curved texts are obtained accurately from scene images.

Tang et al. (2019) [12] implement a seglink++ an arbitrary oriented text detection method is proposed using instance-aware text component grouping technique. In this multi-level features are extracted to predict text components and estimate links between text components and then a grouping algorithm is applied based on estimated links to generate the final result.

Baek et al. (2019)[13] proposed a unique framework that localizes individual characters within text instances and subsequently group them based on affinity scores between adjacent characters to detect the entire text instance. Model is trained with character level ground truths generated from synthetic and real images. This model is more robust to text scales, as model localizes individual characters rather than whole text instance.

Li et al. (2018)[14] implemented a progressive scale expansion network (PSENet) as a segmentation-based shape arbitrary text detector, where text instances are initially shrunk to different scales and then progressively expands the kernels to enclose the entire text instance.

Long et al. (2018a)[15] proposed a novel text detection method for multi-oriented and curved texts using geometric feature estimation technique is reported in Here text instances are predicted using geometric properties of text regions.

Tang and Wu (2018) [16] proposed a lightweight segmentation framework consisting of two modules namely feature pyramid enhancement module (FPEM) and feature fusion module (FFM), where FPEM generates scale-wise feature maps from input image and FFM aggregates those multi-scale feature maps to generate final feature map. Then, a pixel aggregation method is applied to predict text instances on final feature map, where true pixels of text instances are aggregated with appropriate text kernels nearest to corresponding text instances. This method yields high accuracy and efficiency due to its low-cost segmentation process.

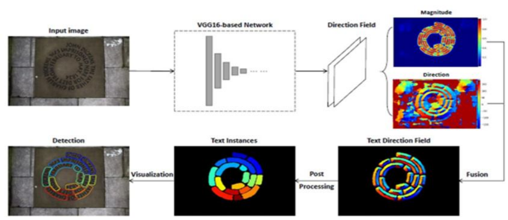

Figure 2 Pipeline and working principle of text detection model for curved text [16]

III. RELATED CHALLENGES

There are various types of challenges that arise in both machine learning and deep learning techniques. Language challenges, text detection in curved and multi-directional text images, and the size of the bounding box are among these issues. Since the previous decade, the research community has nearly pushed itself into a deep learning environment, as the reliability of DNN models and has improved significantly. However, significant challenges must be addressed in addition to making deep learning models more resilient and efficient which will benefit both the academic community and our society.

Scene texts provide more challenges than document image analysis because of complex backdrop, uneven illumination and noise, and the existence of multiple scripts. These combined issues are the most significant obstacle to researchers creating a full scene text detection algorithm. In complicated scene scenarios, rigid picture capture, along with a diverse set of scene texts, presents a number of obstacles for reliable scene text localization. [24]. These challenges are discussed below.

Text heterogeneous in nature In terms of design, arrangement, orientation, and other characteristics, text in scene photos has more variety and unpredictability. Furthermore, natural scene photographs include text from many scripts, making text identification more difficult owing to the uneven pattern.

Backgrounds with an abundance of images The backgrounds of scene photographs are frequently quite complex in nature, making it difficult to locate the inscriptions in the images. For suitable text localization, scene texts that emerged as miscellaneous items in unconstrained contexts may necessitate complex and costly segmentation approaches. Also, items that appear to be letters and occlusion owing to noise, as well as uneven lighting effects, might cause confusion and incorrect detection.

Improper image acquisition Due to uncontrolled circumstances, picture acquisition in the outdoors cannot be flexible, and consequently scene photography and text instances may not be of high quality. Improper image capture, lens distance from the subject, and shooting angle might result in a lower and fuzzy text occurrences in the obtained image.

(4) **Algorithms customized to text** Over the last few years, scene text identification algorithms have gotten more object specific in terms of text alignment and style. Several published datasets are specifically designed for specific types of texts, such as multi-directional, blurred, and curved text, prompting researchers to construct text-specific algorithms instead of general text detectors.

IV. STANDARD CURVED DATASETS FOR TEXT DETECTION

Text with a curved is one of the most prominent observations in real-world scenarios. Despite its performance in multi-directional texts, scene text detection has yet to be completely explored in curved texts.

Total-Text [17] It allows us to create curved text along with horizontal and multi-oriented text. The dataset contains 1255 training images and 300 testing images. At the word level, a polygon with 2N vertices (N 2, 15) is implemented to annotate the images.

CTW-1500[18] It is a 1500-image collection (1000 train and 500 test). At the text line stage, each text detection annotation is a 14-vertex polygon that represents the text area. In text examples, both angled and horizontal texts are presented. Images are mostly captured using camera phones in both indoor and outdoor locations, as well as through the internet, digital image collections, containing English and Chinese languages.

CUTE80 [**19**] he first scene text dataset with a curve orientation is now available. There are a total of 80 curved text line photos, each with a unique backdrop, viewpoint deformation appearance, and small impact (in S, Z shaped text lines and circle). Images are captured with digital camera phones in both indoor and outdoor settings. With altering font size, color, perspective distortion, and so on, the backdrop of a text is complicated. On a line-by-line and word-by-word basis, this dataset has been annotated. It can handle curved character.

ICDAR'19-ArT dataset [89] is currently the largest

ICDAR'19-ArT[20] It's now the most extensive collection of scene text that may be changed in any way. Total-text and SCUT-CTW1500 are merged and expanded in a single package. Each of the new images has at least one arbitrary-shaped sentence. When it comes to text orientations, there is a lot of room for creativity. The ArT dataset is divided into two sections: a training set of 5,603 images and a testing set of 4,563 images. All English and Chinese text occurrences have been placed with tight polygons.

SCUT-CTW1500[21]. It is a difficult dataset to recognize curved text on natural images. It contains 1,000 training and 500 test images, with the majority of the text instances being in English and Chinese.

ICDAR 2019-ReCTS [22] The images in this dataset were captured from a street view Chinese signboard, the letters were largely set against a complex background with varied font styles and orientations. Cell phone cameras capture images in an unrestricted area. Because of its aesthetic approach, the layout of Chinese letters adds greater intricacy to the signboard.

MSRA TD 500[23] The MSRA Text Detection 500 Database (MSRA-TD500) is a collection of 500 natural images captured using a pocket camera in both indoor and outdoor (street) settings. Outside images typically consist of guide boards as well as billboards with a complicated backdrop, while indoor images mostly consist of signs, doorplates, and caution plates. The image sizes range from 1296x864 to 1920x1280 pixels.

KAIST [24] comprises of photographs taken in a variety of settings (Lee et al. 2010). The majority of the photographs are of indoor/outdoor scenes with uneven lighting effects. Multilingual text pictures are included in the collection. This dataset's GT is created at the pixel level.

DAST1500 [25] is a text detection dataset that aggregates commodities images from the Internet in a dense and arbitrary-shaped format, along with extensive descriptions of the goods on tiny wrinkled packet. There are 1,038 images for training and 500 images for testing. Polygons contain annotations only at text-line level.

The dataset is described using the various parameters. The following parameters are described in detail:

Various parameters used to represent the dataset are mentioned below:

a) Train is a tool that shows how many training pictures are in a dataset.

b) Test collection of test images for a given dataset is represented by test.

c) The text type in the dataset describes the type of text to be examine. These styles of text might be horizontal, multi-directional or curved text.

d) Script defines the form of script used in the dataset.

e) Train Instances is the amount of text instances examine for training the model.

f) Test Instances is the amount of text instances examine for testing the model.

g) The region from where the images are collected is known to as the source such as natural scenes, websites and image libraries (Google Open Image).

h) Environment Considered The characteristics considered while taking pictures from the surroundings are referred to as environmental circumstances.

TABLE I. COMMON CURVED DATASETS FOR NATURAL SCENIC TEXT DETECTION

Datasets	Number of Training set	Images Testing set	Text type	Script
Total text	1255	300	Horizontal and multi-directional text	English
CTW-1500	1000	500	Inclined texts and horizontal texts	English and Chinese texts.
SCUT-CTW1500	1000	500	Horizontal as well as multi-directional text.	English
CUTE 80	80	80	Multi-oriented	English
ICDAR'19-ArT	5,603	4,563	Arbitrary-shaped text.	Chinese and Latin scripts
ICDAR 2019-ReCTS	20,000	5,000	Chinese character test images	Chinese script
MSRA TD 500	300	200	Arbitrary orientations.	Chinese, English or mixture of both
KAIST	3000	-NA-	Multi-oriented	Multi-lingual
DAST1500	1038	500	Arbitrary-shaped text	English

V. PERFORMANCE EVALUATION

To show the validity of computer vision-based approaches, they must be verified on the technique. Every new approach is compared and assessed on the basis of a set of standard parameters known as performance matrices. Various criteria including as F-Measure, Precision and Recall are utilized to assess the text detection demonstration.

Precision (P) It is refers as the model's capacity to predict the text region with accuracy.

Recall (R) Recall (R) refers to the model's capacity to identify all true positive and recognized text items between all ground truth.

Intersection over Union (IoU) The approach may be used to measure accuracy and recall by directly employing between the truth and detection frames.

The mathematical expression to these Equation (1) and Equation (2) are used to calculate accuracy and recall:

$$P(D, G) = \frac{\sum_i \mathrm{IF}(\max(I_O U Mat_{i,j}) > t)}{|D|} \quad (1)$$

$$R(G, D) = \frac{\sum_j \mathrm{IF}(\max(I_O U Mat_{i,j}) > t)}{|G|} \quad (2)$$

Where G stands for truth boxes, D for detection boxes, t for threshold value, which is usually 0.5, and IF for logic function. Equation (3) is used to compute the IoU matrix for both the truth and detection frames

$$\mathrm{IoU\ Mat}\ i, j = \frac{\sum i\ \mathrm{IF}\ (area(intersection(Gi, Gj))}{\mathrm{Area\ (union}\ (G_i, G_j))} \quad (3)$$

The cumulative average of precision P and recall R is provided by: Overall assessment metric (F).

$$F = (2 * P * R)/(P + R)$$

VI. CONCLUSIONS

Curved scene text detection is a challenging research area in computer vision and pattern detection that has both practical and theoretical applications. This paper examines current deep learning-based algorithms, as well as their performance measures and benchmarked curved datasets and issues in scenic images recognition. As the study of computer vision and deep learning increases, natural scene text detection technology will begin to expand and research for even improved technical advances.

REFERENCES

[1] S. Mahajan and R. Rani, "Text detection and localization in scene images: a broad review," *Artificial Intelligence Review*, vol. 54, no. 6, pp. 4317–4377, Apr. 2021, doi: 10.1007/s10462-021-10000-8.

[2] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306.

[3] Y. Zhu, C. Ma, and J. Du, "Rotated cascade R-CNN: A shape robust detector with coordinate regression," *Pattern Recognition*, vol. 96, p. 106964, Dec. 2019, doi: 10.1016/j.patcog.2019.106964.

[4] Y. Liu, L. Jin, S. Zhang, C. Luo, and S. Zhang, "Curved scene text detection via transverse and longitudinal sequence connection," *Pattern Recognition*, vol. 90, pp. 337–345, Jun. 2019, doi: 10.1016/j.patcog.2019.02.002.

[5] C. Xue, S. Lu, and S. Hoi, "Detection and rectification of arbitrary shaped scene texts by using text keypoints and links," *Pattern Recognition*, vol. 124, p. 108494, Apr. 2022, doi: 10.1016/j.patcog.2021.108494.

[6] Z. Raisi, M. A. Naiel, P. Fieguth, S. Wardell, and J. Zelek, "Challenges of Deep Learning-based Text Detection in the Wild," *Journal of Computational Vision and Imaging Systems*, vol. 6, no. 1, pp. 1–5, Jan. 2021, doi: 10.15353/jcvis.v6i1.3543.

[7] X. Gironés and C. Julià, "Real-time localization of multi-oriented text in natural scene images using a linear spatial filter," *Journal of Real-Time Image Processing*, Sep. 2019, doi: 10.1007/s11554-019-00911-9.

[8] X. Wang, S. Zheng, C. Zhang, R. Li, and L. Gui, "R-YOLO: A Real-Time Text Detector for Natural Scenes with Arbitrary Rotation," *Sensors*, vol. 21, no. 3, p. 888, Jan. 2021, doi: 10.3390/s21030888.

[9] Richardson E, Azar Y, Avioz O, Geron N, Ronen T, Avraham Z, Shapiro S," It's all about the scale– efficient text detection using adaptive scaling".,2019, In: arXiv:1907.12122.

[10] Xu Y, Wang Y, Zhou W, Wang Y, Yang Z, Bai X ," TextField: learning a deep direction feld for irregular scene text detection. IEEE Trans Image Process 28(11):5566–5579,2019.

[11] Tang J, Yang Z, Wang Y, Zheng Q, Xu Y, Bai X SegLink++: detecting dense and arbitrary-shaped scene text by instance-aware component grouping. In: Pattern recognition, vol 96, pp 106954.

[12] Baek, Youngmin, et al. "Character region awareness for text detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[13] Long, Shangbang, Xin He, and Cong Yao. "Scene text detection and recognition: The deep learning era." International Journal of Computer Vision 129.1 (2021): 161-184.

[14] Wu, Tianyi, et al. "Cgnet: A light-weight context guided network for semantic segmentation." IEEE Transactions on Image Processing 30 (2020): 1169-1179.

[15] C. K. Ch'ng and C. S. Chan, "Total-text: A comprehensive dataset for scene text detection and recognition," in International Conference on Document Analysis and Recognition (ICDAR), 2017, pp. 935–942.

[16] Liu Yuliang, Jin Lianwen, Zhang Shuaitao, and Zhang Sheng. 2017. Detecting curve text in the wild: New dataset and new solution. arXiv:1712.02170

[17] Risnumawan, A., Shivakumara, P., Chan, C. S., & Tan, C. L. (2014). A robust arbitrary text detection system for natural scene images. Expert Systems with Applications, 41(18), 8027-8048.

[18] C.-K. Chng, Y. Liu, Y. Sun, C. C. Ng, C. Luo, Z. Ni, C. Fang, S. Zhang, J. Han, E. Ding, et al., "ICDAR2019 Robust Reading Challenge on Arbitrary-Shaped Text (RRC-ArT)," Proc. IAPR Int. Conf. Document Analysis Recog., 2019

[19] Liu Yuliang, Jin Lianwen, Zhang Shuaitao, and Zhang Sheng. 2017. Detecting curve text in the wild: New dataset and new solution. arXiv:1712.02170.

[20] Liu X, Zhang R, Zhou Y, Jiang Q, Song Q, Li N, Zhou K, Wang L, Wang D, Liao M, Yang M (2019g) ICDAR 2019 robust reading challenge on reading chinese text on signboard. In: arXiv:1912.09641

[21] Yao C, Bai X, Liu W, Ma Y, Tu Z (2012) Detecting texts of arbitrary orientations in natural images. In Proc. CVPR, pp. 1083–1090.

[22] Lee S, Cho MS, Jung K, Kim JH (2010) Scene text extraction with edge constraint and text collinearity. In: 20th international conference on pattern recognition, pp 3983–3986

[23] Lee S, Cho MS, Jung K, Kim JH (2010) Scene text extraction with edge constraint and text collinearity. In: 20th international conference on pattern recognition, pp 3983–3986

[24] Seglink++: detecting dense and arbitrary-shaped scene text by instance-aware component grouping.

[25] Wang, S., Liu, Y., He, Z., Wang, Y. and Tang, Z., 2020. A quadrilateral scene text detector with two-stage network architecture. Pattern Recognition, 102, p.107230.

[26] Lyu, P., Liao, M., Yao, C., Wu, W. and Bai, X., 2018. Mask textspotter: An end-to-end trainable neural network for spotting text with arbitrary shapes. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 67-83).

[27] Dai Y, Huang Z, Gao Y, Xu Y, Chen K, Guo J, Qiu W ," Fused Text Segmentation Networks for Multi oriented Scene Text Detection. In: Proceedings" - International Conference on Pattern Recognition, IEEE, vol 2018-Augus, pp 3604–3609,2018.

[28] Tang Y, Wu X ," Scene text detection using superpixel-based stroke feature transform and deep learning based". IEEE Trans Multimedia 20(9):2276–2288,2018.
[29] Tang Y, Wu X , "Scene text detection and segmentation based on cascaded convolution neural networks:. IEEE Trans Image Process 26(3):1509–1520,2017.
[30] Jaderberg M, Simonyan K, Vedaldi A, Zisserman," A Reading text in the wild with convolutional neural networks. Int J Compution Vis 116(1):1–20,2016.
[31] Liao, Minghui, et al. "Real-time scene text detection with differentiable binarization and adaptive scale fusion." IEEE Transactions on Pattern Analysis and Machine Intelligence 2022.