

Detection and Identification of Text-based Traffic Signs

Xiuyuan Chi,¹ Dean Luo,¹ Qice Liang,² Junxing Yang,¹ and He Huang^{1*}

¹School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture,
No. 15, Yongyuan Road, Huangcun Town, Daxing District, Beijing 102616, China

²Beijing Engineering Co. Ltd., No. 1 Dingfuzhuang West Street, Chaoyang District, Beijing 100024, China

(Received November 18, 2022; accepted January 19, 2023)

Keywords: textual traffic signs; improved Advanced EAST; sign plate detection; text recognition

The detection and recognition of text-based traffic signs are important in the field of automatic driving, but these tasks pose problems in practical applications, such as low accuracy in text detection and extraction, poor long-text extraction, and a lack of datasets. To solve these problems and to improve the detection and recognition accuracy of text-based traffic signs so that they can better serve automated driving, we propose an improved Advanced efficiency and accuracy scene test (EAST) model and fixed-size prediction to enhance the capability of extracting features. The text recognition stage features a text preprocessing method that trains convolutional recurrent neural network (CRNN) models using synthetic datasets of Chinese strings. Experimental results show that the improved Advanced EAST model and fixed-size prediction enabled the detection of text on traffic signs to achieve a 96% recall rate and an 88.5% accuracy rate; we also saw better results in the case of dense text and obscuration. Thus, in the absence of targeted datasets, the designed text image preprocessing method can realize print text recognition in different scenarios only by training models using synthetic data, thereby eliminating the need for a large amount of work on training dataset labeling while still meeting requirements of detection and recognition.

1. Introduction

Automated driving technology needs to sense the environment and make decisions in real time to ensure safety. Traffic signage is an important road element, and it is crucial that we achieve automatic detection and recognition of such signage. According to the type of semantic information provided by traffic signs, they can be divided into text-based traffic signs and symbol-based traffic signs. Among these, text-based traffic signs use text and directional schematics to provide *a priori* information about the next road segment, mostly for such tasks as wayfinding, tourism, and safety at road construction sites. Symbolic traffic signs use patterns, numbers, and arrows combined with colors to convey *a priori* information, mostly for directionals, warnings, and signs indicating prohibitions. In the natural setting, text-based traffic signs are more difficult to detect and identify than symbol-based traffic signs.

*Corresponding author: e-mail: huanghe@bucea.edu.cn
<https://doi.org/10.18494/SAM4253>

Detection methods for text-based traffic signs are usually divided into traditional methods and deep learning methods. Traditional methods include the connectivity domain-based method and the sliding window-based method. The principle of the connectivity-based method is to roughly determine the text area using the edge or stroke information of the text and then realize the text localization based on connectivity analysis. In prior research, the Sobel operator was used to extract text edge information, the K-means clustering algorithm was used to determine the approximate text area, and the final text range was determined by the project profile.⁽¹⁾ In another study, a new stroke width variation operator was designed from the text stroke width and combined with the edge extraction operator and clustering algorithm to determine the text area.⁽²⁾ Chen *et al.*⁽³⁾ proposed an adaptive threshold travel smoothing algorithm to achieve Tibetan text extraction and also using K-means mean adaptive acquisition of travel thresholds to achieve text extraction using the project profile. The text detection method based on a sliding window aims to traverse the image from top to bottom through an artificially set sliding window. The regions within the window are extracted with features during the traversal process, and the text regions are determined by combining confidence and thresholds. Chen and Yuille⁽⁴⁾ extracted the text regions and analyzed them to determine the effective features that represent the text regions, then constructed a weak classifier based on the text features, and finally trained a strong classifier with the weak classifier to achieve text detection. In one study, an irregular polygonal sliding window was designed to handle the effects of changes in perspective, text size, and shape ratios. The experimental results showed a large degree of improvement in the recall rate for irregular text extraction.⁽⁵⁾ However, the traditional detection method has high image requirements and is only applicable to cases involving relatively simple backgrounds, which are less universal, and also has difficulty handling complex and changing actual scenes. With the development of computer vision technology, deep-learning-based methods have gradually become the primary solution for text detection and recognition. In previous research, considering the characteristics of the shape of text information, the original Single Shot MultiBox Detector (SSD) was improved by replacing the original square convolution with rectangular convolution to increase the capabilities of linear text feature extraction.⁽⁶⁾ Moreover, Zhou *et al.*⁽⁷⁾ proposed an anchor-free-based one-stage text detection algorithm, the efficient and accurate scene text (EAST), which processes the pixels on the image pixel-by-pixel using a convolutional network, and achieves text detection via pixel-by-pixel determinations. The EAST algorithm achieves irregular text extraction, but it has poor extraction ability for long texts. In addition, a segmentation-based text detection method was proposed, in which the network generates 10 channels of a feature tensor as output, two of which are used to determine whether the pixel is text, and the remaining 8 channels of which determine the association of the pixel with its immediate 8 pixels and then determine the connected domain to achieve text extraction.⁽⁸⁾ After text detection and extraction, the extracted text must be recognized. In the literature, a text recognition algorithm called a convolutional recurrent neural network (CRNN), which is a combination of a convolutional neural network (CNN) and a recurrent neural network (RNN), was proposed to realize text recognition from an arbitrary length of input.⁽⁹⁾ Shi *et al.*⁽¹⁰⁾ proposed the Attentional Scene Text Recognizer with Flexible Rectification (ASTER) method to solve the problem of text distortion. Moreover, Hamming optical character recognition (OCR)

was proposed on the basis of the transformer structure, which solves the problem of too many categories leading to too many model parameters.⁽¹¹⁾ In recent years, most of the research on text recognition has been carried out simultaneously with text detection, and end-to-end detection recognition models such as fast oriented text spotting (FOTS) and (Towards Efficient and Accurate End-to-End Spotting of Arbitrarily-Shaped Text) PAN++ have been proposed, both of which have achieved good results.^(12,13) Named entity recognition (NER) methods aim at recognizing entities of interest in text such as location, organization, and time.

The lattice structure proposed by Zhang and Yang⁽¹⁴⁾ employed a dictionary to match the words in a sentence, which reduced the incorrect word cut in the NER to some extent. Lai *et al.*⁽¹⁵⁾ used a label expansion strategy to implement label migration learning in the few-shot NER as a way to improve the performance of the few-shot learning. These algorithms produce good experimental results in the specified dataset, but their low recall rate makes it difficult for them to be widely used.

In summary, detection and recognition of text-based traffic signs are limited because detection itself is difficult and the ability to extract long text is poor. To solve these problems, we propose an improved Advanced EAST and image preprocessing method to enhance the accuracy and recall rate of text detection. To address the problem of poor extraction of long text, we propose to carry out a series of preprocessing steps, such as adjusting the coordinates to intercept the text area and binarizing the text before CRNN recognition, to achieve the text-based traffic sign recognition required for automatic driving.

2. Methods of text-based traffic signage detection and identification

Extraction of text regions directly in the natural environment is prone to confusion about textual information, and occurrences of mis-extraction of license plate photo numbers of vehicles near billboard text on distant buildings may impact later text recognition and increase the workload in terms of later manual checks and calibrations. Therefore, a single-category the detection model is trained by You Only Look Once, version 3 (YOLOv3) in one stage before detection of text signage, which is then used to extract regions containing text-based traffic signage in initial photos. After the region containing text class traffic signage is extracted, the experimental process is divided into two steps: (1) text detection of traffic signage and (2) text recognition. The main work in the detection of traffic signage is to improve the image preprocessing method with the improvement of Advanced EAST; text recognition is performed by training the CRNN model with the synthetic Chinese dataset and then preprocessing the image with binarization and color flip for recognition.

2.1 Traffic signage text detection

Advanced EAST uses weighted head and tail text points to find the mean value to directly calculate the four corner points of the text box, which effectively improves the detection ability of the model for long text. However, the feature extraction ability of the model is general, and the detection effect is considerably reduced when the signage is obscured and shadowed. Therefore,

we propose an improved feature extraction network and method for preprocessing predicted images for Advanced EAST, aiming to enhance the accuracy and recall rate of text extraction.

(1) Method of improved image preprocessing

Advanced EAST sets the maximum training size and maximum prediction size before training. Given the GPU video memory and the training batch size setting, the pre-set maximum training size is generally used for the training input size. When the method is combined with the detection of intercepted signs in natural scenes, the width and height of some small signs are adjusted downward to 32 pixels, but it is difficult to detect text on signs that are too small, even if the text can be extracted accurately and completely. To solve this problem, we adopted fixed size in prediction, in which the fixed size is a predetermined maximum predicted size consistent with the maximum training size to ensure that the size of the input image is the same during model training and prediction.

(2) Improvement of Advanced EAST

Advanced EAST uses Visual Geometry Group 16 (VGG16) as the network for feature extraction, but the last three fully connected layers are removed in the actual model, and only 13 convolutional layers are used, making it difficult to meet the requirements for feature extraction under complex conditions. To compensate for and improve the ability for feature extraction of a shallow convolutional network, we replaced VGG16 with Resnet50 as the feature extraction network in Advanced EAST, and the residual structure of Resnet50 allows the convolutional network to go deeper. The overall feature extraction capability is substantially improved over that of VGG16. In this study, the last layer of Resnet50 was eliminated, and the first 49 layers were used to replace the original VGG16 to form the feature extraction network of the improved model, whose structure is shown in Fig. 1.

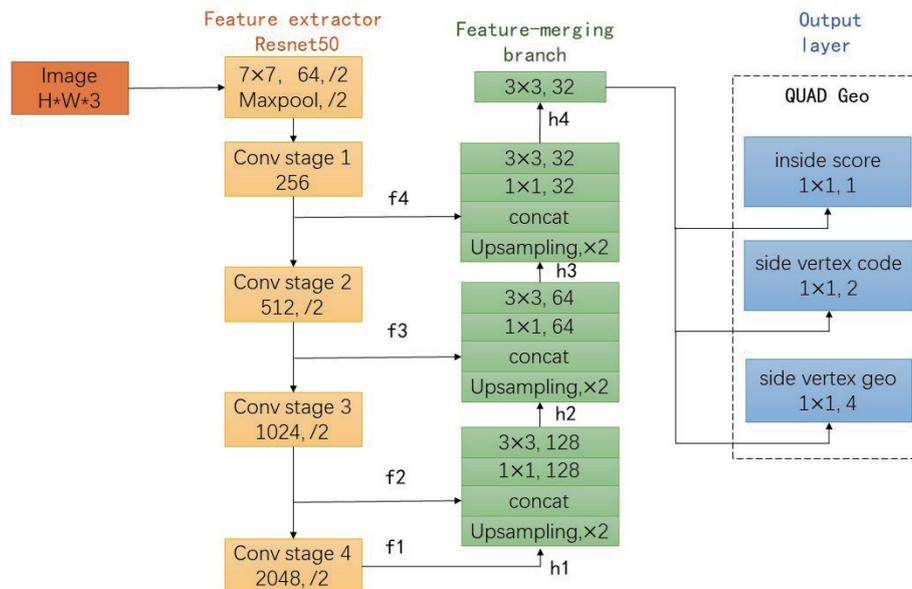


Fig. 1. (Color online) Network structure diagram of the improved Advanced EAST.

2.2 Recognition of text on traffic signs

The extracted textual information on traffic signage text is recognized using the CRNN text recognition model. Considering the number of commonly used Chinese characters, we used a dataset of synthetic printed text for CRNN training. As the text images extracted from the signage and the synthetic dataset differ considerably in color, direct recognition is difficult to guarantee. To ensure that the information can be correctly recognized by CRNN, we proposed to preprocess the extracted text before text recognition. Preprocessing was divided into two steps: size adjustment and binarization. The text images after the above steps reached the approximate printed text state. The overall process of text recognition is shown in Fig. 2.

(1) Text region cropping

After improving the Advanced EAST model for the detection of text, we needed to crop the text region area for extraction. Because Advanced EAST predicts the four corner points of the text box directly, the output text is in an irregular quadrilateral, meaning that the detected text boxes are not strictly rectangular, and the text boxes do not cover the text area completely. Therefore, it is necessary to adjust the coordinates of the corner points of the text box, so we adopted a method of finding the minimum rectangular box and then adjusting the positions of the corners. At the same time, to determine the binary map, the coordinates of the four corner points of the minimum enclosing rectangle are expanded by 10 pixels; that is, the horizontal and vertical coordinates of the upper left corner point are reduced by 10, the horizontal and vertical coordinates of the lower right corner point are increased by 10, and the two corner points determine a rectangle, as shown in Fig. 3.

The final cropped area is a rectangle with a black outer border. The primary purpose of adding the black and red outer expansion areas is to ensure that the text area is surrounded by the rectangular box; the secondary aim is to provide a basis for determining the subsequent color flip, the purpose of which is described in detail in the following.

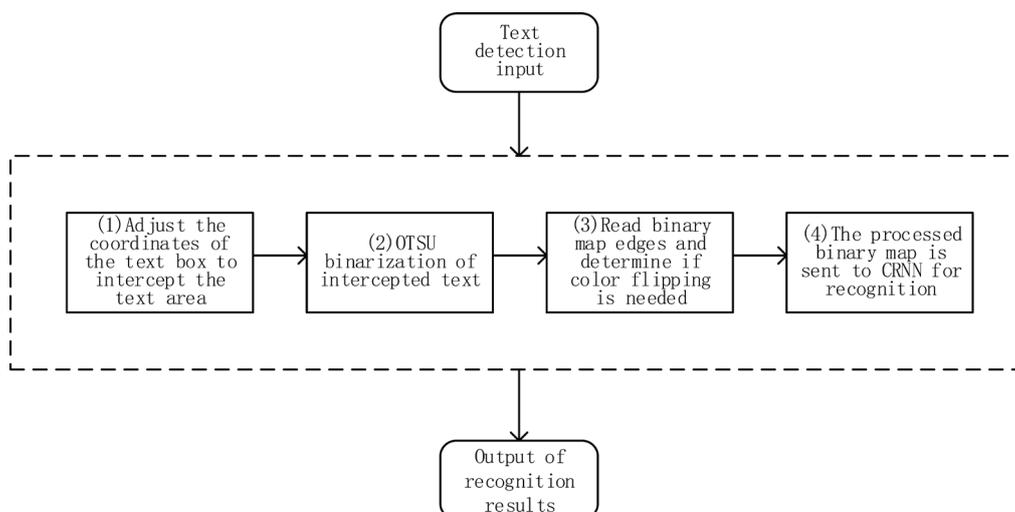


Fig. 2. Flow chart of text recognition.



Fig. 3. (Color online) Schematic diagram of text area cropping

(2) Binarization of intercepted text

In this study, the synthetic Chinese dataset is used, the text font of the whole dataset adopts a printing font, and the overall color distribution of the photo is black with a gray background. The text extracted from the signage has a large gap that contains an image of the dataset, and the model trained directly with this dataset cannot be directly used for text recognition on actual traffic signage. To improve the accuracy of recognition by the model, we binarized the cropped text images by OTSU.⁽¹⁶⁾ The results are shown in Fig. 4.

For example, the binarized text image of “米东区” has a white text and a black background, which is not conducive to the subsequent text recognition, so the binarized image must be color flipped. The binary map can perfectly separate the text and the background, so we can judge whether the image needs to be flipped on the basis of the pixel value of the expansion area in the text box.

(3) Color flipping

The color flip of the binary image is determined by pixel counting, and to avoid the statistics to the text area affecting the correct rate of determination, we only counted the area of 5 pixels around the text box. The method is to count the number of pixels in the peripheral area with a zero pixel value and calculate the proportion of black pixels to the total number of peripheral pixels. The text image after flipping is shown in Fig. 5.

(4) CRNN

After carrying out the steps described, the line-by-line extraction of information on text-based traffic signs is completed, and the text needs to be recognized. Considering that this study mainly focuses on Chinese signage, we adopted the CRNN text recognition model, which is the most used in real-life applications and performed stably.

3. Experimental datasets and analysis of results

3.1 Datasets

To extract traffic signs, we used the dataset named CSUST Chinese Traffic Sign Detection Benchmark (CCTSDB)⁽¹⁷⁾ and Tsinghua-Tencent100k⁽¹⁸⁾ datasets to train the YOLOv3 model, from which a total of 800 photos containing text-based signs were selected. CCTSDB is an open-source traffic sign detection dataset labeled and published by Changsha University of



Fig. 4. Results of Otsu binarization.



Fig. 5. Schematic diagram of color flip.

Technology, and it contains a total of more than 15000 photos of different scenes. Tsinghua-Tencent100k was jointly produced by Tsinghua University and Tencent and was derived from 100000 Tencent street view photos from maps. The traffic signage categories were also refined to include prohibitions, indicating warning categories.

The text detection experiments used the public dataset from the 2018 International Conference on Pattern Recognition (ICPR) Text Detection Challenge, which included a total of more than 10000 images of product information from online shopping malls. The information in the images in the dataset is complex and variable, and the text is dense, rotated, or contains complex typography, so it is sufficiently close in style to traffic signage cropped and extracted from natural scenes to meet the experimental requirements of this study.

The text recognition experiment for traffic signage used a synthetic Chinese string dataset as training data, which contains more than 3.6 million text images covering many categories, such as Chinese and English characters, punctuation marks, and special symbols. The data met the training requirements of the text recognition model. The image size in the dataset is 280×32 pixels, and each image contains only a single line of text. The overall color distribution of the images is black text on a gray background, and color differences may be observed between different images. The details of these datasets are shown in Table 1.

3.2 Traffic signage extraction

The default hyperparameters were used to train the YOLOv3 model. After the model training had reached convergence, photos were selected from outside the training set for testing and also verifying the generalizability of the model, and photos were taken with cell phones on the Beijing University of Civil Engineering and Architecture campus for detection and recognition. The result of the recognition of a sign is shown in Fig. 6.

As shown in Fig. 6, the single-category text signage detection model trained in this study effectively achieved the detection and extraction of text signage in a natural scene, and it could also accurately recognize photos collected by cell phones, which indicates that the model is generalizable.

Table 1
(Color Online) Details of the datasets used.

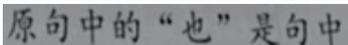
Dataset Name	Dataset Detail	Picture
CCTSDB	More than 15000 sheets	
Tsinghua-Tencent100k	100000 sheets	
ICPR	More than 10000 sheets	
Synthetic Chinese dataset	More than 3.6 million sheets, uniform size of 280×32	



Fig. 6. (Color online) Extraction of a text from a sign in a natural scene.

3.3 Detection of traffic signage text

To verify the performance of the improved Advanced EAST, we conducted a comparison experiment between EAST, Advanced EAST, and improved Advanced EAST using the exact same data set and parameters. A test set of 100 traffic signs in different environments was intercepted; all images in the test set presented a combined total of 500 text messages. The experimental results are shown in Table 2; accuracy and recall were used to compare the performance of the model, and the predicted text box was considered a successful detection if the text area was completely framed.

Table 2 shows that replacing VGG16 with Resnet50 improved the accuracy of the model, but the model had a serious loss of recall. When using the same fixed size prediction as the training size, the improved model with image sizes of 128×128 achieved a 90.8% accuracy and a recall

Table 2
Comparison of results of text detection.

Model and training size	Number of detections	Check the number of error	Accuracy (%)	Recall Rate (%)
EAST 256	309	38	87.7	61.8
Advanced EAST +VGG16 256	413	74	82.1	82.6
Resnet50+Advanced EAST 256	349	39	88.8	69.8
Resnet50+Advanced EAST Fixed Size 128	384	35	90.8	76.8
Resnet50+Advanced EAST Fixed Size 256	480	55	88.5	96

rate close to that of the Advanced EAST level, while the improved model with image sizes of 256×256 achieved a 96% recall rate and 88.5% accuracy. From the analysis of experimental data, we found that replacing VGG16 with Resnet50 improved the accuracy of recognition by the model, but lost some of the recall rate. This occurs because the improved feature extraction network has deeper convolutional layers, which compensates for and improves on the feature extraction ability of the original shallow convolutional network. Using the same fixed size prediction as the training size improves the recall rate, and the comparison shows that the model with the 256×256 image size had the best overall performance, because the fixed size solved the problem that it is more difficult to detect text on signs that are too small.

To further demonstrate the actual performance of the improved model in this study, we selected images of signage with higher detection difficulty for detection and recognition and compared the results from them with other models. The results of the comparisons are shown in Figs. 7 and 8.

As Fig. 7 shows, neither EAST nor Advanced EAST could detect the location of the text correctly under occlusive conditions; EAST used the rectangular box method to frame the text area, while Advanced EAST based on Resnet50 extracted the text area on the signage more completely. From Fig. 8, it can be seen that both EAST and Advanced EAST extracted the text information on the signage completely, but the detection of longer text was general; for example, in the text “南浦大桥”, neither EAST nor Advanced EAST could use one text box to detect it completely, which is not good for the text recognition and integration with semantic information. This result is unfavorable for later text recognition and semantic information integration, while the improved Advanced EAST model was more capable of detecting long text and could achieve complete text detection using only one text box. Thanks to the improved residual structure, the convolutional network can go to deeper structures, and the ability to extract features is enhanced. Unlike the EAST model that resizes the input image into a fixed size input and then maps the input result back to the original size image, and Advanced EAST, which takes the input size down 32-fold, the model in this study fixes the prediction size, and the model detection capability is thereby enhanced.

3.4 Recognition of traffic signage text

The text recognition experiment in this study used a synthetic Chinese string dataset as training data, which contained more than 3.6 million text images and covered many categories, such as Chinese and English characters, punctuation marks, and special symbols, thereby

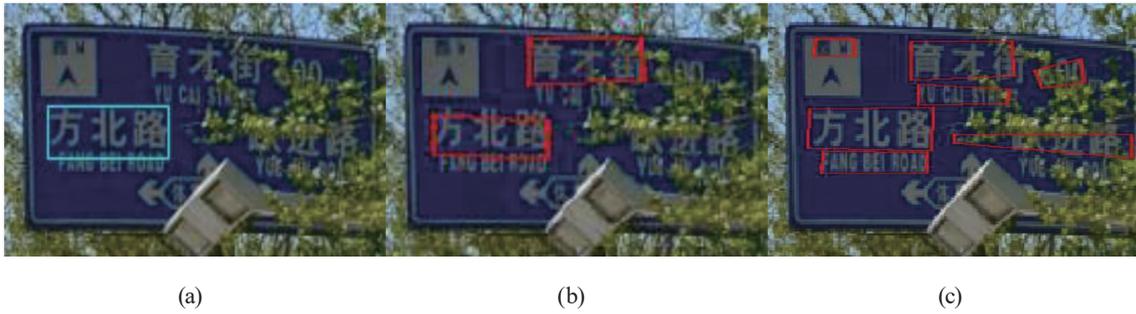


Fig. 7. (Color online) Comparison of the detection by each model under conditions of occlusion: (a) EAST, (b) Advanced EAST+VGG16 256, and (c) Resnet50+Advanced EAST Fixed Size 256.



Fig. 8. (Color online) Comparison of detection by models under conditions with dense text: (a) EAST, (b) Advanced EAST+VGG16 256, and (c) Resnet50+Advanced EAST Fixed Size 256.

meeting the training requirements of the text recognition model. The image size in the dataset was 280×32 , and each image contained only a single line of text. The overall color distribution of the images was black text on a gray background, and color differences were observable between different images. There are also differences in the fonts and text angles between different images in the dataset, and the sample diversity of the dataset met the experimental requirements of this study.

This experiment used the open-source CRNN project based on the PyTorch framework, the training input size was 160×32 , the input batch size was set to 256, the optimizer was Adam, the initial learning rate was set to 0.0001, and the rest of the parameters used were default parameters. The model was trained for a total of 30000 iterations, and the loss function was reduced to 0.0113.

The extracted text images required image preprocessing before they were sent to CRNN for recognition. After the experiments, it was found that the rate of sending the cropped extracted text directly to the model for recognition was low and was due to a problem of font and background color differences. The effect of the direct recognition of cropped extracted text is shown in Fig. 9.

From the results, it can be seen that the recognition results for images with white fonts do not correlate with the actual results, while the recognition results of text images with black font are correct. On the basis of the analysis of the experiments, we conclude that the font color of the text is the key to the recognition accuracy of the CRNN model, and the text image preprocessing



Fig. 9. (Color online) Effect of the direct recognition of samples.



Fig. 10. Results of text image recognition after preprocessing.

algorithm designed in this paper can approximate the input text images as synthetic text. To verify the effectiveness of the text image preprocessing algorithm, we used the same input text image for comparisons; the results are shown in Fig. 10. As can be seen from Fig. 10, the text images processed by the text image preprocessing method designed in this study basically achieved correct recognition.

In summary, the experimental results indicate that the text preprocessing method designed in this study can effectively improve the accuracy of text recognition in the absence of datasets containing targeted text, and CRNN training and recognition can be achieved using only synthetic Chinese string datasets, which reduces the workload of dataset annotation while still meeting the requirements of text recognition.

4. Conclusions

This study may be divided into two parts: one part is the detection of text-based traffic signage, which used an improved Advanced EAST model to enhance the feature extraction capability of the model and also incorporated fixed size prediction. The experimental results show a higher accuracy and recall than other models, with a 96% recall rate and an 88.5% accuracy rate. In the second part, in comparative experiments, the algorithm in this paper achieved better results for both dense text and signage involving occluded objects. In addition, the text recognition stage used a synthetic Chinese string dataset to train the CRNN model. The experimental results in this part indicate that the text image preprocessing method designed and evaluated in this study achieved recognition of printed text in different scenarios using a model trained with synthetic data without targeted datasets, thereby eliminating a large amount of work on the annotation of training datasets and achieving the expected text recognition requirements. These results provide an important reference for the development of autonomous driving technology. More new algorithms have been developed, such as YOLOv7,⁽¹⁹⁾ and we expect to adopt and improve these new algorithms in the future so that text-based traffic signage can be recognized more quickly and accurately.

References

- 1 C. Liu, C. Wang, and R. Dai: Proc. 8th Int. Conf. Document Analysis and Recognition (ICDAR, 2005) 610–614.
- 2 B. Epshtein, E. Ofek, and Y. Wexler: Proc. 2010 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (IEEE, 2010) 2963–2970.
- 3 Y. Chen, W. Wang, H. Liu, Z. Cai, and P. Zhao: Laser & Optoelectron. Prog. **58** (2021) 1410006. <https://doi.org/10.3788/LOP202158.1410006>
- 4 X. Chen and A. L. Yuille: Proc. 2004 IEEE Computer Society Conf Computer Vision and Pattern Recognition (CVPR, 2004) 2:II-II.
- 5 Y. Liu and L. Jin: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, 2017) 454–3461.
- 6 M. Liao, B. Shi, and X. Bai: IEEE Trans Image Process. **27** (2018) 3676. <https://doi.org/10.1109/TIP.2018.2825107>
- 7 X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, 2017) 2642–2651.
- 8 D. Deng, H. Liu, X. Li, and D. Cai: Proc. Thirty-Second AAAI Conf. Artificial Intelligence (AAAI, 2018) 1.
- 9 B. Shi, X. Bai, and C. Yao: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 2298. <https://doi.org/10.1109/TPAMI.2016.2646371>
- 10 B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai: IEEE Trans. Pattern Anal. Mach. Intell. **41** (2019) 2035. <https://doi.org/10.1109/TPAMI.2018.2848939>
- 11 B. Li, X. Tang, X. Qi, Y. Chen, and R. Xiao: ArXiv abs. **2009** (2020) 10874. <https://doi.org/10.48550/arXiv.2009.10874>
- 12 X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, and J. Yan: Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition (IEEE/CVF, 2018) 5676–5685.
- 13 W. Wang, E. Xie, X. Li, X. Liu, D. Liang, Z. Yang, T. Lu, and C. Shen: IEEE Trans. Pattern Anal. and Mach. Intell. **44** (2021) 5349. <https://doi.org/10.1109/TPAMI.2021.3077555>
- 14 Y. Zhang and J. Yang: Proc. 56th Ann. Meeting Association for Computational Linguistics (ACL, 2018) 1554–1564. <http://dx.doi.org/10.18653/v1/P18-1144>
- 15 P. Lai, F. Ye, L. Zhang, Z. Chen, Y. Fu, Y. Wu, and Y. Wang: Proc. The 29th Int. Conf. Computational Linguistics (COLING, 2022) 2199–2209.
- 16 N. Otsu: IEEE Trans. Syst. Man Cybern. **9** (1979) 62. <https://doi.org/10.1109/TSMC.1979.4310076>
- 17 J. Zhang, M. Huang, X. Jin, and X. Li: Algorithms. **10** (2017) 127. <https://doi.org/10.3390/a10040127>
- 18 Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, 2016) 2110–2118.
- 19 C. Wang, A. Bochkovski, and H. Liao: arXiv preprint 2207.02696 (2022). <https://doi.org/10.48550/arXiv.2207.02696>

About the Authors

Xiuyuan Chi graduated from Shandong University of Technology, China, in 2021 and is now pursuing her master's degree at Beijing University of Civil Engineering and Architecture, China. Her research interests are deep learning and target detection. (1330955031@qq.com)

Dean Luo received his B.S. degree from Wuhan University, China, in 1990 and his M.S. and Ph.D. degrees from Southwest Jiaotong University, China, in 1997 and 2002, respectively. Since 2004, he has been a lecturer and professor at Beijing University of Civil Engineering and Architecture, China. His research interests are in GNSS, geodetic technology, and deformation-monitoring technology.

Qice Liang received his B.S. degree from Beijing University of Civil Engineering and Architecture, China, in 2022, and he is now working at Beijing Engineering Co.

Junxing Yang graduated from the School of Surveying and Mapping of Wuhan University, China, in December 2021 with a Ph.D. degree in engineering. He is mainly engaged in research in 3D reconstruction, computer vision, autonomous driving, image stitching, and other related fields.

He Huang received his B.S. degree from Wuhan University, China, in 2000 and his M.S. and Ph.D. degrees from Sungkyunkwan University, South Korea, in 2004 and 2010, respectively. Since 2010, he has been a lecturer and associate professor at the Beijing University of Civil Engineering and Architecture, China. His research interests are in autonomous driving, high-precision navigation maps, and visual navigation and positioning. (huanghe@bucea.edu.cn)