SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
VOL. 5, NO. 2, AUGUST 2021
DOI: 10.35377/saucis.05.02.1073355

Research Article

An Implementation of Traffic Signs and Road Objects Detection Using Faster R-CNN

Emin Güney[1], Cüneyt Bayılmış[2]
[1]Corresponding Author; Department of Computer Engineering,
Sakarya University of Applied Sciences, Sakarya, Turkey; eminguney@subu.edu.tr
[2]Department of Computer Engineering, Sakarya University, Sakarya, Turkey; cbayilmis@sakarya.edu.tr

Received 14 February 2022; Revised 1 March 2022; Accepted 22 July 2022; Published online 31 August 2022

Abstract

Traffic signs and road objects detection is a significant issue for driver safety. It has become popular with the development of autonomous vehicles and driver-assistant systems. This study presents a real-time system that detects traffic signs and road objects in the driving environment with a camera. Faster R-CNN architecture was used as a detection method in this study. This architecture is a well-known two-stage approach for object detection. Dataset was created by collecting various images for training and testing of the model. The dataset consists of 1880 images containing traffic signs and objects collected from Turkey with the German Traffic Sign Recognition Benchmark (GTSRB) dataset. These images were combined and splitted into the training and testing sets with a ratio of 80/20. The model's training was carried out in the computer test environment for about 8.5 hours and approximately 10000 iterations. The experimental results achieve a total loss rate of 0.220 and the best accuracy of 88.99% for the real-time performance. Therefore, the proposed system can be easily used for robust traffic signs and objects detection.

Keywords: traffic sign detection and recognition (TSDR), faster R-CNN, object detection, deep learning.

1. Introduction

Traffic sign and road objects detection and recognition (TSDR) play a fundamental role in keeping vehicle-driver safe in traffic and improving the driving experience. With the correct detection of traffic signs, autonomous vehicles can be developed, which have recently been an essential study field. In addition, accidents can be reduced by ensuring safe driving. Therefore, TSDR systems have been a challenging problem for many years and have become an indispensable task for driver assistance systems in addition to autonomous vehicles. Developing solution systems for this critical task is crucial for safe driving in challenging road and traffic conditions. The effective operation of recognition will also help reduce traffic accident risks [1]. Li et al. developed a system that uses color segmentation and shape matching based Pyramid Histogram Directed Gradient (PHOG) to detect traffic signs. The results were obtained by analyzing the studies tested on the original dataset in various weather conditions [2].

On the other hand, Yin et al. discussed a quick and robust system that recognizes traffic signs to increase driving safety [3]. The consisting of three stages system work using Hough and SIFT transformations and Artificial Neural Network (ANN) structure. The experimental results stated that the study was superior in training and recognition speed in traffic sign recognition. Qian et al. designed a deep Convolutional Neural Networks (CNN) system to detect traffic signs. They used colour space thresholding to identify the candidate region in the input image, while multitasking CNN was used for detection. The system has been tested on different traffic signs and texts, reaching an accuracy level of nearly 90% [4]. Changzhen et al. proposed a traffic sign detection algorithm based on Deep Convolutional Neural Network (D-CNN) using the Region Proposal Network (RPN). A Chinese traffic sign dataset was created by collecting the seven main traffic sign categories and their subclasses. Afterward, the trained network was tested on 33 videos and the real-time detection rate was 99% [5]. On the other hand, Zhang et al. used the enhanced YOLO v2 model for detecting and recognizing traffic signs in real-time [7]. Xu et al. have proposed a method using the Adaptive Color Threshold and Shape Symmetry based on Cumulative Histogram Distribution Function. In a comprehensive experimental study on the GTSRB [6] dataset, the detection accuracy was increased up to 94%. They reached detection accuracy and time efficiency in complex traffic environments [13].

Table 1 Literature studies for traffic signs detection and recognizing task

Article	**Application environment**	**Traffic Sign Application**	**Category**	**Technology**	**Datasets**
Li et al. 2015 [2]	Video Series	Detection	Color segmentation and Shape matching based	Pyramid histogram of directed gradients	Private to work dataset
Yin et al. 2015 [3]	Video Series	Recognition	Hough and SIFT transform, ANN	Feature based fixed binary pattern rotation	GTSRB [6] and STS
Qian et al. 2015 [4]	Video Series	Detection	Based on deep convolutional neural networks	Edge detection and Connected component analysis	GTSRB [6], MNIST [8] , CASIA [9]
Changzhen et al. 2016 [5]	Video Series	Detection	Based on deep convolutional neural networks	Region Proposal Network (RPN)	Belgium TS [10] and GTSRB
Zhang et al. 2017 [7]	Real-Time	Detection and Recognition	Based on enhanced YOLOv2	Technique with grid cells in one step	GTSRB [6] and CTSD [7]
Xu et al. 2019 [13]	Video Series	Detection	Based on Adaptive Color Threshold and Shape Symmetry	Cumulative Histogram Distribution and Shape symmetry detection	GTSRB [6]
Özturk et al. 2020 [14]	Video Series	Recognition	Based on SSD and Faster R-CNN models	Transfer learning and grid cells	COCO [11] and GRAZ [12] datasets
Zhou et al. 2021 [16]	Video Series	Recognition	Region-Based Attention Network	Parallel Attention Network (PFANet)	ITSDB and ITSRB datasets
This study (2022)	Real-Time	Detection and Recognition	Based on Faster R-CNN	Detection method with Grid cells in two stages	GTSRB [6] and Private to work dataset

Öztürk et al. tested the SSD and Faster RCNN models on the image and video of traffic signs. They have reached high speed with SSD and high accuracy with Faster RCNN [14]. Han et al. achieved approximately 13% improvement in mAP by proposing an Online Hard Examples Mining (OHEM) based system for real-time small traffic sign detection with the revised Faster-RCNN [15]. Zhou et al. proposed the two-block PFANet architecture, which can extract more valuable features, to detect traffic signs even in harsh weather conditions. They have produced an innovative study, significantly increasing the accuracy rate [16]. Shao et al. presented a TSD using VGG16 and Faster R-CNN. The experimental results significantly accelerated the detection accuracy by exceeding 98% for various traffic signs [17]. Dai et al. developed the ResNet-50 architecture and presented a multi-tasking Faster R-CNN detector that simultaneously performs distance estimation and pedestrian detection [18].

The literature studies are detailed given in Table 1. This study uses the Faster R-CNN architecture to detect and recognize traffic signs and road objects. Training and testing processes were carried out in the test computer environment, and the results were obtained. Afterthat a real-time working system was obtained. The study performs the TSDR task using the Faster R-CNN architecture to recognize 23 different traffic signs and other objects. In order to increase the diversity of the dataset, in addition to the GTSRB, a job-specific data set was also collected for the study where the signs were intense in Turkey. This enriched study can assist researchers in analyzing TSDR tasks.

The remaining paper is structured as follows: Section 2 describes the used datasets and object detection model in detail. Section 3 elaborates on the comparative analysis of numerical results obtained from the experimental evaluation of the Faster R-CNN object detection algorithm for the TSDR task. Finally, Section 4 is the conclusion and discussion of the realized study.

2. Materials And Methods

This section presents the development stages of the created model for detecting non-vehicle traffic signs. Deep learning applications require high computing power and processing speed. To ensure learning, the number of hidden layers must be increased, and the parameters must be adjusted by constantly updating the weights. The flow chart of the implemented TSDR system is shown in Figure 1.

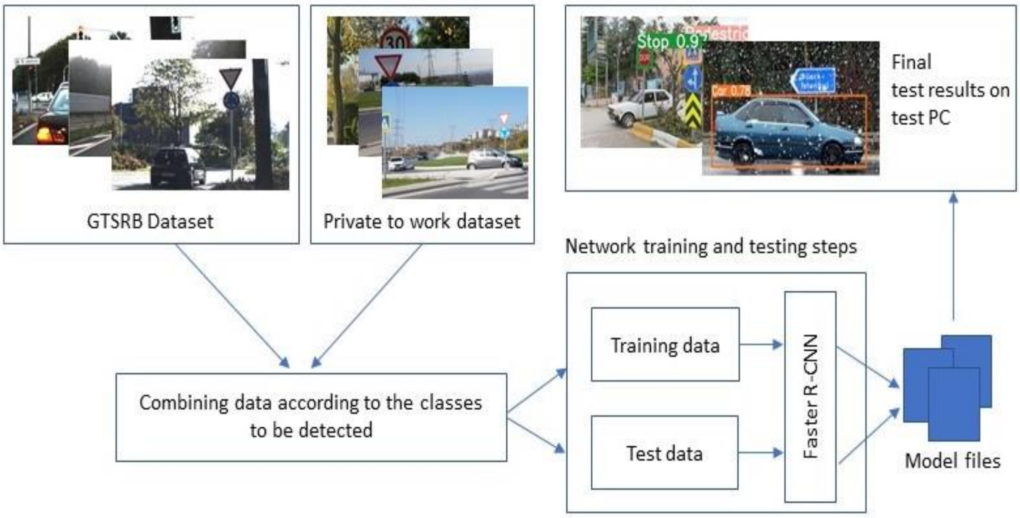

Figure 1 The implemented TSDR system architecture based on Faster R-CNN

2.1 Datasets

The dataset used in this article includes the German Traffic Sign Recognition Benchmark [6] (GTSRB) dataset, which is frequently used for TSDR studies in the literature. In addition to the original data, the dataset was expanded according to the classes to be detected with the data collected under various light and weather conditions. Figure 2 shows the images taken from the dataset used in the study.

Figure 2 Different images from our dataset

The dataset consists of 1880 images containing different traffic signs, vehicles and pedestrians. The data is primarily divided into various classes for detection and recognition. Afterwards, labelling processes were performed on these image data. The online software tool makesense.ai [19] was used while labelling all of the images in the datasets. In this program, the images were examined sequentially and the existing objects were selected with bounding boxes. A sample labeled image is given in Figure 3.

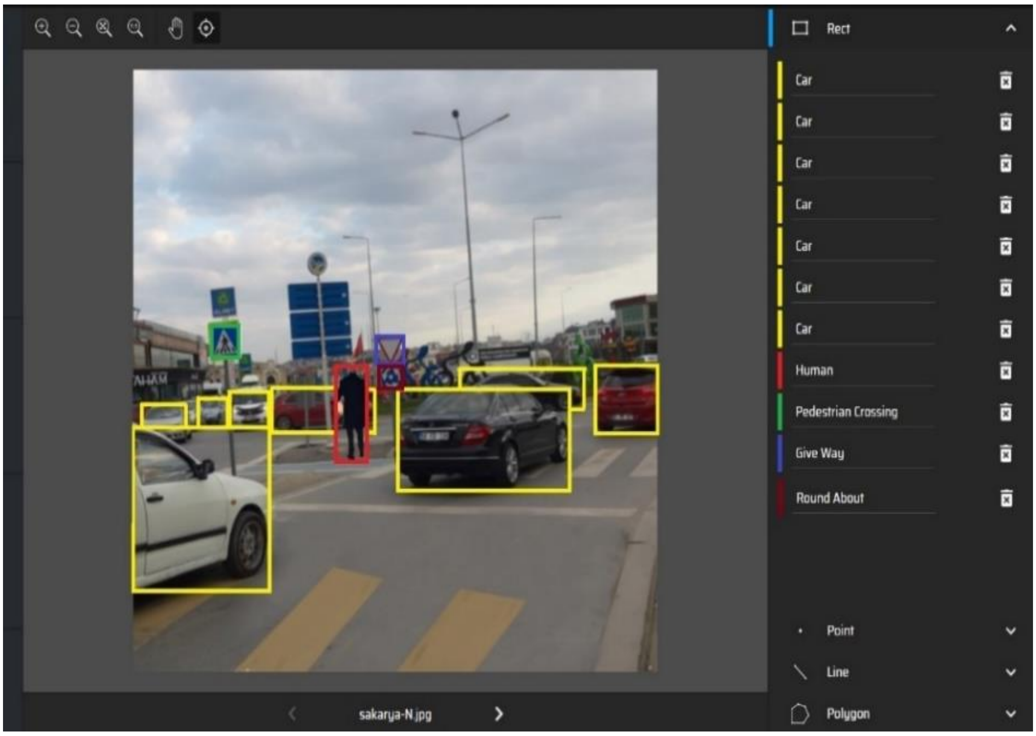

Figure 3 A sample image from the labeling process

Makesense.ai is a free-to-use online tool for labeling images. It makes the dataset preparation process easy and fast. Since the study performance is directly related to these images to be trained, it is critical to choose the corrected objects in this process. However, images in datasets contain multiple objects. Additionally, class names and distributions used in this study are shown in Table 2. After all these data processes, the images are made ready to train the model to be created and test the results by measuring.

Table 2 Brief description of our dataset. It contains information about the used classes and their amounts

Sign name	**No. of Data**	**Sign name**	**No. of data**
Speed Limit 20	40	Slippery Road	50
Speed Limit 30	114	Road Narrows	53
Speed Limit 50	71	Construction	50
Speed Limit 100	50	School Crossing	34
No Overtaking	50	Go Right	52
Priority Road	77	Go Left	42
Give Way	126	Go Straight	30
Stop	50	Roundabout	58
No Entry	50	Pedestrian Crossing	91
Danger	50	Human	267
Bend	51	Bicycle	106
Uneven Road	52	Car	220
Priority at Next Intersection	46		

2.2 Object Detection Models

Object detection is finding and classifying objects in the images. There are many approaches for object detection in terms of accuracy and speed. These approaches for the detection can be divided into two groups, and Figure 4 shows this two-stage and one-stage detection. Two-stage sensors such as Mask R-CNN [20] or Faster R-CNN use a RPN network to generate regions of interest (RoI) in the first step. It sends region proposals for object classification and bounding box regression in a pipeline. Although these kinds of models have high accuracy, they generally run slower than single-stage object detection models [21]. On the other hand, single-stage object detectors, in which an input image is predicted with a single pass through the neural network, have a higher prediction speed. Single-stage models such as YOLO [22] and SSD [23]achieve lower accuracy rates.

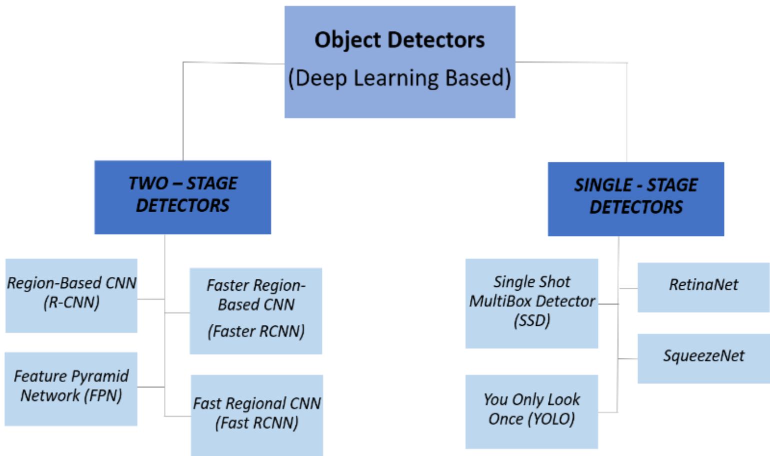

Figure 4 Classification of object detection approaches

2.3. Faster R-CNN

Faster R-CNN [24] was presented by Girshick et al. in 2015. This object detection model uses Convolutional Neural Networks (CNN) to classify the image in two stages. This region-based model takes a more state-of-the-art approach to accelerate region proposals similar to R-CNN [25] and Fast R-CNN [21]. Previous models used the Selective Search Algorithm[26] to find region proposals. However, this algorithm is a time-consuming process that negatively affects network performance. For this reason, the Faster R-CNN algorithm has been developed, which improves the performance of the network by eliminating Selective Search Algorithm and allows it to learn region recommendations. A schematic image from the RCNN, Fast RCNN, and Faster R-CNN models is given in Figure 5. Figure 5 shows the working principle of the Faster R-CNN approach, similar to Fast R-CNN. Region suggestions estimate offset values to classify the image within the proposed region.

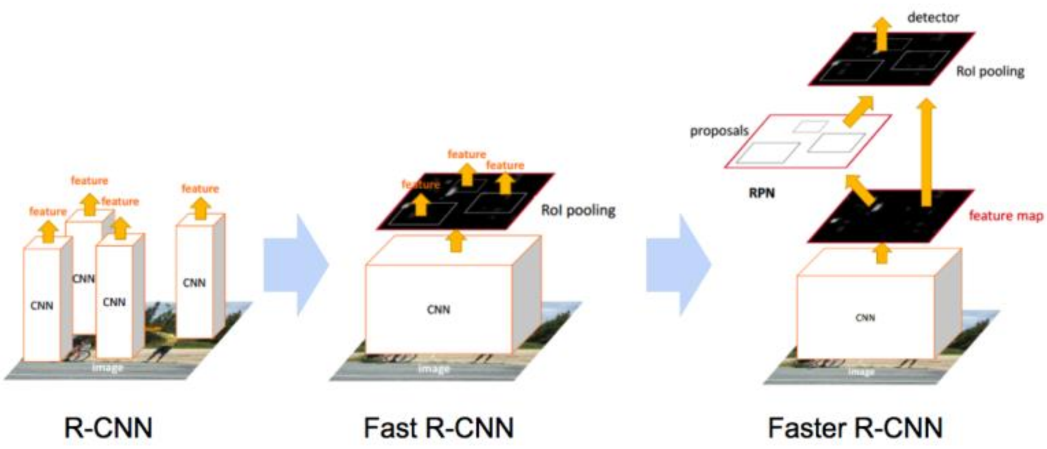

Figure 5 An illustration of R-CNN, Fast R-CNN and Faster R-CNN models [27].

3. Experimental Results

The experimental results was performed on the test computer given in Table 3. The dataset, consisting of 1880 images, is divided 80% to 20% for training and validation processes. Test procedures were carried out in a computer environment for 8.5 hours and approximately 10000 iterations. The training continued until the threshold value reached the appropriate value for the detection accuracy. The results show that the study with Faster R-CNN is suitable for the real-time detection of traffic signs and objects.

Table 3 Characteristics of the test environment used for the model performance

Name	Properties
Operation System (OS)	Windows 10
Central Processing Unit (CPU)	Intel Core i7 6700 HQ
Graphics Processing Unit (GPU)	Nvidia Geforce GTX 950 M 2 GB
Memory (RAM)	8 GB DD4
Harddisk (HDD)	1 TB HDD

The various loss graphics data are recorded in Table 4. As can be easily seen from Table 4, classification, localization, and total loss, etc., both training and validation rates appear to decrease consistently on a large scale. As the number of iterations increases, the loss rate decrease clearly shows the model's performance. Also, this decrement means that the model measurements are correct. In this study, the model's performance was measured according to the loss rates from loss graphics. For example, a loss rate of 0.144 for classification ($L_{class.}$) and a total loss rate (L_{total}) of 0.220 were recorded in 10000 iterations, as given in Table 4.

Table 4 Loss rates for classification and other parameters according to number of iterations

		Number of iterations									
		1k	**2k**	**3k**	**4k**	**5k**	**6k**	**7k**	**8k**	**9k**	**10k**
Loss rates	$L_{class.}$	0.605	0.376	0.337	0.253	0.447	0.257	0.329	0.245	0.228	**0.144**
	$L_{local.}$	0.760	0.642	0.202	0.259	0.264	0.191	0.208	0.152	0.349	**0.090**
	$L_{object.}$	0.028	0.050	0.009	0.008	0.027	0.015	0.010	0.040	0.017	**0.003**
	L_{clone}	1.346	1.037	0.562	0.463	0.651	0.706	0.496	0.518	0.507	**0.231**
	L_{total}	1.325	1.028	0.555	0.456	0.645	0.712	0.504	0.317	0.495	**0.220**

The comparison of the achievements obtained from the study with the literature is shown in Table 5. The first of the compared studies was presented by Qian et al. [4], a deep Convolutional Neural Networks (CNN) system to detect traffic signs. The other is the SSD and Faster RCNN based system presented by Öztürk et al. [14]. When we look at the other studies in the literature, compared to the accuracy values obtained for the TSDR task performed with datasets with three classes and 10 class labels, a success value of 88.99% was calculated at 1500 iterations from test images with 25 classes in the presented study.

Table 5 The detection accuracies for traffic sign detection and recognition task obtained by other studies

Study	**Qian et al. [4]**	**Öztürk et al. [14]**	**Our study**
Accuracy	95%	85.1%	**88.99%**
Used dataset	GTSRB, MNIST, CASIA	COCO and GRAZ	GTSRB and Private dataset
Class no.	3 objects	10 objects	**25 objects**

The dataset was analyzed by combining the data with the study-specific dataset added to reproduce the GTSRB dataset for the test process. Also, the data that was previously reserved at the rate of 20% was used in this process. The dataset used in this study was created concerning the classes to be specified. Some of the results regarding the detection process from the test results are shown in Figure 6. While some of the test results from the dataset were estimated correctly, some estimates were inconsistent as the images of the traffic signs were similar and the data diversity was insufficient.

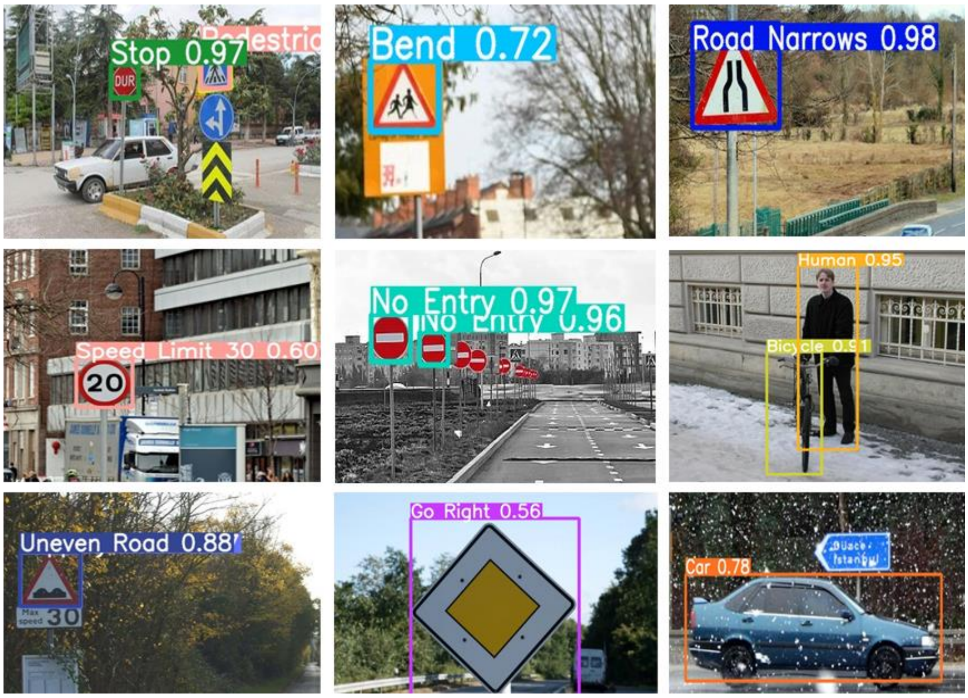

Figure 6 Test images from the developed model

4. Conclusion and Discussion

Traffic sign detection and recognition systems have been a challenging problem for many years. Several methods and technologies have been developed for its solution in the past. This paper proposes a two-stage deep learning method to detect and recognize traffic signs and specific objects. For this purpose, Faster R-CNN, one of the two-stage detection methods, was used in the study. This method is frequently used in the literature for real-time object detection.

We detected 22 different traffic signs, people, bicycles and vehicles. Model was created by training the system with the dataset specifically created for working in the computer environment and the GTSRB dataset. Then, this model was tested with different images. The performance analysis of the proposed method was performed and the results were obtained. The training continued until the threshold value reached the appropriate value for the detection accuracy. Also, testing results achieve the best accuracy of 88.99% at 1500 iterations for the real-time TSDR performance. The results show that Faster R-CNN can be efficiently used in the real-time detection of traffic signs and objects.

In future studies, it is planned to develop some TSDR methods for enhance the system performance. In addition, the dataset used can be expanded to increase accuracy and performance. Also, this article focuses on traffic signs and object detection. However, considering the applications in traffic, lane violations can be added to this system. In addition, the system can be tested on an embedded device and contribute to mobility.

Acknowledgments

We want to thank our hardworking teammates Neslihan ÇAKIRBAŞ, Havva Selin ÇAKMAK, Ali Göktuğ YALÇIN and Dilara KOCA. They helped us prepare the datasets and the development of the system. This study was carried out with Sakarya University Scientific Research Projects Coordination Unit (Project No: 2021-7-24-20).

References

[1] A. Ruta, Y. Li, and X. Liu, “Real-time traffic sign recognition from video by class-spec. discriminative features,” Pattern Recognition, vol. 43, no. 1, pp. 416–430, 2010.

[2] H. Li, F. Sun, L. Liu, and L. Wang, “Neurocomputing A novel traffic sign detection method via color segmentation and robust shape matching,” Neurocomputing, vol. 169, pp. 77–88, 2015.

[3] S. Yin, P. Ouyang, L. Liu, Y. Guo, and S. Wei, “Fast Traffic Sign Recognition with a Rotation Invariant Binary Pattern Based Feature,” pp. 2161–2180, 2015.

[4] R. Qian, B. Zhang, Z. Wang, and F. Coenen, “Robust Chinese Traffic Sign Detection and Recognition with Deep Convolutional Neural Network,” pp. 791–796, 2015.

[5] X. Changzhen, W. Cong, M. Weixin, and S. Yanmei, “A Traffic Sign Detection Algorithm Based on Deep Convolutional Neural Network,” pp. 6–9, 2016.

[6] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, “The German Traffic Sign Recognition Benchmark: A multi-class classification competition,” Proceedings of the International Joint Conference on Neural Networks, pp. 1453–1460, 2011.

[7] J. Zhang, M. Huang, X. Jin, and X. Li, “A Real-Time Chinese Traffic Sign Detection Algorithm Based on Modified YOLOv2,” pp. 1–13, 2017.

[8] Y. Bengio and P. Haffner, “Gradient-Based Learning Applied to Document Recognition,” vol. 86, no. 11, 1998.

[9] C. Liu, F. Yin, D. Wang, and Q. Wang, “Chinese Handwriting Recognition Contest,” 2010.

[10] M. Mathias, R. Timofte, R. Benenson, and L. van Gool, “Traffic sign recognition - How far are we from the solution?,” Proceedings of the Int. Joint Conference on Neural Networks, 2013.

[11] T.-Y. Lin et al., “LNCS 8693 - Microsoft COCO: Common Objects in Context,” 2014.

[12] “INRIA Annotations for Graz-02 (IG02).” https://lear.inrialpes.fr/people/marszalek/data/ig02/ (accessed Nov. 20, 2021).

[13] X. Xu, J. Jin, S. Zhang, L. Zhang, S. Pu, and Z. Chen, “Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry,” Future Generation Computer Systems, vol. 94, pp. 381–391, 2019.

[14] G. Ozturk, R. Koker, O. Eldogan, and D. Karayel, “Recognition of Vehicles, Pedestrians and Traffic Signs Using Convolutional Neural Networks,” Oct. 2020.

[15] C. Han, G. Gao, and Y. Zhang, “Real-time small traffic sign detection with revised faster-RCNN,” Multimedia Tools and Applications, vol. 78, no. 10, pp. 13263–13278, May 2019.

[16] K. Zhou, Y. Zhan, and D. Fu, “Learning region-based attention network for traffic sign recognition,” Sensors (Switzerland), vol. 21, no. 3, pp. 1–21, 2021.

[17] F. Shao, X. Wang, F. Meng, J. Zhu, D. Wang, and J. Dai, “Improved faster R-CNN traffic sign detection based on a second region of interest and highly possible regions proposal network,” Sensors (Switzerland), vol. 19, no. 10, May 2019.

[18] X. Dai et al., “Multi-task faster R-CNN for nighttime pedestrian detection and distance estimation,” Infrared Physics and Technology, vol. 115, Jun. 202.
[19] “Make Sense,” Victorian Studies, vol. 48, no. 3. pp. 395–438, 2006.
[20] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN.”
[21] R. Girshick, “Fast R-CNN,” Proceedings of the IEEE International Conference on Computer Vision, vol. 2015 Inter, pp. 1440–1448, 2015.
[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object sdetection,” IEEE Computer Vision and Pattern Recognition, pp. 779–788, 2016.
[23] W. Liu et al., “SSD: Single shot multibox detector,” Lecture Notes in Computer Science, vol. 9905 LNCS, pp. 21–37, 2016.
[24] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”.
[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation.” pp. 580–587, 2014.
[26] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, “Selective search for obj recog,” International Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, 2013.
[27] “Flagly.” https://www.flagly.org/project/projects/4/sections/42/ (Accessed Dec. 15, 2021).