

# *CCTSDB 2021: A more comprehensive traffic sign detection benchmark*

Article

Published Version

Creative Commons: Attribution-Noncommercial 3.0

Open Access

Zhang, J., Zou, X., Kuang, L.-D., Wang, J., Sherratt, R. S. ORCID: <https://orcid.org/0000-0001-7899-4445> and Yu, X. (2022) CCTSDB 2021: A more comprehensive traffic sign detection benchmark. *Human-centric Computing and Information Sciences*, 12. 23. ISSN 2192-1962 doi: <https://doi.org/10.22967/HCIS.2022.12.023> Available at <https://centaur.reading.ac.uk/106129/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Identification Number/DOI: <https://doi.org/10.22967/HCIS.2022.12.023>  
<<https://doi.org/10.22967/HCIS.2022.12.023>>

Publisher: Springer Berlin Heidelberg

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Human-centric Computing and Information Sciences

May 2022 | Volume 12



[www.hcisjournal.com](http://www.hcisjournal.com)



# CCTSDB 2021: A More Comprehensive Traffic Sign Detection Benchmark

Jianming Zhang<sup>1</sup>, Xin Zou<sup>1</sup>, Li-Dan Kuang<sup>1</sup>, Jin Wang<sup>1</sup>, R. Simon Sherratt<sup>2</sup>, Xiaofeng Yu<sup>3,\*</sup>

## Abstract

Traffic signs are one of the most important information that guide cars to travel, and the detection of traffic signs is an important component of autonomous driving and intelligent transportation systems. Constructing a traffic sign dataset with many samples and sufficient attribute categories will promote the development of traffic sign detection research. In this paper, we propose a new Chinese traffic sign detection benchmark, which adds more than 4,000 real traffic scene images and corresponding detailed annotations based on our CCTSDB 2017, and replaces many original easily-detected images with difficult samples to adapt to the complex and changing detection environment. Due to the increase of the number of difficult samples, the new benchmark can improve the robustness of the detection network to some extent compared to the old version. At the same time, we create new dedicated test sets and categorize them according to three aspects: category meanings, sign sizes, and weather conditions. Finally, we present a comprehensive evaluation of nine classic traffic sign detection algorithms on the new benchmark. Our proposed benchmark can help determine the future research direction of the algorithm and develop a more precise traffic sign detection algorithm with higher robustness and real-time performance.

## Keywords

Intelligent Transportation Systems, Traffic Sign Detection Benchmark, Object Detection, Traffic Weather

## 1. Introduction

With the accelerated urbanization and rapid growth in the number of motor vehicles in China, urban road traffic throughput is increasing, and various traffic problems have emerged. Intelligent transportation systems (ITS) [1] can effectively use existing transportation facilities to ensure traffic safety and improve transportation efficiency, and is the main direction for the development of future transportation systems. Traffic sign detection, an important part of intelligent driving system, possesses high requirements for real-time, precision and robustness. Traffic sign detection is to predict whether a given image contains a traffic sign or not, and to perform coarse classification and localization of the

\* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

\*Corresponding Author: Xiaofeng Yu ([xiaofengyu@nju.edu.cn](mailto:xiaofengyu@nju.edu.cn))

<sup>1</sup>Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

<sup>2</sup>Department of Biomedical Engineering, School of Biological Sciences, University of Reading, Reading, UK

<sup>3</sup>School of Business, Nanjing University, Nanjing, China

sign. Though traffic sign detection has been studied for decades and has been made some progress in recent years, but it has been still a challenging problem to achieve autonomous driving [2].

In order to fully evaluate the algorithm performance, it is crucial to collect representative datasets [3, 4]. There are many factors that affect the performance of traffic sign detection, including lighting variations, complex backgrounds, object occlusions, weather variations, and different countries, and there is no single traffic sign dataset that can successfully encompass all scenarios [5–7]. At present, there are several datasets for traffic sign detection including GTSD (German Traffic Sign Detection Benchmark) [8], LISATSD (Laboratory for Intelligent and Safe Automobiles Traffic Sign Dataset) [9], BTSD (Belgian Traffic Signs Dataset) [10], STSD (Swedish Traffic Signs Dataset) [11], CTSD (Chinese Traffic Sign Dataset) [12], T-T 100K (Tsinghua-Tencent 100K) [13], CCTSDB 2017 (CSUST Chinese Traffic Sign Detection Benchmark) [14], etc. However, most of these datasets suffer from small data volume, insignificant weather changes, incomplete annotation information, single image style, no dedicated test dataset, and some of them are not publicly available [15, 16]. Most existing traffic sign detection datasets do not pay enough attention to these problems or do not fully address them.

Many typical deep learning methods and their extended algorithms have been applied to traffic sign detection, such as R-CNN [17–21], YOLO [22–24], SSD [25]. The detection performance of these algorithms varies for different attributes of the images. Besides many studies use different and few performance metrics, and thus cannot fully reflect the detection performance of the algorithms [26, 27]. Therefore, it is necessary to unify the metrics for performance evaluation, to comparatively evaluate performance of algorithms.

As such, based on our previous benchmark CCTSDB 2017, we aim to generate a new benchmark CCTSDB 2021 for traffic sign detection in China. CCTSDB 2021 contains different image attributes (e.g., different categories, different sizes, weather variations, etc.) so as to restore the real environment of the detection scene as much as possible. The contributions of our paper are threefold:

(1) On the basis of CCTSDB 2017, we add and annotate 5,268 new images of real traffic scenes including 3,268 training set images and 2,000 test set images. While expanding the amount of data, a part of easy samples from the old benchmark dataset is replaced to make the trained neural network more robust. The data and code are available at <https://github.com/csust7zhangjm/CCTSDB2021>.

(2) We generate a new comprehensive and dedicated test set which is categorized according to three dimensions: category meanings (three types), weather conditions (six types), and sign sizes (five types). By virtue of our new test set, the experimental comparison can be fairer.

(3) Nine different algorithms are evaluated on the new benchmark to show the strengths and weaknesses of the algorithms and to promote the development of new traffic sign detection algorithms. We use a unified performance evaluation metric, with six groups including precision, recall rate, miss rate, mAP,  $F_1$  and FPS (frames per second). These metrics allow the algorithms to be compared across the board in new dataset tests.

## 2. Related Work

### 2.1 ITS and Autonomous Driving

ITS aim to make effective use of existing traffic facilities, analyze and process various traffic information, and transmit effective traffic information among vehicles, drivers, pedestrians and various traffic facilities, so as to reasonably plan traffic routes, reduce traffic load and environmental pollution, ensure traffic safety and improve transportation efficiency [28–30]. In September 2019, China issued “The Outline for Building China’s Strength in Transport,” emphasizing to accelerate infrastructure construction and improving the capacity of the transportation system.

Autonomous driving is an indispensable step in ITS and the main development direction of global automobile manufacturers and transportation field at present and in the future [31–33]. Google obtained the first self-driving vehicle license in the United States in May 2012, and the original autonomous

driving team of Google was split into a subsidiary named Waymo at the end of 2016. In December 2015, Baidu launched an autonomous driving road test and announced the Apollo plan in April 2017. In February 2020, China's Ministry of Industry and Information Technology and other 10 ministries and commissions jointly issued "The national intelligent vehicle innovation and development strategy" to promote the construction of industry clusters for autonomous vehicles key parts.

Traffic sign detection is an important part of the autonomous driving system [34, 35]. It requires comprehensive application of various technologies such as machine vision, artificial intelligence, image processing, and so on. It has high requirements for real-time, accuracy and robustness [36–38]. A vehicle-based traffic sign detection system uses the camera mounted on the vehicle to collect the surrounding real traffic scene, and accurately predict the location and coarse category of traffic signs. These detection results will directly affect the fine-grained classification of traffic signs, and thus affect the vehicle control of autonomous driving system [39–41].

## 2.2 Traffic Sign Detection Algorithms

At present, the traffic sign detection technology is mainly divided into two kinds: traffic sign detection based on traditional methods [42–44] including HOG+SVM [45], RBD [46], SRM [47], ICF [48], etc., and another kind of traffic sign detection based on deep learning [49, 50]. The traditional detection methods are mainly based on the inherent physical characteristics of object being detected, including color and shape-based detection [51, 52]. Based on the image color and shape information, these methods select the features of the regions, and then output the regions of interest that may contain traffic signs. However, the traditional detection method is a slightly cumbersome process and lacks in real-time. The models of artificial intelligence and deep learning that have emerged in recent years can use pixels directly as input to the model, without pre-processing operations on the image. These methods can achieve automatic extraction of object features, and predict the output to know the presence or absence of the object and to get information about its location.

The success of deep learning-based object detection [53–55] can be attributed to the robustness of detection models, increased computational power, and the availability of large amounts of labeled data. Various novel convolutional approaches [56, 57] have been exploited. More and more effective neural network structures [58–60] have being explored. The fusion of multi-scale features [61, 62] can take full advantage of the small object feature [63–65]. The emergence of residual networks [66] solves the problem of network degradation. The authors of [67] proposed SPPNet to solve the problem of requiring a fixed input image size for feature extraction by CNN networks. [68] feeds the topmost layer of feature images in the network layer by layer and fuses them with the feature maps of the previous layer. [69] is able to describe the shape of an object by modeling the relative geometric positions of points and thus capturing local shape features. [70] obtains better voxel feature encoding methods by mixing voxel feature encoders of different scales at the point level, which results in speed and accuracy improvements.

## 2.3 Traffic Sign Detection Datasets

The GTSDB was proposed in 2013 and is one of the most widely-used benchmark for evaluating the performance of traffic sign detection algorithms. At the same time, GTSDB is also an internationally recognized measurement data set with high credibility. However, the data volume of this benchmark is small, which is not conducive to deep learning training. The LISATSD is a set of videos and annotated frames containing US traffic signs. It includes 47 US sign types with a total of 6,610 images, with traffic sign sizes ranging from  $6 \times 6$  to  $167 \times 168$ , some images are in color and some are in grayscale. However, the category of this dataset is not well defined and the image resolution is too low. The STSD was created by recording a total of 350 km of roads and cities in Sweden, creating a dataset of about 20,000 images, but with only 20% of the data labeled. The CTSD was produced by the machine vision group of the Institute of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, and is an earlier

recognized dataset in China, but there are problems such as small amount of data and inaccurate labeling of individual images. The high-resolution dataset T-T 100K was proposed by Tsinghua University, the number of datasets is at the level of 100,000, 10 regions in five different cities in China were selected and 100,000 panoramas were obtained from Tencent's data center. However, the panorama has distortion problems, the street view does not have extreme weather, and the percentage of positive samples is less than 10%. The CCTSDB 2017 was produced by a team from Changsha University of Technology in 2017. It is a relatively new benchmark in China, and there is an adequate number of benchmarks with the presence of multiple resolution images. However, there are still problems such as insufficient attribute categories and no dedicated test set.

With the development of the times, computer hardware is constantly being updated and the computing power of GPU hardware has increased significantly. However, the relatively slow development of traffic sign datasets has hindered the development of traffic sign detection to a certain extent, and most of the dataset production teams do not continuously update and adjust the datasets. Therefore, we produced a data-rich detection benchmark of Chinese traffic sign with a large amount and variety of data. Meanwhile, the new benchmark is comprehensively evaluated with classical algorithms in object detection to lay a solid foundation for the development of traffic sign detection technology.

### 3. Our Traffic Sign Detection Benchmark

In this section, we present a large-scale traffic sign detection benchmark called CCTSDB 2021. CCTSDB 2021 is an expansion of CCTSDB 2017 in which we capture, process and label new images. We remove the images with incomplete annotation information in CCTSDB 2017 and added 5,268 images, and generate a dedicated test set. Next, we describe the process of creating the benchmark dataset and the statistical information related to the benchmark dataset.

#### 3.1 Image Selection and Annotation

We randomly collected more than one thousand car recorder videos that have been publicly available online for different time periods, locations, and speeds. Thus, the coverage and diversity of our benchmark are well ensured. As the vehicle is equipped with different models of car recorders, so the length of each video captured varies from 15 seconds to 10 minutes, respectively, and the resolution of the videos is  $860 \times 480$ ,  $1280 \times 720$ ,  $1920 \times 1080$ , etc. The frame rate varies from 25 to 30 frames, and the bit rate format of the original videos varies between 2,000 kbps and 10,000 kbps. We retain original resolution and bit rate of the original videos to keep the diversity of the data.

Due to data quality, diversity, and sample issues, some videos are not suitable for use in the study, and the line-of-sight videos need to be manually screened. We get a total of 423 videos containing traffic signs, and the videos containing traffic signs are divided into frames with a frame skip interval of 5, i.e., one image is saved every 5 frames. The images saved by the split-frame operation are then manually filtered once again, keeping only those images that contain traffic signs. The final filtered traffic sign images are labeled and positioned by sign category. For the quality of labeling, all our images are manually labeled and positioned on LabelImg software. Finally, we restore all the annotated coordinates to the image and manually proofread them one by one to prevent any annotation or processing errors. As shown in Fig. 1, we standardize the benchmark production process into six steps.

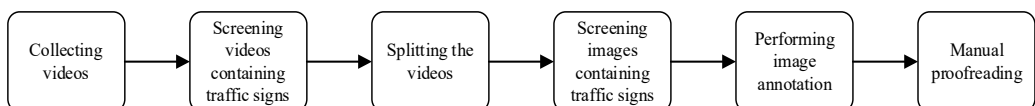


Fig. 1. Production process of CCTSDB 2021.

The annotation information format is shown in Fig. 2. The information in the red box in Fig. 2 is the annotation information that we use when conducting experiments, and we explain some of the important information. The *annotation* tag indicates the annotation file of an image, and the *filename* tag information indicates the file name. When there are  $n$  traffic signs in an image, there are  $n$  *object* tags. The *name* tag indicates the category meaning of the traffic sign, the *bndbox* tag indicates the position of the traffic sign in the image. The *xmin* tag indicates the horizontal coordinate of the top left corner of the traffic sign bounding box. The *ymin* tag indicates the vertical coordinate of the top left corner of the traffic sign bounding box. The *xmax* tag indicates the horizontal coordinate of the top left corner of the traffic sign bounding box. The *ymax* tag indicates the vertical coordinate of the lower right corner of the traffic sign bounding box.

```

▼ <annotation verified="no">
  <folder>images</folder>
  <filename>00022</filename>
  <path>J:\360\Foggy\images\00022.png</path>
  ▶ <source>...</source>
  ▶ <size>...</size>
  <segmented>0</segmented>
  ▼ <object>
    <name>mandatory</name>
    <pose>unspecified</pose>
    <truncated>0</truncated>
    <Difficult>0</Difficult>
    ▼ <bndbox>
      <xmin>38</xmin>
      <ymin>270</ymin>
      <xmax>64</xmax>
      <ymax>316</ymax>
    </bndbox>
  </object>
</annotation>

```

Fig. 2. Annotation information of the image.

## 3.2 Training Set Statistics

CCTSDB 2021 is an expansion of CCTSDB 2017, with a total of 16,356 images. All training images are located in the Train folder in JPG format, where the first 13,087 images are all from CCTSDB 2017 and the last 3,269 images are added samples. To enrich the training set information, we acquire images from six different weather conditions. There were 22 images acquired from foggy days, 60 images from snowy days, 204 images from rainy days, 518 images at night, 1,201 images in cloudy weather, and 1,264 images in sunny days. There are three types of traffic signs in the training set, including 13,876 prohibition signs, 4,598 warning signs and 8,363 mandatory signs. As shown in Fig. 3, four sample images in the training set are displayed.

## 3.3 Test Set Statistics

In CCTSDB 2021, we produce a dedicated test set. More specifically, there are 2,000 images in total, and the first 1,500 images are positive sample images and the last 500 are negative sample images. All test images are located in the Test folder in JPG format. In addition, we divide the positive samples in the test set in more detail according to three dimensions: category meanings, sizes, and weather conditions, as detailed below.

### 3.3.1 Classification based on meaning of traffic signs

The meaning of traffic signs varies from country to country, and the correct classification of traffic sign meanings is the most basic requirement for detection algorithms. According to the definition of common traffic signs in road traffic signs and markings, we divide the signs appearing in the benchmark dataset into the following three categories according to their meanings, as shown in Fig. 4.





Fig. 3. Some examples of training set in CCTSDB 2021.



Fig. 4. Some Chinese traffic signs with meaningful classification.

A prohibition signs prohibit or restricts a certain traffic behavior of vehicles or pedestrians. The main colors of the prohibition sign are red, black and white, with a small amount of dark blue, and the shape of the sign is usually circular or square octagonal. Warning signs warn vehicles and pedestrians of dangerous locations ahead of the road. The main color scheme of the sign is yellow and black, and the shape of the sign is usually a square triangle. Mandatory signs indicate the movement of vehicles and pedestrians. The main color scheme of the sign is blue and white, and the shape of the sign is usually rectangular or circular. As shown in Fig. 5, there are 3,228 traffic signs in the whole test set, including 2,177 prohibition signs, 718 mandatory signs and 333 warning signs according to their meanings.

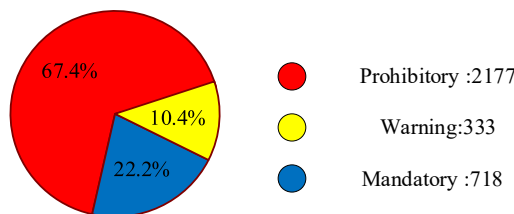


Fig. 5. Proportion of three types of traffic signs.

### 3.3.2 Classification based on size of traffic signs

In general, the performance of the detection algorithm decreases as the size of the inspected object becomes smaller. For traffic participants, however, it is often the smaller traffic signs at a distance that are more suggestive. If a traffic sign is close, the driver is often too late to react or already knows the meaning of the sign when driving according to it. Therefore, the traffic sign detection algorithm needs to correctly detect more and more traffic signs that are smaller and further away under limited conditions.

We counted the sizes of all flags in the full test set sample, divided them equally into four intervals, and then divided the interval with the largest size into two intervals. Because the larger the size of traffic signs can be easier to detect them relatively, we set smaller number of samples for the larger the size. The final sizes of the CCTSDB 2021 were classified into five categories: access small (XS), small (S), medium (M), large (L), and extra-large (XL). There were 813 XS size signs, 807 S size signs, 828 M size signs, 408 L size signs, and 372 XL size signs in the test set according to the traffic sign size. T-T 100K divides the size of traffic signs into three intervals: small, medium, and large. Their corresponding pixel area sizes are shown in Table 1. From the table, we can see that we subdivide three more size categories within the small object range corresponding to T-T 100K, so the CCTSDB 2021 focuses more on the detection of small objects.

**Table 1.** Cropped image size comparison between T-T 100K and CCTSDB 2021

T-T 100K	CCTSDB 2021 (number of traffic signs)
small, area $\leq 32^2$ pixels	XS, area $\leq 210$ pixels (813)
	S, area $> 210$ pixels and area $\leq 400$ pixels (807)
	M, area $> 400$ pixels and area $\leq 1000$ pixels (828)
medium, area $> 32^2$ pixels and area $\leq 96^2$ pixels	L, area $> 1000$ pixels and area $\leq 2000$ pixels (408)
large, area $> 96^2$ pixels	XL, area $> 2000$ pixels (372)

**Table 2.** Characteristics of various types of weather

Weather classification	Number of images	Image characteristics
Sunny	400	There is strong sunlight, the direction of the light source is opposite to the direction of travel, the traffic signs are directly illuminated by sunlight, and the object light is clear.
		There is strong sunlight, the direction of the light source is the same as the direction of travel, the front of the traffic sign is not directly illuminated by sunlight, and the target is dark and blurred.
Cloud	300	No obvious direct sunlight, no obvious light source direction, the object is relatively clear.
Night	500	Since the traffic signs are coated with special fluorescent materials, the direct illumination of vehicle lights on the traffic signs at night will make a clear contrast between the traffic signs and the background.
		Due to the different vehicle angles, the vehicle lights are scattered on the traffic signs and the light intensity is insufficient, making the overall traffic signs dark and blurred.
Snow	100	There is snow on the ground, the image is easily overexposed, and the traffic signs are white.
Foggy	40	There is water mist in the air, overall white, and blurred traffic signs.
Rain	160	The overall environment is dark, the ground is easily reflective, and the front windshield and traffic signs appear blurred with rainwater.

### 3.3.3 Classification based on weather and environment

After checking some car recorder videos, we found some traffic sign detection problems related to weather and light environment. Extreme weather conditions (such as rain, snow or fog) can temporarily degrade the image quality of a car recorder, and dim light, overexposure and glare can have a negative

effect on the visibility of traffic signs. Therefore, we consider the impact of changes in weather conditions on traffic sign detection at specific times and in specific areas.

We divided the weather lighting conditions of all sample images in the test set into six categories: foggy, snow, rain, night, sunny, and cloud, for a total of 1,500 images. It is worth noting that the weather environment attribute is only present in the test set images, and is not present in the training set. As shown in Table 2, the test images will show different characteristics under different weather as well as lighting conditions, which is a great challenge for the detection algorithm. Examples of the data set is shown in Fig. 6.



Fig. 6. Examples of weather environment classification.

In Fig. 7, we counted the number of each sign in the test set data for different weather environment attributes. There are 579 sunny signs, 655 cloudy signs, 1,279 night signs, 488 rainy signs, 61 foggy signs, and 166 snowy signs according to the weather environment classification. The proportion of all kinds of weather is different in real life, and the proportion of all kinds of images in the dataset is also different.

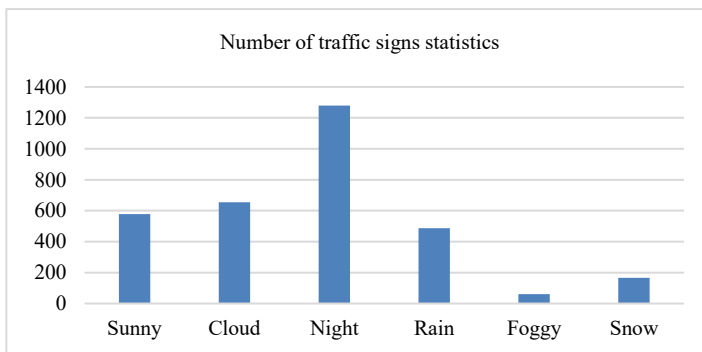


Fig. 7. Weather environment classification statistics of traffic signs.

## 4. Experiments

The task of our neural network is to detect the location [71] of the traffic sign in the image and then discriminate which of the prohibition signs, mandatory signs, and warning signs the object belongs to. We evaluated nine representative detection algorithms on the CCTSDB 2021, all of which were network trained on the training set and evaluated on a dedicated test set. Among the detected algorithms, there are R-CNN related detection algorithms including Faster R-CNN [18], Libra R-CNN [72], Dynamic R-CNN [19] and Sparse R-CNN [73], YOLO related detection algorithms including YOLOv3 [23], YOLOv4 [24] and YOLOv5 [74], SSD related detection algorithms including SSD [25], and we also evaluated the first stage detection algorithms in RetinaNet [75].

### 4.1 Experimental Details

All code for training and testing the models was run in a Linux environment with Ubuntu 16.04, CUDA version 10.1, and the framework adopted for the experiments was PyTorch. The processor model is Intel Xeon CPU E5-2640 2.40 GHz, the graphics card model is GeForce RTX 2080 Ti, the graphics memory size is 11 G, and the memory size is 16 G. Among them, YOLOv4 and YOLOv5 were not tested on this platform due to the non-support of MMDetection, and all the remaining algorithms extract features from the dataset in JSON format [76] through the MMDetection platform to test the algorithm performance. Empirically the hyperparameters of various algorithms are set as follows. The initial learning rate of YOLOv4 is 0.001 and the weight decay parameter is 0.0005 for a total of 100 training batches, and the learning rate is adjusted to 0.001 again for the 51st training batch. YOLOv5 has an initial learning rate of 0.001, a momentum size of 0.98, and a weight decay parameter of 0.001, and is trained for a total of 50 batches. All algorithms running on MMDetection have a momentum size of 0.9. The Libra R-CNN, YOLOv3, and RetinaNet algorithms have an initial learning rate of 0.1, and the remaining algorithms have a momentum value of 0.2. The weight decay parameter for YOLOv3 is 0.0005, and the weight decay parameter for the remaining algorithms running on MMDetection is 0.0001. The Faster R-CNN, Libra R-CNN, Dynamic R-CNN, and Sparse R-CNN algorithms were trained in 12 batches uniformly, and the 8th and 11th batches reduced the learning rate to one-tenth of the original rate. The SSD algorithm was trained for 24 batches, and the 16th and 22nd batches reduced the learning rate to one-tenth of the original rate. YOLOv4 and YOLOv5 use the Adam optimization algorithm, and other algorithms tested on the MMDetection platform adopt the SGD strategy as the optimization algorithm.

### 4.2 Evaluation Metrics

Since the majority of algorithms refer to a single performance metric that is insufficient to fully reflect the detection performance of the algorithm, there is also a need to unify the metrics for performance evaluation on a completely new benchmark. TP denotes the number of samples that predicted positive samples correctly, FP denotes the number of samples that predicted negative samples as positive samples, and FN is the number of samples that predicted positive samples incorrectly. In the evaluation metric, FPS indicates the number of images processed per second. We use precision (P) [77], recall rate (R) [77], miss rate (MR) [77], mAP [77],  $F_1$  [77] and speed FPS to measure the performance of the proposed algorithm.  $P(R)$  is a function with R as the parameter, and classes is the number of meaning categories of the dataset. Therefore, the metric can be calculated according to the following formula:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$MR = 1 - \frac{TP}{TP + FN} \quad (3)$$

$$mAP = \frac{1}{classes} \sum_{i=1}^{classes} \int_0^1 P(R) dR \quad (4)$$

$$F_1 = \frac{2PR}{P + R} \quad (5)$$

### 4.3 Experimental Results

In order to comprehensively evaluate the dataset, we select six common metrics for evaluation. In the meanwhile, the selected algorithm is also a classical detection algorithm in one-stage and two-stage detection networks. In our experimental evaluation of traffic sign detection, a dedicated test benchmark dataset is used. This test set greatly increases the number of samples and attribute categories for each type of traffic signs and is able to measure them with the most demanding performance metrics. In the experiments, the threshold value of IOU (intersection over union) is selected as 0.5, and the experimental results are shown in Table 3, algorithms are sorted by time from top to bottom.

**Table 3.** Comprehensive test results of CCTSDB 2021

Method	P (%)	R (%)	MR (%)	mAP (%)	F <sub>1</sub>	FPS
Faster R-CNN [18]	84.43	54.98	45.02	56.58	0.60	4.87
SSD [25]	86.47	27.74	72.26	49.20	0.42	22.33
RetinaNet [75]	86.70	52.88	47.12	57.78	0.65	8.88
YOLOv3 [23]	84.63	42.71	57.29	50.48	0.54	20.34
Libra R-CNN [72]	83.72	60.04	39.96	61.35	0.70	8.81
YOLOv4 [24]	76.16	52.50	47.50	51.69	0.59	16.55
Dynamic R-CNN [19]	86.98	58.33	41.67	60.01	0.69	9.03
Sparse R-CNN [73]	94.12	52.58	47.42	59.65	0.67	8.45
YOLOv5 [74]	90.80	69.20	30.80	76.30	0.78	123.46

From the corresponding data measured from the nine models on the CCTSDB 2021 listed in Table 3, we learn that the results of each metric detected by this benchmark dataset are relatively low due to the addition of many difficult samples in CCTSDB 2021. Overall, the two-stage algorithms are more highly precise but slower than single-stage algorithms. The more outstanding value of precision is the two-stage detection algorithm Sparse R-CNN, with a P value of 94.12%. The fastest is the one-stage detection algorithm YOLOv5, which has an FPS value of 123.46. This is because the two-stage object detection algorithms first extract the candidate frame for the image, and then conduct a secondary correction based on the candidate region to get the detection point result, with higher detection precision, but slower detection speed. However, the single-stage detection algorithms directly calculate the image to generate the detection results, fast detection speed, but low detection precision. Besides, the leakage rate and recall rate are interrelated, and the sum of the two is always 1. Therefore, the lowest leakage rate is the YOLOv5 algorithm with the highest recall rate value. In addition, mAP is the average of the area under the curve drawn for each category using a combination of points with precision and recall rate, so the highest mAP value is 76.30% for the YOLOv5 algorithm, which has relatively high precision and recall rate. The F1 combines the results of precision and recall rate, and when the F<sub>1</sub> is higher, it indicates that the method is more effective. The highest F<sub>1</sub> value among the above algorithms is YOLOv5, which is 0.78.

In order to enrich the experimental results, we set different thresholds of IOU in the test and measured the corresponding values of mAP. We set the IOU thresholds in the range of 0.1 to 0.9, and the interval between the thresholds is 0.2. The experimental results are shown in Table 4, and from the results we can learn that for each detection model, the detected mAP values become smaller to some extent when the IOU threshold increases. For the network model, when the IOU threshold increases, the obtained mAP values decrease. This is because the increased IOU threshold makes the bounding box filtered by the model less, thus retaining the correct bounding box to a great extent. However, the stability of this algorithm is higher for traffic sign detection because the mAP of YOLOv5 decreases relatively slowly. When the IOU threshold is taken as 0.7, Libra R-CNN, Dynamic R-CNN, Sparse R-CNN and YOLOv5 algorithms still have some validity with mAP values greater than 50%. The YOLOv5 does not cause significant differences in the detected mAP values due to changes in the IOU threshold, and all values measured are greater than 70%. A large part of the reason is due to the auto learning bounding box anchors in the YOLOv5, which automatically learns the size of the anchor frame to improve detection performance to some extent since the object detection framework often requires scaling the original image size.

**Table 4.** Detection results of CCTSDB 2021 in different IOU thresholds (unit: %)

Method	IOU <sub>0.1</sub>	IOU <sub>0.3</sub>	IOU <sub>0.5</sub>	IOU <sub>0.7</sub>	IOU <sub>0.9</sub>
Faster R-CNN [18]	58.17	57.96	56.58	47.06	3.07
SSD [25]	61.56	58.99	49.20	29.54	3.19
RetinaNet [75]	58.61	58.45	57.78	49.46	5.24
YOLOv3 [23]	55.89	55.45	50.48	31.92	0.81
Libra R-CNN [72]	61.52	61.51	61.35	55.58	5.83
YOLOv4 [24]	70.80	68.45	51.69	11.20	0.13
Dynamic R-CNN [19]	60.10	60.07	60.01	55.86	6.06
Sparse R-CNN [73]	60.38	60.15	59.65	52.87	4.41
YOLOv5 [74]	75.40	76.90	76.30	76.50	72.10

Tables 5–7 show the experimental results in various cases after dividing the test set, but mAP represents the average value of the average accuracy of all categories, so it is more appropriate to use the pair of metrics P and R in Tables 5–7. The CCTSDB 2021 provides a coarse classification of traffic sign detection, and we classify all traffic signs into three major categories: prohibitory, warning, and mandatory. Among them, 67.4% were prohibition signs, 10.4% were warning signs, and 22.2% were mandatory signs. The data results of precision and recall rate of each category measurement when the IOU threshold was taken as 0.5 are shown in Table 5.

**Table 5.** Detection results of CCTSDB 2021 in different meaning categories (unit: %)

Method	Prohibitory		Warning		Mandatory	
	P	R	P	R	P	R
Faster R-CNN [18]	90.60	55.51	83.63	67.93	79.05	41.49
SSD [25]	80.75	24.84	86.15	26.60	92.50	31.79
RetinaNet [75]	93.68	52.46	81.96	63.66	84.47	42.53
YOLOv3 [23]	88.15	42.31	82.37	54.39	83.37	31.44
Libra R-CNN [72]	92.24	57.82	80.65	71.26	78.26	51.03
YOLOv4 [24]	75.85	50.11	76.20	59.40	76.42	47.99
Dynamic R-CNN [19]	95.44	57.53	84.86	70.55	80.65	46.91
Sparse R-CNN [73]	97.12	50.22	90.82	68.17	94.43	39.35
YOLOv5 [74]	90.90	69.80	90.40	82.00	91.10	55.80

As can be seen from Table 5, among the nine detection models, the Sparse R-CNN algorithm has the highest detection precision for prohibition signs with a P value of 97.12%. The YOLOv5 algorithm has

the highest detection recall rate for warning signs with a R value of 82.00%. The better overall performance of the Sparse R-CNN algorithm for detection results is due to the fact that the algorithm introduces learnable proposal features that combine with the coarse region of interest information extracted from proposal boxes to better represent some details of the object. During network model training, YOLOv5 transfers each batch of training data through the data loader and performs data augmentation simultaneously. Since the data loader performs three kinds of data augmentation: scaling, color space adjustment and mosaic data augmentation, the high R value is obtained by YOLOv5.

Generally speaking, the smaller the traffic sign, the less likely it is to be detected. However, during actual road travel, detecting relatively distant traffic signs allows traffic participants to have more reaction time. In CCTSDB 2021, we classify the sizes of the test set into five categories based on the pixel sizes of the traffic signs in the image: extremely small (XS), small (S), medium (M), large (L), and extremely large (XL). When we perform size classification, we do not classify an image into any of the five categories if there are traffic signs of two or more sizes in the image. We counted the detection results for five different sizes of traffic signs at an IOU threshold of 0.5, and the precision and recall rate of each algorithm's detection are shown in Table 6.

**Table 6.** Detection results of CCTSDB 2021 in different object sizes (unit: %)

Method	XS		S		M		L		XL	
	P	R	P	R	P	R	P	R	P	R
Faster R-CNN [18]	77.14	48.67	83.62	78.08	88.97	79.23	85.06	88.09	85.26	81.31
SSD [25]	74.84	16.61	72.92	25.44	89.48	32.68	97.74	54.60	99.29	82.65
RetinaNet [75]	77.64	47.00	86.67	64.77	91.59	78.03	85.60	88.50	86.04	84.15
YOLOv3 [23]	86.76	39.24	86.10	66.33	92.88	68.79	80.68	60.68	89.21	71.39
Libra R-CNN [72]	70.19	52.75	79.88	81.25	87.91	81.26	85.60	88.66	76.19	84.03
YOLOv4 [24]	62.44	36.96	70.16	46.47	77.36	59.97	91.55	96.55	96.09	97.43
Dynamic R-CNN [19]	81.24	51.42	83.48	78.87	91.38	80.25	83.28	90.04	83.44	84.31
Sparse R-CNN [73]	91.19	53.97	94.55	74.69	95.56	75.62	95.56	81.10	93.47	78.23
YOLOv5 [74]	75.60	55.90	88.60	75.70	94.70	88.30	97.30	89.00	96.90	91.30

The precision and recall rate of each algorithm are generally smaller to some extent when the size of traffic signs becomes smaller. However, Sparse R-CNN still has good performance in detecting traffic signs at very small-scale sizes, with a P value of 91.19% and the best recall rate result is the YOLOv5 algorithm with a R value of 55.90%. The Sparse R-CNN with better detection performance is because the features of the region of interest are solved to obtain the final features. In this way, those bounding boxes that contain most of the foreground information have an impact on the final object location and classification. In addition, the self-attentive module also facilitates the detection of small objects. SSD extracts six feature maps with different scales, so the network model has better performance in the detection of large traffic signs, and the P values are 97.74% and 99.29% respectively. Due to mosaic data augmentation, the R results of YOLOv4 in L scale and XL scale are higher, i.e., 96.55% and 97.43% respectively.

In real life, weather conditions are complex and variable, and the performance of the detection algorithm varies with the weather environment in which the inspected object is located. We classify the weather conditions in the CCTSDB 2021 into six categories. The detection results of the test set under six weather conditions at an IOU threshold of 0.5 are shown in Table 7.

From Table 7, we can learn that the precision and recall rate of the detection algorithm are relatively high under sunny, snowy and cloudy conditions, indicating that the algorithm is more effective in detecting without interference such as rain and fog. The precision and recall rate of the detection algorithm are relatively low in rain and fog and at night, indicating that rain and fog will have some influence on the detection of traffic signs, and also the visibility is relatively low at night, which is not conducive to the detection of traffic signs. In addition, it can be seen from Table 7 that the impact of traffic sign detection in rainy days is greater than that in snow and fog days. This is because in the image,

only a small amount of water droplets in fog days are attached to the windshield, and most of the snow in snow days is on the ground. However, YOLOv3 have high precision even in the case of blurred vision and certain occlusion. The YOLOv3 still has good detection performance in rain and fog because the algorithm makes extensive use of residual structures for cross-layer connectivity, and to reduce the negative effects of pooling, the algorithm uses a convolution with a step size of 2 for subsampling in the network structure.

**Table 7.** Detection results of CCTSDB 2021 in different weather conditions (unit: %)

Method	Sunny		Cloud		Night		Rain		Foggy		Snow	
	P	R	P	R	P	R	P	R	P	R	P	R
Faster R-CNN [18]	85.47	77.42	92.74	57.61	76.89	47.87	61.35	34.61	77.00	67.09	96.27	91.12
SSD [25]	90.56	32.65	84.45	21.77	85.22	24.59	57.88	27.53	85.42	32.99	95.65	28.10
RetinaNet [75]	90.71	75.37	93.43	53.92	81.09	43.81	67.98	39.55	69.45	64.86	90.18	88.49
YOLOv3 [23]	92.01	64.03	87.12	44.65	75.98	34.81	91.17	31.55	88.66	56.39	87.54	70.59
Libra R-CNN [72]	82.08	78.93	94.47	58.84	80.07	52.52	66.67	48.25	69.24	71.74	90.39	91.12
YOLOv4 [24]	83.83	53.95	74.24	52.92	67.65	32.47	22.43	13.41	85.00	37.43	64.32	40.84
Dynamic R-CNN [19]	86.26	78.92	93.87	58.40	83.70	52.26	64.21	41.13	70.57	69.52	96.25	89.48
Sparse R-CNN [73]	96.56	73.27	96.72	55.97	91.48	44.15	69.69	34.07	92.11	81.11	95.01	88.49
YOLOv5 [74]	95.90	85.10	94.00	81.20	86.10	60.60	47.9	46.7	64.80	81.30	96.10	80.70

We tested the last 500 negative samples of the test set, which has a special experimental procedure. We used the trained network model for batch testing of the negative sample images, which do not have any traffic signs in them, and if there is a detection result in the image, then it means that the image is misdetections. The false detection rate is the ratio of the number of false detected images to the number of negative sample images, and we denote the false detection rate as F. Because we only need to know whether there is a test result on the negative samples, we can know whether it is wrongly detected, so we only need to use F. We set the threshold to 0.5, and the experimental results are shown in Table 8.

**Table 8.** Detection results of negative samples

Method	F (%)
Faster R-CNN [18]	17.60
SSD [25]	0.40
RetinaNet [75]	4.40
YOLOv3 [23]	2.60
Libra R-CNN [72]	37.20
YOLOv4 [24]	7.40
Dynamic R-CNN [19]	8.80
Sparse R-CNN [73]	6.60
YOLOv5 [74]	7.00

In the experimental results, the SSD has the lowest false detection rate for negative samples with a value of 0.04%. This is because the SSD algorithm utilizes six feature maps of different sizes for both classification and regression. Low-level features have higher resolution and contain more location and detail information, but they are less semantic and more noise due to less convolution undergone. High-level features have stronger semantic information, but have very low resolution and poor perception of details [78–80]. The fusion of features at different scales is an important reason why the SSD has the lowest false detection rate for negative samples [81–83].

## 5. Conclusion



In this paper, we expand the CCTSDB 2017 to address the problems of that benchmark and propose a large Chinese traffic sign detection benchmark named CCTSDB 2021. This benchmark contains a total of 17,856 images of real traffic scenes and corresponding detailed annotations, which is a much larger amount of data, and in CCTSDB 2021, we generate a more difficult and dedicated test set, which contains as many scenes as possible. In the test data, we divided Chinese traffic signs into three categories according to their meaning categories: prohibition signs, warning signs and mandatory signs, five categories according to the sizes of traffic signs in the images: extremely small, small, medium, large and extremely large, and six categories according to the weather categories: sunny, rain, night, foggy, snow and cloud. We also selected nine existing classical detection algorithms to be evaluated on the Chinese traffic sign detection benchmark. The evaluation metrics we selected are more comprehensive, with six sets of data: P, R, MR, mAP,  $F_1$  and FPS. For negative samples, we measured the results of their false detection rate. In addition, we evaluate three dimensions for traffic sign category meaning, weather category and size, allowing the algorithm to be evaluated comprehensively on the CCTSDB 2021.

### Author's Contributions

Conceptualization, JZ, XZ, XY, RSS. Investigation and methodology, JZ, XZ. Writing of the original draft, XZ. Writing of the review and editing, JZ, XZ, LK. Data Curation, XZ, JW, XY. Validation, XZ, JW, XY. Supervision, JZ. All the authors have proofread the final version.

### Funding

This research was supported in part by the National Natural Science Foundation of China (Grant No. 61972056 and 61901061), the Science Fund for Creative Research Groups of Hunan Province (Grant No. 2020JJ1006), the Natural Science Foundation of Hunan Province of China (Grant No. 2020JJ5603), the Postgraduate Training Innovation Base Construction Project of Hunan Province (Grant No. 2019-248-51), the Basic Research Fund of Zhongye Changtian International Engineering Co. Ltd. (Grant No. 2020JCYJ07), the Scientific Research Fund of Hunan Provincial Education Department (Grant No. 19C0028 and 19C0031), and the Young Teachers' Growth Plan of Changsha University of Science and Technology (Grant No. 2019QJCZ011).

### Competing Interests

The authors declare that they have no competing interests.

### References

- [1] J. Li and Z. Wang, "Real-time traffic sign recognition based on efficient CNNs in the wild," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 975-984, 2019.
- [2] Y. Yuan, Z. Xiong, and Q. Wang, "VSSA-NET: vertical spatial sequence attention network for traffic sign detection," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3423-3434, 2019.
- [3] A. Gudigar, S. Chokkadi, and U. Raghavendra, "A review on automatic detection and recognition of traffic sign," *Multimedia Tools and Applications*, vol. 75, no. 1, pp. 333-364, 2016.
- [4] Z. Liang, J. Shao, D. Zhang, and L. Gao, "Traffic sign detection and recognition based on pyramidal convolutional networks," *Neural Computing and Applications*, vol. 32, no. 11, pp. 6533-6543, 2020.
- [5] R. Ayachi, M. Afif, Y. Said, and M. Atri, "Traffic signs detection for real-world application of an advanced driving assisting system using deep learning," *Neural Processing Letters*, vol. 51, no. 1, pp. 837-851, 2020.
- [6] J. A. Khan, Y. Chen, Y. Rehman, and H. Shin, "Performance enhancement techniques for traffic sign recognition using a deep neural network," *Multimedia Tools and Applications*, vol. 79, no. 29, pp. 20545-20560, 2020.
- [7] X. Chen, H. Li, Q. Wu, K. N. Ngan, and L. Xu, "High-quality R-CNN object detection using multi-path detection calibration network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 715-727, 2021.

- [8] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: the German traffic sign detection benchmark," in *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-8.
- [9] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484-1497, 2012.
- [10] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localization," in *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV)*, Snowbird, UT, 2009, pp. 1-8.
- [11] F. Larsson and M. Felsberg, "Using Fourier descriptors and spatial models for traffic sign recognition," in *Image Analysis*. Heidelberg, Germany: Springer, 2011, pp. 238-249.
- [12] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," in *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems*, Qingdao, China, 2014, pp. 87-92.
- [13] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 2110-2118.
- [14] J. Zhang, M. Huang, X. Jin, and X. Li, "A real-time Chinese traffic sign detection algorithm based on modified YOLOv2," *Algorithms*, vol. 10, no. 4, article no. 127, 2017. <https://doi.org/10.3390/a10040127>
- [15] F. Fang, L. Li, H. Zhu, and J. H. Lim, "Combining faster R-CNN and model-driven clustering for elongated object detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 2052-2065, 2020.
- [16] Z. Zhao, X. Li, H. Liu, and C. Xu, "Improved target detection algorithm based on Libra R-CNN," *IEEE Access*, vol. 8, pp. 114044-114056, 2020.
- [17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 580-587.
- [18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.
- [19] H. Zhang, H. Chang, B. Ma, N. Wang, and X. Chen, "Dynamic R-CNN: towards high quality object detection via dynamic training," in *Computer Vision – ECCV 2020*. Cham, Switzerland: Springer, 2020, pp. 260-275.
- [20] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant Haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10070-10087, 2019.
- [21] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 751-755, 2019.
- [22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 779-788.
- [23] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018 [Online]. Available: <https://arxiv.org/abs/1804.02767>.
- [24] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020 [Online]. Available: <https://arxiv.org/abs/2004.10934>.
- [25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: single shot multibox detector," in *Computer Vision – ECCV 2016*. Cham, Switzerland: Springer, pp. 21-37.
- [26] X. Wang, H. Ma, X. Chen, and S. You, "Edge preserving and multi-scale contextual neural network for salient object detection," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 121-134, 2018.
- [27] H. Ji, Z. Gao, T. Mei, and Y. Li, "Improved faster R-CNN with multiscale feature fusion and homography augmentation for vehicle detection in remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 11, pp. 1761-1765, 2019.
- [28] D. Tabernik and D. Skocaj, "Deep learning for large-scale traffic-sign detection and recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1427-1440, 2020.

- [29] J. Guo, R. You, and L. Huang, "Mixed vertical-and-horizontal-text traffic sign detection and recognition for street-level scene," *IEEE Access*, vol. 8, pp. 69413-69425, 2020.
- [30] C. G. Serna and Y. Ruichek, "Traffic signs detection and classification for European urban environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4388-4399, 2020.
- [31] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, "Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1467-1479, 2019.
- [32] H. Zhang, L. Qin, J. Li, Y. Guo, Y. Zhou, J. Zhang, and Z. Xu, "Real-time detection method for small traffic signs based on YOLOv3," *IEEE Access*, vol. 8, pp. 64145-64156, 2020.
- [33] D. Temel, M. H. Chen, and G. AlRegib, "Traffic sign detection under challenging conditions: a deeper look into performance variations and spectral characteristics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3663-3673, 2020.
- [34] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic sign recognition using a multi-task convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1100-1111, 2018.
- [35] J. Cao, J. Zhang, and W. Huang, "Traffic sign detection and recognition using multi-scale fusion and prime sample attention," *IEEE Access*, vol. 9, pp. 3579-3591, 2021.
- [36] A. Boukerche and Z. Hou, "Object detection using deep learning methods in traffic scenarios," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1-35, 2021.
- [37] K. K. Santhosh, D. P. Dogra, and P. P. Roy, "Anomaly detection in road traffic using visual surveillance: a survey," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1-26, 2020.
- [38] J. V. Gomes, P. R. Inacio, M. Pereira, M. M. Freire, and P. P. Monteiro, "Detection and classification of peer-to-peer traffic: a survey," *ACM Computing Surveys (CSUR)*, vol. 45, no. 3, pp. 1-40, 2013.
- [39] Y. Wu, Z. Li, Y. Chen, K. Nai, and J. Yuan, "Real-time traffic sign detection and classification towards real traffic scene," *Multimedia Tools and Applications*, vol. 79, no. 25, pp. 18201-18219, 2020.
- [40] L. Yu, X. Xia, and K. Zhou, "Traffic sign detection based on visual co-saliency in complex scenes," *Applied Intelligence*, vol. 49, no. 2, pp. 764-790, 2019.
- [41] 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, 2019.
- [42] P. Dollar, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532-1545, 2014.
- [43] Z. Zhao, Z. Zhang, X. Xu, Y. Xu, H. Yan, and L. Zhang, "A lightweight object detection network for real-time detection of driver handheld call on embedded devices," *Computational Intelligence and Neuroscience*, vol. 2020, article no. 6616584, 2020.
- [44] C. Li, Z. Chen, Q. J. Wu, and C. Liu, "Deep saliency with channel-wise hierarchical feature responses for traffic sign detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 7, pp. 2497-2509, 2019.
- [45] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on viola-jones and HOG+ SVM from UAV images," *Sensors*, vol. 16, no. 8, article no. 1325, 2016. <https://doi.org/10.3390/s16081325>
- [46] J. Hosang, R. Benenson, P. Dollar, and B. Schiele, B. (2015). What makes for effective detection proposals?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 814-830, 2016.
- [47] T. Wang, A. Borji, L. Zhang, P. Zhang, and H. Lu, "A stagewise refinement model for detecting salient objects in images," in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 2017, pp. 4039-4048.
- [48] A. Mogelmoose, D. Liu, and M. M. Trivedi, "Detection of US traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3116-3125, 2015.
- [49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.
- [50] Y. Zeng, X. Xu, D. Shen, Y. Fang, and Z. Xiao, "Traffic sign recognition using kernel extreme learning machines with deep perceptual features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1647-1653, 2017.

- [51] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," in *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-5.
- [52] Y. Yuan, Z. Xiong, and Q. Wang, "An incremental framework for video-based traffic sign detection, tracking, and recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1918-1929, 2017.
- [53] Z. Zhong, L. Sun, and Q. Huo, "An anchor-free region proposal network for Faster R-CNN-based text detection approaches," *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 22, no. 3, pp. 315-327, 2019.
- [54] B. Riyaz and S. Ganapathy, "A deep learning approach for effective intrusion detection in wireless networks using CNN," *Soft Computing*, vol. 24, pp. 17265-17278, 2020.
- [55] L. Chen, X. Hu, T. Xu, H. Kuang, and Q. Li, "Turn signal detection during nighttime by CNN detector and perceptual hashing tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3303-3314, 2017.
- [56] J. Zhang, X. Jin, J. Sun, J. Wang, and A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15095-15115, 2020.
- [57] J. Zhang, X. Jin, J. Sun, J. Wang, and K. Li, "Dual model learning combined with multiple feature selection for accurate visual tracking," *IEEE Access*, vol. 7, pp. 43956-43969, 2019.
- [58] J. Zhang, C. Lu, X. Li, H. J. Kim, and J. Wang, "A full convolutional network based on DenseNet for remote sensing scene classification," *Mathematical Biosciences and Engineering*, vol. 16, no. 5, pp. 3345-3367, 2019.
- [59] S. Zhou, W. Liang, J. Li, and J. U. Kim, "Improved VGG model for road traffic sign recognition," *Computers, Materials & Continua*, vol. 57, no. 1, pp. 11-24, 2018.
- [60] J. Zhang, C. Lu, J. Wang, X. G. Yue, S. J. Lim, Z. Al-Makhadmeh, and A. Tolba, "Training convolutional neural networks with multi-size images and triplet loss for remote sensing scene classification," *Sensors*, vol. 20, no. 4, article no. 1188, 2020. <https://doi.org/10.3390/s20041188>
- [61] J. Zhang, Z. Xie, J. Sun, X. Zou, and J. Wang, "A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection," *IEEE Access*, vol. 8, pp. 29742-29754, 2020.
- [62] J. Zhang, Y. Wu, W. Feng, and J. Wang, "Spatially attentive visual tracking using multi-model adaptive response fusion," *IEEE Access*, vol. 7, pp. 83873-83887, 2019.
- [63] R. Qian, B. Zhang, Y. Yue, Z. Wang, and F. Coenen, "Robust Chinese traffic sign detection and recognition with deep convolutional neural network," in *Proceedings of 2015 11th International Conference on Natural Computation (ICNC)*, Zhangjiajie, China, 2015, pp. 791-796.
- [64] J. Zhang, Y. Liu, H. Liu, and J. Wang, "Learning local-global multiple correlation filters for robust visual tracking with Kalman filter redetection," *Sensors*, vol. 21, no. 4, article no. 1129, 2021. <https://doi.org/10.3390/s21041129>
- [65] J. Zhang, Y. Liu, H. Liu, J. Wang, and Y. Zhang, "Distractor-aware visual tracking using hierarchical correlation filters adaptive selection," *Applied Intelligence*, vol. 52, pp. 6129-6147, 2022. <https://doi.org/10.1007/s10489-021-02694-8>
- [66] J. Zhang, J. Sun, J. Wang, and X. G. Yue, "Visual object tracking based on residual network and cascaded correlation filters," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8427-8440, 2021.
- [67] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.
- [68] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 936-944.
- [69] J. Chen, B. Lei, Q. Song, H. Ying, D. Z. Chen, and J. Wu, "A hierarchical graph network for 3D object detection on point clouds," in *Proceedings of the IEEE/CVF Conference On Computer Vision And Pattern Recognition*, Seattle, WA, 2020, pp. 389-398.

- [70] M. Ye, S. Xu, and T. Cao, "Hvnet: hybrid voxel network for Lidar based 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 2020, pp. 1631-1640.
- [71] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7, pp. 369-379, 2020.
- [72] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: towards balanced learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2019, pp. 821-830.
- [73] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, et al., "Sparse R-CNN: end-to-end object detection with learnable proposals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Virtual Event, 2021, pp. 14454-14463.
- [74] ultralytics/yolov5: v4.0 [Online]. Available: <https://zenodo.org/record/4418161#.YcQsLWBBxPY>.
- [75] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, 2020.
- [76] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft coco: common objects in context," in *Computer Vision – ECCV 2014*. Cham, Switzerland: Springer, 2014, pp. 740-755.
- [77] X. Yuan, J. Guo, X. Hao, and H. Chen, "Traffic sign detection via graph-based ranking and segmentation algorithms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1509-1521, 2015.
- [78] G. Cicceri, F. De Vita, D. Bruneo, G. Merlino, and A. Puliafito, "A deep learning approach for pressure ulcer prevention using wearable computing," *Human-centric Computing and Information Sciences*, vol. 10, article no. 5, 2020. <https://doi.org/10.1186/s13673-020-0211-8>
- [79] S. K. Singh, A. E. Azzaoui, T. W. Kim, Y. Pan, and J. H. Park, "DeepBlockScheme: a deep learning-based blockchain driven scheme for secure smart city," *Human-centric Computing and Information Sciences*, vol. 11, article no. 12, 2021. <https://doi.org/10.22967/HCIS.2021.11.012>
- [80] D. Cao, Z. Chen, and L. Gao, "An improved object detection algorithm based on multi-scaled and deformable convolutional neural networks," *Human-centric Computing and Information Sciences*, vol. 10, article no. 14, 2020. <https://doi.org/10.1186/s13673-020-00219-9>
- [81] Y. D. Zhang, Z. Dong, S. H. Wang, X. Yu, X. Yao, Q. Zhou, et al., "Advances in multimodal data fusion in neuroimaging: overview, challenges, and novel orientation," *Information Fusion*, vol. 64, pp. 149-187, 2020.
- [82] S. Wang, M. E. Celebi, Y. D. Zhang, X. Yu, S. Lu, X. Yao, et al., "Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects," *Information Fusion*, vol. 76, pp. 376-421, 2021.
- [83] Y. Zhang, S. Wang, Y. Sui, M. Yang, B. Liu, H. Cheng, et al., "Multivariate approach for Alzheimer's disease detection using stationary wavelet entropy and predator-prey particle swarm optimization," *Journal of Alzheimer's Disease*, vol. 65, no. 3, pp. 855-869, 2018.