Research Article

Intelligent Recognition of Traffic Signs Based on Improved YOLO v3 Algorithm

Zhonglai Yang (ID)

Shanxi Police College, Shanxi, Taiyuan 030000, China

Correspondence should be addressed to Zhonglai Yang; yangzhlai@chd.edu.cn

In recent years, assisted driving and autonomous driving technology have been paid more attention to by the public. Road sign recognition is of great practical significance for the realization of auto-driving technology. In the actual traffic environment, the traffic signs have the problems of small detectable volume, low resolution, unclear characteristics, and easy to be disturbed by the environment. In order to better realize road traffic sign recognition, this paper improves and optimizes the YOLO v3 network derived from YOLO v3 structure algorithm, enhances the data of the traffic signs by using color enhancement and other technologies, and improves the original FPN structure of the YOLO v3 network algorithm to 52×52. Then, the secondary sampling output characteristic diagram 108 in the YOLO v3 network is used $\times$ 108 solutions to solve these difficulties of picture size and image distortion. Use 5, 9, and 13 fixed-size pools in front of the surface of the control architecture, then the output characteristics are associated with the original characteristics of the picture so that inputs of different sizes can obtain the same output. Finally, we use the intermediate class K algorithm to group the TT100K landmark data set, reconsider the original network parameters, and compare the TT100K data set with the small target determination algorithm, such as YOLO v3 network model and improved YOLO v3 network model. The results show that compared with the traditional YOLO v3 algorithm, the optimized YOLO v3 road sign recognition algorithm has a significant improvement in sign recognition accuracy, sign recognition speed, and learning cost. When the change of FPS is very small, the recall rate and accuracy will be greatly improved. At the same time, compared with other small target detection algorithms, the improved YOLO v3 algorithm has more accurate and faster detection accuracy.

1. Introduction

Road traffic signs play an irreplaceable role in traffic order and safety. They gather road information, warnings, prohibitions, and other information into a simple sign to guide and restrict drivers to drive safely. The setting of traffic signs maintains the safety and smoothness of road traffic to a great extent. And as a sign to assist road safety, traffic signs also provide a simple and clear breakthrough for the development of intelligent transportation. Road traffic signs are usually composed of some simple words or symbols and have color characteristics that form a sharp contrast with the surrounding environment so that road traffic signs can better attract the attention of drivers. Traffic signs with different symbols and colors represent different traffic information [1].

For now, driverless technology has been widely concerned. Hope to improve the safety factor of driverless vehicles on the road, it needs to improve the car's perception of surrounding things, real-time and accurate detection of all targets on the road surface is an important part of environmental perception [2]. Nowadays, the traffic sign recognition system mainly samples the road traffic signs, further detects and recognizes the collected sample information, outputs the recognized results, compares the original image with the traffic sign database to give the results, and finally, sends out warnings and other information through the control center. Because the traffic sign recognition system is usually used in high-speed motor vehicles, the input signal needs to be processed in the embedded equipment of motor vehicles [3]. Facing such complicated steps, how to make the traffic sign recognition

system in the embedded equipment, is higher real-time accuracy of the common difficulty, we face. At the same time, driverless driving has attracted more and more global attention. It is very important to ensure the safety of driverless vehicles in the situation, it is necessary to perceive the surrounding environmental information. Among them, real-time and accurate detection of all targets on the road surface is an important part of environmental perception. By identifying long-distance targets, more time can be provided for vehicle decision-making and control. Usually, long-distance targets (such as traffic signs) are smaller in size, occupy few pixels, and have no obvious features in the image, which makes it difficult to detect and recognize them in real-time [4]. Therefore, how to accurately recognize traffic signs, while ensuring real-time performance, is the key problem to be solved. Up to now, the recognition of traffic signs is mainly divided into two directions: conventional feature extraction and deep learning.

1.1. Basics Feature Extraction Method for Traffic Standard Recognition.
There are generally three traditional traffic sign recognition schemes: color-based traffic sign recognition method, shape-based traffic sign recognition method, and machine learning-based traffic sign recognition method. See Figure 1 for the traditional traffic sign recognition process.

1.1.1. Color-Based Traffic Sign Recognition Method.
Road signs generally contain red, yellow, and blue colors. This bright color leads to strong separability of feature information in the image, and color space threshold segmentation is relatively easy. Up to now, many scholars have reached a color-based traffic sign recognition method. The recognition method based on the color of road signs adopts the method of dividing the color spatial distribution of road signs to realize the detection and recognition of road signs, then extracts the feature information of the segmented image, and finally, classifies the extracted feature information through SVM classifier. RGB color space model proposes an image segmentation algorithm, which improves the operation speed of the algorithm [5]. Yang and Wu [6] proposed a two-stage algorithm for road traffic sign detection. The algorithm first calculates the color probability and then converts the image into a probability model for feature extraction. The extracted feature information is passed through the integral channel to reduce the error. Yuan et al. [7] used edge information to detect color changes in local areas of traffic signs.

1.1.2. Detection Scheme Using Road Sign Shape Recognition.
Because the shape of road signs with different meanings varies greatly, we can recognize traffic signs by recognizing their shapes. We call this recognition algorithm based on the shape of traffic signs. This kind of algorithm first extracts the feature information of traffic sign shape and then classifies the extracted feature information through different shapes. Moreno et al. [8] and others detect traffic signs by limiting the hough transform of geometry in a certain area, which

Figure 1: Schematic diagram of traditional detection algorithm.

improves the robustness of the detection system. Boumediene et al. [9] proposed a coding gradient detection scheme for road sign damage and occlusion, which improved the poor detection effect of traffic signs that detect damage and occlusion. Pei et al. [10] proposed a low-rank matrix recovery architecture with a detection model to solve the problem that the relativity of characteristic information in traffic signs is easy to be ignored, which can better use the relativity of traffic signs to identify road signs.

1.1.3. Recognition Scheme Based on the Shape of Road Signs.
The road sign detection scheme based on machine learning usually uses the moving window method to detect the given traffic sign images in turn, and the researchers manually select and extract the image feature information. In the research of target detection based on machine learning, Dalal [11] proposed the HOG algorithm in 2005. The working principle of the algorithm is to use the gradient direction distribution histogram in the image to describe the location-specific data of the feature information in the image and normalize it. This algorithm can effectively detect the local data of target features in the image, and then the HOG + SVM [12] structure has been continued, which also has a great adverse effect on road sign recognition. Because traffic signs have distinct color information, Huan et al. [13] added color information to the HOG algorithm to expand and achieved good results in traffic sign distinction. According to Lecun et al. [14], research findings a variant gradient direction histogram feature based on HOG algorithm, and trained a single classifier to detect traffic signs through a limit learning machine, which improved the detection efficiency without reducing the detection accuracy.

1.2. Road Sign Recognition Algorithm Based on Deep Learning.
The computer technology research of road sign detection schemes using deep learning methods for recognition is also gradually maturing. The rise of convolutional neural networks makes the deep learning method using deep neural networks combined with different training methods shine in the field of computer vision. Since Geoffrey Hinton [15] proposed the research of artificial intelligence in 2006, deep learning has rapidly swept all research fields of computer technology, among which the most representative algorithm is a convolutional neural network (CNN). In computer vision, the convolution neural network solves the problems of difference recognition accuracy and slow recognition speed at the current stage and can extract the feature information in the image more efficiently and accurately. With the development of CNN, two-stage network structures such as RCNN [16], VGG [17], and AlexNet [18] for image classification and one-stage network structures based on

SSD [19] and YOLO [20] series algorithms for target detection have been successively extended. The algorithm first extracts the features of the target image, then generates candidate regions through the extracted feature information, and finally, uses convolution neural network to classify. In contrast with the traditional target detection algorithm, two-stage algorithm solves the shortcomings of more feature information, a large amount of data, slow detection rate, poor generalization ability, and so on. Single-phase architecture is also known as the identification framework of application regression. It mainly uses the idea of regression theory to give the area, information directly through the backbone network, and discards the candidate areas and RPN network, respectively. Compared with two-stage algorithms, this algorithm can recognize faster, but the recognition accuracy is not as good as two-stage algorithm. With the deepening of the research on deep learning algorithms, more and more scholars study the use of deep learning algorithms to identify road signs. Zuo et al. [21] proposed cascaded RCNN algorithm, which has a detection accuracy of 99% on CCTSDB data set, but the detection rate is relatively slow. Jianming et al. [22] used faster CNN to detect traffic signs and optimized the detection performance. Jianming et al. [22] reduced the amount of calculation and parameters of the algorithm by clipping the network on the basis of YOLO v2 [23] and enhanced the detection performance of small target traffic signs by meshing the input characteristic image.

2. Main Problems of Traditional Traffic Sign Detection Algorithm

2.1. Main Problems of Color-Based Traffic Compilation Detection Algorithm. Different types of traffic signs have different colors. For example, red traffic signs generally indicate prohibited behaviors. Different color combinations of traffic signs also convey different information. The identification of traffic signs can effectively read the meaning of traffic signs. With the deepening of the research on the color of road signs, the color-based road sign detection architecture has greatly improved the detection speed and accuracy. However, traffic signs are often on open and exposed roads, sometimes facing the influence of illumination, fading, occlusion, and bad weather, which makes the results obtained by the color-based detection algorithm unstable, resulting in wrong detection results and missed detection.

2.2. Main Problems of Road Sign Shape Recognition Architecture. The shape of traffic signs is an important feature of traffic sign information. For example, triangles often indicate reminders, and circles indicate prohibition or release of prohibition. Effective identification of traffic sign shapes can solve the initial reading of traffic sign information. For the detection algorithm of road signs shape, although the recognition accuracy of traffic signs has been greatly improved after continuous improvement research, due to the complexity of the road environment, the detection

results when the traffic signs face occlusion, deformation and other situations are unsatisfactory [24]. In addition, the amount of calculation required to extract the shape feature information of traffic signs is large, it increases the calculation time of the model and requires higher computing power of the machine. Although many scholars are also studying the detection algorithm of unifying the color and shape of road signs, it also models size reduction and improves the real-time performance of the algorithm, but the reliability and real-time performance of this traditional traffic sign detection algorithm are still difficult to meet people's requirements for safe driving [25].

2.3. There Are Main Problems in Traffic Sign Recognition Algorithm Using Machine Learning. Although the traffic sign detection framework based on machine learning has a great improvement in the detection accuracy compared with the traffic sign detection algorithm based on color and shape, this kind of detection algorithm has higher requirements for feature extraction. In addition, the detection algorithm based on machine learning usually needs to manually select the region of feature information, which makes this kind of algorithm have a high workload and poor real-time performance. For traffic sign detection, the target detection algorithm based on machine learning still has some limitations.

3. Basic Principle of YOLO v3 Algorithm

YOLO v3 is a target detector. Its backbone architecture uses Darknet-53 instead of Darknet-19. There are 53 convolution layers in total. The network structure is shown in Figure 2.

Darknet-53 uses RESNET's residual idea for reference to form a residual structure, which can well control the spread of gradients, avoid situations that are not conducive to training, such as gradient disappearance or explosion, and greatly reduce the difficulty of training deep networks. The main part of the network is composed of five other debris. Multiple residual units form a residual block and each residual unit is constituted of two DBL modules and quick links. The deep separable convolution model is shown in Figure 3.

Darknet-53 minimum weight DBL module is composed of convolution, packet standardization, and leakage recovery firing. YOLO v3 divides the forecast into 13×13, 26×26, and 52×52. These three parameters push the three performance graphs to the test level. In particular, the features of low-level mapping have a small sensitive field and strong small target detection ability, while the features of depth mapping have a large sensing range and improving the performance of detecting large targets [26]. Therefore, YOLO v3 has obvious advantages in determining the size of detection targets. Because YOLO v3 network has high learning efficiency and strong adaptability to different task scales in complex traffic scenes, TT100K [18] signaling data set is used to improve, train and test YOLO v3 network.
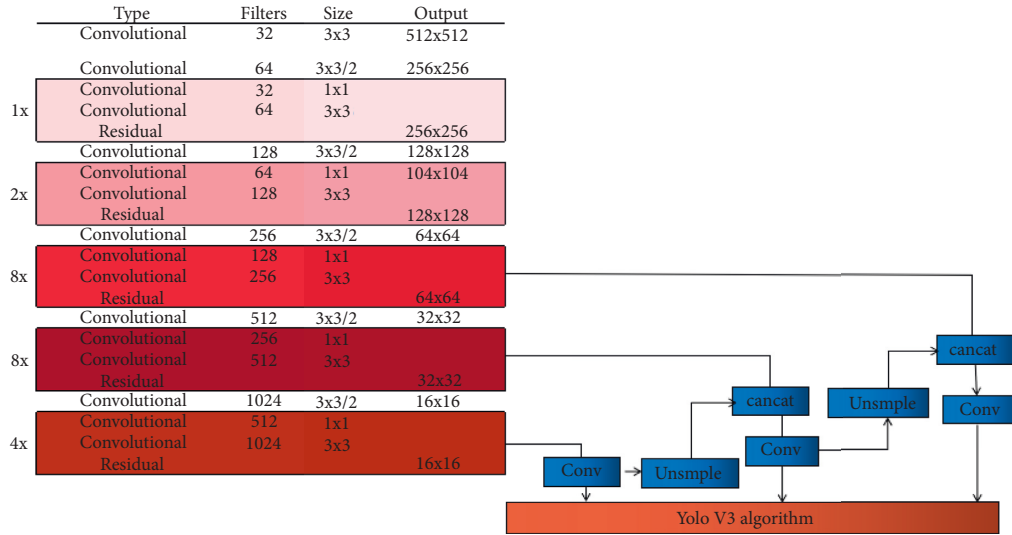
	Type	Filters	Size	Output
	Convolutional	32	3x3	512x512
	Convolutional	64	3x3/2	256x256
1x	Convolutional	32	1x1	
	Convolutional	64	3x3	
	Residual			256x256
	Convolutional	128	3x3/2	128x128
2x	Convolutional	64	1x1	104x104
	Convolutional	128	3x3	
	Residual			128x128
	Convolutional	256	3x3/2	64x64
8x	Convolutional	128	1x1	
	Convolutional	256	3x3	
	Residual			64x64
	Convolutional	512	3x3/2	32x32
8x	Convolutional	256	1x1	
	Convolutional	512	3x3	
	Residual			32x32
	Convolutional	1024	3x3/2	16x16
4x	Convolutional	512	1x1	
	Convolutional	1024	3x3	
	Residual			16x16

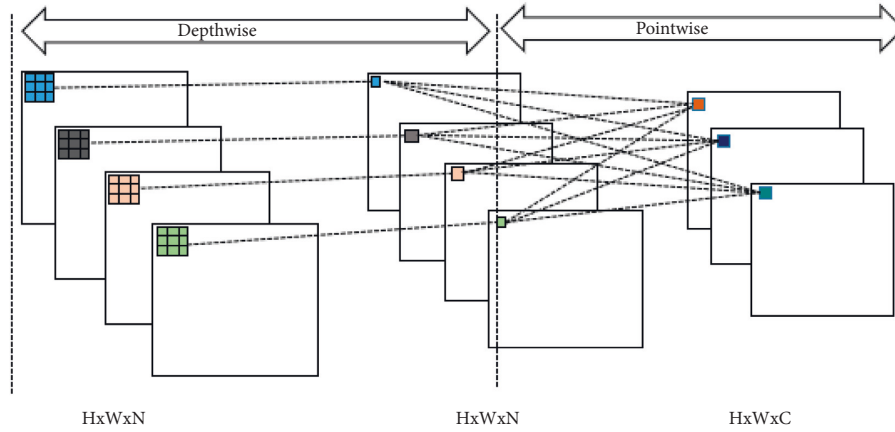Figure 2: Basic model of YOLO v3 framework.

Figure 3: Deep separable convolution model.

4. Improvement of 4 YOLO v3 Algorithm

Aiming at the low accuracy of identification of the original YOLO v3 neural network for long-distance lower target objects, this paper improves the algorithm composition, K-means network structure, and loss function.

4.1. Improvement of Network Structure. Since the deep network of the original deep structure of YOLO v3 is conducive to the detection of large targets, and the shallow structure is convenient for the detection of small targets because the shallow algorithm structure passes through small convolution layers [27], it lacks deep semantic features, contains less semantic information, and has weak feature representation ability, these features affect the detection of small targets, which depends on the shallow algorithm structure. In order to improve the feature extraction ability of the detection algorithm structure, this paper uses Inception architecture that can enrich the features of the shallow network for reference.

As shown in Figure 4, inception the neural network operation and pool operation are performed on the identified image, and the output results are spliced into the deep marking feature image of different convolution kernel sizes such as 1×1, 3×3, 3×3, or 5×5. The information of different perception domains can be obtained from the input picture data, these operations can be combined, and all the structure can be combined to improve the image quality representation. Inspired by the concept architecture, a concept redefinition module structure is proposed and applied to the shallow layer network of YOLO v3. Compared with the traditional YOLO v3 network, the recognition algorithm of the shallow layer network has a stronger ability to extract the specific representation of the picture, and the information extraction abundance of the recognition system is improved. The improved YOLO v3 algorithm is also more closely combined with the feature points of the neural network, the recognition and perception efficiency of the image is higher, and the recognition ability of small traffic signs is improved.
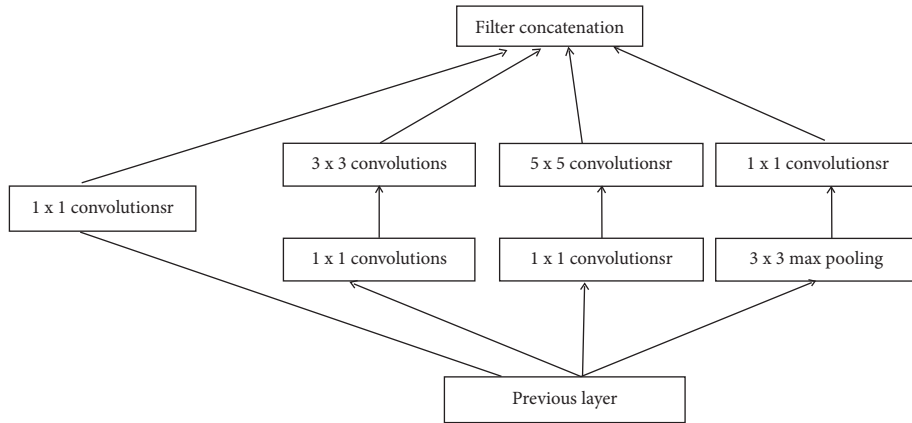
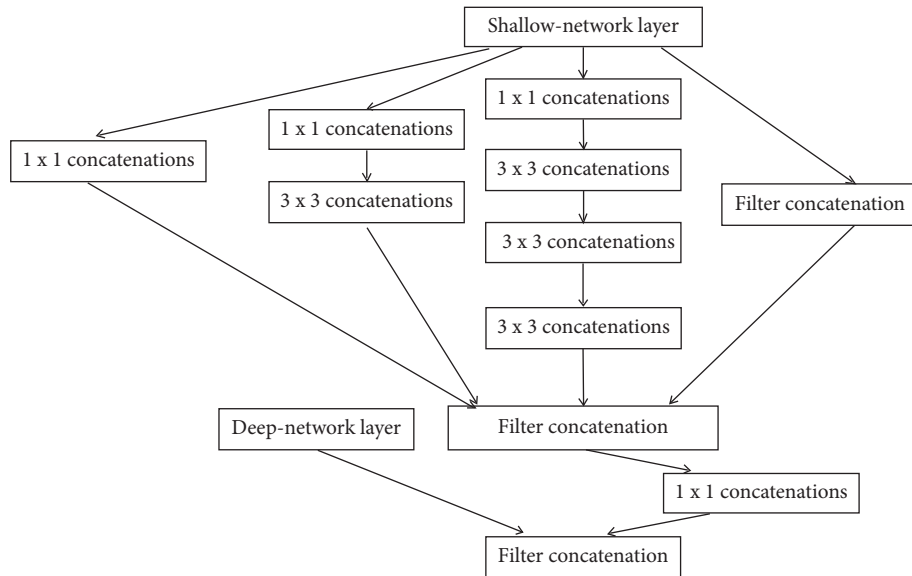FIGURE 4: Structure diagram of inception.

FIGURE 5: Structure diagram of inception-redefined module.

Figure 5 shows the structure of the initial redefinition module. The two ends of the structure are shallow network layer and deep network layer, respectively, which are used to connect to the network of YOLO v3. The structure consists of four substructures: the first substructure is 1×1 volume accumulation; the second branch is composed of a convolution of 1×1 and then a convolution of 3×3; the third branch consists of the convolution of one 1×1 followed by the convolution of three 3×3; and the fourth branch consists of the maximum pooling layer. A 7×7 convolution effectively extracts basic information from various small pictures. In this paper, using three 3×3 convolutions instead of one 7×7 convolution can save $7 \times 7(3 \times 3 \times 3) = 1.81$ times the calculation amount, which can improve the calculation speed. Front 3×3 convolution is 1×1. The convolution layer can reduce the number of input channels, effectively reduce the number of input parameters, and increase the parallel ability of the architecture. Benefit from 1×1 convolution passes through the ReLU activation function [28], so the

TABLE 1: Test results of different picture sizes.

Picture size	Test model	mAP	FPS
416×416	YOLO v3	0.431	54.4
	Improved YOLO v3	0.514	50.7
608×608	YOLO v3	0.691	34.2
	Improved YOLO v3	0.752	31.3
1024×1024	YOLO v3	0.791	18.7
	Improved YOLO v3	0.834	16.4

generalization performance of neural network can be improved through the introduction of nonlinearity data pool for extraction image features. The four branches extract features of different scales that increase the adaptability of the network to different scales and obtains the information from multiple scales, respectively, [29]. Then the feature maps under the four branches are fused, and finally, the number of output access is reduced through 1×1
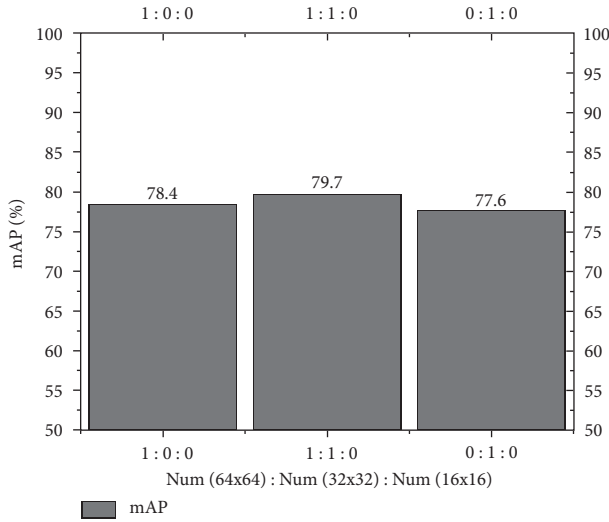
FIGURE 6: Comparison of map values of inception redefined module under different distribution proportions 4.2 k-means algorithm improvement.

convolution layer. The channel ratio of the characteristic image output from the left 1×1 convolution to the right 3×3 convolution will affect terminal identification accuracy. In this paper, different proportions were tested, and the quantity ratio of $1:1$ was finally selected with the highest accuracy. Table 1 shows the mAP values for different proportions.

Introduce the concept-redefined module structure into the output 64 in Figure 1, 64×64 (and 32×32) between the characteristic diagram and concatenation to form the YOLO v3 improvement network. The introduction method is shown in Figure 4. By connecting the shallow-network layer to the deep-network layer, the combination of deep information and shallow information is more conducive to the prediction of small targets. For output 64×64 in terms of a characteristic graph, the channel flux capacity and size of deep network characteristic graph are $64 \times 64 \times 128$. The size and surface channels network characteristic map are $64 \times 64 \times 256$, then the fused feature map are $64 \times 64 \times 384$.

As shown in Figure 6, this paper attempts to apply several distributions of the inception redefined module structure. Finally, this paper selects the distribution with the best result. Compared with shallow information, deep information can provide more image features. The multidimensional and multilevel convolution kernel in the improved YOLO v3 algorithm also provides convenience for the perception of visual field information in different ranges, and the sentence information abundance improves the perception ability of small targets.

In order to improve the accuracy of the traffic sign recognition algorithm for image proportion recognition, this paper improves the disadvantages of the original Yolo v3 and K-means clustering algorithm that lack filtering function and proposes an improved k-means clustering algorithm. Based on this, this paper proposes an improved K-means algorithm. First, the invalid data in the data set are eliminated by calculating the width-height ratio of the object

coordinates, and the valid data are retained. Next, use k-means architecture operation to classify the obtained data. The mother is to obtain the size and proportion of the anchor. Finally, the classification results are added to the YOLO layer for training and recognition. The execution order of the improved k-means algorithm is as follows: Input: dimension file in the data set. Output: width, height, and proportion of anchor box. Where I is the number that marks the target.

(1) Eliminate significance of data annotation in data set.

  (1) for $i = 1$ to total do
  (2) Write coordinate data from the dimension file of the data set.
  (3) Mark according to the following rules. In the upper left corner of the annotation box, the standard x-axis is X_{min}, the coordinate of the lower left corner of the annotation box on the x-axis is X_{min}, the coordinate of the small left corner on the y-axis is Y_{min}, the x-axis of the lower right corner of the coordinate axis is marked as X_{max}, and the upper right corner of the y-axis is marked as Y_{max}.
  (4) $Dx = X_{max} - X_{min}$, $dy = Y_{max} - Y_{min}$, if $Dx = 0$ or $dy = 0$, the mark data correspondence to Dx and Dy is meaningless.
  (5) $Q = Dx - Dy$, if $0.3 < Q < 1$, next, make data annotation by comparison to DX and Dy is valid, and else annotation data are invalid.
  (6) Filter all meaningful comments in the information set.
  (7) end for

(2) effective annotation data are clustered

  (1) choose k clusters intentionally, and choose the central initial cluster of k aimlessly.
  (2) do
  (3) Calculate the IOU value and cluster center of all valid annotation data.
  (4) Data points with large IOU values will be divided into clusters location of cluster center.
  (5) The new family center point is generated from the central set of each cluster data point selected.
  (6) While (the cluster center moves).

The optimized K-means clustering algorithm can effectively ignore the adverse impact of invalid annotations on the clustering center, significantly improve the fit between the anchor box and traffic signs, and significantly improve the recognition accuracy of YOLO v3 network model.

4.2. Optimization Loss Algorithm. The data lost in YOLO v3 algorithm are mainly divided into coordinate regression loss, confidence loss, and clustering loss. For the loss function of coordinate regression calculated by mean square error, the size of the target can directly resulting in decreased accuracy of coordinate regression, using IOU as the target scale in YOLO v3 will bring two problems: first, when IOU (A, B) is equal to 0 $(A, B$ are the forecast

boundary box and the real boundary box, respectively), that is, when A and B do not overlap, it is impossible to know whether a and B are adjacent to each other or far away, that is, it cannot reflect the distance between them, and its gradient will be zero, so it cannot be optimized; the second is that IOU (A, B) is not 0, that is, when a and B overlap, the specific overlap of the two cannot be reflected, and in the case of different distances, different proportions, and different aspect ratios, using IOU as a loss, the regression situation is usually incomplete. Compared with IOU, it only focuses on the areas where clusters overlap, unlike IOU, which focus only on overlapping areas, GIOU not only pays attention to the specific overlap of superposition areas but also at the same time, there is enough computing power to match in other nonoverlapping areas, which can better feedback the matching degree between objects. As shown in Figure 5, the IOU values are all 0.33, but there are three different overlaps, that is, the GIOU values are 0.33, 0.24, and 0.1 from left to right, respectively. GIOU is defined as follows:

$$\text{GIOU} = \text{IOU} - (C - S_{\text{union}})/C. \quad (1)$$

In this formula, C is the minimum superposition area of a and B. The value range of IOU is [0, 1], and the value range of GIOU is [−1, 1]. For GIOU, when the predicted value completely coincides with the actual value, the value is 1. When the two do not overlap and approach infinity, GIOU takes the minimum value 1. Then, GIOU is the preparation of expressing the measurement accuracy, which can accurately reflect the difference between the predicted value and the true value. Therefore, this paper selects GIOU to replace the coordinate regression loss, and the formula is given as follows:

$$L_{\text{Got}} = 1 - \text{GIOU}. \quad (2)$$

The confidence loss function is given as follows:

$$L_{\text{coaf}} = -\sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{\text{oljj}} \left[\widehat{C}_i \log(C_i) + (1 - \widehat{C}_i) \log(1 - C_i) \right]$$
$$- \lambda_{\text{nooblij}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{\text{objj}} \left[\widehat{C}_i \log(C_i) + (1 - \widehat{C}_i) \log(1 - C_i) \right]. \quad (3)$$

The first term in this formula is the confidence error of the prediction frame containing the target. The second item is the confidence error of the prediction frame without targets. S^2 is the number of grids of markers in the input image; I_{ij}^{olji} indicates whether the jth anchor box of the ith grid catches the target, which is 1 or 0; C_i is the confidence score of the real box; $\widehat{C}_i$ is the confidence score of the prediction box.

$$L_{\text{cls}} = -\sum_{i=0}^{s^2} I_{ij}^{\text{obj}} \sum_{C \subset \text{Class}} \left[\widehat{p}_i(C) \log(p_i(C)) + (1 - \widehat{p}_i(C)) \log(1 - p_i(C)) \right]. \quad (4)$$

When the first anchor box of the ith grid captures the target, the bounding box generated by the anchor box will calculate the classification loss function. In Equation (4), C represents the detected target category, and $p_i(C)$, $\widehat{p}_i(C)$ express the probability of the real box and the prediction box belonging to category C in grid i, respectively. The final improved loss function is given as follows:

$$\text{Loss} = L_{\text{Glov}} + L_{\text{coof}} + L_{\text{cls}}. \quad (5)$$

5. Experimental Results and Analysis of Recognition Algorithm

In order to test the accuracy of the optimized YOLO v3 algorithm designed in this paper to recognize traffic signs, this paper carries out two inspections, which are the comparative experiment of different improved YOLO v3 algorithms and the use of three image input sizes (416×416, 608×608, and 1024×1024), and verify these two tests from the aspects Average detection accuracy (map), number of detected pictures per second (FPS) and accuracy recall (P-R) curve.

5.1. Road Sign Sample Collection and Experimental Platform. In this experiment, traffic sign data set tt100 is used, which provides 100000 2048×2048 images, of which 30000 traffic sign instances are small targets. There are 45 types of objects in the data set, representing the corresponding traffic signs, respectively: *i*2, *i*4, *i*5, *il*100, *il*60, *il*80, *io*, *ip*, *p*10, *p*11, *p*12, *p*19, *p*23, *p*26, *p*27, *p*3, *p*5, *p*6, *pg*, *ph*4, *ph*4.5, *ph*5, *pl*100, *pl*120, *pL*20, *pl*30, *pl*40, *pl*5, *pl*50, *pl*60, *pl*70, *pl*80, *pm*20, *pm*30, *pm*55, *pn*, pne, po, pr40, w13, w32, w55, w57, w59, and wo. First of all, we should put forward the pictures without landmark files in the traffic sign data set, select 6105 pictures for algorithm practicing, and 3070 pictures for accuracy and speed detection.

Experimental hardware: software system ubuntu22.04, *d* Neural network learning framework pytorch1.4, CPU AMD-r5 5600g, 32 GB memory, NvidiageForceRTX3080Ti ∗ 2 GPU, 24 GB video memory.

5.2. Algorithm Training and Detection. In this experiment, we trained the YOLO v3 architecture and the improved YOLO v3 architecture and used the conventional parameter optimization method of YOLO v3 to optimize the parameters. The initial learning rate is 0.001, and the maximum iterations are 300 cycles. The training rate is set to decay 10 times when the number of iterations is 75 epochs, 150 epochs, and 250 epochs, respectively. The data set is enhanced by flipping, translation transformation, and other methods. At the same time, multiscale training is adopted to make the scale float up and down in the set range, so as to achieve a better training effect. First, test the two models after training and calculate their precision and recall. The formulas for calculating precision and recall are given as follows:
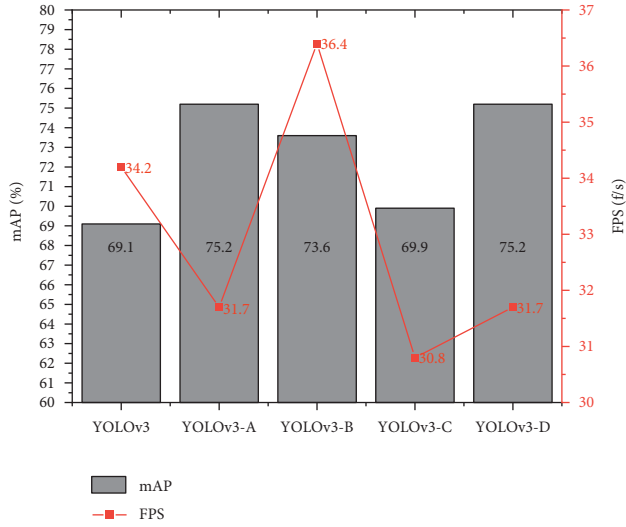
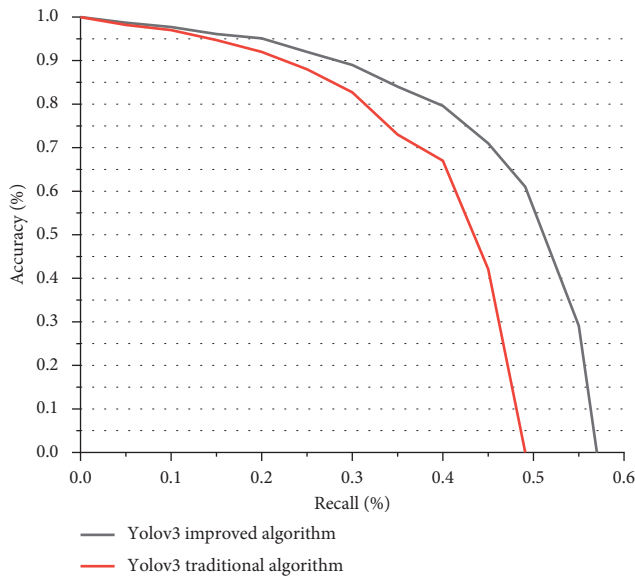FIGURE 7: Improved detection results of YOLO v3 and YOLO v3.

FIGURE 9: *P-R* curve (picture size 608×608).

FIGURE 8: *P-R* curve (picture size 416×416).
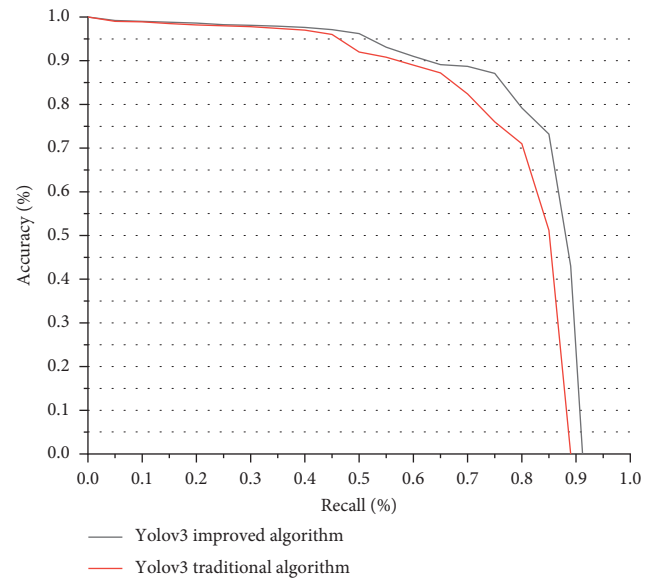
FIGURE 10: *P-R* curve (picture size 1024×1024).

$$p = \frac{\text{M}P}{\text{M}P + \text{O}P},$$

$$R = \frac{\text{M}P}{\text{M}P + \text{O}N}.$$

(6)

In the formula, MP is the number of positive classes predicted by positive classes, OP is the number of negative classes predicted by positive classes, and ON is the number of positive classes predicted by negative classes. By setting a fixed threshold, the prediction results of the detector are arranged in descending order of confidence, and positive prediction samples are generated respective, the P and *R* values can be calculated and the P-R curve can be drawn.
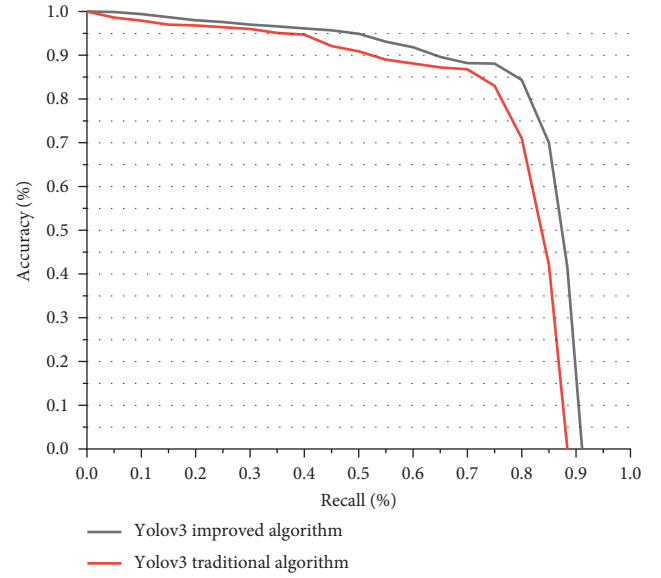
5.3. Comparison Results and Analysis

5.3.1. The Contrast Experiment of Improved. YOLO v3 is based on traffic signs (named YOLO v3-A), improved YOLO v3 network detection layer and FPN structure (YOLO v3-B), added spatial pyramid module (YOLO v3-C), and added to the above three improved YOLO v3-D networks. The input image size is 608×608. These four optimized models have been tested on TT100K photo acquisition and conventional YOLO v3 network. The experimental results are shown in Figure 7.

Figure 7 shows that the detection rate and accuracy of the optimized YOLO v3 architecture on the TT100K traffic sign data set are higher than that of the conventional YOLO v3
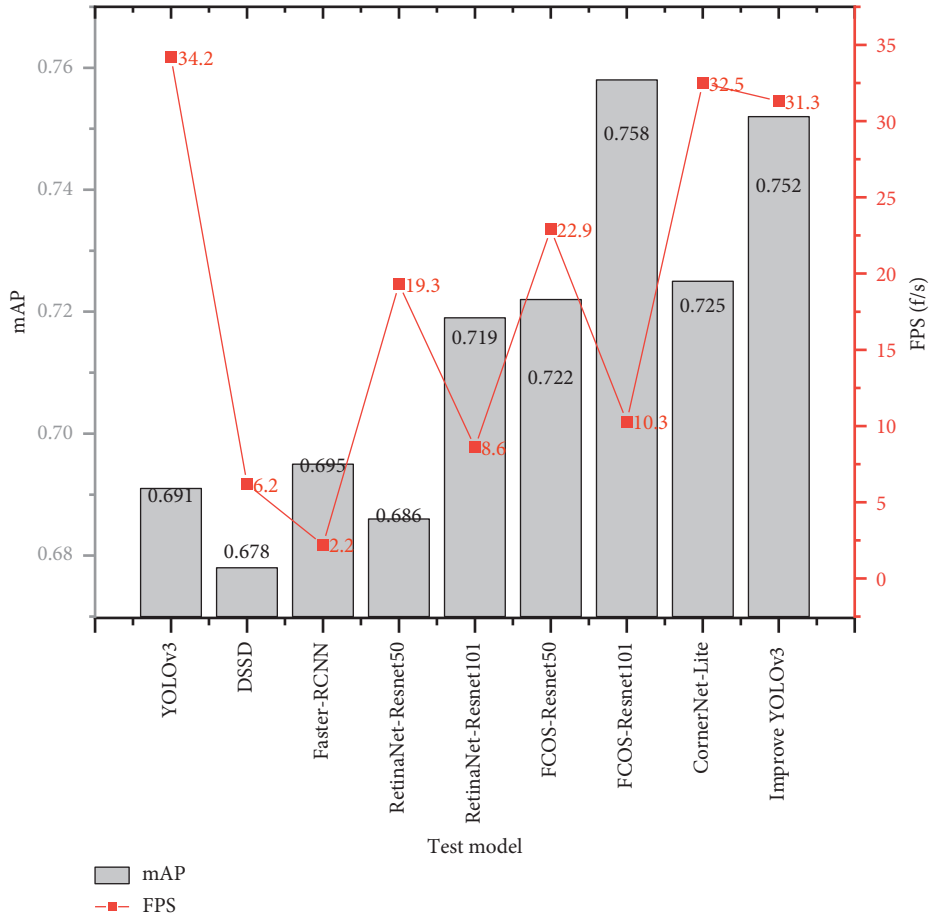
FIGURE 11: Comparison of recognition results of different recognition architectures.

architecture. Among the four optimized YOLO v3 models, the mapping of YOLO v3-D model reaches 75.2%, and the effect is the best among the four networks. Although its FPS has been reduced to 31.3 f/s, it can still meet the needs of implementing traffic sign detection and recognition.

5.3.2. Improved YOLO v3 Experiment with Different Picture Sizes. Furthermore, test the effectiveness of the optimized YOLO v3 algorithm, and carry out comparative analysis experiments on the optimized YOLO v3 neural network algorithm (YOLO v3-D) and the original YOLO v3 algorithm when the image input sizes were 416, 608, and 1024, respectively. When the input sizes are 416, 608, and 1024, respectively, the P-R curve comparison diagrams of the YOLO v3 network and the improved YOLO v3 network model are shown in Figures 8–10. Accuracy and recall rate of optimized YOLO v3 algorithm model are higher than that of the YOLO v3 network to a certain extent.

There are 3,070 pictures and 7,700 targets in the test set. Two network models are tested, respectively. Their map and FPs are shown in Table 1.

The input dimensions are 416×416, 608×608, and 1024×1024, the map of optimized YOLO v3 algorithm

model increased by about 8.3%, 6.1%, and 4.3%, respectively, while FPS did not decrease significantly; at size 416×416 and 608×608, it has the characteristics of fast detection and can meet the needs of road standard identification in reality.

5.3.3. Comparative Analysis of the Optimized Algorithm and the Original Algorithm. In order to further test the recognition effect of the optimization algorithm, the input image size is 608×608. The improved YOLO v3 algorithm is compared with RetinaNet, FCOS, CornerNet, and other advanced small target detection algorithms. The final comparative analysis is shown in Figure 11.

It can be seen that (Figure 10)) only FCOS algorithm is higher and is better than the optimized YOLO v3 framework in recognition accuracy. However, its FPS is far lower than the optimized YOLO v3 architecture; the detection speed of CornerNet algorithm is similar to that of the improved YOLO v3 algorithm, but its mAP is 2.7% lower than that of the improved YOLO v3 algorithm. Experiments show that the optimized YOLO v3 algorithm proposed in this paper can achieve good results in traffic sign recognition. When recognizing small traffic signs in TT100K traffic sign big data set and traffic signs with small range occlusion and long

distance, the improved YOLO v3 algorithm has also significantly improved its recognition efficiency and accuracy in the experimental results.

6. Conclusion

This article introduces an optimization model utilized on YOLO v3, which aims to solve the e problem of low accuracy of road traffic sign recognition, in the task of road sign recognition, the detection mode needs to deal with many parameters and slow speed. Aiming at the shortcomings of YOLO v3, the algorithm architecture, K-means clustering algorithm, and loss function are optimized, which greatly improves the accuracy and speed of the detection framework. The simulation results show that the optimized YOLO v3 framework has more advantages for small traffic standard recognition. The detection experiments on three different resolution photos show that compared with conventional YOLO v3 algorithm, recognition accuracy improvement of optimization algorithm 8.1%, 5.9%, and 4.6%, respectively. Under the premise of ensuring that the gap between FPS is small, the recall rate and accuracy rate have been significantly improved. In general, the main advantage of the optimized YOLO v3 algorithm in road sign detection and recognition is that the recognition efficiency is improved and the recognition accuracy is higher. In particular, the recognition rate is higher for small and distant traffic signs and traffic signs that are covered by foreign matters in a small range. It can be seen that the improved YOLO v3 has higher usability in actual road traffic.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. Saleh and H. Fleyeh, "Factors affecting night-time visibility of retroreflective road traffic signs: a review," *International Journal of Traffic and Transportation Engineering*, vol. 11, no. 1, pp. 115–128, 2021.

[2] H. Liu, *Train Unmanned Driving Algorithm Based on Reasoning and Learning strategy*Elsevier, Netherlands, 2021.

[3] P. P. . Thakur, "Detection and classification of traffic signs for driverless cars," *Bioscience Biotechnology Research Communications*, vol. 13, no. 14, pp. 517–521, 2020.

[4] F. Franzen, C. Yuan, and Z. Li, "Traffic sign recognition with neural networks in the frequency domain," *Journal of Physics: Conference Series*, vol. 1576, no. 1, Article ID 012015, 2020.

[5] A. W. Utoyo, H. D. Aprilia, R. A. D. R. I. Kuntjoro, and A. Kurniawan-Jakti, "Visual communication analysis the effect of signs and colors on traffic safety in Jakarta," *IOP Conference Series: Earth and Environmental Science*, vol. 729, no. 1, p. 6, Article ID 012087, 2021.

[6] Y. Yang and F. Wu, *Real-Time Traffic Sign Detection via Color Probability Model and Integral Channel Features*, Springer, Berlin Heidelberg, 2014.

[7] X. Yuan, L. F. Liu, C. H. Li, and Y. Y. Qu, "Unifying visual saliency with hog feature learning for traffic sign detection," in *Proceedings of the Intelligent Vehicles Symposium*, IEEE, Xi'an, 2009.

[8] D. . Moreno, F. Roberto Pichler, and A. Quesada Arencibia, "Fast road sign detection using hough transform for assisted driving of road vehicles," in *Proceedings of the 10th International Conference on Computer Aided Systems Theory*, Springer, Berlin Heidelberg, 2005.

[9] M. Boumediene, C. Cudel, M. Basset, and A. Ouamri, "Triangular traffic signs detection based on RSLD algorithm," *Machine Vision and Applications*, vol. 24, no. 8, pp. 1721–1732, 2013.

[10] D. Pei, F. Sun, and H. Liu, "Supervised low-rank matrix recovery for traffic sign recognition in image sequences," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 241–244, 2013.

[11] N. Dalal, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005.

[12] I. M. Creusen, R. G. J. Wijnhoven, and E. Herbschleb, "Use of artificial neural network in pattern recognition," *Color Exploitation in Hog-Based Traffic Sign detection[C]//IEEE International Conference on Image Processing*, Springer, Berlin Heidelberg, 2020.

[13] Z. Huang, Y. Yu, J. Gu, and H. Liu, "An efficient method for traffic sign recognition based on extreme learning machine," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 920–933, 2017.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–44, 2015.

[15] R. Girshick, J. Donahue, T. Darrell, and M. Jitendra, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Computer Society*, vol. 580–587, 2013.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, vol. 1409, 2014.

[17] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.

[18] W. Liu, D. Anguelov, and D. Erhan, *SSD: Single Shot MultiBox Detector*, pp. 21–37, European conference on computer vision, Springer, Cham, 2016.

[19] J. Redmon, S. Divvala, and R. Girshick, "You only look once: unified, real-time object detection," in *Proceedings of the Computer Vision & Pattern Recognition*, IEEE, NV, USA, 2016.

[20] J. Zhang, Z. Xie, J. Sun, Z. Xin, and W. Jin, "A cascaded R-CNN with multiscale Attention and imbalanced samples for traffic sign detection," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[21] Z. Zuo, Y. Kai, Z. Qiao, W. Xu, and L. Ting, "Traffic signs detection based on faster R-CNN," in *Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, IEEE, Atlanta, GA, USA, 2017.

[22] Z. Jianming, H. Manting, J. Xiaokang, and L. Xudong, "A real-time Chinese traffic sign detection algorithm based on modified YOLOv2," *Algorithms*, vol. 10, no. 4, p. 127, 2017.

[23] S. U. Kim and J. W. Lee, "Traffic sign recognition, and tracking using RANSAC-based motion estimation for autonomous

vehicles," *Journal of Institute of Control, Robotics and Systems*, vol. 22, no. 2, pp. 110–116, 2016.

[24] X. Yang, W. Liu, S. Zhang, L. Wei, and T. Dacheng, "Targeted Attention Attack on Deep Learning Models in Road Sign Recognition," *IEEE*, vol. 8, 2020.

[25] Y. Sun and L. Chen, "Traffic sign recognition based on CNN and twin support vector machine hybrid model," *Journal of Applied Mathematics and Physics*, vol. 9, no. 12, p. 21, 2021.

[26] Z. Zhu, S. Zhang, H. Xiaolei, L. Baoli, H. Shimin, and L. Dun, "Traffic-signdetectionandclassificationinthewild," in *Proceedings of the 2016IEEEConferenceonComputerVisionand PatternRecognition(CVPR)*.

[27] A. Hechri and A. Mtibaa, "Two-stage traffic sign detection and recognition based on SVM and convolutional neural networks," *IET Image Processing*, vol. 14, no. 5, pp. 939–946, 2020.

[28] R. Ayachi, M. Afif, Y. Said, and A. B. Abdelali, "Real-time implementation of traffic signs detection and identification application on graphics processing units," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, 2021.

[29] W. Sun, Y. Du, X. Zhang, and G. Zhang, "Detection and recognition of text traffic signs above the road," *International Journal of Sensor Networks*, vol. 35, no. 2, p. 69, 2021.