https://doi.org/10.3311/PPtr.41491
Creative Commons Attribution ⓘ

Periodica Polytechnica Transportation Engineering

Automatic Traffic Sign Recognition Algorithm Based on Attention Mechanism and YOLOv4

Yuke Han[1*]

[1] School of Automotive Engineering, Shaanxi Vocational and Technical College, Xi'an, 710038, China
[*] Corresponding author, e-mail: hanhykk@outlook.com

Received: 02 July 2025, Accepted: 07 November 2025, Published online: 13 March 2026

Abstract

Image recognition, a key technique in deep learning within the realm of computer vision, has found extensive application in the transportation sector in recent years. However, traditional image recognition technologies suffer from low efficiency and weak analytical capabilities. This research proposes a traffic sign recognition model that embeds a reconstructed squeeze and excitation network channel attention mechanism into the You Only Look Once Version 4 framework. Specifically, depthwise separable convolution is adopted to reconstruct the attention module, and a soft threshold denoising module is integrated before multi-scale feature fusion. The model also utilizes a soft threshold denoising module for feature extraction of complex semantic information. Experimental results show that when the attention mechanism fusion algorithm iterates five times, the accuracy reaches 98.5%. The highest recognition accuracy, prediction recall rate, and harmonic mean of recall rate are 96.35%, 95.88%, and 95.12%, respectively. The evaluation of the fusion model shows that the model has the highest recognition accuracy of 0.98 for different types of traffic signs. Compared with the highest accuracy of 0.89 of Faster Region-based Convolutional Neural Network and the highest accuracy of 0.90 of Field-Programmable Gate Array, the research method has significantly higher recognition accuracy. These results suggest that the improved traffic sign recognition model can effectively identify real-world road traffic signs for autonomous vehicles, with excellent feature-capturing performance. This research contributes to the future development of road traffic and autonomous driving fields.

Keywords

SENet, autonomous driving, sign recognition, YOLOv4 algorithm, depthwise separable convolution

1 Introduction

In recent years, the research and application of autonomous driving systems and assisted driving systems have continued to develop and innovate and have become a hot topic for researchers and companies in the automotive field. Among them, the automatic detection and recognition of traffic signs occupies an extremely critical position. The application of intelligent algorithms in the field of autonomous driving traffic sign recognition can not only improve the recognition accuracy but also enhance the real-time performance of recognition (Megalingam et al., 2023). Therefore, the research on traffic sign recognition algorithms is of great significance for technological breakthroughs and industrial upgrading in the field of autonomous driving. As one of the most revolutionary technologies in recent years, artificial intelligence algorithms are widely used in the field of image recognition (Simran et al., 2022). The current mainstream traffic sign recognition methods include convolutional neural networks, Transformer-based target detection algorithms, feature extraction algorithms, etc. However, the above methods generally have limitations such as complex feature extraction process, the need for a large amount of labeled data for training, and insufficient robustness. Therefore, a method that can accurately and efficiently recognize traffic signs is needed. Compared with other recognition algorithms, the You Only Look Once Version 4 (YOLOv4) algorithm adopts a single-stage architecture with faster recognition speed and focuses on the fusion of multi-scale features, so the recognition accuracy is also higher. Since the YOLO series of algorithms usually adopt rich data enhancement strategies, they also have unique advantages in generalization (Terven et al., 2023; Diwan et al., 2023). At the same time, the introduction of the attention mechanism Squeeze and Excitation Network (SENet) can also enable

Cite this article as: Han, Y. (2026) "Automatic Traffic Sign Recognition Algorithm Based on Attention Mechanism and YOLOv4", Periodica Polytechnica Transportation Engineering, https://doi.org/10.3311/PPtr.41491

deep learning models to process input data more flexibly (Lescoat et al., 2023). Therefore, this study proposes a new traffic sign recognition algorithm that integrates the attention mechanism with YOLOv4, hoping that the algorithm can help autonomous driving to efficiently and accurately identify traffic signs. The research innovatively employs depthwise separable convolution to reconstruct the SENet channel attention mechanism, significantly reducing computational complexity. It also introduces a soft threshold denoising module to suppress noise interference before feature fusion. These two elements are deeply integrated with the multiscale detection architecture of YOLOv4 to form a new model that provides a high-precision, low-latency solution for traffic sign recognition in complex scenarios.

2 Related works

As an advanced target detection algorithm, YOLOv4 has undergone a lot of improvements and optimizations based on YOLOv3. By adopting more advanced network structures, training techniques and algorithm optimization, YOLOv4 has achieved significant performance improvements in various target detection tasks. Domestic and foreign researchers have conducted extensive discussions on this algorithm. For example, Gai (2023) proposed an improved YOLOv4 deep learning algorithm to detect cherry fruits for the problem of fast and accurate detection of cherries. The comparative experimental results show that the average accuracy value obtained by the proposed improved YOLOv4 model network is 0.15 higher than that of the comparison model, and it can detect cherries of different maturity in the same area. Wu (2023) proposed an optimized model based on the YOLOv4 network to solve the problem of difficulty in distinguishing crops and weeds in farmlands. Compared with the comparison model, the AP value of small target weeds increased by 15.1%, the mAP value increased by 4.2%, and the size of model parameters and training weight files decreased by 34%, indicating that the accurate detection of small target weeds can be improved. Kumar et al (2023) proposed a new mask detection system based on an improved miniature YOLOv4 object detector to address the issue of detecting mask-wearing people under surveillance cameras. Experimental results showed that the proposed miniature algorithm network achieved a 64.31% mAP on the dataset, 6.6% higher than miniature YOLOv4. Astuti et al. (2023) introduced a YOLOv4-based training model for dental detection, specifically designed to identify tooth sensitivity in the field of oral medicine. Experimental results showed that the true positive and true negative values for YOLOv4 were 1.534 and 1.568, respectively, with sensitivity and specificity of 99.42% and 87.06% on panoramic X-ray images, proving the effectiveness of the model. Li et al. (2023) proposed a fusion algorithm combining channel attention mechanisms with YOLO Head to address the low detection accuracy of traffic sign detection algorithms. Using the TT100K dataset for evaluation, the proposed method achieved real-time and accurate performance in complex backgrounds compared to existing methods.

With the continuous development of image recognition technology, new theories and methods are constantly being proposed and applied. Scholars from various countries have conducted in-depth research in this field. For example, Dewi et al (2023) proposed a method using the principle of Spatial Pyramid Pooling to enhance YOLOv3 and Densenet backbone networks for feature extraction, aimed at addressing practical problems in traffic sign recognition. According to experimental results, the method achieved an accuracy of 87.8% for small objects, 98.0% for medium-sized objects, and 98.6% for large objects within the BTSD dataset. Wei et al. (2023) introduced a traffic sign recognition algorithm combining ResNet and Convolutional Neural Network, aimed at enhancing road safety and minimizing accidents. Experimental results demonstrated that the ResNet-based model achieved a recognition accuracy of 99% on the test set, while the convolutional neural network-based model reached 98%. In order to solve the problem of difficulty in identifying traffic signs in bad weather, Dang et al. (2023) proposed to build a model using deep learning technology to help with identification. The experimental results showed that the accuracy of YOLOv7 was 78%, the accuracy of YOLOv5s+SE attention module was 78.4%, and the accuracy of YOLOv5s+C3GC was 79.2%. The YOLOv5s+C3GC model significantly improved the recognition accuracy of blurred and distant targets. Wang et al. (2023) proposed an improved feature pyramid model named adaptive feature fusion feature pyramid network to solve the problem of traffic sign detection in unmanned driving systems. The adaptive attention module and feature enhancement module were used to reduce the information loss in the feature map generation process. The experimental results showed that compared with several advanced methods, this method is more versatile and superior. In order to prevent traffic accidents and ensure the safety of all road users, Soylu et al. (2024) proposed a traffic sign recognition system based on YOLOv8. The experimental results showed that the system can assist drivers by providing real-time alerts and warnings about potential dangers, speed limits and other important information.

In summary, existing research has made significant progress in traffic sign recognition, but challenges such as low efficiency and insufficient accuracy remain. YOLOv4, with the integration of attention mechanisms, can address these issues. Therefore, the proposed model that combines the attention mechanism and YOLOv4 for traffic sign recognition has practical potential and aims to improve the efficiency and accuracy of traffic sign recognition to meet the demands of autonomous driving.

3 Traffic sign recognition strategy design based on attention mechanism and YOLOv4

3.1 Attention mechanism optimization based on SENet

Given the ongoing advancements in autonomous driving and deep learning technologies in recent years, information technology can efficiently recognize and process traffic sign image information (Tan et al., 2024). However, traditional image processing techniques still heavily rely on manual work and struggle to accurately identify certain regional features (Alimova, 2024). To address issues such as complex computation of traffic image data and difficulty in detecting small visual targets, the attention mechanism SENet is introduced (Rousselot et al., 2025). This method enables the algorithm to focus more on important feature channels and the model structure is shown in Fig. 1.

As shown in Fig. 1, after the image data is input into the attention module, a global spatial pooling operation is performed on each channel. Then, using two fully connected layers and a Sigmoid function with nonlinear characteristics, each channel is assigned a corresponding weight value. However, the number of parameters is directly correlated with the increasing complexity of the network model, further expanding the parameter count also amplifies the computational load. To address the issue of channel dependency, SENet reduces the feature map containing global information of size $W \times H \times C$ to a $1 \times 1 \times C$ feature vector. The channel-level statistical operation process z is shown in Eq. (1).

$$z_c = F_{sq}\left(u_c\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c\left(i, j\right) \tag{1}$$

In Eq. (1), z_c and u_c represent the feature map channels of the c-th elements of the feature map, respectively, and F_{sq} represents the compression operation. The weight calculation S obtained through the Sigmoid activation is shown in Eq. (2).

$$s = F_{ex}\left(z, W\right) \tag{2}$$

In Eq. (2), W represents the weight matrix of the fully connected layer, and F_{ex} represents the activation operation. The reweighting operation multiplies the learned channel weights S by the corresponding channels in the original feature map u, weighting the feature map to highlight important channels and suppress unimportant ones. The output feature map u^* is calculated as shown in Eq. (3).

$$u_c^* = F_{\text{scale}}\left(u_c, s_c\right) = s_c \cdot u_c \tag{3}$$

In Eq. (3), u_c^* represents the c-th channel of the output feature map u^*, and F_{scale} represents the Reweight operation. However, SENet generates a large computational load and parameter size when processing high-resolution images or feature maps with many channels, slowing down model training and inference speed. In tasks requiring fine local feature extraction, the performance may not be ideal. Therefore, the study proposes using depthwise separable convolutions to independently perform convolution operations on each channel. This not only improves the model's running efficiency but also captures image details and local structures more effectively. The structure of the depthwise separable convolution model is shown in Fig. 2.

As illustrated in Fig. 2, depthwise separable convolution breaks down standard convolution into two stages: Depthwise Convolution (DC) and Pointwise Convolution (PC). In convolution calculation, each time, M convolution kernels of size $L \times L$ are used, and the output is generally 1. PC, on the other hand, uses M convolution kernels of size 1×1 in each

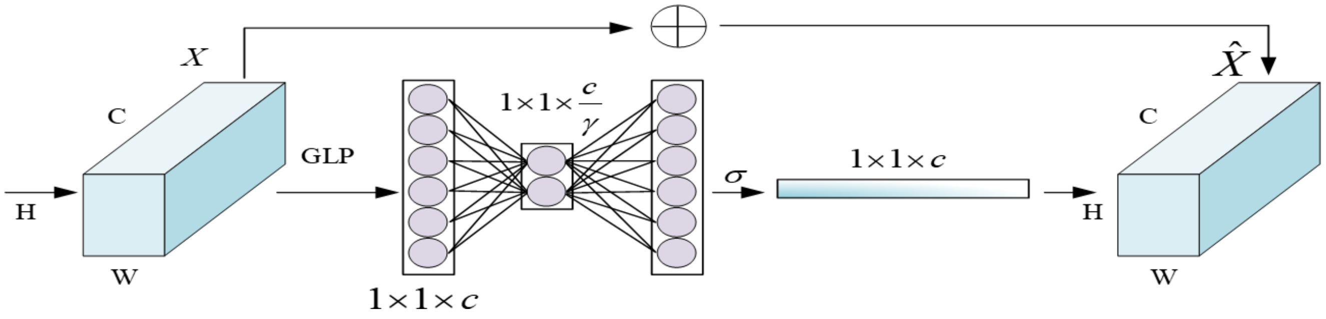

Fig. 1 Structure of SENet attention channel model

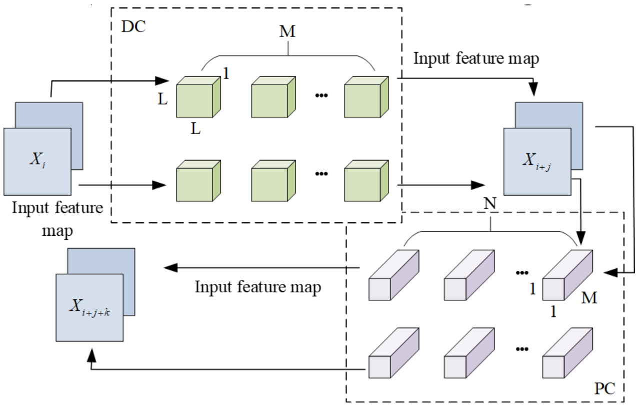

Fig. 2 Structure diagram of the depthwise separable convolution model

operation for convolution filtering. DC and PC together form a standard convolution with a kernel size of $L \times L$ and M channels. The computational load of DC $FLOP_{sdw}$ is calculated as shown in Eq. (4).

$$FLOP_{sdw} = C_{\text{in}} \times H_{\text{out}} \times W_{\text{out}} \times k \times k \tag{4}$$

In Eq. (4), C_{in} represents the number of input channels of the feature map, H_{out} and W_{out} represent the height and width of feature map after DC, and k is the size of the convolution kernel. The computational load of PC $FLOP_{sqw}$ is computed as demonstrated in Eq. (5).

$$FLOP_{spw} = C_{\text{in}} \times C_{\text{out}} \times H_{\text{out}} \times W_{\text{out}} \tag{5}$$

In Eq. (5), C_{in} represents the number of channels of PC, H_{out} and W_{out} are the height and width of the output feature map, and C_{out} is the number of output channels. The total computational load $FLOP_{sds}$ of depthwise separable convolution is the sum of the computational loads of DC and PC, as given by Eq. (6).

$$\begin{aligned} FLOP_{sds} &= C_{\text{in}} \times C_{\text{out}} \times H_{\text{out}} \times W_{\text{out}} \\ &+ C_{\text{in}} \times H_{\text{out}} \times W_{\text{out}} \times k \times k \end{aligned} \tag{6}$$

Subsequently, the study adopts SENet optimized with Deep Separable Convolution, named Depthwise Separable SENet (D-SENet). The deep convolution in the Deep Separable Convolution in this new attention mechanism extracts features for each channel separately and better captures the spatial features of the image. This enables SENet to provide richer spatial information when learning the channel attention mechanism, which helps SENet more accurately judge the importance of different channel features. The operation process of D-SENet is shown in Fig. 3.

As shown in Fig. 3, D-SENet first inputs the image to be recognized, then optimizes it through deep separable convolution to reduce the number of parameters, then inputs it into SENet and assigns it to network training, and finally uses the discriminator to determine whether the image recognition is completed, and outputs the recognized image after completion. The calculation process of the computational amount $FLOP_{sstd}$ of the standard convolution is shown in Eq. (7).

$$FLOP_{sstd} = C_{\text{in}} \times C_{\text{out}} \times H_{\text{out}} \times W_{\text{out}} \times k \times k \tag{7}$$

Combining Equations (6) and (7), when C_{in}, C_{out}, H_{out}, W_{out}, and k are large, the computational complexity of depthwise separable convolution is significantly smaller than that of standard convolution, indicating its advantage in computational efficiency.

3.2 Design of traffic sign recognition model combining attention mechanism and YOLOv4

Although the D-SENet recognition algorithm can effectively recognize traffic sign image data, multiple types and sizes of traffic signs often appear simultaneously in traffic scenes. D-SENet is more focused on image classification tasks and has significant limitations in the accuracy and efficiency of multi-object detection. YOLOv4 is an end-to-end object detection algorithm that directly outputs detection results from the input image, making both training and inference processes more efficient (Song et al., 2023). Therefore, the study proposes using the advantages of YOLOv4, such as fast detection speed, strong multi-object detection capability, and efficient detection of small objects, to tackle the aforementioned challenges. The YOLOv4 structure is shown in Fig. 4.

As illustrated in Fig. 4, the YOLOv4 algorithm is primarily composed of the backbone network, neck network, and detection head. CSPDarknet53 serves as the backbone network, with each grid unit on every feature map responsible for predicting multiple bounding boxes. Each bounding box contains the class information, position information, and confidence level of the object. In object detection, the predicted bounding boxes need to be regressed to make them as close as possible to the ground truth bounding boxes. Let the pre-set box be b and the ground truth box be b^{gt}, and their positions are shown in Eq. (8).

$$\begin{cases} b = (x, y, w, h) \\ b^{gt} = \left(x^{gt}, y^{gt}, w^{gt}, h^{gt}\right) \end{cases} \tag{8}$$

In Eq. (8), (x, y) and (w, h) represent the center coordinates and the width and height of the bounding box, respectively. Taking into account factors such as the distance between the center of the bounding boxes and the aspect ratio, the bounding box regression loss is computed as shown in Eq. (9).

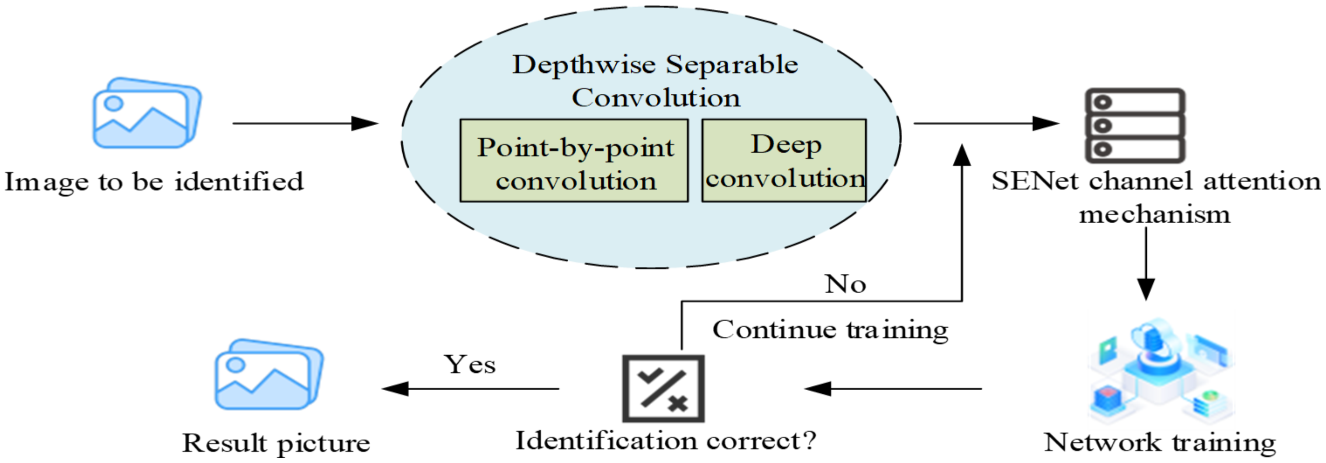

Fig. 3 Operation flow chart of D-SENet

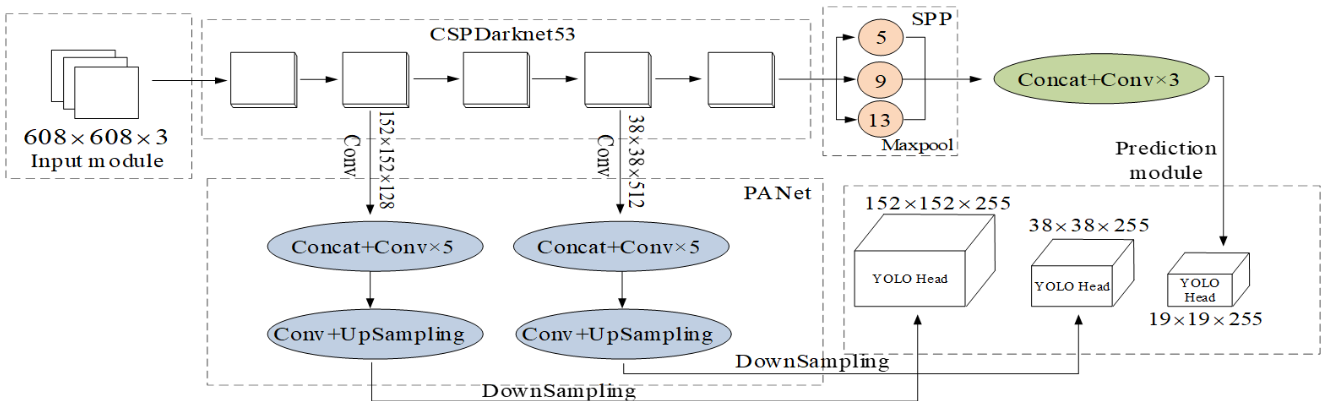

Fig. 4 Structure diagram of the YOLOv4 algorithm

$$L_{CIoU}=1-IoU+\frac{\rho^2\left(b,b^{gt}\right)}{c^2}+\alpha\upsilon \tag{9}$$

In Eq. (9), IoU is the Intersection-over-Union of the predicted and ground truth boxes, $\rho^2(b, b^{gt})$ is the Euclidean distance squared between the centers of the predicted and ground truth boxes, c is the diagonal length of the minimum enclosing rectangle, α is the weight coefficient, and υ is the parameter measuring the consistency of the aspect ratio. In object detection, there is often an imbalance between positive and negative samples. To solve this problem, YOLOv4 uses classification loss to adjust the weights of different samples. The classification loss is calculated as shown in Eq. (10).

$$L_{Focal}=-\alpha_t\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{10}$$

In Eq. (10), α_t represents the weight coefficient that balances positive and negative samples, while denotes the adjustment factor. The object confidence score reflects the probability that the predicted bounding box contains an object. YOLOv4 employs binary cross-entropy loss to compute the object confidence loss, as shown in Eq. (11).

$$L_{\text{Confidence}}=-C^{gt}\log\left(C\right)-\left(1-C^{gt}\right)\log\left(1-C\right) \tag{11}$$

In Eq. (11), C is the predicted object confidence, and C^{gt} is the ground truth object confidence. In real image data, noise interferes with YOLOv4's ability to extract and recognize object features, leading to a decrease in detection accuracy (Sun et al., 2023). Therefore, the study proposes using a soft thresholding module, which shrinks the feature values corresponding to noise near zero, to process the feature map and improve detection accuracy. The structure of the soft thresholding module is shown in Fig. 5.

As illustrated in Fig. 5, the input image A has a size of $c \times h \times w$. A changes the structure of the ReLU activation layer and convolution layer to generate image B. Global average pooling is then applied to obtain the initial threshold S. The initial threshold is used to obtain a scaling vector M, which is then combined with M and S to calculate the final threshold. Finally, based on this threshold, the noise in the feature map B is filtered, and the final result O is output. The soft thresholding decision process is shown in Eq. (12).

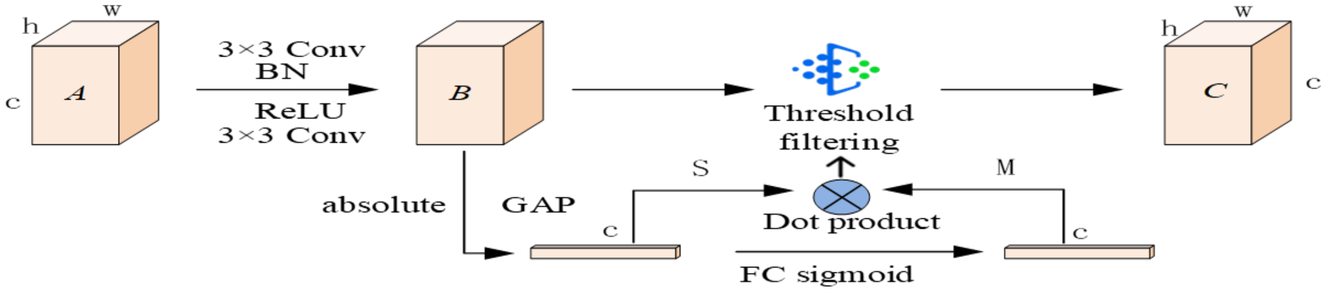

Fig. 5 Structure diagram of the soft thresholding module structure diagram

$$y = \begin{cases} x-\tau, \text{if } x>\tau \\ 0, \text{if } -\tau \le x \le \tau \\ x+\tau, \text{if } x>-\tau \end{cases} \quad (12)$$

In Eq. (12), x and y represent input and output respectively, and τ represents threshold. Subsequently, the soft threshold processing is performed on the coefficients in the transform domain. The operation of the soft threshold processing is shown in Eq. (13).

$$W_{j,k}{}' = \begin{cases} \operatorname{sgn}\left(W_{j,k}\right)\left(\left|W_{j,k}\right|-\tau\right), \left|W_{j,k}\right| \ge \tau \\ 0, \left|W_{j,k}\right| < \tau \end{cases} \quad (13)$$

In Eq. (13), $W_{j,k}$ is the original transform domain coefficient, and $W_{j,k}$ is the coefficient after soft threshold processing. In summary, the soft threshold denoising module ensures the removal of noise before feature fusion, avoiding noise interference in subsequent target positioning and classification, especially for small target detection. This soft threshold denoising module enhances semantic feature extraction through adaptive noise suppression. Its complex semantic extraction mechanism utilizes a global average pooling layer to capture the global context of the feature map and generate an initial threshold. The final threshold is dynamically adjusted using a learnable scaling vector, enabling the module to distinguish between noise and subtle semantic features. Noise is filtered from shallow, high-resolution feature maps to prevent noise from being amplified during subsequent feature pyramid propagation. The denoised feature maps are then fed into the attention module, which focuses channel weight allocation on valid semantic regions, forming a cascaded "denoising-attention enhancement" optimization process. The new traffic sign recognition model after integrating D-SENet channel attention mechanism, soft threshold denoising module and YOLOv4 is named Depthwise Separable SENet YOLOv4 (DSS-YOLOv4). The model's processing and recognition process for traffic signs is shown in Fig. 6.

In Fig. 6, after the input feature map is processed by CSPDarknet53 for feature extraction, the soft threshold denoising module removes noise from these input feature maps. CSPDarknet53 outputs feature maps at multiple scales, and D-SENet is applied to these multi-scale feature maps. The DSS-YOLOv4 backbone network combines the characteristics of CSPNet and Darknet, while the neck network uses Path Aggregation Network (PANet) and spatial pyramid pooling modules. The detection head is improved using multi-scale detection.

4 Performance evaluation of traffic sign recognition method improved by YOLOv4

4.1 Effect verification of improved D-SENet channel attention mechanism

To verify the superiority of the D-SENet channel attention mechanism, the study compared it with three attention mechanisms for image recognition: Efficient Channel Attention Network (ECA-Net), Convolutional Block Attention Module (CBAM), and Dual Attention Network (DANet). The experimental system was set up with Windows 10, Ubuntu 10.08 operating system, PyTorch 1.7.0deep learning framework, Adam optimizer, Python 3.8programming language, NVIDIA TITAN XP Graphics Processing Unit, and 48GB of RAM. To ensure the authenticity and reliability of the experiment, the TT100K dataset and CCTSDB 2021 dataset were selected for testing and training, covering a variety of traffic scenarios and annotating 128 types of traffic signs. D-SENet, ECA-Net, CBAM, and DANet were tested for accuracy on both datasets. The test results are shown in Fig. 7.

As shown in Fig. 7 (a), when D-SENet was trained on the TT100K dataset, the accuracy reached 98.5% when the number of iterations was 5. The overall accuracy curve gradually stabilized after 10 iterations. The DANet accuracy curve fluctuated significantly. When the number of iterations was 0–20, the accuracy fluctuated greatly, reaching a maximum of 80.2%. The accuracy curve of CBAM was relatively flat when the number of iterations was 0–20, and there was a significant fluctuation when the number of iterations was 25–40,

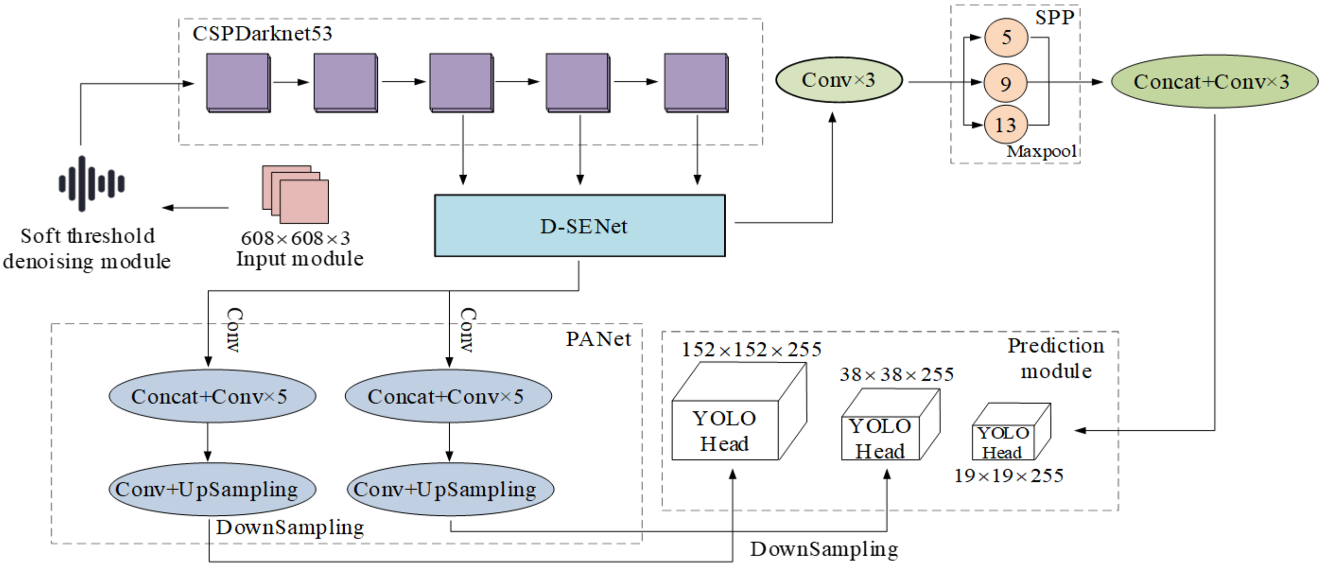

Fig. 6 Flowchart of traffic sign recognition using the proposed model

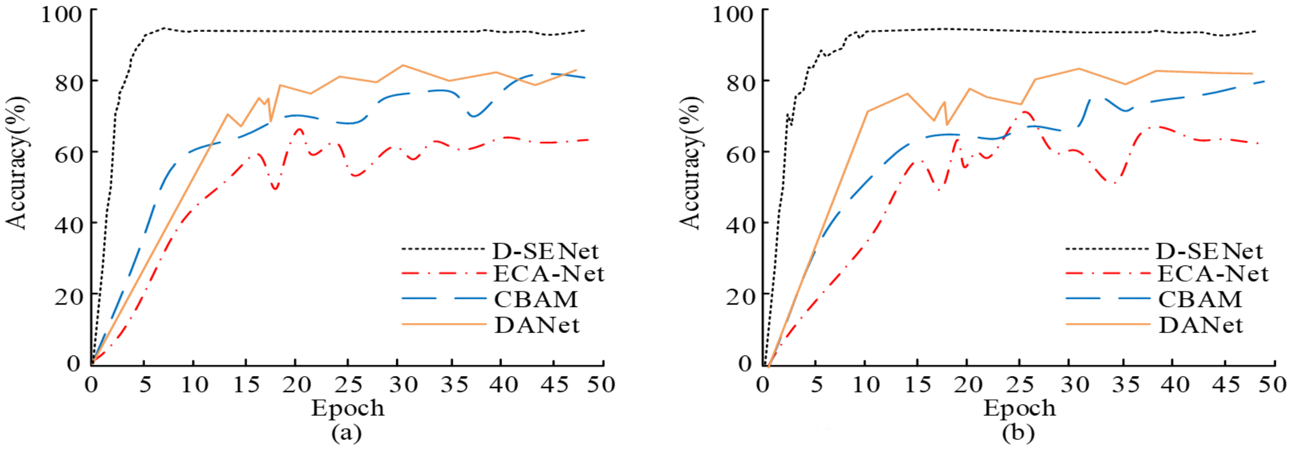

Fig. 7 Accuracy test results comparison chart: (a) Accuracy curve in the TT100K dataset; (b) Accuracy curve in the CCTSDB 2021 dataset

with a maximum accuracy of 79.8%. The average accuracy of ECA-Net was the lowest among the comparison algorithms, and it only maintained around 60% after stabilization. As shown in Fig. 7(b), when D-SENet was trained on the CCTSDB 2021 dataset, the overall accuracy curve gradually stabilized after 10 iterations, with a maximum accuracy of 98.4%. The accuracy of DANet fluctuated greatly, reaching a maximum of 81.4%. The accuracy curve of CBAM showed a gentle rise, with a maximum accuracy of 80.5%. The accuracy of ECA-Net dropped significantly when the number of iterations was 25–35, and then stabilized at around 60%. In summary, D-SENet showed higher accuracy, stronger convergence, and did not decrease with the increase in the number of iterations, which was significantly better than the comparison algorithms. At the same time, in order to demonstrate the feature prediction performance of D-SENet, the study selected 180 traffic sign images for prediction evaluation, and the test results are shown in Fig. 8.

As shown in Fig. 8, when the training time for D-SENet was 43.6 sec, the number of recognized predictions closely matched the real number, with 165 traffic sign images successfully identified. When the training time was between 20 s and 30 s, the prediction performance for traffic signs was weaker, with 80 images not successfully predicted. At 40.5 s of training time, 120 images were correctly predicted, but due to the complexity and blurriness of some traffic signs, the predictions did not match the real images accurately. In summary, the D-SENet traffic sign image recognition algorithm was capable of predicting various types of traffic signs, with overall good recognition accuracy. To further demonstrate D-SENet's recognition performance, the study continued to test its precision, recall ratio, and the harmonic mean of recall (F1-score) against ECA-Net, CBAM, and DANet. The test results are shown in Table 1.

As shown in Table 1, when tested on the TT100K dataset, D-SENet achieved a precision of 96.35%, recall ratio

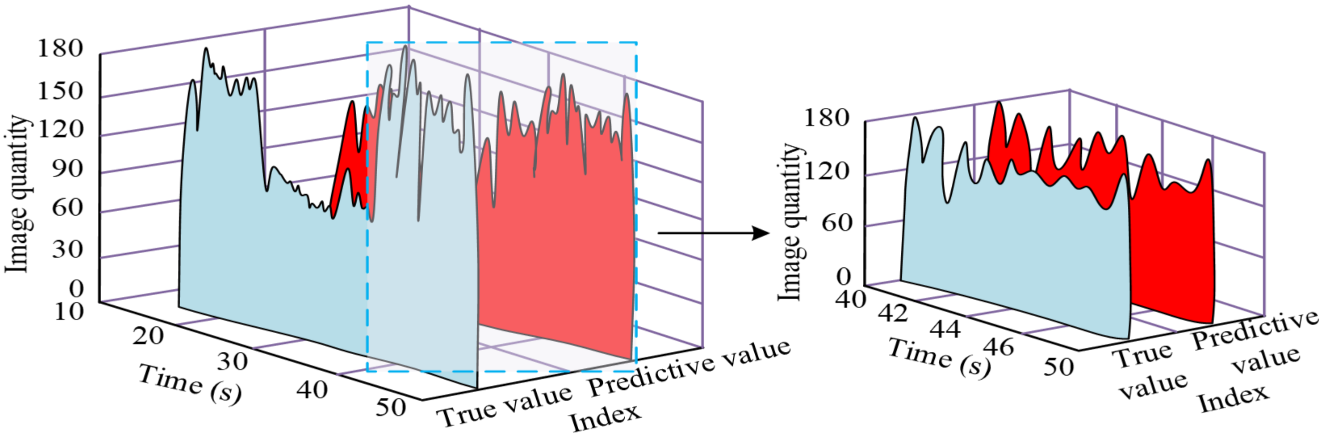

Fig. 8 Feature prediction performance comparison results

Table 1 Precision, recall ratio and F1-score experimental results

z	Algorithm	Precision (%)	Recall (%)	F1-score (%)
TT100K	D-SENet	96.35	95.88	95.12
	ECA-Net	92.14	90.23	91.45
	CBAM	91.66	90.25	92.41
	DANet	84.83	89.68	88.12
CCTSDB 2021	D-SENet	97.45	98.23	97.65
	ECA-Net	89.23	90.66	89.46
	CBAM	90.11	88.95	92.13
	DANet	85.98	87.96	90.17

of 95.88%, and F1-score of 95.12%. ECA-Net had a precision and recall ratio of 92.14% and 90.23%, respectively, which were 4.21% and 5.65% lower than those of D-SENet. On the CCTSDB 2021 dataset, CBAM's recall ratio and F1-score were 88.95% and 92.13%, respectively, both lower than those of D-SENet. For both datasets, DANet's three indicators did not exceed 90%, with a recall ratio of 87.96%, which was 10.27% lower than those of D-SENet. The introduction of the SENet channel attention mechanism significantly enhances the model's ability to weight key color channels. This mechanism adaptively enhances discriminative color information while suppressing interfering channels, effectively reducing confusion between similarly colored landmarks (as evidenced by the high precision and recall in Table 1). However, it should be noted that the model may still struggle to distinguish between landmarks in extreme lighting conditions (such as strong backlight or dusk) or in scenes with extremely high color similarity due to severe fading and color distortion.

4.2 Evaluation and analysis of the improved YOLOv4 image classification model

After validating the performance of the D-SENet traffic sign recognition algorithm, the study further evaluated the performance of the DSS-YOLOv4 image classification model, which was based on D-SENet and YOLOv4 improvements, by comparing it with Faster Region-based Convolutional Neural Network (Faster R-CNN), Field-Programmable Gate Array (FPGA), and transfer learning ensemble algorithms. The experiment used PyTorch as the core deep learning framework, with the development environment set to Anaconda 3. The hardware configuration included an Intel Xeon Gold 6230R CPU, an NVIDIA GeForce RTX 3080 GPU, and the traffic sign classification recognition datasets CTSDB and CCTSDB. The study tested the accuracy of four image classification recognition models–DSS-YOLOv4, Transfer learning ensemble, Faster R-CNN, and FPGA–on four main recognition tasks. The results are shown in Fig. 9.

As shown in Fig. 9 (a), the accuracy of DSS-YOLOv4 in the recognition test of traffic prohibition signs was 0.95, the accuracy in the recognition test of traffic warning signs was 0.97, and the accuracy in the recognition test of traffic

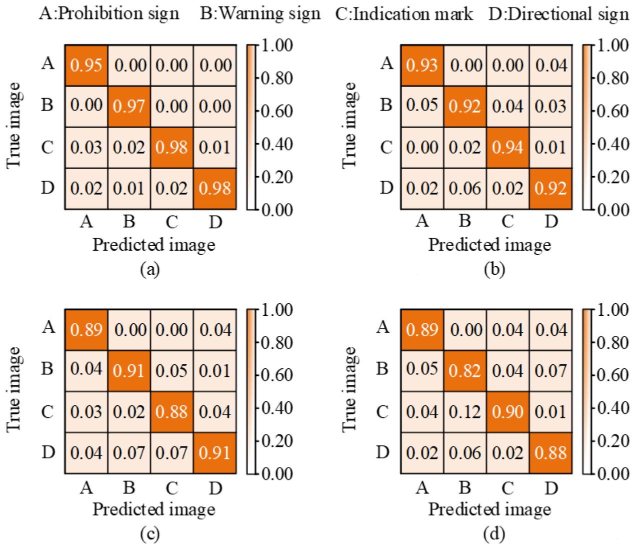

Fig. 9 Traffic sign recognition accuracy experimental results: (a) Recognition result of DSS-YOLOv4 model; (b) Recognition result of Transfer learning ensemble midel; (c) Recognition result of Faster R-CNN model; (d) Recognition result of FPGA model identification result

instructions and road signs was 0.98. Only some images were recognized as the wrong type. As shown in Fig. 9 (b), the accuracy of the transfer learning integrated image classification recognition model in the recognition test of traffic prohibition signs was 0.93, and the accuracy in the recognition test of traffic warning signs and road signs was 0.92. As shown in Fig. 9 (c), the recognition accuracy of Faster R-CNN for traffic prohibition signs and road signs was low, with recognition accuracies of 0.89 and 0.88, respectively. As shown in Fig. 9 (d), the recognition accuracy of FPGA for the four types of traffic signs was low, with the highest being 0.90. In summary, DSS-YOLOv4 performed good recognition of different types of traffic signs and was significantly better than the comparison model. In order to further verify the recognition effect of DSS-YOLOv4 on different types of traffic signs on actual traffic sections, the recognition accuracy visualization effect of it and the other three comparison models is shown in Fig. 10.

As shown in Fig. 10 (a), DSS-YOLOv4 achieved the best recognition performance for unobstructed traffic signs, with a detection accuracy of up to 0.99. The recognition performance slightly decreased for signs with occlusions or at a distance. As shown in Fig. 10 (b), the transfer learning ensemble model had poor recognition for small traffic signs, with the lowest accuracy being 0.81. As shown in Fig. 10 (c), Faster R-CNN achieved a maximum detection accuracy of 0.95 for unobstructed traffic signs and a minimum of 0.86. As shown in Fig. 10 (d), FPGA's detection accuracy for various traffic signs was low, with the highest accuracy being only 0.89. The research method's recognition performance for complex-shaped traffic signs benefits from the D-SENet attention mechanism optimized with deep separable convolutions. The model excels at extracting local details and spatial features. This enables the model to effectively capture subtle features of non-standard shapes and slender structures (such as some warning signs), significantly improving recognition accuracy for such objects. However, for complex-shaped signs that are extremely rare or lack sufficient training data, the model's generalization ability may still be limited, resulting in reduced recognition confidence or misclassification. Therefore, to test the robustness of the algorithm model to noise interference, −10dB Gaussian Signal-to-Noise Ratio (SNR) noise was added for comparison. The Average Precision (AP) comparisons of the four models are shown in Fig. 11.

As shown in Fig. 11, after adding −10 dB Gaussian signal-to-noise ratio noise, the average precision of the four algorithm models showed a general decline. However, compared with Faster R-CNN, transfer learning integration, and FPGA recognition models, the decline of DSS-YOLOv4 was smaller, dropping from the original 95% to nearly 90%. The experimental results showed that DSS-YOLOv4 had better robustness.

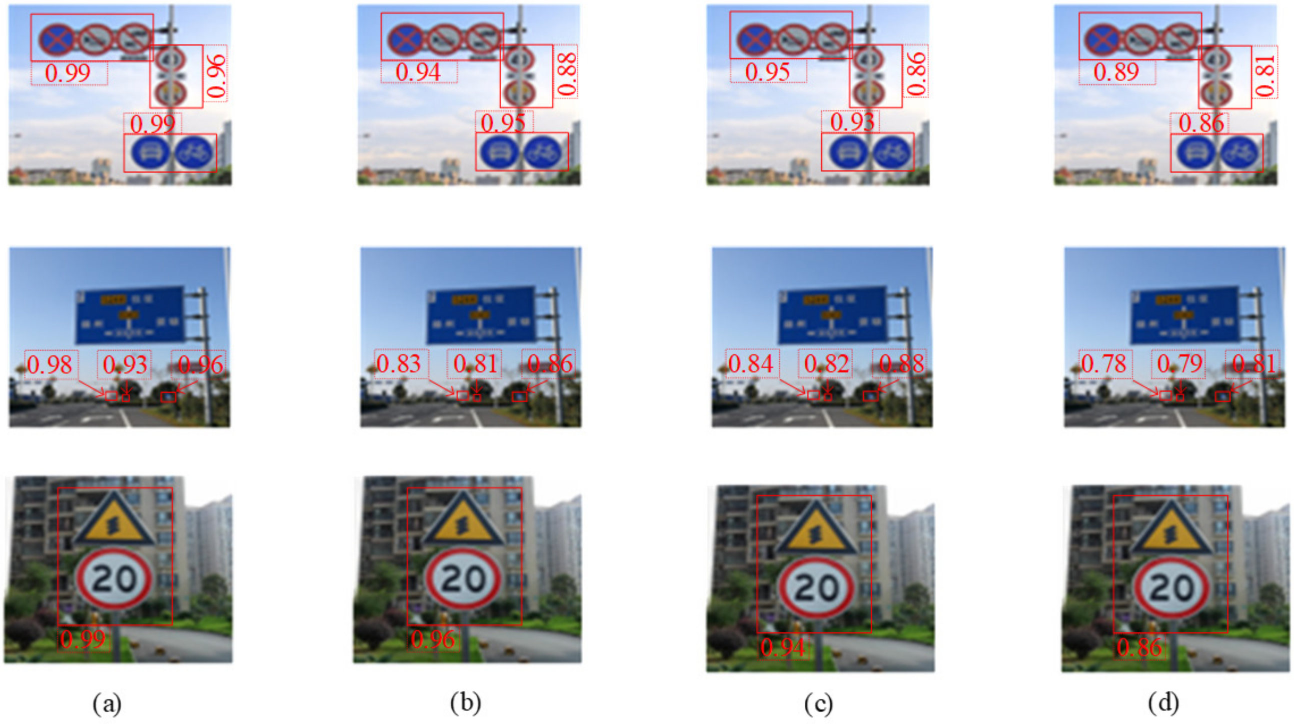

Fig. 10 Visualization of recognition accuracy : (a) Target recognition result of DSS-YOLOv4; (b) Target recognition result of Transfer learning ensemble; (c) Target recognition result of Faster R-CNN; (d) Target recognition result of FPGA

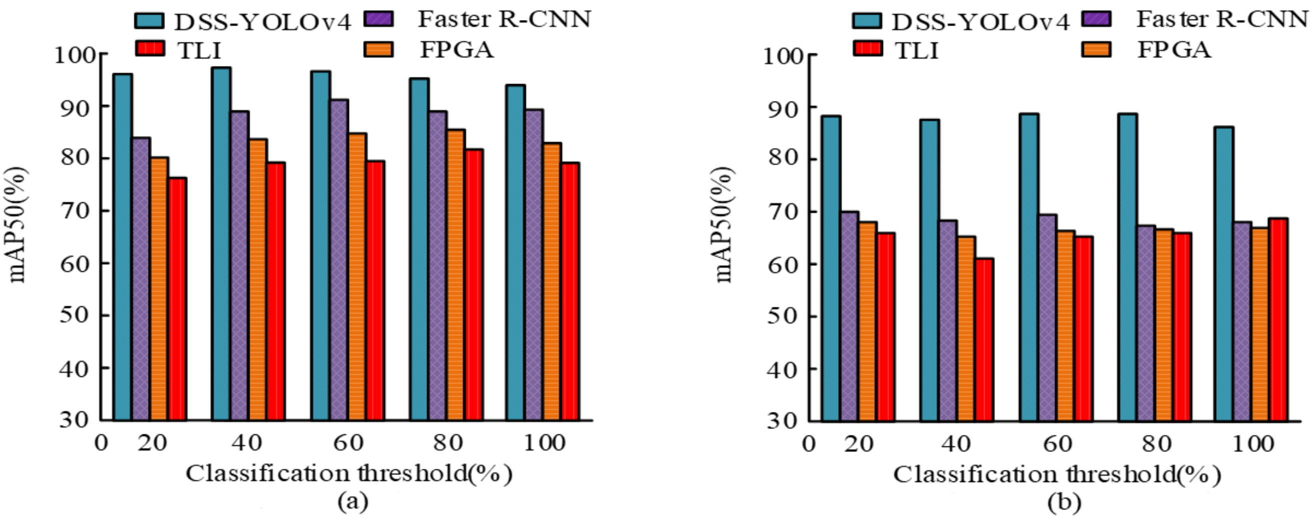

Fig. 11 AP results of the four models examined: (a) Primary noise; (b) −10dB Gaussian white noise

5 Conclusion

In order to solve the problems of untimely feature extraction and inaccurate recognition in traffic sign recognition models, a DSS-YOLOv4 traffic sign recognition model was designed. The model used a soft thresholding module to process the feature map and shrink the eigenvalues corresponding to the noise to near zero, thereby reducing the impact of noise on the subsequent detection process and improving the recognition accuracy. The experimental results showed that when the number of D-SENet iterations was 5, the accuracy reached 98.5%. The harmonic mean of recognition accuracy, prediction recall rate, and recall rate reached 96.35%, 95.88%, and 95.12% respectively. The DSS-YOLOv4 traffic sign recognition model was evaluated and found to have good recognition accuracy for different types of traffic signs, reaching a maximum of 0.98. In addition, the average accuracy of DSS-YOLOv4 under −10 dB Gaussian signal-to-noise ratio noise interference was maintained at 90%, which was much higher than that of the comparison models. In summary, the DSS-YOLOv4 traffic sign recognition model effectively classified and recognized different types of traffic signs. It should be noted that this study did not systematically test the model's real-time inference speed on embedded platforms. Although the YOLOv4 architecture itself is highly efficient, the introduction of the depthwise separable SENet attention mechanism and soft threshold denoising module increases computational complexity, potentially impacting real-time performance on low-computing automotive devices. Future work requires in-depth exploration of hardware acceleration (such as TensorRT deployment) and model lightweighting (such as channel pruning or quantization) to further enhance the model's applicability in real-time autonomous driving scenarios. Furthermore, recognition robustness under extreme lighting conditions needs to be enhanced through methods such as adversarial training and multimodal fusion.

References

Alimova, S. (2024) "The role of information technology in the personnel management system", Modern Science and Research, 3(2), pp. 385–390. https://doi.org/10.5281/zenodo.10647479

Astuti, E.R., Putra, R.H., Putri, D.K., Ramadhani, N.F., Ahmad, T.N.E.B.T., Putra, B.R., Djajadiningrat, A.M.P. (2023) "The Sensitivity and Specificity of YOLO V4 for Tooth Detection on Panoramic Radiographs", Journal of International Dental and Medical Research, 16(1), pp. 442–446.

Dang, T. P., Tran, N. T., To, V. H.,Thi, M. K. T. (2023) "Improved YOLOv5 for real-time traffic signs recognition in bad weather conditions", The Journal of Supercomputing, 79(10), pp. 10706–10724. https://doi.org/10.1007/s11227-023-05097-3

Dewi, C., Chen, R. C., Yu, H., Jiang, X. (2023) "Robust detection method for improving small traffic sign recognition based on spatial pyramid pooling", Journal of Ambient Intelligence and Humanized Computing, 14(7), pp. 8135–8152. https://doi.org/10.1007/s12652-021-03584-0

Diwan, T., Anirudh, G., Tembhurne, J. V. (2023) "Object detection using YOLO: challenges, architectural successors, datasets and applications", Multimedia Tools and Applications, 82(6), pp. 9243–9275. https://doi.org/10.1007/s11042-022-13644-y

Gai, R., Chen, N., Yuan, H. (2023) "A detection algorithm for cherry fruits based on the improved YOLO-v4 model", Neural Computing and Applications, 35(19), pp. 13895–13906. https://doi.org/10.1007/s00521-021-06029-z

Kumar, A., Kalia, A., Sharma, A., Kaushal, M. (2023) "A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system", Journal of Ambient Intelligence and Humanized Computing, 14(6), pp. 6783–6796. https://doi.org/10.1007/s12652-021-03541-x

Lescoat, A., Huang, S., Carreira, P. E., Siegert, E., de Vries-Bouwstra, J., …, Allanore, Y. (2023) "Cutaneous Manifestations, Clinical Characteristics, and Prognosis of Patients With Systemic Sclerosis Sine Scleroderma: Data From the International EUSTAR Database", JAMA Dermatology, 159(8), pp. 837–847. https://doi.org/10.1001/jamadermatol.2023.1729

Li, Y., Li, J., Meng, P. (2023) "Attention-YOLOV4: a real-time and high-accurate traffic sign detection algorithm", Multimedia Tools and Applications, 82(5), pp. 7567–7582. https://doi.org/10.1007/s11042-022-13251-x

Megalingam, R. K., Thanigundala, K., Musani, S. R., Nidamanuru, H., Gadde, L. (2023) "Indian traffic sign detection and recognition using deep learning", International Journal of Transportation Science and Technology, 12(3), pp. 683–699. https://doi.org/10.1016/j.ijtst.2022.06.002

Rousselot, M., Mahé, E., Senet, P., Rousselot, P., Baudot, N., …, Tella, E. (2025) "Chemotherapy-induced leg ulcers: a case series", Journal of Wound Care, 34(3), pp. 250–254. https://doi.org/10.12968/jowc.2020.0128

Song, L., Liu, M., Liu, S., Wang, H., Luo, J. (2023) "Pest species identification algorithm based on improved YOLOv4 network", Signal, Image and Video Processing, 17(6), pp. 3127–3134. https://doi.org/10.1007/s11760-023-02534-x

Soylu, E., Soylu, T. (2024) "A performance comparison of YOLOv8 models for traffic sign detection in the Robotaxi-full scale autonomous vehicle competition", Multimedia Tools and Applications, 83(8), pp. 25005–25035. https://doi.org/10.1007/s11042-023-16451-1

Simran, Sristi, T., Shilpi, K., Radhey, S. (2022) "Detection of traffic sign using CNN", Recent Trends in Parallel Computing, 9(1), pp. 14–23. https://doi.org/10.37591/rtpc.v9i1.269

Tan, K., Wu, J., Zhou, H., Wang, Y., Chen, J. (2024) "Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems", Journal of Theory and Practice of Engineering Science, 4(1), pp. 41–48. https://doi.org/10.53469/jtpes.2024.04(01).06

Terven, J., Córdova-Esparza, D. M., Romero-González, J. A. (2023) "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS", Machine Learning and Knowledge Extraction, 5(4), pp. 1680–1716. https://doi.org/10.3390/make5040083

Wang, J., Chen, Y., Dong, Z., Gao, M. (2023) "Improved YOLOv5 network for real-time multi-scale traffic sign detection", Neural Computing and Applications, 35(10), pp. 7853–7865. https://doi.org/10.1007/s00521-022-08077-5

Wei, Y., Gao, M., Xiao, J., Liu, C., Tian, Y., He, Y. (2023) "Research and Implementation of Traffic Sign Recognition Algorithm Model Based on Machine Learning", Journal of Software Engineering and Applications, 16(6), pp. 193–210. https://doi.org/10.4236/jsea.2023.166011

Wu, H., Wang, Y., Zhao, P., Qian, M. (2023) "Small-target weed-detection model based on YOLO-V4 with improved backbone and neck structures", Precision Agriculture, 24(6), pp. 2149–2170. https://doi.org/10.1007/s11119-023-10035-7

Sun, Y., Li, S., Gao, H., Zhang, X., Lv, J., Liu, W., Wu, Y. (2023) "Transfer learning: A new aerodynamic force identification network based on adaptive EMD and soft thresholding in hypersonic wind tunnel", Chinese Journal of Aeronautics, 36(8), pp. 351–365. https://doi.org/10.1016/j.cja.2023.03.024