DTSR-YOLO: Traffic Sign Detection Based on Improved YOLOv8

Bozhang Liu,Wenhua Cui*,Ye Tao

Abstract—**With the continuous development of artificial intelligence, traffic sign recognition plays a crucial role in intelligent vehicle perception systems. However, due to complex environmental changes such as weather, lighting, and occlusions, traffic sign recognition remains highly challenging. This paper proposes a traffic sign detection model DTSR-YOLO based on the improved YOLOv8n model to address these issues. Firstly, the DAT (Deformable Attention) module is introduced into the backbone network to adjust the shape and size of the attention model dynamically, enhancing the model's feature extraction capability. Secondly, the RFAConv convolution is introduced into the C2f module to construct a new C2f_RFA module, simplifying the calculation process and improving the detection speed and accuracy of the model. To further enhance accuracy and robustness, ELAN and SPPF are combined to form a new SPPELAN module, improving computational efficiency and feature extraction ability. Additionally, a small object detection head is added to integrate multi-scale features better and improve detection performance for small objects and complex scenes. Experimental results show that compared with the original YOLOv8n model, the proposed method enhances the mAP values by 8.4% and 2.8% on the Tsinghua-Tencent 100K (TT 100K) and CUST Chinese Traffic Sign Detection Benchmark 2021 (CCTSDB 2021) datasets, respectively. Test results in real complex scenarios indicate that the detection performance of this algorithm is superior to that of the YOLOv8n algorithm, and the DTSR-YOLO algorithm can accurately detect traffic signs that the YOLOv8n algorithm cannot. Therefore, the algorithm proposed in this paper can effectively improve the detection accuracy of traffic signs, is suitable for complex scenarios, and has good detection performance for small objects.**

Index Terms—**Traffic sign detection, YOLOv8, Attention mechanism, SPPELAN**

I. Introduction

IN recent years, With the in-depth research of artificial intelligence and computer vision[1] , object recognition technology has been rapidly developed in the field of automatic driving. However, small object recognition in a complex background has been a difficult problem in the industry. With the development of autonomous driving technology and the increase in car ownership, the safety of autonomous driving has attracted more and more attention. The detection and recognition system of traffic signs is one of the critical components of automatic driving. In the current road scene, many factors lead to missing and wrong detection of traffic signs, such as trees, lighting, and so on. These factors may affect the recognition of traffic signs. Traffic sign detection is one of the core technologies in intelligent driving, which is directly related to the safety of drivers and the regular operation and commuting order of the city. However, in practical applications, especially in driverless cars, misjudgment of traffic signs often occurs and may even lead to serious traffic accidents. Therefore, it is imperative to improve the accuracy of traffic sign detection.

As one of the core technologies in intelligent transportation System (ITS), traffic sign detection has been widely studied and applied in recent years. Domestic and foreign scholars in this field have experienced the transformation from traditional image processing methods to deep learning technology, and have made a lot of progress, but still face some challenges.

In foreign countries, the research on traffic sign detection started earlier, especially with the promotion of automatic driving and advanced driver assistance systems (ADAS), and related technologies have been rapidly developed [2]. Initially, foreign research mainly relied on traditional image processing methods, such as colour, shape, and texture feature extraction. It carried out mark detection based on template matching, edge detection, Hough transform, and other methods. However, these methods are more sensitive to environmental changes (such as light, shadows, weather changes, etc.) and are difficult to handle complex traffic scenes. With the rapid development of computer vision technology and deep learning, traffic sign detection methods based on convolutional neural networks (CNN) and deep neural networks (DNN)[3] have gradually become mainstream. Especially since AlexNet won the ImageNet competition in 2014, the advantages of deep learning in the field of image recognition have gradually emerged, and more and more traffic sign detection methods have adopted end-to-end models based on deep learning. Examples include Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector) [4]. These methods can automatically learn compelling features from a large amount of data, significantly improving the accuracy and robustness of mark detection, especially in complex environments. Foreign research institutions, such as Waymo and Tesla in

Manuscript received December 24, 2024; revised May 26,2025 .

This work was supported by National Natural Science Foundation of China (62272093), the Economic and Social Development Research Topics of Liaoning Province (2025-10146-244),and Postgraduate education and teaching reform research project of Liaoning Province (LNYJG2024092).

Bozhang Liu is a postgraduate student majoring in computer technology at School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. (e-mail:liubz_1998@163.com)

Wenhua Cui is a professor at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan,114051, China. (corresponding author to provide phone: +8613304224928; e-mail: taibeijack@126.com)

Ye Tao is an Associate Professor at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, China. (e-mail: taibeijack@163.com)

the United States, have also successfully applied traffic sign detection technology to autonomous driving systems to achieve efficient identification in different traffic scenarios.

In China, the research of traffic sign detection started relatively late, but in recent years, with the rapid development of intelligent transportation systems and autonomous driving technology, the research intensity and application demand have also increased sharply. There are many kinds of traffic signs in China; their forms are complex, and they involve different cultural and linguistic backgrounds. Realizing efficient and accurate sign detection has become the focus of domestic research. Domestic scholars' research on traffic sign detection mainly focuses on the following aspects: first, the robustness of the algorithm, especially its performance in complex environments such as different weather, illumination, Angle, and occlusion; Second, multi-modal and multi-scale detection methods are combined with different image features and deep learning models to improve the generalization ability of the model. The third is real-time and computational efficiency, especially in the process of real-time detection and reaction framework, combined with algorithms such as YOLO[5] and Faster R-CNN[6], and improved the auction time. Domestic researchers proposed an improved object literacy and real-time performance of traffic sign detection through multi-task learning, data enhancement, and transfer learning. In addition, with the rise of autonomous driving technology, domestic technology companies and universities are also engaged in cross-field cooperation to develop sign detection systems suitable for China's complex traffic environment.

At present, some significant challenges in domestic and international research include: 1) Robustness in complex environments, especially in extreme weather, night driving, intense light or shadow, and the existing detection methods still have certain limitations; 2) Classification and recognition of multi-category signs, especially in the case of apparent diversity and heterogeneity of traffic signs, how to efficiently and accurately distinguish different types of signs is still a challenge; 3) Data sets and labelling problems. Due to the inconsistent labelling standards of traffic sign data and the lack of large-scale and high-quality labelling data, using a small amount of labelled data for practical training is still a research hotspot.

In general, traffic sign detection technology has made remarkable progress at home and abroad, and the introduction of deep learning has greatly improved the detection accuracy and real-time performance. However, there are still challenges in adapting to complex environments, effectively recognizing multiple signs, and acquiring and labelling large-scale data sets. With the continuous progress of technology and the gradual maturity of intelligent transportation systems, the research and application prospects of traffic sign detection will be broader.

With the application and development of deep learning in the field of target detection, the advantage of fast target acquisition and accurate detection is achieved by using the target detection algorithm. In the automatic driving scenario, the captured traffic sign information is fed back to the vehicle system and the driver, and the subsequent driving guidance is given. However, the recognition of traffic signs is easily affected by trees, weather, light changes, and other factors. At present, most traffic sign detection can only be carried out in a typical environment. If the environment is complex, the detection of traffic signs will have certain inaccuracies. Inspired by the YOLO[7] series of object detection algorithms, this study introduces an improved traffic sign detection algorithm, DTSR-YOLO. The experimental results on the benchmark data set show that DTSR-YOLO significantly improves the performance of traffic sign small target detection.

The main contributions of this paper are as follows:

- Incorporating the deformable Attention (DAT)[8] mechanism into the backbone tail of YOLOv8[9]. In this mechanism, the attention is dynamically sparse, only a few key positions are calculated, and the efficiency is improved. At the same time, the deformable attention window is used to adjust the area of attention to the data content dynamically. Effectively address the challenges posed by complex backgrounds common in traffic scenarios.
- An efficient C2f_RFA module constructed by RFAConv is further introduced to replace the original C2f module, enhance the capability of the receptive field of the convolutional layer, improve the efficiency of large-size convolutional nuclei, and thus improve the detection efficiency.
- The SPPELAN[10] module then extends the model's receptive field, enhances its robustness, and improves the integration of features at different scales.
- Finally, a small target detection head is added to the head to enhance the model's ability to detect small targets.

II. RELATED WORK

A. Traffic sign detection

Traffic sign detection is an important field of application in computer vision. With the rapid development of deep learning technology, this field has experienced significant technical evolution. From the early traditional methods to the current advanced methods based on deep learning, the technology of traffic sign detection has been continuously improved, improving the accuracy, robustness, and real-time detection.

1)Traffic sign detection based on colour

Before the wide application of deep learning, the detection of traffic signs was mainly through traditional image processing methods. The colour-based traffic sign detection method mainly includes converting images from RGB space to other colour Spaces and extracting candidate areas by using the three unique colours of traffic signs (red, yellow, and blue). The edge information in the image is extracted by the Canny compilation and detection algorithm to match the common shape of traffic signs, such as circles and rectangles.

2)Traffic sign detection based on shape

With the development of machine learning methods such as vector machine (SVM), the performance of traffic sign detection has been improved. Through the research of scholars, feature extraction and classification are derived by combining machine learning algorithms, such as HOG features and SVM, AdaBoost, and Cascade Classifiers. In 2012, the success of AlexNet's model heralded the widespread application of deep learning, including convolutional neural networks (CNNS) in computer vision.

In terms of traffic sign detection, from traditional algorithms to CNN-based automatic feature learning methods, deep learning can learn more robust features from large data sets without the need for manual design. The convolutional neural network can be used to extract and classify traffic signs. CNN can extract complex features from images, with the development of deep learning frameworks and the emergence of large-scale datasets (such as CCTSDB[11] and GTSDB). With the development of YOLO[12] (You Only Look Once) and other end-to-end object detection frameworks, the research on traffic sign detection has entered a new stage.

B. YOLOv8

YOLOv8 is the latest iteration in the YOLO series of object detection algorithms, combining cutting-edge technologies for enhanced performance. Building on the advantages of YOLOv5 and YOLOv7, YOLOv8 offers a balanced approach to accuracy and speed. Key features include an improved feature pyramid network (FPN)[13] and path aggregation network (PANet), providing scalability for diverse applications. The model is composed of several components, such as input modules, backbone networks, neck layers, and detection heads, each optimized for better performance. YOLOv8 comes in five versions (n, s, m, l, x), varying in depth and width to cater to different requirements, from lightweight to high-precision models. Compared to previous versions, YOLOv8 reduces computational load while boosting accuracy for more complex tasks[14]. The input module uses adaptive scaling to adjust image size dynamically and incorporates the Mosaic data augmentation technique to improve robustness. The backbone network includes convolutional layers[15], the C2f module, and SPPF (Space Pyramid Pooling Fast Edition), which enhance feature extraction and information flow. The C2f module optimizes gradient flow and maintains network efficiency, while the SPPF module improves multi-scale feature fusion for better overall performance. YOLOv8 exemplifies the latest advancements in object detection, offering both efficiency and accuracy for complex scenarios.

III. IMPROVEMENTS

This paper presents the DTSR-YOLO network, significantly improving object detection compared to traditional YOLOv8n. Specifically, the DTSR-YOLO integrates DAT, SPPELAN modules, and a custom layer designed explicitly for small object detection. Together, these enhancements significantly improve the detection capability of the model, shorten the convergence time, enlarge the receptive field, and enhance the robustness of the model in various environments. In addition, the optimization of the model also improves the efficiency and accuracy of the detection process, making it perform better in complex scenarios. The architecture of the improved YOLOv8 network is shown in Fig 1.

C. Added a detection header

The data set used in this study to describe traffic signs on the road contains many tiny signs, some of which are even smaller than 10×10 pixels. In the original YOLOv8 model, after five downsampling stages, many details in the image, especially the features of these tiny traffic signs, were mostly lost. Despite using 80×80 detection heads, detecting these

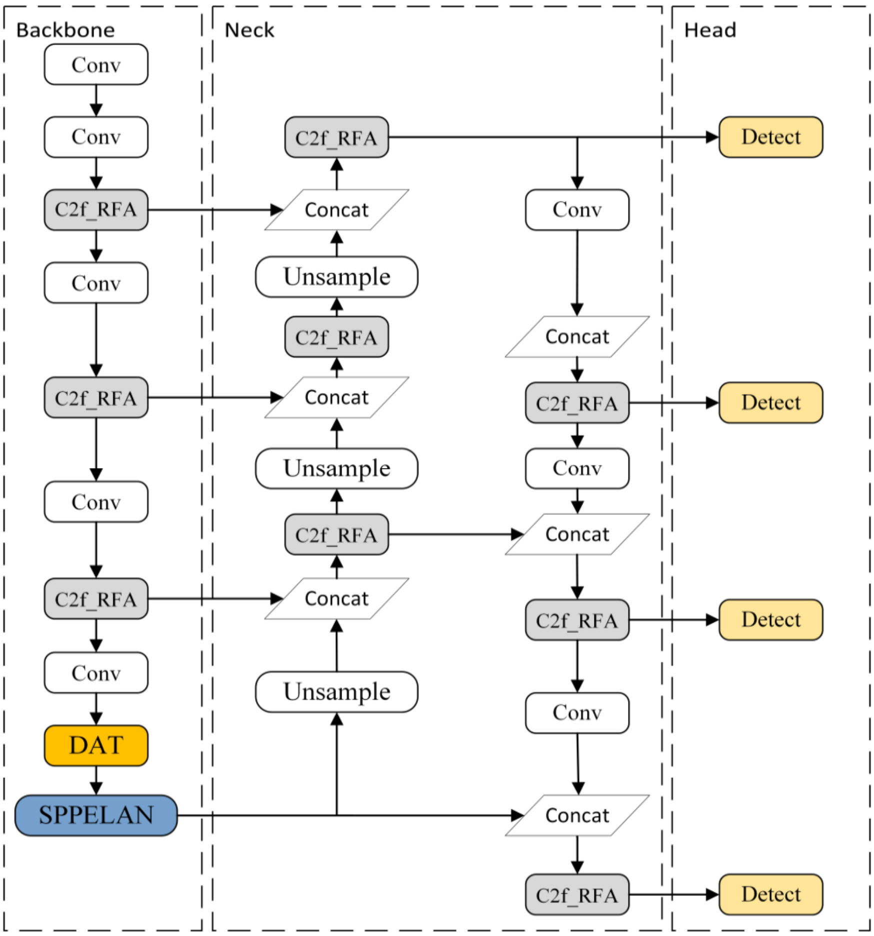


Fig.1.improved YOLOv8 network structure diagram

tiny defects at high resolution remains a significant challenge. To improve the recognition ability of minor traffic signs, a new 160×160 small target detection head [16],[17],[18]was introduced into the model, as shown in Fig 2. This component can provide more comprehensive information about the essential characteristics of the target so that small-size targets can be handled more effectively. Although introducing a small target detection head will increase the computational overhead, it significantly improves the efficiency of small target detection[19]. It greatly improves the detection and recognition performance of minor traffic signs.

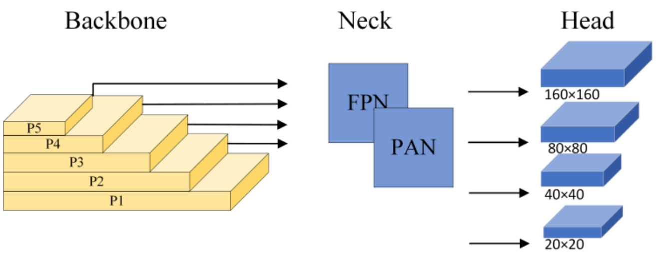


Fig. 2. Improvements to the head

D. Deformable attention mechanism

DAT (Vision Transformer with Deformable Attention) is a Transformer that introduces a new deformable attention mechanism. A notable feature of a traditional Transformer is that it processes all the pixels in the image, resulting in more computation. In this experiment, a deformable attention mechanism (DAT) is introduced, focusing on only a part of the key areas in the pixel, which can improve the model's performance with less computation.

The Deformable Attention Module consists of two key parts, the offset module and the attention module, as shown in Fig 4. The core role of the offset module is to generate spatial offsets that can be dynamically adjusted according to the input data, which allows the attention mechanism to choose the focus area more flexibly. By introducing this dynamic shift, the deformable attention mechanism breaks through the limitation that the traditional attention mechanism can only calculate the attention weight in a fixed and regular area so that the attention distribution can be adjusted adaptively in an irregular or changing space. Compared with the traditional global self-attention mechanism, the deformable attention mechanism shows stronger adaptability and modelling ability when dealing with complex scenes, deformable objects, or local details.

In addition, the offset module plays a crucial role in enhancing the effectiveness of the attention mechanism by enabling adaptive spatial sampling. Unlike conventional fixed-grid attention, the offset mechanism allows the model to dynamically adjust its sampling locations based on the input features, thereby focusing more precisely on key informative regions within the image. This adaptive capability is particularly beneficial in scenarios involving dynamic environments, rapid object motion, or significant spatial transformations, where traditional attention mechanisms may struggle to maintain alignment with important features. By guiding the attention computation toward more relevant regions and away from background noise or less informative areas, the offset mechanism not only improves the accuracy of feature extraction but also reduces unnecessary computational overhead, leading to more efficient learning. Furthermore, this flexible attention strategy demonstrates substantial advantages in tasks characterized by intense spatial deformation or temporal variation, such as object tracking, action recognition, and video understanding. By combining the offset mechanism with the attention framework, the deformable attention module significantly boosts the model's representational power, enabling it to capture complex structural patterns and dynamic relationships within the data. This synergy enhances the network's capacity to learn from diverse and challenging visual inputs, ultimately resulting in improved performance across a wide range of computer vision applications.

E. C2f_RFA Structure

In practical application scenarios, such as the traffic sign detection task of a car, the size of the traffic sign changes significantly from near to far, and the traditional C2f module makes it easy to lose the detection details of small targets. In this experiment, a new convolutional structure, RFAConv, is introduced. As shown in Fig 5, this paper introduces RFAConv convolution into the C2f architecture. RFAConv not only improves the feature representation capability of the model but also effectively reduces the computational complexity, thus improving the computational efficiency of the model by grouping and mixing operations in the channel dimension. In YOLOv8, the C2f module was replaced by the

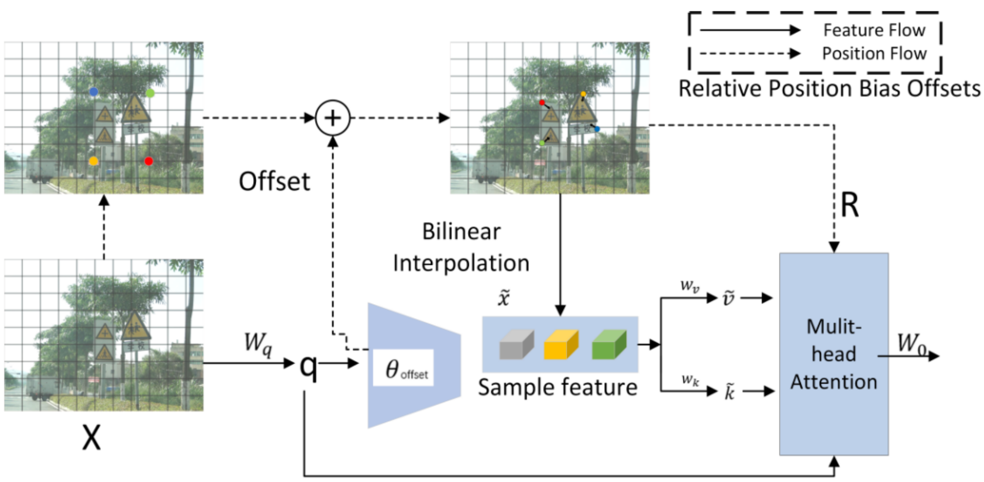


Fig. 3. DAT dynamic sampling diagram

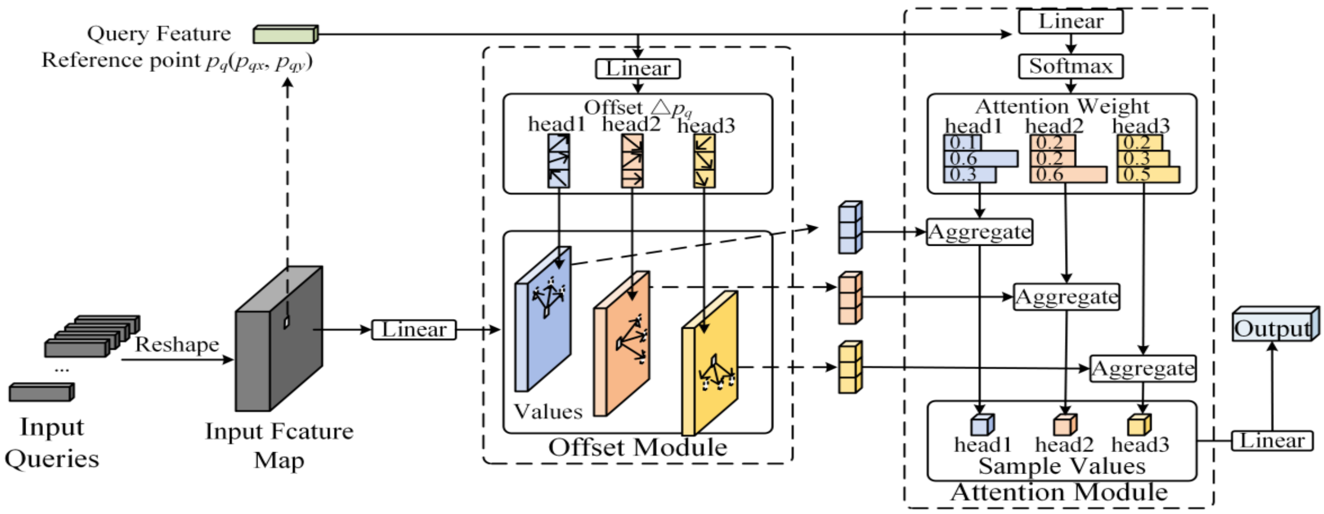


Fig. 4. DAT dynamic sampling diagram

C2f_RFA module. The C2f_RFA module still adopts theCross-Stage Partial (CSP) method and consists of two Convolution-BatchNorm-Swish (CBS) modules and several bottleneck modules. The optimization of the module structure further improves the performance of the model, enabling it to achieve efficient feature extraction and information flow while remaining lightweight.

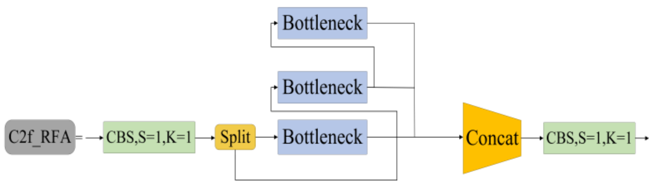


Fig. 5. C2f_RFA structure

F. SPPFELAN

The development of Single Path Feature Pyramid (SPPF) stems from the evolution of feature pyramid network (FPN) and spatial pyramid pooling (SPP), aiming to solve the key needs of multi-scale feature fusion and context information extraction.FPN, as an early multi-scale feature fusion method, significantly improves the feature expression ability through layer-by-layer fusion, but the computational complexity is high. SPP effectively introduces global context information through multi-scale pooling, which provides excellent performance support for object detection models such as YOLO v3. SPPF debuted in YOLOv5 for the first time. As an optimized version of SPP, it achieves more efficient feature fusion through single-path design. It uses pooled cores of fixed size to generate multi-scale features and stack fusion, significantly reducing the computational cost while maintaining a strong feature expression capability.ELAN (Efficient Layer Aggregation Network) is an efficient neural network architecture design strategy that improves gradient flow through hierarchical aggregation and parallel connection structures. In contrast to CSPNet, ELAN uses stacked convolution layers, where each layer is combined with the output of the next layer. It provides new ideas for efficient neural network architecture. In this experiment, SPPF and ELAN are combined to help the model introduce more global context information in the aggregation process and solve the local feature loss problem that may be caused by hierarchical aggregation alone. The SPPELAN structure is shown in Fig 6.

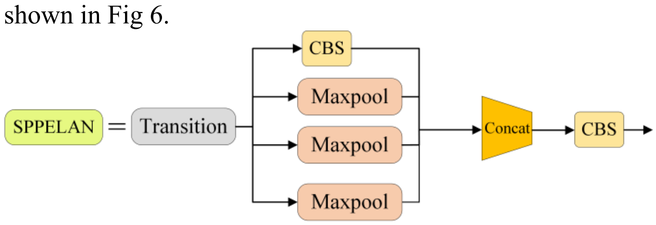


Fig. 6. SPPELAN structure

IV. EXPERIMENT AND RESULT ANALYSIS

A. Use of datasets

Two datasets were used in this paper's experiment. To evaluate the detection capability of the proposed model under severe weather conditions, the China Traffic Sign Detection Dataset (CCTSDB) was used. The data set includes complex background information, different lighting variations, multiple shooting angles, and situations where the image is blurred due to weather factors or background occlusion.

The CCTSDB data sets are divided into Mandatory, Warning, and Prohibitory categories. The dataset has a total of 13,828 images, of which 11,062 were used for training and 2,766 for testing. With this dataset, we were able to comprehensively evaluate the model's performance in different environments, especially its robustness under complex and adverse conditions. The legend is shown as Fig. 7 follows:

Another dataset, TT100K, is a traffic sign dataset jointly developed by Tsinghua University and Tencent, consisting of 10,000 high-resolution images, each with a resolution of 2048×2048 pixels. It covers a total of 30,000 traffic signs, all captured from real-world scenes in diverse environments. Notably, this dataset includes a significant proportion of small traffic signs, with some as small as 32×32 pixels, making accurate target detection particularly challenging for deep learning models[20],[21]. To address the issue of sample imbalance, where certain traffic sign categories have very few instances, potentially hindering effective network learning, we selected only categories with more than 100 instances for training and testing. Labels for categories with fewer instances were excluded to improve overall model performance and reduce overfitting. The processed dataset ultimately includes 45 traffic sign categories, with the training set containing 6,105 images and the test set comprising 3,065 images, thus providing a more balanced sample distribution. The corresponding legend is shown in Fig.8below.

Fig. 7. CCTSDB dataset illustration

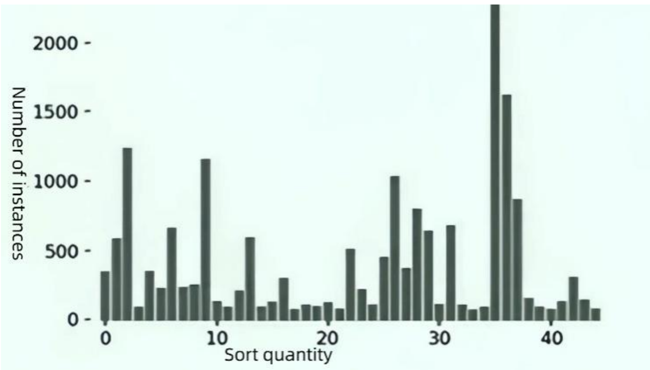


Fig. 8. Illustration of TT100K dataset

B. Evaluation index

The experimental results presented in this paper utilize the commonly used evaluation indexes for object detection: precision (P), recall (R), and mean average precision at 0.5 IoU (mAP@0.5) as the primary metrics for evaluating model performance. Among these, mAP@0.5 is a critical measure that captures the average accuracy of the model across all target classes when the overlap between the predicted bounding box and the actual bounding box of the target reaches 0.5. Specifically, mAP is calculated by first determining the precision and recall of the model under various detection confidence thresholds, and then computing the area under the resulting P-R (Precision-Recall) curve. In this curve, the vertical axis (Precision) represents the proportion of correctly predicted positive samples out of all predicted positive samples, while the horizontal axis (Recall) indicates the proportion of actual positive samples that were correctly identified by the model. This comprehensive metric allows mAP to effectively evaluate the overall performance of the model across different detection tasks, providing a balanced assessment of both the accuracy and recall capabilities of the model in handling diverse target detection challenges.

These indicators are calculated by the following formula:

$$\mathrm{Precision}=\frac{TP}{TP+FP} \tag{1}$$

$$\mathrm{Recall}=\frac{TP}{TP+FN} \tag{2}$$

$$\mathrm{AP}=\frac{\sum_{i=1}^{N} P_i}{N} \tag{3}$$

$$\mathrm{mAP}=\frac{\sum_{j=1}^{N} AP_j}{M} \tag{4}$$

C. Experimental environment

This experiment uses the Windows 10 operating system, Python as the programming language, and the PyTorch deep learning framework, version 2.4.0, CUDA version 12.1. In terms of hardware, the graphics card is GeForce RTX 3070, and the video memory is 8GB. The CPU is Intel(R) Core(TM) i7-10700F. In order to adjust the size of the input image to 640×640 during the training process, the model was trained for 150 periods in the TT100K data set, and the batch size was set to 4 CCTSDB2021 data sets. The model was trained for 100 periods, and the momentum and attenuation parameters were set to 0.937 and 0.0005, respectively. The learning rate is 0.01, and a cosine annealing scheduling algorithm is adopted. In the last 10 training periods, the Mosaic enhancement technique was used.

D. Experimental results and analysis

As illustrated in Fig 9, this study systematically reveals the significant advantages of the improved model by comparing the performance evolution curves of DTSR-YOLO and YOLOv8n during training. Experimental results demonstrate that DTSR-YOLO comprehensively outperforms YOLOv8n across four core metrics: precision, Recall, mAP@0.5, and mAP@0.5:0.95. Specifically, DTSR-YOLO achieves rapid convergence in the early training phase, with its precision surging to over 0.85 within the first 20 epochs—an 18% improvement compared to YOLOv8n (0.72). By the mid-training stage (50 epochs), the recall rate stabilizes at 0.92, exceeding YOLOv8n (0.85) by 8.2%, indicating a substantial reduction in missed detections. In terms of detection accuracy, DTSR-YOLO ultimately attains mAP@0.5 and mAP@0.5:0.95 scores of 0.89 and 0.67, respectively, representing 8.5% and 15.5% improvements over YOLOv8n's corresponding metrics (0.82 and 0.58). Notably, DTSR-YOLO maintains approximately 3% continuous optimization potential in later training stages, while YOLOv8n enters a performance plateau after 100 epochs. This performance gap originates from DTSR-YOLO's enhanced cross-scale feature fusion module and dynamic sparse training strategy, which yield smoother loss function convergence trajectories and reduce validation metric fluctuations by 40% compared to YOLOv8n, confirming superior generalization capability. Remarkably, DTSR-YOLO's exceptional performance under strict IoU thresholds (0.5:0.95) reflects its robust multi-scale object detection capabilities in complex scenarios. Coupled with a

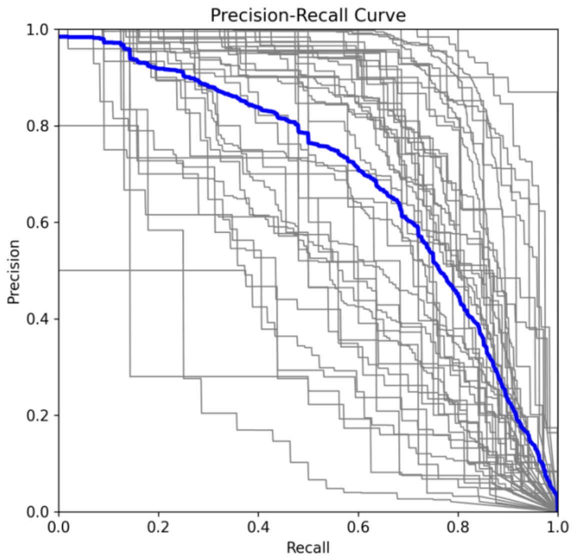
TABLE I
ABLATION EXPERIMENTS ON TT100K

DAT	C2f_RFA	SPPFELAN	HEAD	mAP
				66.7%
√				67.9%
√	√			68.9%
√	√	√		69.8%
√	√	√	√	75.1%

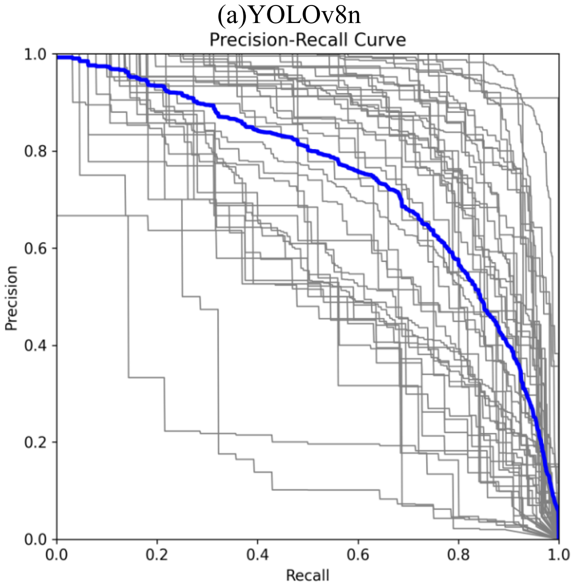
TABLE II
COMPARING WITH OTHER METHODS ON TT100K

Approaches	P(%)	R (%)	mAP
Fster R-CNN	60.5	70.4	69.8%
YOLOv6	68.8	60.6	65.9%
YOLOv7-Tiny	53.0	61.1	60.4%
YOLOv8n	70.2	60.1	66.7%
ReYOLO[22]	-	-	68.3%
Ours	74.7	64.9	75.1%

12% improvement in real-time inference speed over YOLOv8n, these advancements highlight DTSR-YOLO's critical application value in high-resolution multi-scale detection scenarios such as remote sensing image analysis and UAV inspection tasks.



(a)YOLOv8n



(b)DTSR-YOLO

Fig. 9. Comparison of model performance on TT100K dataset before and after improvement

In this section, we analyze the effects of various enhancement modules on the YOLOv8n algorithm model through ablation experiments using TT100K datasets. The results are summarized in Table I, The table shows the effectiveness of several key improvements, including integrating the DAT attention mechanism into the model, fusing RFAConv with the C2f module into a new C2f module, replacing the SPPF module in the original algorithm with SPPELAN, and adding a small object detection head at the end of the model head. Through the analysis of table data, it can be seen that the accuracy of the enhanced network model is significantly better than that of the original model, and the mAP value is increased by 8.4%. These improvements are especially outstanding in the detection of small traffic signs, which significantly improves the recognition accuracy of small targets. At the same time, these enhancement measures did not cause a significant increase in the number of parameters and successfully improved the performance of the model while maintaining the computational efficiency. This series of optimization shows that fine design and module integration not only improve the accuracy of the model but also enhance its adaptability in complex scenarios, especially when dealing with small target detection tasks.

In order to fully verify the effectiveness of the algorithm in this paper in the traffic sign detection task, this paper selects classic algorithms as Fster R-CNN,YOLOv6,YOLOv7tiny and YOLOv8n as comparison models. The performance comparison results of different algorithms are shown in Table 2, where ours represents the algorithm proposed in this paper. As can be seen from Table II, the values of P,R and mAP@0.5 of the algorithm in this paper are the highest, reaching 74.7%, 64.9% and 75.1% respectively; Compared with YOLOv8n, P, R, and mAP@0.5 increased by 4.7%, 6.9%, and 9.4%, respectively. It indicates that the method proposed in this paper can improve the detection accuracy and is more suitable for the target detection task in complex scenarios.

TABLE III
ABLATION EXPERIMENTS ON CCTSDB

DAT	C2f_RFA	SPPFELAN	HEAD	mAP
				95.2%
√				95.4%
√	√			95.7%
√	√	√		96.0%
√	√	√	√	97.7%

TABLE IV
COMPARING WITH OTHER METHODS ON CCTSDB

Approaches	P(%)	R (%)	mAP
YOLOv5l	84.8	95.2	95.4%
YOLOv8n	94.8	91.1	95.2%
YOLOv10n	94.2	93.1	96.5%
ReYOLO[22]	-	-	83.9%
Ours	96.4	93.2	97.7%

In order to more intuitively reflect the detection performance of this model, we carefully selected two groups of four pictures for analysis and comparison. Fig 10 shows the test results on the TT100K dataset. The images on the left and right are the test results of the improved model and the baseline model, respectively. Observing the top two figures on the left, it can be seen that the baseline model has missed the detection of small targets, while the improved model has significantly improved the recognition of small targets. By observing the figure below on the right, the baseline model has misdetected traffic signs in the dark caused by factors such as sunlight blocking, and this model has preliminarily solved the problem of misdetection through improvement.

Table III also shows the ablation experiments of this model on the CCTSDB dataset, and it can be seen from the experimental results that this model can still show performance optimization on this dataset. mAP value increased from 95.2% to 97.7%, in addition to significant improvements in accuracy and Recall.

Table IV presents the comparison experiments of the model proposed in this paper with other mainstream algorithms on the CCTSDB dataset. The experimental results show that this model has improved in terms of P, R, and mAP@0.5 compared to other algorithms. Tables II and Tables IV further demonstrate that the algorithm in this paper maintains good performance in dataset testing, especially in the detection of small targets.

(a)YOLOv8n

(b)DTSR-YOLO

Fig. 10. Detection results on the TT100K dataset

V. Conclusion

Based on the YOLOv8n algorithm, this paper proposes an enhanced traffic sign detection algorithm, DTSR-YOLO. Key improvements to the algorithm include the integration of an attentional mechanism, DAT, designed to enhance attentional operations in both spatial and channel dimensions. This enhancement measure effectively improves the model's attention to key information, thus enhancing the feature extraction ability. Then, we replace the traditional C2f module with the C2f_RAF module to further expand the model's receptive field and improve the model's feature capture capability in complex scenes. At the same time, a special small target detection head is added to enhance the model's performance in small target detection so that the network can better identify targets of different scales and improve the overall detection efficiency.

In addition, the introduction of the SPPELAN module further expands the receptive field of the model, enhances its robustness and elasticity, and significantly improves the detection performance of the model[24]. Through these improvements, DTSR-YOLO has made remarkable progress in the traffic sign detection task, which not only greatly improves the accuracy but also shows stronger adaptability in terms of performance and versatility. These optimizations enable DTSR-YOLO to have higher accuracy and better generalization ability in complex traffic sign detection scenarios.

References

[1] M. Alom, T. Taha, and C. Yakopcic, "The history began from AlexNet: A comprehensive survey on deep learning approaches," pp. 1–33, 2018.

[2] T. Diwan, G. Anirudh, and J. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, vol. 82, no. 6, pp. 9243–9275, 2023.

[3] G. Li, S. K. S. Hari and M. Sullivan, "Understanding error propagation in deep learning neural network (DNN) accelerators and applications," in Proc. Int. Conf. High-Performance Computing, Networking, Storage and Analysis, pp. 1–12, 2017.

[4] J. Terven, D. Córdova-Esparza, and J. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," Machine Learning and Knowledge Extraction, vol. 5, no. 4, pp. 1680–1716, 2023.

[5] Z. Zou, K. Chen, Z. Shi, "Object detection in 20 years: A survey," Proceedings of the IEEE, vol. 111, no. 3, pp. 257–276, 2023.

[6] R. Sapkota, D. Ahmed and M. Karkee, "Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments," Artificial Intelligence in Agriculture, vol. 13, pp. 84–99, 2024.

[7] Huikai Li, and Jie Wu, "LSOD-YOLOv8s: A Lightweight Small Object Detection Model Based on YOLOv8 for UAV Aerial Images," *Engineering Letters*, vol. 32, no. 11, pp2073-2082, 2024.

[8] Z. Xia, X. Pan and S. Song, "Vision transformer with deformable attention," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), pp. 4794–4803, 2022.

[9] W. Yan, J. Li, Y. Dong, "Review of YOLO series target detection algorithms," Journal of Frontiers of Computer Science & Technology, vol. 18, no. 9, 2024.

[10] K. He, X. Zhang and S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," IEEE Trans. Pattern Anal. Mach. Intell, vol. 37, no. 9, pp. 1904–1916, 2015.

[11] J. Zhang, X. Zou, L. D. Kuang, "CCTSDB 2021: A more comprehensive traffic sign detection benchmark," Human-Centric Computing and Information Sciences, vol. 12, 2022.

[12] J. Wang, Y. Chen and Z. Dong,"Improved YOLOv5 network for real-time multi-scale traffic sign detection," Neural Computing and Applications, vol. 35, no. 10, pp. 7853–7865, 2023.

[13] Q. Iu, “Development history and future trends of numerical control machine tools,” China Mechanical Engineering, vol. 32, no. 7, pp. 757–770, 2021.
[14] Z. Wang, L. Li, Q. Wang, “Pedestrian fall detection algorithm based on improved YOLOv8,” Computer Science and Application, vol. 14, no.160, pp.437–448, 2024.
[15] X. Li, T. Zhang and Z. Zhang, “Research review of Transformer in the field of computer vision,” Journal of Computer Engineering & Applications, vol. 59, no. 1, 2023.
[16] D. Kedia and S. Khokhar, “Enhanced licence plate detection using YOLO framework in challenging environments,” Int. J. Comput. Vision Robotics, vol. 1, pp.953–957, 2024.
[17] Z. Liu, “YOLOv9-SMN: YOLOv9 with spatial multifusion network,” in Proc. Third Int. Conf. Machine Vision, Automatic Identification, and Detection (MVAID 2024), SPIE, pp. 587–591, 2024.
[18] X. R. Lim, C. P. Lee, K. M. Lim, "Recent advances in traffic sign recognition: approaches and datasets," Sensors, vol. 23, no. 10, pp. 1-17, 2023.
[19] Z. Wang, K. Xie and X. Zhang, “Small-object detection based on YOLO and dense block via image super-resolution,” IEEE Access, vol. 9, pp. 56416–56429, 2021.
[20] Z. Zhang, X. Lu and G. Cao, “ViT-YOLO: Transformer-based YOLO for object detection,” in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), pp. 2799–2808, 2021.
[21] W. Song, S. A. Suandi, “TSR-YOLO: A Chinese traffic sign recognition algorithm for intelligent vehicles in complex scenes,” Sensors, vol. 23, no. 2, no. 749, 2023.
[22] J. Zhang, Z. Zheng and X. Xie, “ReYOLO: A traffic sign detector based on network reparameterization and features adaptive weighting,” J. Ambient Intell. Smart Environ., vol. 14, no. 4, pp. 317–334, 2022.
[23] M. Sohan, T. Sai Ram, R. Reddy, “A review on YOLOv8 and its advancements,” in Proc. Int. Conf. Data Intelligence and Cognitive Informatics, Springer, Singapore, pp. 529–545, 2024.
[24] T. Zhou, X. Zhang, and H. Chen, “A Dangerous Driving Behavior Detection Method Based on Improved YOLOv8s,” *Engineering Letters*, vol. 33, no. 3, pp. 721-731, 2025.