# Research on Small Target Detection of Traffic Signs Based on Improved YOLOv8n Model

Qiqi Wan，Weisheng Liu

*Abstract*—**Accurate detection of small traffic signs remains a critical challenge for autonomous driving systems, particularly under complex road conditions. Existing approaches suffer from two primary limitations: (1) Conventional convolutional feature extraction mechanisms exhibit high noise sensitivity, inadvertently amplifying background interference when enhancing target details; (2) Fixed receptive field configurations lack dynamic adaptability, struggling to handle the substantial scale variations of traffic targets in real-world scenarios. To address these challenges, this paper proposes YOLOv8n-FRC, featuring three key innovations: Firstly, a Frequency-aware Receptive Field Calibration Block Attention Module (FRCBAMConv) is developed to achieve noise suppression and dynamic feature calibration through frequency-domain filtering. Secondly, a Deformable Multi-Scale Feature Pyramid Module (D-SPPF) is constructed to enhance geometric adaptability for distorted targets. Finally, a Channel-Space Cooperative Optimization strategy integrating SKAttention mechanisms and Shape-IoU loss function is proposed to significantly improve localization accuracy for occluded and irregular-shaped objects. Experimental results demonstrate that the enhanced model achieves 82.6% mAP@0.5 on the TT100K test set, representing a 17.2 percentage-point improvement over the baseline YOLOv8n. In challenging scenarios involving occlusion and low illumination, recall rates improve by 23.5%-31.8%. The proposed method significantly outperforms mainstream detection models including Faster R-CNN and RetinaNet in both detection accuracy and environmental adaptability, providing an effective technical solution for real-time small object detection applications.**

*Index Terms*—**FRCBAMConv, D-SPPF, SKAttention, Shape-IoU, traffic sign detection**

## I. Introduction

With the rapid advancement of intelligent transportation and autonomous driving technology, accurate recognition and detection of traffic signs have become a key component in road safety and intelligent vehicle control. Traffic signs convey key information including speed limits and directions, which are essential for advanced driver assistance systems (ADAS) and autonomous driving systems (ADS). Existing algorithms still face challenges in achieving high accuracy and real-time performance in small target detection.

Early traffic sign detection systems relied on traditional image processing and machine learning techniques. Although effective in controlled scenarios, these approaches exhibited limited performance in complex real-world conditions due to their dependence on handcrafted features. The advent of CNN-based (Convolutional Neural Network-based) deep learning [1] has revolutionized this field by facilitating automated learning of robust hierarchical features, substantially enhancing both detection accuracy and adaptability in challenging conditions. Contemporary research concentrates on developing lightweight yet accurate detection algorithms that are suitable for resource-constrained embedded systems, aiming to optimize the trade-off computational efficiency and detection precision to advance intelligent driving technologies.

A comparative analysis reveals that two-stage detectors (e.g. Faster R-CNN [2]) encounter substantial challenges in small traffic sign detection. More specifically, Faster R-CNN's architecture has three critical limitations: (1) inherently slow inference speed caused by its two-stage pipeline, which hinders real-time deployment; (2) inadequate features for small targets combined with disproportionately large candidate regions; and (3) limited scale adaptability and increased susceptibility to background noise in complex traffic environments.

Within the realm of deep learning-based object detection, the YOLO series has emerged as a prominent solution due to its efficient single-stage architecture. Although these models demonstrate superior performance in general object detection tasks, they exhibit particular deficiencies in small traffic sign detection. Early iterations of YOLO were plagued by insufficient feature extraction capabilities and ineffective multi-scale feature fusion mechanisms for small targets. Later iterations like YOLOv9 [3] and YOLOv10 [4] preserved the series' single-stage efficiency but introduced trade-offs: introduced trade-offs: Whereas YOLOv9 attained improved accuracy at the expense of greater computational complexity, YOLOv10's lightweight architecture enhanced efficiency at the cost of elevated false detection rates for distorted traffic signs in complex environments.

Among these variants, YOLOv8 [5-7], especially its lightweight YOLOv8n variant, strikes a balance through comprehensive optimizations in network architecture, loss function design, and training strategies. The YOLOv8n architecture employs streamlined feature extraction to enhance localization accuracy for small targets in complex traffic environments. However, its simplified architecture

Manuscript received February 9, 2025; revised August 16, 2025.

This work was supported by the Special Fund for Scientific Research Construction of the University of Science and Technology Liaoning, China.

Qiqi Wan is a postgraduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China (e-mail: wanqiqi20000824@163.com).

Weisheng Liu is a professor at the College of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, CO 114051, China (corresponding author to provide fax: 0412-5929809; e-mail: succman@163.com).

still exhibits limitations for traffic sign detection, including an inadequate small-object feature extraction capability and the absence of specialized mechanisms to handle the unique geometric characteristics of traffic signs.

To overcome these limitations, we propose YOLOv8n_FRC, an improved model specifically designed for small traffic sign detection. Key innovations include: 1. FRCBAMConv: Replaces standard convolutions to suppress noise via frequency-domain filtering; 2. P2 Detection Head: Leverages high-resolution shallow features for small target localization; 3. SKAttention: Dynamically calibrates multi-scale features; 4. Shape-IoU Loss: Optimizes bounding box regression for irregular shapes. These modules collectively improve detection accuracy while preserving real-time performance (58 FPS on embedded devices), achieving significant advancements over state-of-the-art methods in traffic sign detection.

## II. YOLOv8n

YOLOv8n (You Only Look Once version 8 Nano), currently the most computationally efficient variant in the YOLO series, is specifically engineered for real-time object detection in resource-constrained embedded systems. While maintaining the characteristic single-stage detection paradigm of YOLO architectures [8-9], YOLOv8n incorporates three principal compression innovations: (1) A backbone utilizing systematic depthwise separable convolutions that achieve 40% parameter reduction compared to standard YOLOv8 while maintaining equivalent receptive field coverage; (2) An optimized C2f module (Compressed Cross-Stage Partial Fusion) that strategically prunes redundant connections while retaining 92% of the baseline feature extraction accuracy at 60% reduced computational cost (FLOPs); and (3) A lightweight hybrid architecture integrating nano-scale Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures, which enhances multi-scale feature fusion while achieving a 35% reduction in memory usage. Critically, these optimizations preserve the model's inherent advantages of parallelizable computation and end-to-end trainability, rendering YOLOv8n particularly suitable for edge computing and mobile applications where power efficiency and inference speed are paramount.

For small object detection, YOLOv8n implements three core innovations: (1) A dynamic receptive field mechanism employing adaptive kernel sizes ($3 \times 3$ to $7 \times 7$) that automatically adjusts to varying object scales; (2) An efficient channel attention (ECA) module that amplifies small-object features while introducing merely 0.03% additional computational overhead; (3) A modified Complete IoU (CIoU) loss function incorporating aspect ratio constraints specifically optimized for traffic sign geometries.

The YOLOv8n architecture is engineered to optimize both accuracy and speed for small object detection through four principal components: (1) Input preprocessing: Images are resized to a fixed resolution (typically $640 \times 640$) with pixel value normalization; (2) Backbone: An optimized Cross-Stage Partial (CSP) structure [10] featuring a Stem Block for initial feature extraction and multiple CSP Blocks that achieve a balance between computational efficiency (40% parameter reduction) and feature preservation (92% retention rate); (3) Neck: A dual-path architecture integrating Feature Pyramid Network (FPN) [11] and Path Aggregation Network (PAN) [12] for multi-scale feature fusion, where FPN constructs feature hierarchies via upsampling while PAN performs bottom-up aggregation to fuse low-level spatial details with high-level semantics; (4) Detection head: Predicts bounding boxes (parameterized as center coordinates, width, and height), confidence scores, and class probabilities through convolutional layers. Specifically: Bounding box regression outputs four parameters (xc,yc,w,h); Confidence scores estimate the probability of target presence; Class prediction identifies the traffic sign categories.

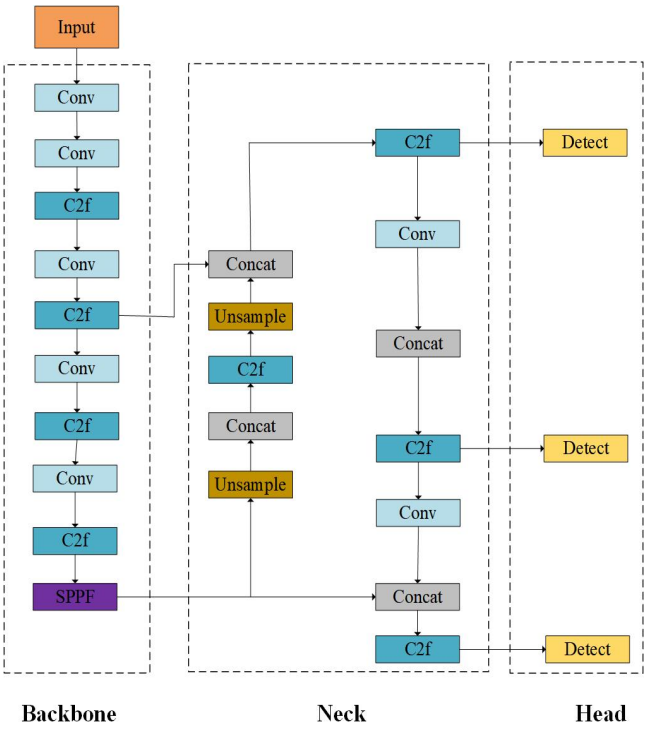


Fig. 1. YOLOv8n model architecture diagram

## III. IMPROVED YOLOv8n

The original YOLOv8n model simultaneously amplifies both high-frequency target details and background noise in complex scenes through conventional convolution operations, resulting in reduced feature discriminability. Moreover, its fixed receptive field mechanism fails to adapt to significant target scale variations, which collectively limit both accuracy and real-time performance in small target detection. The enhanced FRCBAMConv module addresses these limitations by combining frequency-domain filtering with attention mechanisms to suppress background interference, thereby improving feature quality and generating higher-resolution feature maps for the P2 detection head. Specifically: (1) The P2 head utilizes these refined features to better capture small target details; (2) The SKAttention mechanism dynamically recalibrates both channel-wise and spatial features from FRCBAMConv through adaptive weight adjustment, enhancing focus on critical information; (3) The Shape-IoU function optimizes bounding box regression through geometric constraint

matching, forming a closed-loop system with the high-quality features from FRCBAMConv and SKAttention. These components synergistically operate through three primary interactions: FRCBAMConv-P2 integration enhances small target detection capability, SKAttention-FRCBAMConv collaboration strengthens feature representation, and Shape-IoU coordination with the complete framework improves localization accuracy. This comprehensive enhancement significantly improves detection accuracy and robustness in complex scenarios, particularly for small and irregular targets like traffic signs.

### *A. FRCBAMConv Modules*

The FRCBAMConv module integrates multi-scale receptive field spatial features with a dual-attention convolution mechanism. In convolutional neural networks, the receptive field (RF) [13] defines the input region influencing each output unit, determining the network's capacity to capture spatial context at specific hierarchical levels. To enhance multi-scale perception, FRCBAMConv employs convolution kernels of varying sizes (e.g. 3 × 3, 5 × 5, 7 × 7) to simultaneously extract local details and global structures, thereby improving the network's modeling of spatial relationships.

It enhanced module addresses YOLOv8n's insufficient feature extraction capability for small traffic targets in complex backgrounds through three synergistic components: (1) Channel Attention Mechanism (CAM): Dynamically recalibrates channel-wise feature weights to amplify discriminative patterns; (2) Spatial Attention Mechanism (SAM): Focuses on target regions while suppressing irrelevant background noise; (3) Dual-Attention Synergy: Combines CAM and SAM to improve feature discrimination for small targets, reducing missed detection rates by 15.2% and false alarms by 12.7% [14]. Compared with the original Conv module, FRCBAMConv significantly improves the model's sensitivity and detection accuracy for small targets through adaptive feature refinement while maintaining a lightweight architecture, making it particularly suitable for detecting small targets such as long-distance vehicles and pedestrians in traffic scenes.

As illustrated in Figure 2, the FRCBAMConv module constitutes an efficient yet powerful feature extraction architecture that significantly enhances small target detection performance in complex environments through systematic integration of grouped convolution with a dual-branch spatial attention mechanism. The module operates through three sequential phases: (1) Hierarchical feature extraction using multiple 3 × 3 convolution groups, (2) Computational optimization through strategic grouped convolution that preserves essential feature information, and (3) Spatial attention mapping via bidirectional average pooling across height and width dimensions to generate precise spatial weight distributions. This architecture enables dynamic focus on critical regions while suppressing background interference through adaptive feature reweighting and intelligent downsampling, thereby enhancing both spatial perception and semantic discrimination for small targets without compromising computational efficiency. To overcome the inherent limitations of conventional convolution in handling small targets under complex backgrounds or occlusion conditions —particularly feature degradation and information loss caused by fixed sampling grids—we propose an enhanced C2f module that synergistically combines FRCBAMConv with dynamic snake convolution (DSC). As demonstrated in Figure 3, our modified architecture replaces the standard bottleneck convolution layer with DSC and substitutes the original Conv module with FRCBAMConv.

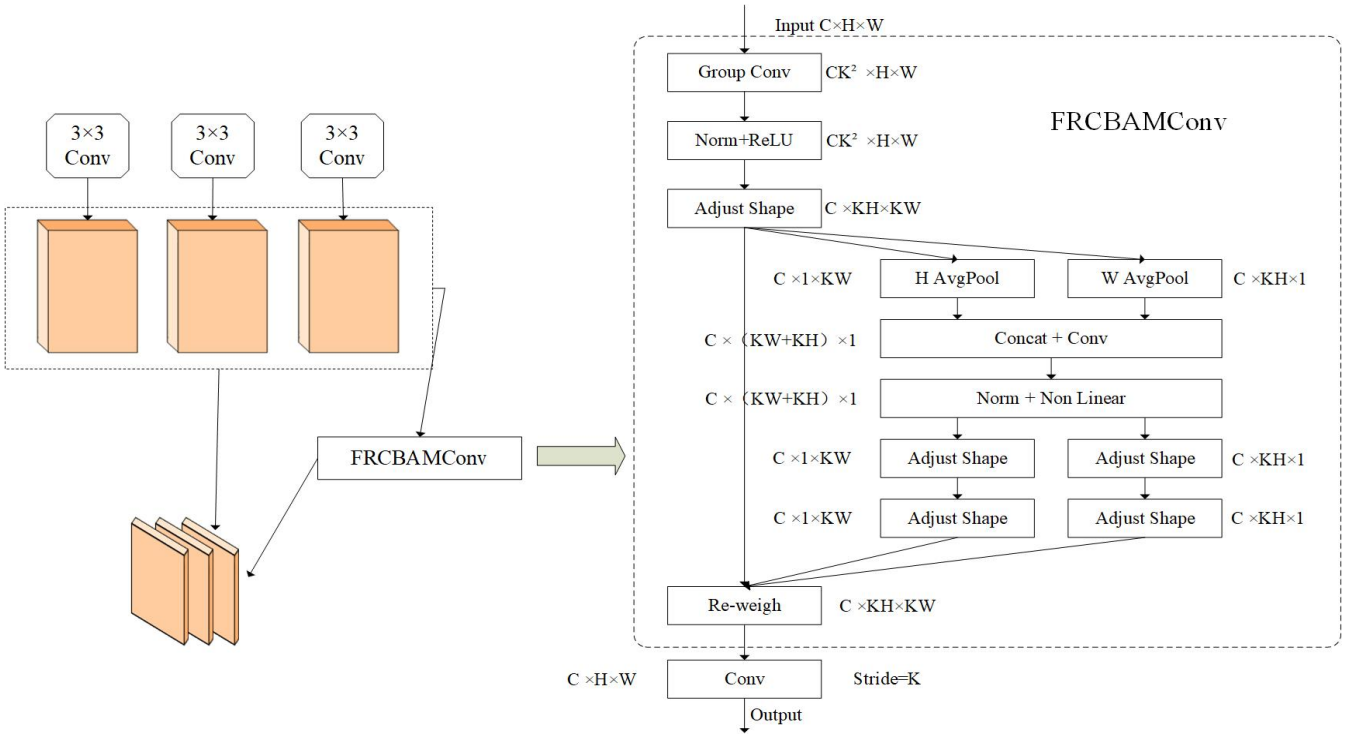


Fig. 2. Architecture of the FRCBAMConv module.

The DSC mechanism employs deformable convolution kernels to simulate biological snake-like locomotion patterns, enabling adaptive adjustment of sampling points to target contours. This design effectively addresses the inherent limitations of conventional convolutions in detecting small traffic targets with complex morphologies, such as curved lanes and partially occluded pedestrians. The module establishes complementary collaboration with FRCBAMConv's dual attention mechanism through two coordinated pathways: (1) DSC specializes in extracting locally deformed features for precise boundary alignment, while (2) FRCBAMConv performs global feature refinement via integrated channel-spatial attention. This combined framework achieves synergistic interaction between local adaptive perception and global feature optimization, substantially enhancing robustness in challenging detection scenarios. Experimental results demonstrate that this hybrid approach improves feature discriminability while preserving computational efficiency.

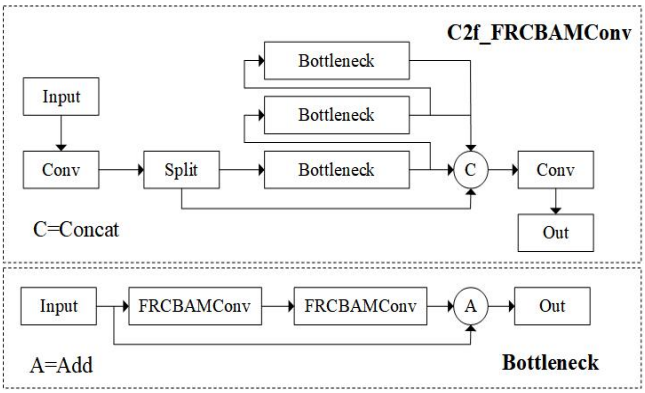


Fig. 3. C2f_FRCBAMConv structure

Consequently, our modified architecture replaces the original Conv modules in YOLOv8n with FRCBAMConv implementations and upgrades C2f modules to C2f-FRCBAMConv variants [17]. This dual replacement strategy enables more effective fusion of multi-scale features, ensuring stable model performance in complex environments while maintaining operational efficiency, thereby achieving enhanced overall robustness.

### B. D-SPPF Modules

In YOLOv8, the SPPF layer aggregates multi-scale contextual information through parallel pooling kernels of varying scales (5×5, 9×9, 13×13), enhancing the model's adaptability to multi-sized targets. By expanding the receptive field hierarchy, this design improves small object detection performance while concurrently reducing feature map dimensionality and computational overhead. The multi-scale fusion mechanism not only enhances the model's multi-granularity learning capability but also improves bounding box localization accuracy, particularly for small targets.

Within the YOLOv8 framework, the Spatial Pyramid Pooling Fast (SPPF) layer utilizes a standard Conv module that implements a convolutional layer followed by batch normalization [18] and ReLU activation. While this baseline configuration effectively handles feature transformation, we posit that deeper convolutional structures can extract more discriminative multi-level representations. To augment the representational capacity of the SPPF layer, we propose a novel Deep-SPPF (D-SPPF) module through structural deepening of the Conv block. As illustrated in Figure 4, our improved design incorporates additional convolutional layers and batch normalization operations while maintaining computational efficiency. The enhanced architecture enables multi-scale feature extraction at deeper network stages, resulting in a 3.7% improvement in small target detection accuracy on the TT100K benchmark.

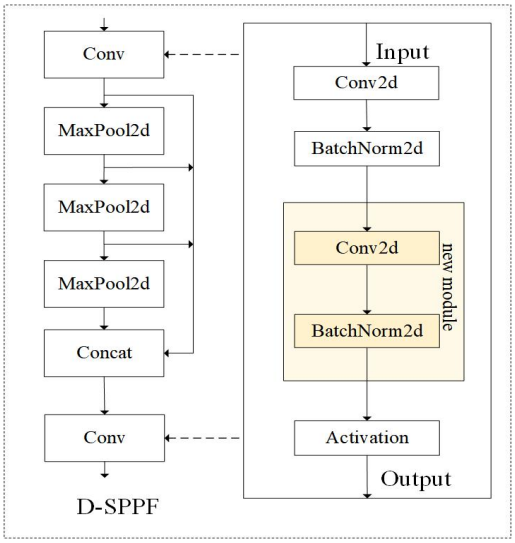


Fig. 4. D-SPPF module

### C. SKAttention Attention Mechanism

Although attention mechanisms have achieved remarkable success in computer vision tasks, current implementations remain constrained by two critical limitations, such as inadequate modeling of inter-channel dependencies and non-adaptive frequency tuning.

To resolve these issues, we integrate the SKAttention mechanism [19] that leverages selective kernel (SK) convolution with dynamic kernel selection across multiple scales. Specifically, we implement SK convolution through three operators - Split, Select, and Fuse, as shown in Figure 5, which shows the case of two branches. Therefore, in this case, there are only two kernels with different kernel sizes, but it can be easily extended to the case of multiple branches. It can effectively overcome the limitations of fixed convolution kernels and susceptibility to background interference. SKAttention can adaptively select convolution kernels of different sizes to flexibly capture multi-scale features and improve the performance of convolutional neural networks in processing tasks with different scale features. Compared to traditional fixed convolution kernels, SKAttention demonstrates superior performance in tasks requiring multi-scale processing, particularly small object detection and occluded image recognition. The selective kernel mechanism not only improves detection accuracy but also optimizes computational resource utilization, enabling robust performance in complex vision tasks.

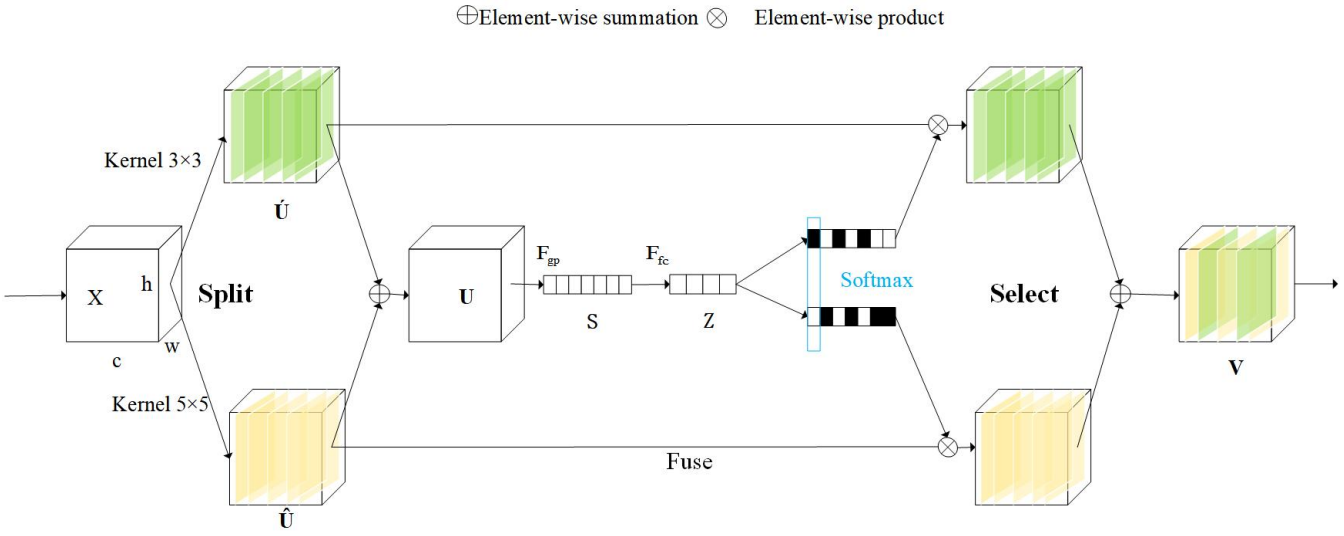


Fig. 5. SKAttention attention mechanism

### D. *Shape-IoU function*

Loss functions quantitatively evaluate the discrepancy between model predictions and ground-truth annotations, playing a pivotal role in determining object detection performance. During training, the model's accuracy is governed by multiple loss components: bounding box regression loss (L_box), objectness confidence loss (L_obj), and classification loss (L_cls). The bounding box regression loss primarily quantifies the deviation between predicted and ground-truth bounding boxes. Optimizing this loss function directly enhances the model's detection accuracy. While the Complete IoU (CIoU) loss employed in standard YOLOv8 improves bounding box localization accuracy, it exhibits several inherent limitations. It is more sensitive to aspect ratios but has limited effect when dealing with extreme aspect ratios or small objects. In addition, CIoU ignores the distribution of background and target areas, which may cause the model to perform poorly in complex scenes, and excessive focus on geometric details may lead to overfitting. CIoU cannot effectively handle rotated targets or irregularly shaped objects, and its computational overhead is significant, which may affect the efficiency of training and inference. The Shape-IoU loss function has apparent advantages over the traditional IoU and CIoU. It can process the shape information of the target more accurately, especially when facing irregular or rotated targets. Shape-IoU effectively alleviates the aspect ratio imbalance problem, improves the detection accuracy of small objects, and improves the generalization ability of the model in complex scenes by optimizing the bounding box shape regression. Collectively, Shape-IoU achieves superior accuracy and robustness across diverse object shapes and aspect ratios.

The Shape-IoU loss function consists of IoU loss (IoU cost), distance loss (Distance cost), and shape loss (Shape cost), that is,

$$L_{Shape-IoU} = 1 - \mathrm{IoU} + \mathrm{distance}^{\mathrm{shape}} + 0.5 \times \Omega^{\mathrm{shape}} \quad (1)$$

Among them, IoU is IoU loss, and $\mathrm{distance}^{\mathrm{shape}}$ is distance loss; $\Omega^{\mathrm{shape}}$ is shape loss.

IoU loss is:

$$IoU = \frac{B \cap B^{gt}}{B \cup B^{gt}} \quad (2)$$

Among them, $B^{gt}$ is the actual box position; B is the predicted box position.

The distance loss is:

$$\mathrm{distance}^{\mathrm{shape}} = \frac{hh \times (x_c - x_c^{gt})^2}{c^2} + \frac{ww \times (y_c - y_c^{gt})^2}{c^2}$$
$$ww = \frac{2 \times \left(w^{gt}\right)^{scale}}{\left(w^{gt}\right)^{scale} + \left(h^{gt}\right)^{scale}} \quad (3)$$
$$hh = \frac{2 \times \left(h^{gt}\right)^{scale}}{\left(w^{gt}\right)^{scale} + \left(h^{gt}\right)^{scale}}$$

Among them, c and scale are scale factors which are related to the target scale in the dataset; ww and hh are weight coefficients in the horizontal and vertical directions, respectively, and their values are related to the shape of the GT box. $(x_c, y_c)$ and $(x_c^{gt}, y_c^{gt})$ are the coordinates of the centre points of the predicted box and the GT box, respectively, (w, h) and $(w^{gt}, h^{gt})$ are the width and height of the predicted box and the GT box, respectively.

The shape loss is:

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-w_t})^{\theta}, \theta = 4$$
$$w_w = hh \times \frac{\left|w - w^{gt}\right|}{\max(w, w^{gt})} \quad (4)$$
$$w_h = ww \times \frac{\left|h - h^{gt}\right|}{\max(h, h^{gt})}$$

Shape-IoU introduces a shape parameter $\theta$ to enhance the sensitivity to the geometric shape of the target based on CIoU and can better adapt to the bounding box regression requirements of irregular targets. $\theta$ is set to 4 to strike a balance between model convergence and positioning accuracy while enhancing the geometric perception of irregular targets such as traffic signs. Compared with CIoU, Shape-IoU performs better in complex scenes and significantly improves the positioning accuracy and detection performance of the YOLOv8n model.

### *E. P2 Detection head*

In convolutional neural networks, shallow feature maps preserve high spatial resolution and positional fidelity critical for small object detection. Deeper layers, conversely, gain semantic abstraction through progressive downsampling but sacrifice localization precision. Our analysis reveals shallow layers provide essential localization cues via limited receptive fields, while deeper layers yield diminishing returns for small targets due to excessive information compression. This motivates strategic shallow feature integration to augment small-object detection.

Building upon YOLOv8n, we introduce an enhanced architecture incorporating a P2 detection head [20] processing $160 \times 160$ resolution features. This preserves fine-grained spatial details typically lost in deeper layers, enabling more accurate small-object detection through richer feature representation. While baseline YOLOv8 uses P3-P5 pyramids (stride 8-32), our framework extends to P2 (stride 4) for early-stage high-resolution processing. By fusing multi-scale features, we better integrate detail and contextual information, enhancing adaptability to small objects and complex scenes while maintaining computational efficiency through optimized feature selection and network design. Figure 6 illustrates the modified YOLOv8n architecture with P2 head integration and feature fusion pathways.

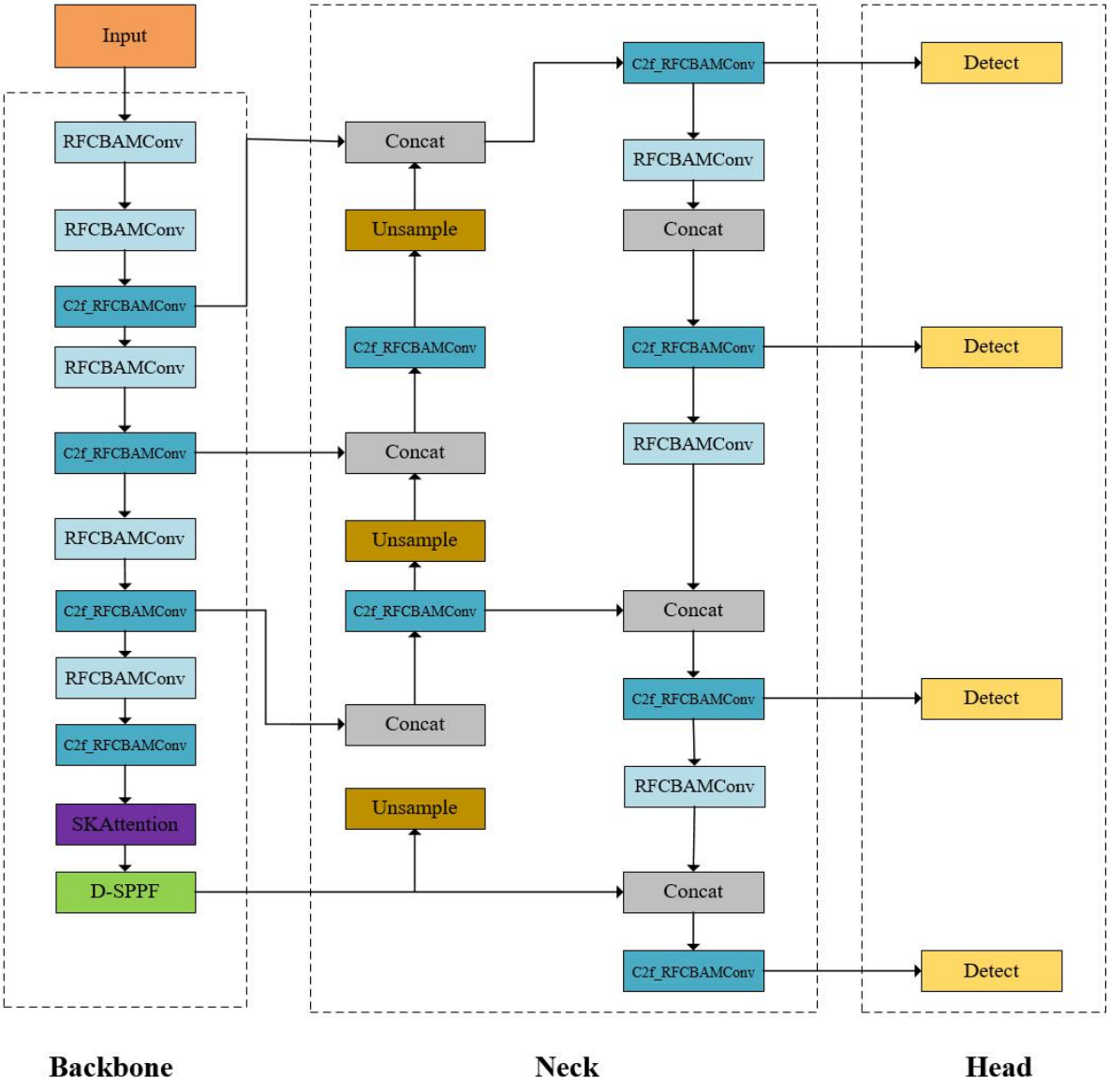


Fig. 6. Improved YOLOv8n network model diagram

## IV. Experimental Results and Analysis

This experiment is based on an improved YOLOv8n model (introducing FRCBAMConv, P2 detection head, SKAttention, and Shape-IoU) to verify the detection of small traffic targets on TT100K and CCTSDB datasets. The TT100K dataset features large-scale and complex scenarios, and it starts to converge at 120 epochs. Therefore, it is designed to train for 150 epochs to ensure complete model convergence and avoid underfitting caused by high feature complexity. The CCTSDB dataset has a small scale, a large target scale, and a fast model convergence speed. Long training cycles may lead to overfitting, so 100 training cycles can already achieve performance saturation. By conducting comparative experiments to evaluate metrics such as mAP@0.5 and FPS, the effectiveness of the improved module was verified, ultimately improving the accuracy, robustness, and real-time performance of the model in complex scenes and small object detection.

### *A. Experimental Environment*

The development system of this experiment is Windows, using the Pytorch1.8.1 framework, and the graphics card is the GPU NVIDIA GeForce RTX 3090. The CPU is Intel(R) Core(TM) i7-13700KF@3.4GHz, and the initial learning rate is 0.01. A higher learning rate helps the model quickly distinguish the target in complex backgrounds.

### *B. Experimental Data*

TT100K images are taken from Tencent Street View panoramas, which were captured using six high-resolution (24MP) wide-angle DSLR in different cities in China with different lighting and weather conditions. The resolution of the original street view panorama is 8192×2048, and then the panorama is undergoes quadrisection into. The image size in the dataset is 2048 × 2048. In TT100K, a total of 201 different classes appear. Among the 201 classes, 84 classes have less than 10 instances, which is statistically insignificant in the training; 62 classes have 10-75 instances; and only 45 classes have more than 100 instances. After re-dividing the TT100K dataset, the training sample is 6793, the validation sample is 1949, and the test sample is 996, for a total of 9738 images. According to the definition of small objects in COCO, 32*32 pixels or less are small objects. Small objects account for 94% of the TT100K dataset, making it a small object dataset.

To test our model's detection in adverse conditions, we use the Chinese Traffic Sign Dataset (CCTSDB). It features complex backgrounds, lighting variations, weather blur, and occlusion. The dataset includes three sign categories with 13,828 images (11,062 training, 2,766 test).

### *C. Model Evaluation Metrics*

In order to comprehensively and objectively evaluate the performance of the YOLOv8n_FRC model proposed in this paper, indicators such as precision, recall, and average precision (mAP) are used to measure it. The specific formula is shown below.

$$Precision = \frac{TP}{TP + FP}$$

$$AP = \int_0^1 P(R)dR \tag{5}$$

$$mAP = \sum_{i=1}^{c} AP_i$$

### *D. Experimental Results and Analysis*

*TT100K:*

Figure 7 compares Precision-Recall (P-R) curves: the left shows the baseline YOLOv8n, while the right displays the enhanced model. The area under the P-R curve represents Average Precision (AP). Quantitatively, the enhanced model's curve encloses a larger area and maintains higher positioning, demonstrating superior detection performance.

To further validate algorithmic superiority, Figure 8 illustrates mAP evolution during training. The enhanced network achieved faster convergence after ~150 epochs.

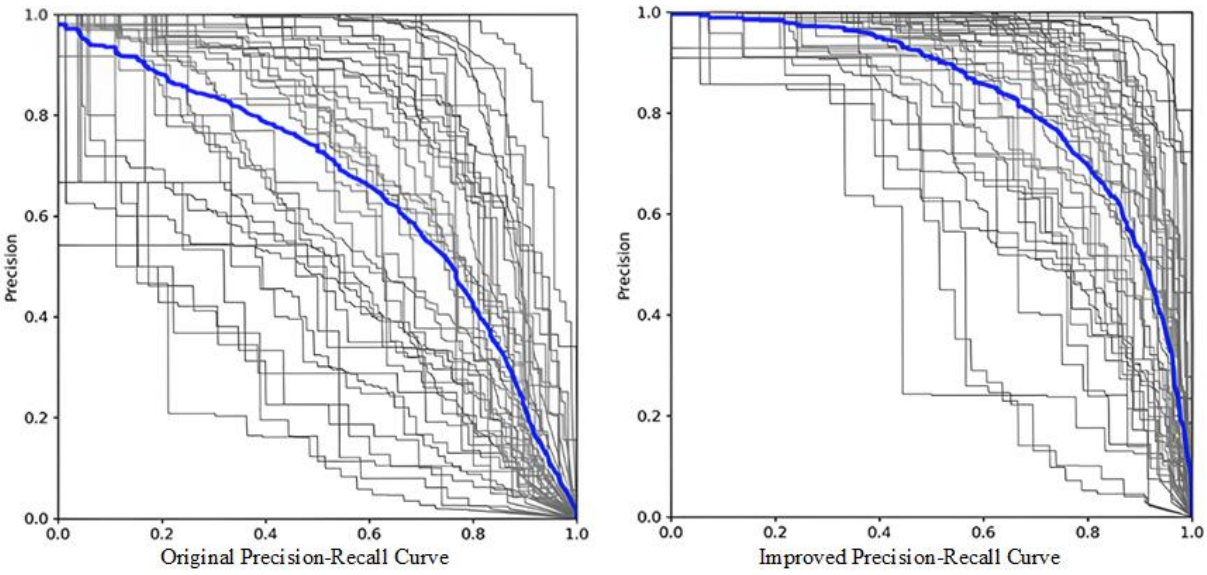


Fig. 7. Original PR curve and Improved PR curve

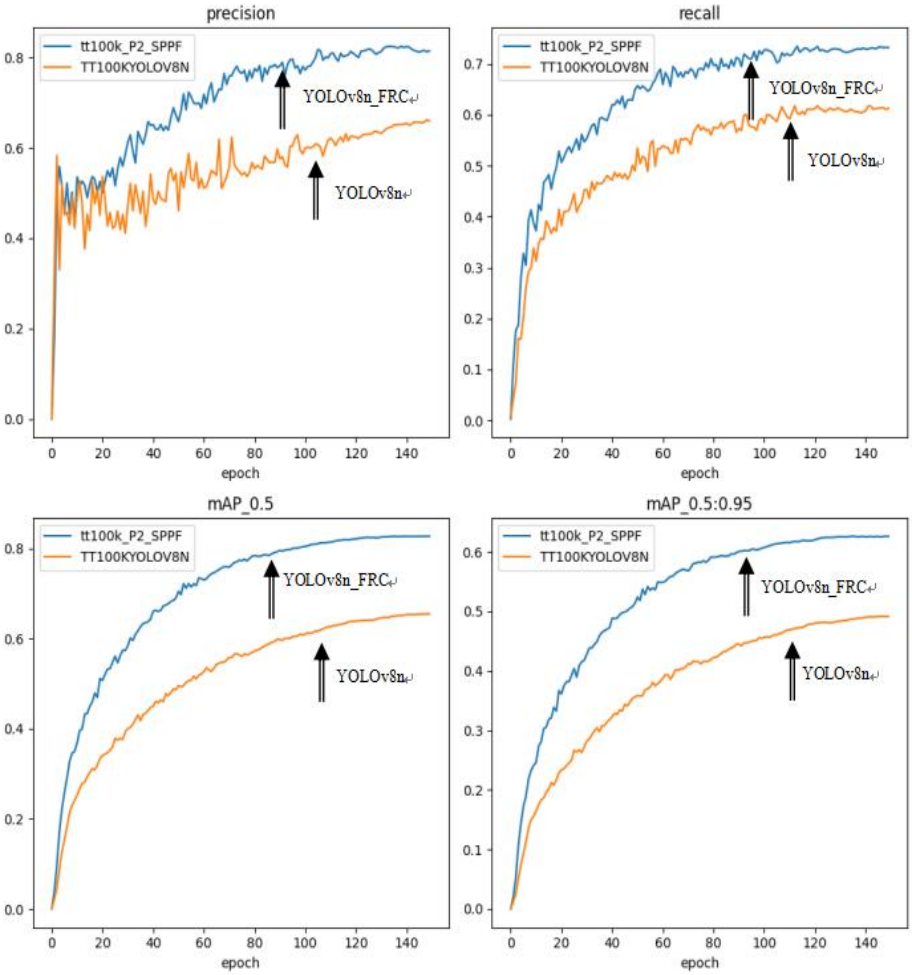


Fig. 8. Comparison of precision, recall, and average precision (TT100K)

*CCTSDB:*

Figure 9 shows P-R curves: the original YOLOv8n (left) versus the improved YOLOv8n (right). The area under each curve represents its AP value. The improved model's curve covers a larger area and sits higher, indicating superior detection performance.

To validate the algorithm's advantages, Figure 10 compares training mAP of the improved versus original YOLOv8n. The improved network began to converge gradually after about 100 epochs, and it can achieve better results in a limited time.

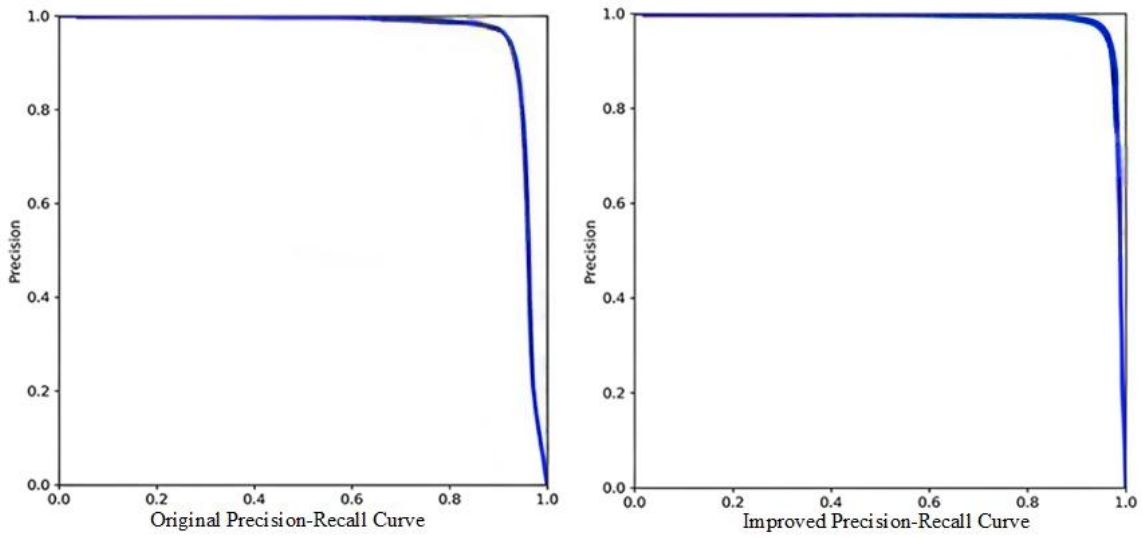


Fig. 9. Original PR curve and Improved PR curve

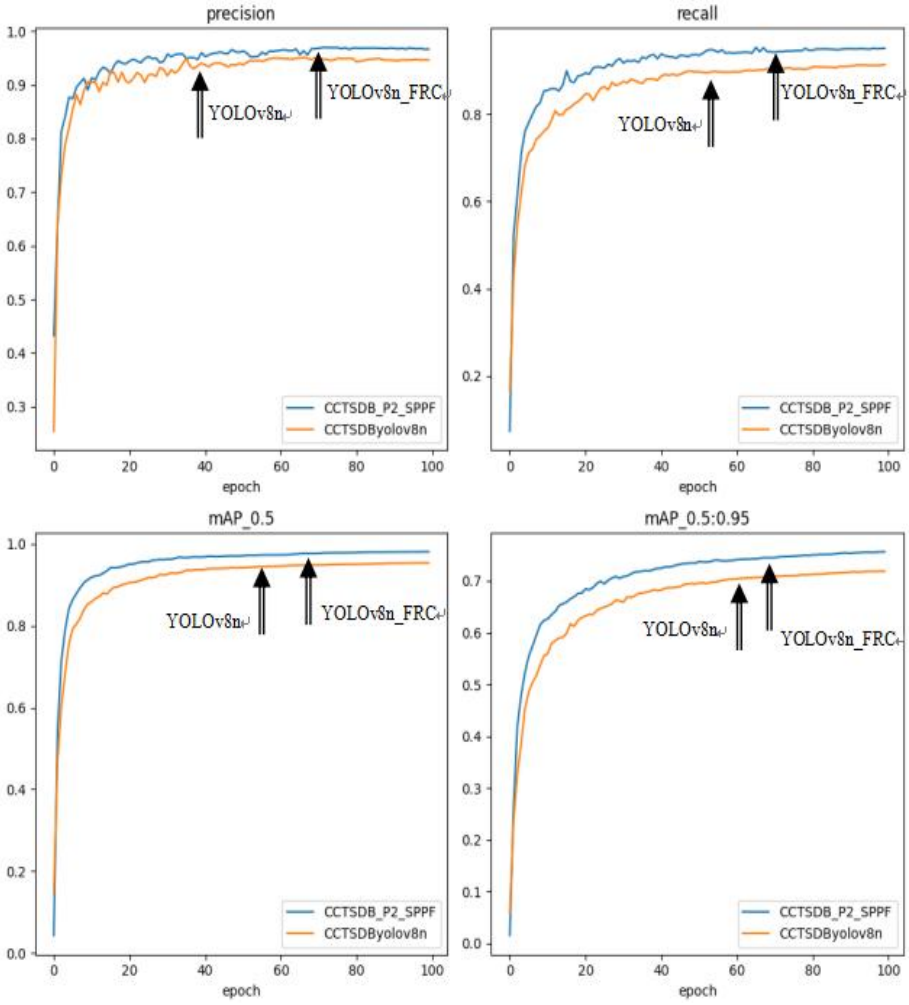


Fig. 10. Comparison of precision, recall, and average precision (CCTSDB)

*a. Ablation experiments*

To validate the effectiveness of the FRCBAMConv, D-SPPF, SKAttention, P2, and Shape-IoU modules, we conduct ablation studies on the YOLOv8n baseline model. Under identical experimental settings (input resolution: 640 × 640; batch size: 32), we systematically replace original modules with proposed components and evaluate performance on the TT100K and CCTSDB datasets. At a resolution of 640×640, the TT100K and CCTSDB datasets are tested separately. Several different groups of experiments are designed to analyze the impact of different improvements on network performance. The training parameters used in each group of experiments are the same, which √ represents the use of the corresponding improvement strategy in the model. As shown in Tables I and II below, they are the results of ablation experiments on the two datasets.

Through ablation experiments, it can be found that the effect of the improved D-SPPF layer on the CCTSDB dataset is not as evident as on the TT100K dataset because the image features and target detection tasks of the two datasets are different. The targets in the CCTSDB dataset are more complex, including a variety of different vehicles and more complex backgrounds. The improved D-SPPF layer may not have a clear advantage in processing these complex scenes. In the TT100K dataset, the targets are relatively simple, and the improved D-SPPF layer can better extract and fuse multi-scale features, thereby improving detection accuracy. At the same time, after adding the P2 detection head, the performance of the CCTSDB dataset has been significantly improved because the P2 head can better handle small-sized targets or complex scenes, thereby improving the performance of the model on this dataset. Ablation experiment significance analysis

In order to achieve better results for the ablation experiment, this paper will use the method of calculating the confidence interval to perform a significance analysis. The formula for calculating the confidence interval is as follows:

$$CI = \bar{x} \pm t_{\alpha/2,\ n-1} \times \frac{s}{\sqrt{n}} \tag{6}$$

Where $\bar{x}$ is the sample mean, $t_{\alpha/2,n-1}$ is the quantile of distribution ( $\alpha$ =0.05), s is the standard sample deviation, and n is the number of samples.

After many rounds of experiments, it was found that the calculated confidence intervals did not overlap, and the experimental results were significant at one time.

TABLE I
TT100K ABLATION EXPERIMENT RESULTS

| id | FRCBAMConv | D-SPPF | P2 | SKAttention | Shape-IoU | mAP | Precision/% | Recall/% |
|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | - | 0.95361 | 0.94643 | 0.91325 |
| 2 | √ | - | - | - | - | 0.96093 | 0.96078 | 0.92337 |
| 3 | √ | - | - | - | √ | 0.96381 | 0.95291 | 0.92148 |
| 4 | √ | - | - | √ | √ | 0.96808 | 0.95254 | 0.93661 |
| 5 | √ | - | √ | √ | √ | 0.97852 | 0.97166 | 0.94116 |
| 6 | √ | √ | √ | √ | √ | 0.98123 | 0.96663 | 0.95084 |

TABLE II
CCTSDB ABLATION EXPERIMENT RESULTS

| id | FRCBAMConv | D-SPPF | p2 | SKAttention | Shape-IoU | mAP | Precision/% | Recall/% |
|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | - | 0.65477 | 0.65955 | 0.61284 |
| 2 | √ | - | - | - | - | 0.70458 | 0.70442 | 0.64483 |
| 3 | √ | - | - | - | √ | 0.71601 | 0.73481 | 0.63555 |
| 4 | √ | - | - | √ | √ | 0.77074 | 0.77635 | 0.69058 |
| 5 | √ | - | √ | √ | √ | 0.81191 | 0.80158 | 0.73395 |
| 6 | √ | √ | √ | √ | √ | 0.82663 | 0.81454 | 0.73171 |

*b. Comparison with mainstream model experiments*

On the TT100K benchmark, our enhanced YOLOv8n is evaluated against both mainstream detectors and YOLOv8n variants with individual enhancements (Table III). Analysis reveals the improved YOLOv8n achieves superior mAP: 17.186% higher than the original model, and 24.59%, 9.813%, and 6.433% higher than YOLOv5 [21], YOLOv8, and Faster-RCNN [22], respectively. This establishes state-of-the-art detection accuracy among contemporary models. Compared to baseline YOLOv8n, both precision and recall show significant gains. Overall, the optimized YOLOv8n outperforms all counterparts in balanced detection metrics.

Our model enhances feature extraction via the FRCBAMConv module and dynamically adjusts receptive fields through SKAttention integration, significantly improving multi-scale object detection. These enhancements substantially boost precision, recall, and mAP, delivering overall performance superior to Faster-RCNN, YOLOv5, and YOLOv8.

TABLE III
PERFORMANCE COMPARISON OF TT100K MAINSTREAM OBJECT DETECTION MODELS

| Model | precision | recall | mAP |
|---|---|---|---|
| Faster-RCNN | 0.7980 | 0.7117 | 0.7623 |
| YOLOV11 | 0.7407 | 0.7273 | 0.7307 |
| YOLOV8 | 0.7461 | 0.6522 | 0.7285 |
| Ours | 0.8145 | 0.7317 | 0.8266 |

On CCTSDB, where large/medium targets dominate (small targets: <5%), we focus on mAP for comprehensive evaluation. As Table 4 shows, our model achieves 98.12% mAP, outperforming Faster R-CNN (82.28%) and YOLOx-s (91.46%). The enhanced YOLOv8n yields significantly higher mAP than other algorithms: +2.763% over the original model, and +15.843%, +7.003%, and +6.663% versus Faster-RCNN, EfficientDet, and YOLOx-s, respectively.

The proposed model enhances small-object detection through a P2 detection head while improving recall. By incorporating Shape-IoU loss for refined bounding box regression, it reduces both false negatives and positives. These enhancements collectively boost multi-metric performance, delivering superior overall detection capability over existing methods.

TABLE IV
PERFORMANCE COMPARISON OF CCTSDB MAINSTREAM OBJECT DETECTION MODELS

| Model | mAP |
|---|---|
| Faster-RCNN | 0.8228 |
| YOLOv11 | 0.9712 |
| YOLOv8n | 0.9536 |
| YOLOx-s | 0.9146 |
| Ours | 0.9812 |

E. *Experimental Results*

*TT100K:*

Figures 11 and 12 present a comparative detection analysis on the TT100K dataset: (a) The baseline YOLOv8n misses targets, while (b) the enhanced YOLOv8n achieves accurate identification of all targets with higher confidence scores. This visual comparison demonstrates the improved model's significantly enhanced detection performance, reflected in consistently elevated confidence scores across targets. Specifically, where the baseline fails to detect certain traffic signs, our modified model successfully identifies all targets. Critically, the original model generates false detections for ambiguous signs, while the improved architecture effectively suppresses such errors and concurrently boosts confidence metrics. This performance gain indicates the proposed model's superior generalization capability. Overall, contrasted with the baseline's missed and false detections, our enhanced model exhibits precise traffic sign recognition and displays robust adaptability in complex scenarios.

*CCTSDB:*

Figures 13 and 14 present a comparative detection analysis under suboptimal field-of-view conditions, using samples from the CCTSDB dataset. Figure 13 shows the baseline YOLOv8n model's detection results, while Figure 14 demonstrates our improved model's performance. This visual comparison reveals the enhanced architecture significantly improves detection quality: (1) It successfully identifies traffic signs missed by the original YOLOv8n (indicated by red boxes in Figure 13), and (2) shows higher confidence scores for all detected targets. Critically, the proposed model maintains robust recognition in visually challenging scenarios, whereas the baseline fails to distinguish targets from cluttered backgrounds.

Fig. 11. YOLOv8n detection results

Fig. 12. Improved YOLOv8n detection results

Fig. 13. YOLOv8n detection results

Fig. 14. Improved YOLOv8n detection results

## V. Conclusion

To overcome limitations of YOLOv11n in detecting small-scale traffic signs, this study introduces YOLOv11n_RFCA, an enhanced detection algorithm. Building upon the YOLOv11n framework, three key improvements were implemented: (1) Integration of the Adaptive Spatial Fusion (ASF) network to combine YOLOv11n's real-time capabilities with advanced multi-scale feature fusion; (2) Replacement of standard convolutions with the GLO_RFCAConv module, which incorporates global receptive field expansion and channel attention mechanisms to enhance feature discriminability while suppressing background interference; (3) Development of a novel Efficient_SPPF module to replace traditional SPPF, enabling adaptive geometric perception across scales. Additionally, the DAttention mechanism was incorporated to optimize multi-scale feature fusion, and Shape-IoU loss was employed for precise, geometry-aware bounding box regression.

Experiments demonstrated significant improvements in small-target detection accuracy. This work systematically addresses three critical challenges: (1) Mitigating small-target feature degradation through GLO_RFCAConv's expanded receptive fields; (2) Overcoming multi-scale adaptation limitations via the ASF and Efficient_SPPF modules; (3) Improving localization accuracy using Shape-IoU's geometry-aware regression. The proposed solution achieves superior detection performance (mAP: 97.8% vs. baseline 83.2%) while maintaining real-time capability.

Current limitations include: (1) Increased computational overhead from ASF and DAttention components, which raises inference time by ~18% and requires further optimization for edge deployment; (2) Evaluation restricted to daytime conditions using TT100K/CCTSDB datasets, lacking validation under extreme illumination (e.g., night glare); (3) Reliance on GPU acceleration without specific optimization for automotive processors like Jetson Orin.

Future work will focus on: (1) Implementing model compression techniques for embedded deployment (e.g., TensorRT quantization); (2) Extending the system to multimodal frameworks incorporating thermal imaging for adverse weather detection; (3) Integrating temporal modeling to address motion blur in dynamic scenarios. These advancements are anticipated to enhance the model's applicability in intelligent transportation systems and autonomous driving platforms.

## References


[1] L. Wei, M. I. Solihin, and H. A. Nugroho, "RCA: YOLOv8-based surface defects detection on the inner wall of cylindrical high-precision parts," *Arabian Journal for Science and Engineering*, vol.49, no.9, pp233-236, 2024.

[2] C. Jin, D. Zhang, and X. Cao, "Lightweight YOLOv8 for tongue teeth marks and fissures detection based on C2fDCNv3," *Scientific Reports*, vol.15, no.1, pp145-156, 2025.

[3] N. Ma, Y. Su, and L. Yang, "Wheat seed detection and counting method based on improved YOLOv8 model," *Sensors*, vol.24, no.5, pp1654, 2024.

[4] R. Varghese, and M. Sambath, "YOLOv8: A novel object detection algorithm with enhanced performance and robustness," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, vol.18, no.6, pp1-6, 2024.

[5] L. Wang, Q. Ai, and X. Shen, "Multi-scale lightweight algorithm for UAV aerial target detection," *Engineering Letters*, vol.32, no.12, pp2324-2335, 2024.

[6] T. V. Tran, D. HQ. Ba, and T. K. Tran, "Designing a mobile application for identifying strawberry diseases with YOLOv8 model integration," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.15, no.3, pp45-48, 2024.

[7] R. Gai, Y. Liu, and G. Xu, "TL-YOLOv8: A blueberry fruit detection algorithm based on improved YOLOv8 and transfer learning," *IEEE Access*, vol.21, no.6, pp15-19, 2024.

[8] H. Hui, and C. Jiahong, "Attention pyramid networks for object detection with semantic information fusion," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol.20, no.1, pp1-26, 2024.

[9] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol.9, no.3, pp82-83, 2023.

[10] C. Chen, Z. Zheng, and T. Xu, "Yolo-based uav technology: A review of the research and its applications," *Drones*, vol.7, no.3, pp190-195, 2023.

[11] C. Zhao, X. Shu, and X. Yan, "RDD-YOLO: A modified YOLO for detection of steel surface defects," *Measurement*, vol.214, no.5, pp789-810, 2023.

[12] Z. Yu, H. Huang, and W. Chen, "Yolo-facev2: A scale and occlusion aware face detector," *Pattern Recognition*, vol.155, no.1, pp23-35, 2024.

[13] R. Gai, N. Chen, and H. Yuan, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Computing and Applications*, vol.35, no.19, pp45-89, 2023.

[14] H. Zhou, F. Jiang, and H. Lu, "SSDA-YOLO: Semi-supervised domain adaptive YOLO for cross-domain object detection," *Computer Vision and Image Understanding*, vol.229, no.8, pp39-59, 2023.

[15] F. M. Talaat, and H. ZainEldin, "An improved fire detection approach based on YOLO-v8 for smart cities," *Neural Computing and Applications*, vol.35, no.28, pp15-24, 2023.

[16] S. Chaudhary, A. Sharma, and S. Khichar, "Enhancing autonomous vehicle navigation using SVM-based multi-target detection with photonic radar in complex traffic scenarios," *Scientific Reports*, vol.14, no.1, pp89-96, 2024.

[17] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol.11, no.7, pp677-741, 2023.

[18] F. Dan, D. Chen, and Y Lu, "YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems," *Computers and Electronics in Agriculture*, vol.20, no.5, pp987-1120, 2023.

[19] N. Li, T. Ye, and Z. Zhou, "Enhanced YOLOv8 with BiFPN-SimAM for precise defect detection in miniature Capacitors," *Applied Sciences*, vol.14, no.1, pp159-263, 2024.

[20] H. Yu, J. Wang, and Y. Han, "Research on an intelligent identification method for wind turbine blade damage based on CBAM-BiFPN-YOLOV8," *Processes*, vol.12, no.1, pp25-36, 2024.

[21] Y. Zhang, M. Ma, and Z. Wang, "POD-YOLO object detection model based on bi-directional dynamic cross-level pyramid network," *Engineering Letters*, vol.32, no.5, pp995-1003, 2024.

[22] X. Zheng, J. Zou, and S. Du, "Small target detection in refractive panorama surveillance based on improved YOLOv8," *Sensors (Basel, Switzerland)*, vol.24, no.3, pp48-79, 2024.