

Article

Not peer-reviewed version

---

# A Traffic Sign Detection Algorithm Based on an Improved YOLOv8n

---

[Yanyan Jia](#)<sup>\*</sup> and Siyi Wang

Posted Date: 14 April 2026

doi: 10.20944/preprints202604.0991.v1

Keywords: traffic sign detection; YOLOv8n; VoVGSCSP module; multi-scale contextual attention



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# A Traffic Sign Detection Algorithm Based on an Improved YOLOv8n

Yanyan Jia \* and Siyi Wang

College of Artificial Intelligence & Computer Science, Xi'an University of Science and Technology,  
Xi'an 710054, China

\* Correspondence: yyjia@xust.edu.cn

## Abstract

Traffic sign detection in autonomous driving faces challenges including multi-scale objects, complex backgrounds, and limited edge-computing power. To address insufficient multi-scale feature representation and high false negatives for small traffic signs in YOLOv8n, this study proposes an improved algorithm integrating the VoVGSCSP module with a Multi-scale Contextual Attention (MCA) mechanism. The original C2f module is replaced with VoVGSCSP, enhancing feature representation through parallel residual branches and cross-stage connections. A lightweight neck, SlimNeck, is designed and combined with MCA, employing multi-branch pooling and dynamic weight fusion to capture geometric features and color semantics. The PAN-FPN path is optimized with cross-level connections and learnable weights for adaptive multi-scale fusion. Experiments on the GTSRB dataset show that the improved model reduces parameters to 2.66 M (an 11.6% decrease) and computational complexity to 7.49 GFLOPs, while mAP@0.5 increases from 94.7% to 96.3% and FPS improves from 82.3 to 90.6. The proposed algorithm achieves comprehensive gains in lightweighting, accuracy, and speed, demonstrating its effectiveness and practical applicability.

**Keywords:** traffic sign detection; YOLOv8n; VoVGSCSP module; multi-scale contextual attention

## 1. Introduction

The ongoing intelligent transformation of the automotive industry has positioned autonomous driving at the forefront of global technological competition. According to industry reports, by 2025, the penetration rate of passenger vehicles equipped with Level 2 or higher driver assistance functions in China is expected to exceed 50%, making intelligent driving a standard feature rather than a high-end option [1]. Traffic sign detection, a critical component of environmental perception in autonomous driving, plays a pivotal role in ensuring vehicle safety. Its accuracy and real-time performance directly influence the vehicle's understanding of traffic regulations and the reliability of its driving decisions [2].

Compared with general object detection tasks, traffic sign detection exhibits unique challenges. First, target scales vary dramatically: distant signs may occupy as few as 10 pixels, while nearby signs may exceed 200 pixels [1]. Second, the category distribution follows a long-tail pattern, with significant sample size disparities across different types of traffic signs. Third, environmental factors such as illumination changes, adverse weather (e.g., rain and fog), and partial occlusion substantially degrade sign visibility [3]. Furthermore, in mass-production autonomous driving solutions, perception algorithms must be deployed on onboard computing platforms with strict constraints on power consumption, heat dissipation, and cost. Given that onboard cameras continuously capture video streams at over 30 fps, real-time and lightweight algorithm design becomes imperative [4].

In recent years, deep learning-based object detectors have achieved significant progress in traffic sign recognition. One-stage detectors, especially the YOLO series, have become mainstream due to their balance between accuracy and speed [5,6]. YOLOv8n, a lightweight version released by Ultralytics in 2023 [7], employs an optimized CSPDarknet53 backbone, a C2f feature extraction

module, and an anchor-free detection head. However, when applied to traffic sign detection, YOLOv8n shows limitations: the C2f module has limited multi-scale feature capture capability, the feature enhancement network inadequately fuses shallow detail features with deep semantic features, leading to high miss rates for small signs, and the model lacks effective background interference suppression [3,8].

To address these issues, researchers have proposed various improvements. Dewi et al. [3] developed a YOLOv4-based model incorporating synthetic training data. Wang et al. [8] modified YOLOv5 for real-time multi-scale detection. Wang et al. [9] proposed YOLO-AEF for complex illumination conditions. He et al. [10] introduced NTS-YOLO for nighttime detection. Ji et al. [11] improved YOLOv8 for small-target detection in complex environments. Du et al. [12] presented ES-YOLO, achieving a 10.4% mAP50 improvement on the TT100K dataset. Despite these contributions, there remains room for improvement in balancing feature representation and computational efficiency [8].

In this paper, we propose an improved algorithm integrating the VoVGSCSP module with a Multi-scale Contextual Attention mechanism. The main contributions are threefold: (1) introducing VoVGSCSP to replace the C2f module in the feature enhancement network, enhancing feature representation through parallel residual branches and cross-stage connections; (2) designing SlimNeck, a lightweight neck embedded with MCA, which uses multi-branch pooling and dynamic weight fusion to strengthen geometric and semantic feature capture while suppressing background interference; and (3) optimizing the PAN-FPN feature fusion path [13] by adding cross-level connections and learnable weight parameters for adaptive multi-scale feature fusion, thereby improving small-target detection performance. Experiments on the GTSRB dataset [2] validate the advantages of the proposed algorithm in accuracy, speed, and model lightweighting.

## 2. Methodology

### 2.1. Analysis of the YOLOv8n Baseline Model

YOLOv8n is a lightweight object detection model whose architecture consists of four components: input, backbone, neck, and head [7]. The backbone adopts an optimized CSPDarknet53 architecture, comprising Conv modules, C2f modules, and an SPPF module. The input image is resized to  $640 \times 640 \times 3$ , and three feature maps are output at scales of  $80 \times 80 \times 256$ ,  $40 \times 40 \times 512$ , and  $20 \times 20 \times 512$ . The C2f module draws on the ELAN design concept from YOLOv7, employing a gradient branching strategy to achieve lightweighting while enhancing feature transfer and fusion efficiency [6]. The SPPF module uses serial and parallel max-pooling layers to enlarge the receptive field with reduced computational cost [8]. The Conv module combines convolution, batch normalization, and the SiLU activation function (Equation 1, Figure 1).

$$\text{SiLU}(x) = x \cdot \text{Sigmoid}(x) \quad (1)$$

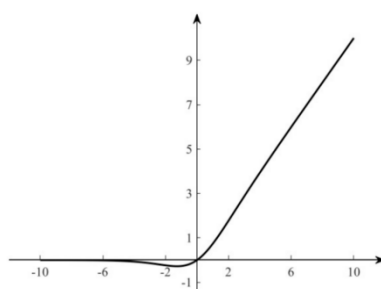


Figure 1. SiLU activation function curve.

The neck adopts a PAN-FPN hybrid structure [13] with a top-down FPN path and a bottom-up PAN path for multi-scale feature fusion. The head uses a decoupled design, with binary cross-

entropy loss for classification and a combination of distribution focal loss (DFL) [14] and CIoU loss [15] for regression. As an anchor-free model, YOLOv8n dynamically assigns samples using a task-aligned assigner (Equation 2).

$$t = s^\alpha \times u^\beta \quad (2)$$

where  $s$  is the classification score,  $u$  is the IoU between the predicted box and the ground truth, and  $\alpha$  and  $\beta$  are hyperparameters.

## 2.2. Improved Network Architecture

### 2.2.1. VoVGSCSP Module

To address the high sensitivity of traffic sign detection to geometric features and color semantics, this study replaces the original C2f module with the VoVGSCSP (Voxel-Guided Visual Geometry Group Shuffling Cross-Stage Partial) module, as shown in Figure 2. Based on the CSP architecture [16], this module incorporates parallel residual and dense connection branches to capture both local detail features and global semantic features of traffic signs, while employing channel grouping strategies to reduce computational complexity [16,17].

The VoVGSCSP module improves gradient flow within the network, ensuring stable training, and leverages the structural advantages of parallel branches to extract complementary features. Compared with C2f, VoVGSCSP improves gradient flow, ensures stable training, and extracts complementary features, making it more suitable for traffic sign detection [3,8].

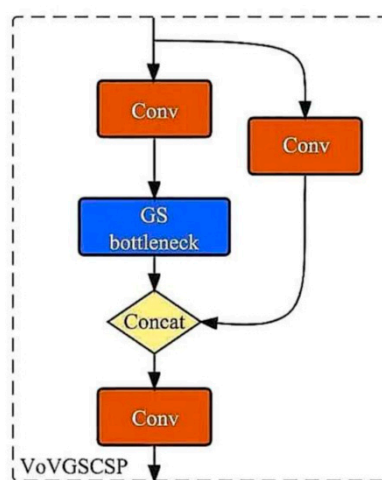


Figure 2. VoVGSCSP module structure.

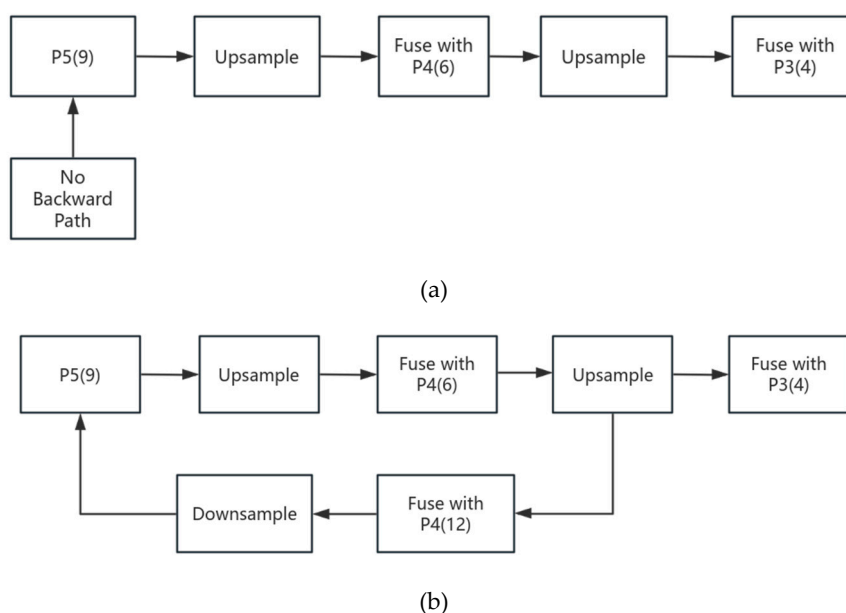
### 2.2.2. SlimNeck and Multi-scale Contextual Attention (MCA)

To improve adaptability to multi-scale traffic signs and robustness against complex backgrounds, we design a lightweight neck structure called SlimNeck and introduce a Multi-scale Contextual Attention (MCA) mechanism. SlimNeck is based on GSConv [16,17] to reduce redundant computation. MCA comprises two submodules: MCALayer and MCAGate. MCALayer performs cross-dimensional context modeling, while MCAGate uses three pooling strategies (average, standard deviation, and max pooling) with learnable weights for adaptive feature fusion [18]. The core computation is defined in Equation 3:

$$\text{Attention} = \sigma(\text{MCALayer}(F) \otimes \text{MCAGate}(F)) \quad (3)$$

### 2.2.3. Feature Fusion Path Optimization

To address the insufficient fusion of small-target features in the original PAN-FPN structure [13], we introduces three optimizations to the feature fusion path. The improved paths are illustrated in Figure 3. First, a direct connection is added between the P3 (80×80) small-target feature layer and the P5 (20×20) deep feature layer, allowing small-target features to directly incorporate high-level semantic information and improving detection accuracy for small objects. Second, a bidirectional enhanced feature fusion path is designed to strengthen information exchange across feature scales while preserving the original top-down and bottom-up pathways, balancing semantic and spatial detail. Third, learnable weight parameters [8] are introduced before feature concatenation to adaptively adjust the contribution of different scale features based on the characteristics of traffic sign detection tasks, avoiding feature imbalance during fusion.



**Figure 3.** Comparison of feature fusion paths before and after optimization: (a) Original structure (one-way fusion); (b) Improved structure (bidirectional fusion).

### 2.3. Model Lightweighting Strategy

To adapt the improved model to the computational constraints of onboard edge platforms, we adopts a multi-dimensional lightweighting strategy tailored to traffic sign detection. The neck structures before and after optimization are compared in Figure 4. First, channel pruning is applied to reduce the number of channels in non-critical feature layers, lowering parameter count and computational cost while maintaining feature extraction capability [19]. Second, modular deployment is implemented, applying the VoVGSCSP module and MCA mechanism only to key layers of the feature enhancement network while retaining lightweight basic modules in non-critical layers, balancing performance and computational efficiency. Third, quantization compression is employed, applying 8-bit quantization to non-sensitive layers [4] to further reduce storage requirements and inference latency, enhancing deployment adaptability on embedded devices.

```

# YOLOv8.0n head
head:
- [-1, 1, nn.Upsample, [None, 2, 'nearest']]
- [[-1, 6], 1, Concat, [1]] # cat backbone P4
- [-1, 3, C2f, [512]] # 12

- [-1, 1, nn.Upsample, [None, 2, 'nearest']]
- [[-1, 4], 1, Concat, [1]] # cat backbone P3
- [-1, 3, C2f, [256]] # 15 (P3/8-small)

- [-1, 1, Conv, [256, 3, 2]]
- [[-1, 12], 1, Concat, [1]] # cat head P4
- [-1, 3, C2f, [512]] # 18 (P4/16-medium)

- [-1, 1, Conv, [512, 3, 2]]
- [[-1, 9], 1, Concat, [1]] # cat head P5
- [-1, 3, C2f, [1024]] # 21 (P5/32-large)

- [[15, 18, 21], 1, Detect, [nc]] # Detect(P3, P4, P5)

```

```

head:
- [-1, 1, nn.Upsample, [None, 2, 'nearest']]
- [[-1, 6], 1, Concat, [1]] # cat backbone P4
- [-1, 3, VoVGSCSP, [512]] # 12

- [-1, 1, nn.Upsample, [None, 2, 'nearest']]
- [[-1, 4], 1, Concat, [1]] # cat backbone P3
- [-1, 3, VoVGSCSP, [256]] # 15 (P3/8-small)

- [-1, 1, Conv, [256, 3, 2]]
- [[-1, 12], 1, Concat, [1]] # cat head P4
- [-1, 3, VoVGSCSP, [512]] # 18 (P4/16-medium)

- [-1, 1, Conv, [512, 3, 2]]
- [[-1, 9], 1, Concat, [1]] # cat head P5
- [-1, 3, VoVGSCSP, [1024]] # 21 (P5/32-large)
- [-1, 1, MCALayer, [1024]] # 21 (P5/32-large)

- [[15, 18, 22], 1, Detect, [nc]] # Detect(P3, P4, P5)

```

**Figure 4.** Comparison of neck structures before and after optimization:(a) Before improvement;(b) After improvement.

### 3. Experimental Design

#### 3.1. Dataset

Experiments are conducted on the German Traffic Sign Recognition Benchmark (GTSRB) dataset [2], a classic public dataset in the field. It contains 51,839 images across 43 categories of common traffic signs, divided into 39,209 training images and 12,630 test images. All images are resized to 640×640 pixels to align with the YOLOv8n input specifications. The dataset covers various real-world scenarios including illumination variations, partial occlusion, motion blur, and complex backgrounds, effectively evaluating algorithm performance in realistic environments [2,3].

#### 3.2. Experimental Setup

A controlled experimental design is adopted, comparing the baseline YOLOv8n model with the improved YOLOv8n-SlimNeck-MCA model. The baseline YOLOv8n model has approximately 3.01 M parameters [7], while the improved model replaces the original C2f modules with VoVGSCSP modules and embeds the MCA attention mechanism in the PANet stage.

Experiments are implemented using the PyTorch deep learning framework with the following training configuration: 50 epochs, batch size 16, input size 640×640; the Adam optimizer with Cosine decay learning rate scheduling, initial learning rate 0.001, final learning rate 0.00001; 16 data loading threads to improve I/O efficiency; mixed precision training (AMP) enabled for acceleration; and an early stopping mechanism with patience of 10 epochs to prevent overfitting [8]. The improved model is configured via a custom YAML file (yolov8-slimneck-mca.yaml) to load the architectural modifications.

#### 3.3. Evaluation Metrics

To comprehensively evaluate the performance of the improved model, mean average precision (mAP), precision, recall, and F1-score are used as detection accuracy metrics, while parameter count, computational cost (GFLOPs), and frames per second (FPS) serve as lightweighting and real-time performance metrics [5,7].

(1) Mean average precision (mAP):

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

where  $AP_i$  is the average precision for class  $i$ , computed as the area under the precision-recall curve, and  $N$  is the total number of traffic sign categories [1].

(2) Precision and recall:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

where TP is the number of true positives (correctly detected signs), FP is the number of false positives (background incorrectly detected as signs), and FN is the number of false negatives (undetected signs) [9].

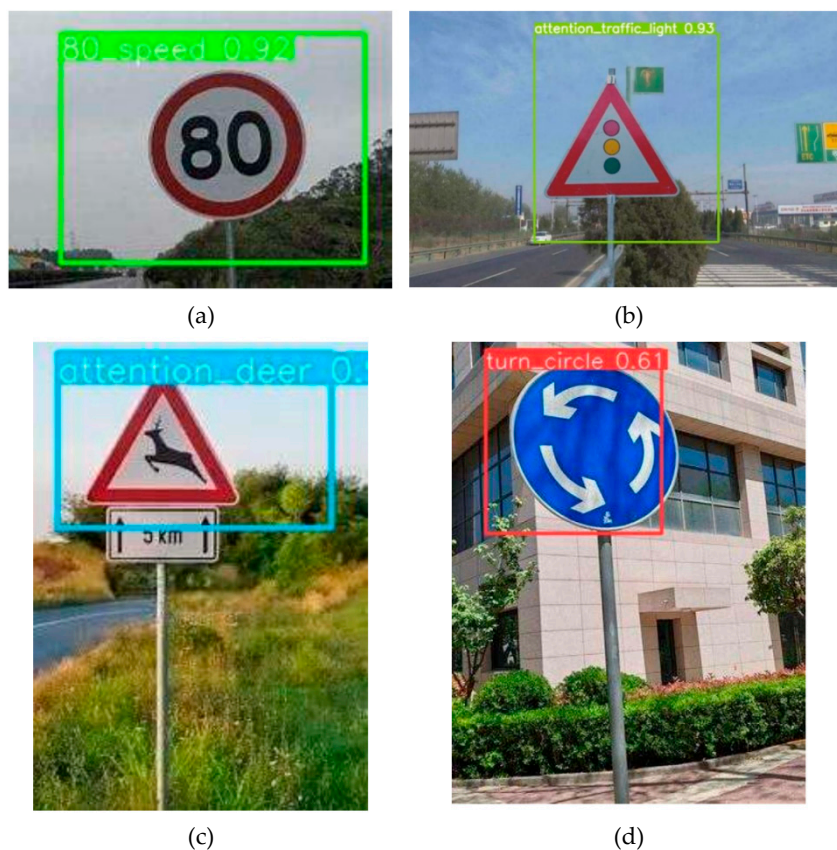
(3) F1-score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

The F1-score is the harmonic mean of precision and recall, providing a comprehensive assessment of detection performance [6].

#### 4. Results and Analysis

To intuitively demonstrate the traffic sign detection capability of the improved YOLOv8n model in real-world scenarios, this section first presents qualitative detection results, followed by quantitative analysis including metrics, loss curves, precision-recall curves, and F1-confidence curves to comprehensively validate the effectiveness of the proposed improvements. Detection results of the improved YOLOv8n model are shown in Figure 5, demonstrating accurate detection and classification across various traffic signs, including small-scale speed limit signs and signs under complex backgrounds, maintaining robust detection performance under varying target scales and background interference.



**Figure 5.** Detection results of the improved model:(a) Speed limit 80; (b) Traffic light ahead;(c) Deer crossing; (d) Roundabout.

#### 4.1. Quantitative Results Analysis

To clearly illustrate performance differences between the improved and baseline models, key performance metrics are summarized in Table 1.

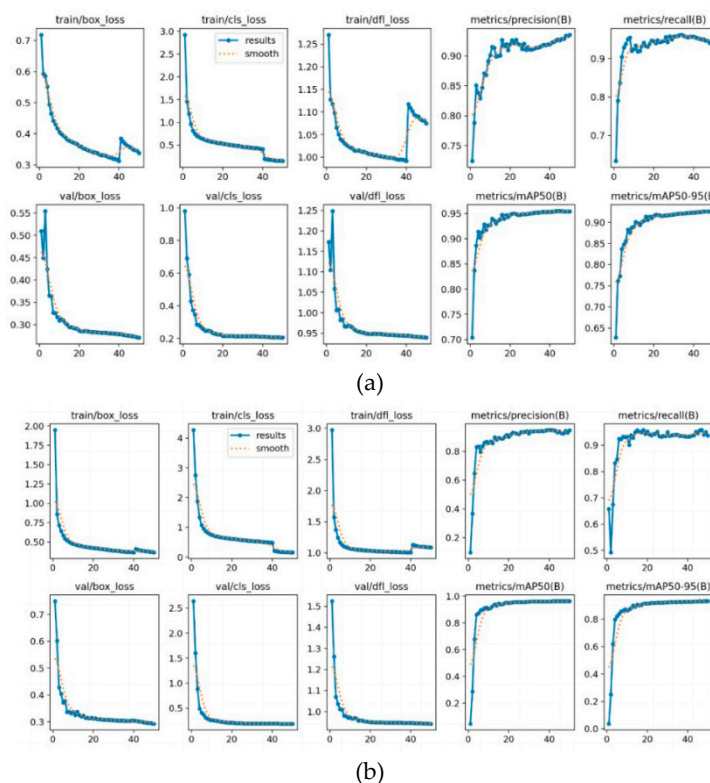
**Table 1.** Comparison of model performance metrics.

Model	Parameters (M)	GFLOPs	mAP@0.5 (%)	FPS
Baseline	3.01	8.25	94.7	82.3
Improved	2.66	7.49	96.3	90.6

As shown in Table 1, the improved YOLOv8n model achieves significant gains across all dimensions: parameters are reduced from 3.01 M to 2.66 M (-11.6%); computational cost drops from 8.25 GFLOPs to 7.49 GFLOPs; mAP@0.5 improves from 94.7% to 96.3%; and FPS increases from 82.3 to 90.6. These results demonstrate that the proposed improvements simultaneously enhance model lightweighting, detection accuracy, and inference speed, achieving superior overall performance compared to the baseline [7].

#### 4.2. Loss Curves and Performance Metrics Analysis

Loss curves and performance metrics for both models are presented in Figure 6. Loss metrics include bounding box loss (box\_loss), classification loss (cls\_loss), and distribution focal loss (dfl\_loss). Performance metrics include precision, recall, and mAP@0.5.



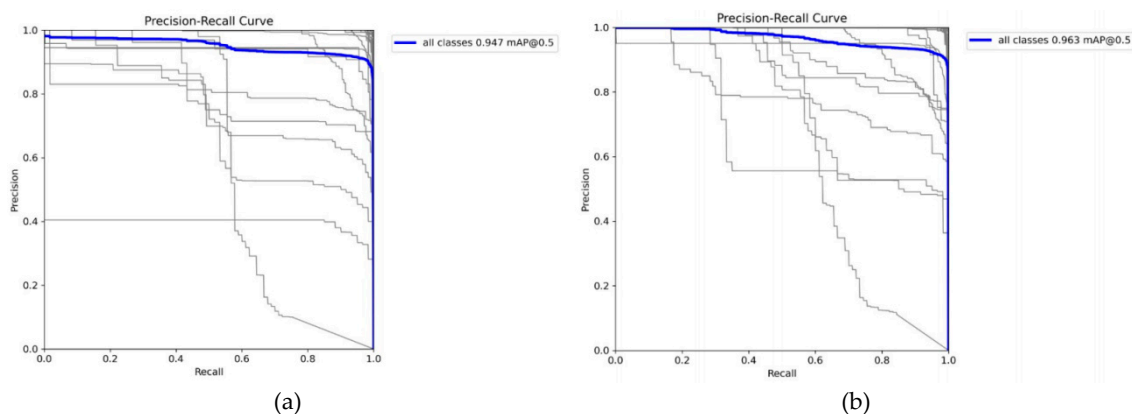
**Figure 6.** Comparison of loss and performance metrics before and after improvement:(a) Training and validation losses and performance metrics of YOLOv8n;(b)Training and validation losses and performance metrics of the improved YOLOv8n.

Analysis of loss curves reveals that the baseline YOLOv8n model exhibits slow loss reduction, with validation loss showing substantial fluctuations during early training, indicating slow convergence and limited generalization capability [7]. In contrast, the improved model achieves faster loss reduction and stabilizes earlier, with validation loss remaining stable throughout training,

demonstrating enhanced convergence and generalization. Analysis of performance metrics shows that while the baseline model achieves rapid initial gains in precision and recall, a significant gap between the two persists, with recall remaining low—indicating a pronounced small-target missed detection issue—and mAP@0.5 showing limited improvement. The improved model achieves faster initial gains in both precision and recall, maintains higher values with a substantially reduced gap between them, effectively mitigating the missed detection problem, and achieves a substantial increase in mAP@0.5, indicating overall detection performance optimization.

#### 4.3. Precision-Recall Curve Analysis

Precision-recall (PR) curves for both models are compared in Figure 7. The area under the PR curve represents average precision (AP), and the mean across all categories is mAP@0.5. The baseline model achieves mAP@0.5 of 0.947, while the improved model achieves 0.963. The PR curve shifts upward and rightward, maintaining stability in high-precision and high-recall regions, indicating that the improved model enhances recall without sacrificing precision, effectively optimizing the precision-recall balance [15]. Additionally, the smoother PR curve of the improved model suggests more balanced detection performance across different traffic sign categories, with significantly reduced missed detections and false positives, enabling more efficient and accurate recognition of various traffic signs.

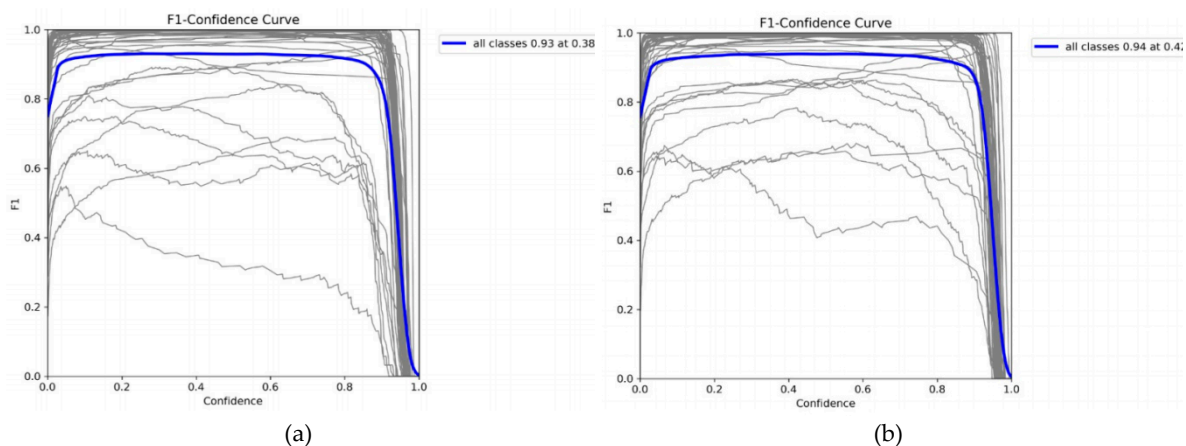


**Figure 7.** Precision-recall curves before and after optimization:(a) Recall curve before improvement;(b) Recall curve after improvement.

#### 4.4. F1-Confidence Curve Analysis

F1-confidence curves for both models are compared in Figure 8. These curves reflect F1-score variations under different confidence thresholds, providing a comprehensive assessment of detection stability [8]. The baseline model achieves a maximum F1-score of 0.93 at a confidence threshold of 0.383, with notable fluctuations across categories, indicating insufficient detection stability and significant performance variations across confidence thresholds [7]. The improved model achieves a maximum F1-score of 0.94 at a confidence threshold of 0.428, with a smoother curve and reduced fluctuations, demonstrating higher consistency across categories.

These results indicate that the improved model achieves better overall performance in balancing precision and recall, with significantly enhanced detection stability, maintaining high detection performance across various confidence thresholds and effectively reducing false positives and missed detections.



**Figure 8.** F1-confidence curves before and after optimization: (a) F1–confidence curve before improvement; (b) F1–confidence curve after improvement.

#### 4.5. Qualitative Results Analysis

Qualitative detection results demonstrate that the improved model accurately detects traffic signs across various scenarios. First, it achieves high detection accuracy for small-target signs such as speed limits and yield signs, addressing the baseline model's small-target missed detection issue and validating the effectiveness of the optimized PAN-FPN path for small-target feature fusion [13]. Second, under challenging conditions including strong illumination, backlighting, and partial occlusion, the model maintains accurate traffic sign recognition, validating the MCA attention mechanism's ability to suppress background interference and capture key features [18]. Third, in scenes with overlapping signs and complex backgrounds, the model achieves precise detection and classification, validating the VoVGSCSP module's strong capability to represent geometric and semantic features of traffic signs.

The qualitative results align with quantitative experimental findings, collectively validating the effectiveness of the three proposed improvements—VoVGSCSP module, MCA mechanism, and PAN-FPN path optimization—which synergistically enhance the model's traffic sign detection capability.

## 5. Conclusion

To address the limitations of YOLOv8n in traffic sign detection—namely insufficient multi-scale feature representation, high false negatives for small targets, and poor robustness in complex environments—this paper proposes an improved lightweight detection algorithm that integrates the VoVGSCSP module with a Multi-Context Attention (MCA) mechanism. Specifically, the original C2f module is replaced with VoVGSCSP to enhance feature representation. A lightweight neck, SlimNeck, embedded with MCA, reduces computational cost while improving the model's ability to capture key traffic sign features and suppress background interference. Furthermore, the PAN-FPN path is optimized with cross-level connections and learnable weight parameters to enable adaptive multi-scale feature fusion, effectively mitigating the fusion deficiency for small targets. Experimental results on the GTSRB dataset demonstrate that the proposed model reduces parameters by 11.6% (to 2.66M), lowers computational cost to 7.49 GFLOPs, increases mAP@0.5 from 94.7% to 96.3%, and improves FPS from 82.3 to 90.6. These results confirm the effectiveness of the proposed improvements in balancing model lightweighting, detection accuracy, and inference speed. Compared with existing YOLO-based traffic sign detectors, the proposed method offers better deployment suitability for onboard edge devices, providing a practical solution for traffic sign detection in autonomous driving.

Future work will focus on enhancing cross-scenario robustness under extreme conditions (e.g., rain, fog, and nighttime) and exploring further model compression and acceleration techniques for low-power embedded platforms.

**Author Contributions:** Conceptualization, Y.Y. J. and S.Y.W.; methodology, Y.Y. J. and S.Y.W.; software, S.Y.W.; validation, Y.Y. J. and S.Y.W.; formal analysis, Y.Y. J.; investigation, Y.Y. J.; resources, Y.Y. J.; data curation, Y.Y. J. and S.Y.W.; writing—original draft preparation, Y.Y. J.; writing—review and editing, Y.Y. J.; visualization, Y.Y. J.; supervision, Y.Y. J.; project administration, Y.Y. J.; funding acquisition, Y.Y. J.. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2110-2118.
2. STALLKAMP J, SCHLIPSING M, SALMEN J, et al. The German traffic sign recognition benchmark: A multi-class classification competition[C]//Proceedings of the International Joint Conference on Neural Networks. 2011: 1453-1460.
3. DEWI C, CHEN R C, LIU Y T, et al. YOLO V4 for advanced traffic sign recognition with synthetic training data[J]. IEEE Access, 2021, 9: 97228-97242.
4. HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
5. REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
6. BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
7. JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLOv8[Z]. 2023.
8. WANG J, CHEN Y, DONG Z, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection[J]. Neural Computing and Applications, 2023, 35(10): 7353-7366.
9. WANG P, MOHAMED R, MUSTAPHA N, et al. YOLO-AEF: Traffic sign detection via adaptive enhancement and fusion[J]. Neurocomputing, 2025, 655: 131430.
10. HE Y, GUO M, ZHANG Y, et al. NTS-YOLO: A nocturnal traffic sign detection method based on improved YOLOv5[J]. Applied Sciences, 2025, 15(3): 1578.
11. JI B, XU J, LIU Y, et al. Improved YOLOv8 for small traffic sign detection under complex environmental conditions[J]. Franklin Open, 2024, 8: 100167.
12. DU S, SU S, LIN C, et al. ES-YOLO: Edge and shape fusion-based YOLO for traffic sign detection[J]. Computers, Materials & Continua, 2026, 87(1): 88.
13. LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2117-2125.
14. LI X, WANG W, WU L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[C]//Advances in Neural Information Processing Systems. 2020: 23902-23912.
15. REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 658-666.
16. WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: Scaling cross stage partial network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13029-13038.
17. LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022.

18. LI H, LI J, WEI H, et al. Slim-neck by GSConv: A lightweight design for real-time detector architectures[J]. *Journal of Real-Time Image Processing*, 2024, 21(3): 62.
19. LIU Z, LI J, SHEN Z, et al. Learning efficient convolutional networks through network slimming[C]//*Proceedings of the IEEE International Conference on Computer Vision*. 2017: 2736-2744.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.