Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 3, 2026

A Training Strategy–Driven YOLOv11s Method for Traffic Sign Detection

Zhuya Qi*, Yiran Han, Gelin Zeng

Shanghai Ocean University, Shanghai, China
* **Corresponding author**: Zhuya Qi (Email: 3525301162@qq.com)

Abstract: Traffic sign detection is a fundamental component of intelligent transportation systems, where both detection accuracy and real-time performance are critical for practical deployment. Although deep learning–based object detectors, particularly single-stage YOLO-based methods, have achieved promising results in traffic sign detection, their performance remains limited when dealing with visually similar categories, small-sized targets, and complex traffic environments. Moreover, many existing approaches rely on architectural modifications or additional modules, which may increase model complexity and hinder real-time applicability. To address these challenges, this paper proposes a training-strategy-driven traffic sign detection method based on YOLOv11s, termed TS-YOLOv11s. Without modifying the original network architecture or loss function formulation, the proposed method improves detection performance by optimizing the training process. Specifically, loss weight adjustment is employed to enhance fine-grained category discrimination, label smoothing is introduced to mitigate overconfidence caused by limited samples, and data augmentation strategies tailored for small objects and complex scenes are applied to improve robustness and generalization. Extensive experiments conducted on the Chinese Traffic Sign Dataset demonstrate that the proposed method achieves high detection accuracy while maintaining real-time inference efficiency. The results indicate that TS-YOLOv11s provides an effective balance between performance and computational cost, highlighting the potential of training strategy optimization as a general and practical approach for traffic sign detection in real-world intelligent transportation systems.

Keywords: Traffic Sign Detection, YOLOv11s, Training Strategy Optimization, Economic Analysis.

1. Introduction

Traffic sign detection plays a critical role in intelligent transportation systems and advanced driver assistance systems, as it provides essential visual information for vehicle perception, navigation, and driving safety. Accurate and real-time detection of traffic signs is particularly important in complex road environments, where variations in illumination, viewing angles, object scale, and background clutter pose significant challenges. As traffic scenes become increasingly diverse, the demand for robust and efficient traffic sign detection methods continues to grow.

In recent years, deep learning–based object detection approaches have achieved remarkable progress in traffic sign detection tasks. Among them, single-stage detectors represented by the You Only Look Once (YOLO) series have been widely adopted due to their end-to-end architecture and high inference efficiency. Numerous studies have demonstrated that YOLO-based frameworks exhibit strong real-time performance and favorable engineering adaptability in traffic sign detection applications. However, existing research also indicates that these methods still face limitations in detecting small-sized traffic signs, discriminating visually similar categories, and maintaining robustness under complex environmental conditions [1].

To address these challenges, various improvements to the YOLO framework have been proposed. For example, scene-specific optimization strategies have been introduced to enhance detection performance in low-light or nighttime environments. Feature enhancement and contextual modeling techniques have been employed to improve fine-grained category discrimination among visually similar traffic signs. In addition, lightweight model designs have been explored to reduce computational complexity and facilitate deployment on resource-constrained devices. While these approaches have achieved performance gains under specific conditions, they often rely on architectural modifications, additional modules, or increased model complexity, which may negatively affect deployment efficiency and real-time performance [2].

Motivated by these observations, this work explores an alternative perspective for improving traffic sign detection performance without increasing model complexity. Instead of modifying the network architecture, we focus on optimizing the training strategy of a lightweight YOLO-based detector. Specifically, a training-strategy-driven method, termed TS-YOLOv11s, is proposed for Chinese traffic sign detection [3]. By adjusting the loss function weights to enhance fine-grained category discrimination, introducing label smoothing to alleviate overconfidence caused by limited samples, and employing data augmentation strategies tailored for small objects and complex traffic scenes, the proposed method aims to improve detection accuracy while preserving the lightweight and real-time characteristics of the baseline model [4].

The main contributions of this work can be summarized as follows:

A training-strategy-driven optimization framework is proposed for traffic sign detection based on YOLOv11s, without modifying the original network architecture or loss function formulation.

A loss weight adjustment strategy is designed to better balance classification and localization objectives for visually similar traffic sign categories.

Extensive experiments conducted on the Chinese Traffic Sign Dataset demonstrate that the proposed method achieves high detection accuracy and robust performance while maintaining real-time inference speed, highlighting its suitability for practical intelligent transportation systems.

2. Training Strategy Optimization Framework Based on YOLOv11s

2.1. Experimental Dataset and YOLOv11s Runtime Environment for Traffic Sign Detection

Experiments were conducted using the Chinese Traffic Sign Dataset. The dataset contains 58 categories of common Chinese traffic signs, covering speed limit, prohibition, warning, and mandatory signs. It exhibits strong scene representativeness and effectively reflects the category distribution characteristics of traffic sign targets in real road environments. Moreover, the large number of categories and the high visual similarity among certain traffic signs pose significant challenges for fine-grained classification and robust detection. The dataset was divided into a training set and a validation set. The validation set consists of 600 images with a total of 611 annotated instances. All annotations follow the YOLO standard format to ensure consistency between the training and evaluation processes.

All experiments were performed on a computing platform equipped with an NVIDIA GeForce RTX 4060 Laptop GPU. The software environment includes Python 3.8, PyTorch 2.4.1, and the Ultralytics YOLOv11 framework. To ensure the comparability of experimental results, all experiments were conducted under identical hardware conditions and software configurations.

2.2. YOLOv11s Baseline Architecture and Training-Strategy-Driven Framework

YOLOv11s was selected as the baseline model. YOLOv11s is a single-stage object detection model that adopts an end-to-end detection architecture, achieving a favorable balance between detection accuracy and inference efficiency. This makes it suitable for traffic scenario applications with high real-time requirements. The overall network architecture consists of a backbone network, a feature fusion network, and a detection head, enabling joint modeling of multi-scale features for effective detection of objects at different scales.

Based on the baseline model, the proposed TS-YOLOv11s introduces no modifications to the original network architecture. All improvements are realized through training strategies. This design choice aims to preserve the lightweight characteristics and deployment efficiency of YOLOv11s, while demonstrating that effective training strategy optimization alone can substantially enhance detection performance in complex traffic scenarios.

2.3. Loss Weight Optimization Method for Fine-Grained Classification in YOLOv11s

The object detection loss function of YOLOv11s mainly consists of bounding box regression loss, classification loss, and distribution regression loss (Distribution Focal Loss, DFL) [5]. The overall loss function is defined as follows:

$$\mathrm{L} = \lambda_{box}\mathrm{L}_{box} + \lambda_{cls}\mathrm{L}_{cls} + \lambda_{dfl}\mathrm{L}_{dfl} \tag{1}$$

where λ_{box}, λ_{cls}, and λ_{dfl} denote the weight coefficients corresponding to the bounding box regression loss, classification loss, and DFL loss, respectively.

Considering the large number of categories and the high visual similarity among different classes in the Chinese traffic sign dataset, the loss function weights were appropriately adjusted during training. Specifically, the classification loss weight was increased to encourage the model to focus more on discriminating fine-grained categories. Meanwhile, the DFL loss weight was slightly adjusted to prevent overfitting during boundary distribution regression. The final loss weight configuration is set as follows:

$$\lambda_{box} = 7.5, \lambda_{cls} = 0.8, \lambda_{dfl} = 1.2 \tag{2}$$

Without altering the original loss function formulation, this strategy guides the model during training to achieve a more reasonable balance between category discrimination capability and target localization accuracy.

2.4. YOLOv11s Training Strategy Integrating Label Smoothing and Data Augmentation

To alleviate overfitting caused by the relatively small number of samples in certain categories, a label smoothing strategy was introduced during model training. This method converts hard label distributions into soft label distributions, thereby reducing excessive confidence in single-category predictions [6]. For a ground truth label y, the smoothed label $\hat{y}$ is defined as:

$$\hat{y} = (1-\varepsilon)y + \varepsilon / K \tag{3}$$

where ε denotes the smoothing factor and K represents the total number of categories. In the experiments, $\varepsilon = 0.05$ was adopted.

In addition, to enhance adaptability to small objects and complex traffic scenarios, data augmentation strategies such as Mosaic, Mix Up, and Copy-Paste were applied during training. These were combined with random scale resizing (scale = 0.9) to increase sample diversity, thereby improving the robustness and generalization capability of the model across different scenarios.

3. Detection Performance Evaluation and Computational Efficiency Analysis of TS-YOLOv11s

To comprehensively evaluate the performance of the proposed TS-YOLOv11s model on the Chinese traffic sign detection task, experiments were conducted from multiple perspectives, including overall detection performance, class-level detection behavior, model robustness under different confidence thresholds, and computational efficiency. The experimental results are visualized using Precision–Recall curves, F1–Confidence curves, and a normalized confusion matrix.

3.1. Overall Traffic Sign Detection Performance of TS-YOLOv11s

Experiments were carried out on the validation set consisting of 600 images with a total of 611 annotated traffic sign instances. Under an IoU threshold of $\mathrm{IoU} = 0.5$, the proposed TS-YOLOv11s achieved an mAP@0.5 of 0.967, with Precision and Recall reaching 0.977 and 0.966, respectively. These results indicate that the model maintains a high detection accuracy while effectively reducing missed detections, demonstrating strong target recall capability. This performance suggests that the proposed training strategy enables the model to achieve a favorable balance between precision and recall without sacrificing localization accuracy.

Under more stringent evaluation criteria, the model still exhibited stable performance. Specifically, the mAP@0.5:0.95 reached 0.952, suggesting that TS-YOLOv11s can reliably perform accurate target localization and bounding box regression even as the IoU threshold increases. This behavior reflects the robustness of the model in handling variations in object scale and localization precision.

The corresponding Precision–Recall and F1–Confidence curves are illustrated in Figure 1. It can be observed that the Precision–Recall curves remain smooth and concentrated near the upper-right region, while the F1–Confidence curve maintains relatively high values over a wide confidence interval. This indicates that the model exhibits stable detection performance across different confidence thresholds.

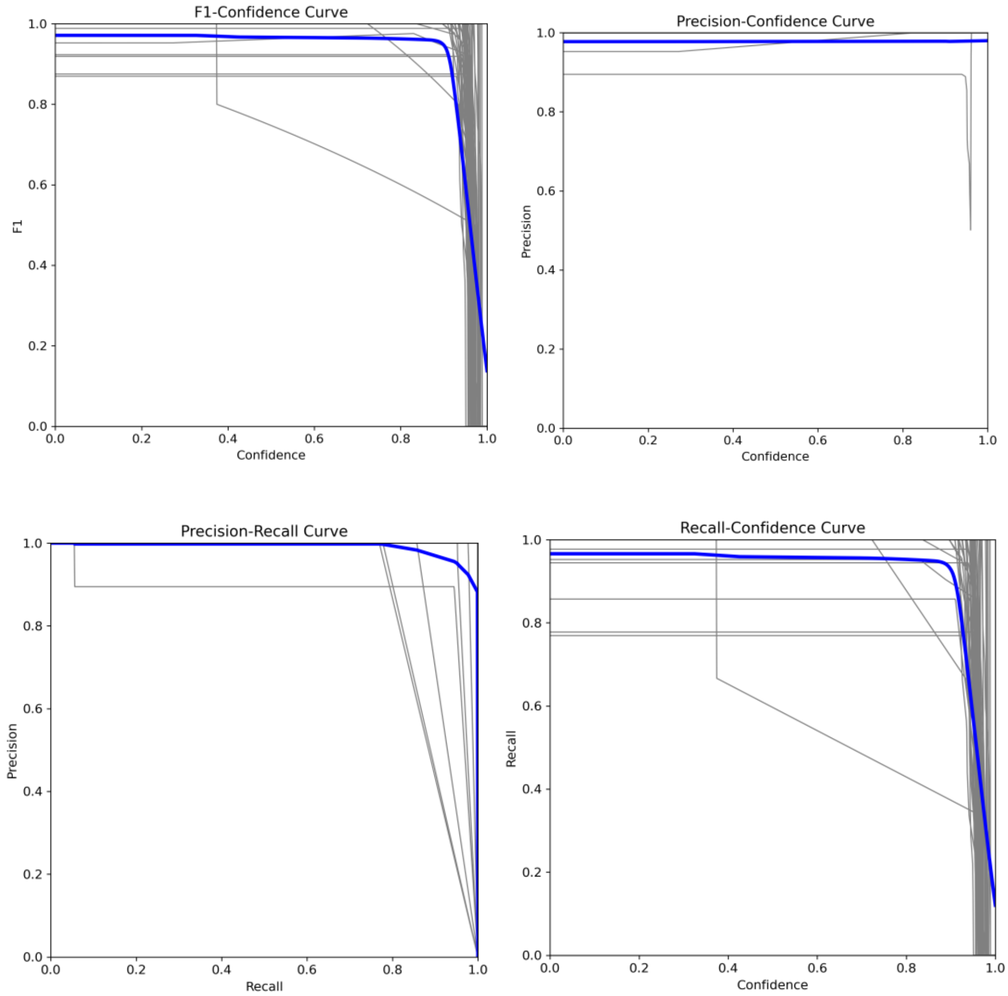

Figure 1. Performance curves of TS-YOLOv11s

3.2. Fine-Grained Class-Level Detection Performance of TS-YOLOv11s

In Figure 2, at the class level, TS-YOLOv11s demonstrates consistently strong detection performance across most traffic sign categories. For speed limit signs (e.g., SpeedLimit5, SpeedLimit15, SpeedLimit30, SpeedLimit40, SpeedLimit50, SpeedLimit60, SpeedLimit70, and SpeedLimit80), both Precision and Recall values are close to 1.0. This indicates that the model is capable of reliably distinguishing traffic signs with highly similar visual structures but different numerical information.

For guidance and warning signs such as Motor Vehicles Permitted, No Parking, Pedestrians Ahead, and Series of Shar p Curves, the model also achieves strong detection results, with most categories exhibiting mAP@0.5 values above 0.97. These results demonstrate that TS-YOLOv11s can effectively extract discriminative features under complex backgrounds and varying object scales.

It is worth noting that a small number of categories with limited samples or highly similar visual characteristics—such as T-Junction (Left), T-Junction (Right), and Stop and Yield—show slight reductions in Precision or Recall. This performance degradation can be attributed to limited training samples and challenging real-world conditions such as viewpoint changes and partial occlusion. Such limitations remain common challenges in real-world traffic sign detection tasks and indicate potential directions for future dataset expansion and training strategy refinement.

Figure 2. Class-wise detection performance

3.3. Category Discrimination Ability and Confusion Matrix Analysis of TS-YOLOv11s

The normalized confusion matrix of the validation results is presented in Figure 3. As shown, most prediction results are concentrated along the main diagonal, indicating that the proposed model has strong discriminative capability across different traffic sign categories. In particular, almost no confusion is observed between speed limit signs and functional traffic signs, further validating the effectiveness of the model in category recognition.

The limited misclassifications mainly occur between semantically related or visually similar categories, such as certain directional signs and warning signs. This phenomenon reflects the inherent visual similarity of some traffic signs in real-world scenarios. However, the proportion of such misclassifications is relatively small and has a negligible impact on the overall detection performance.

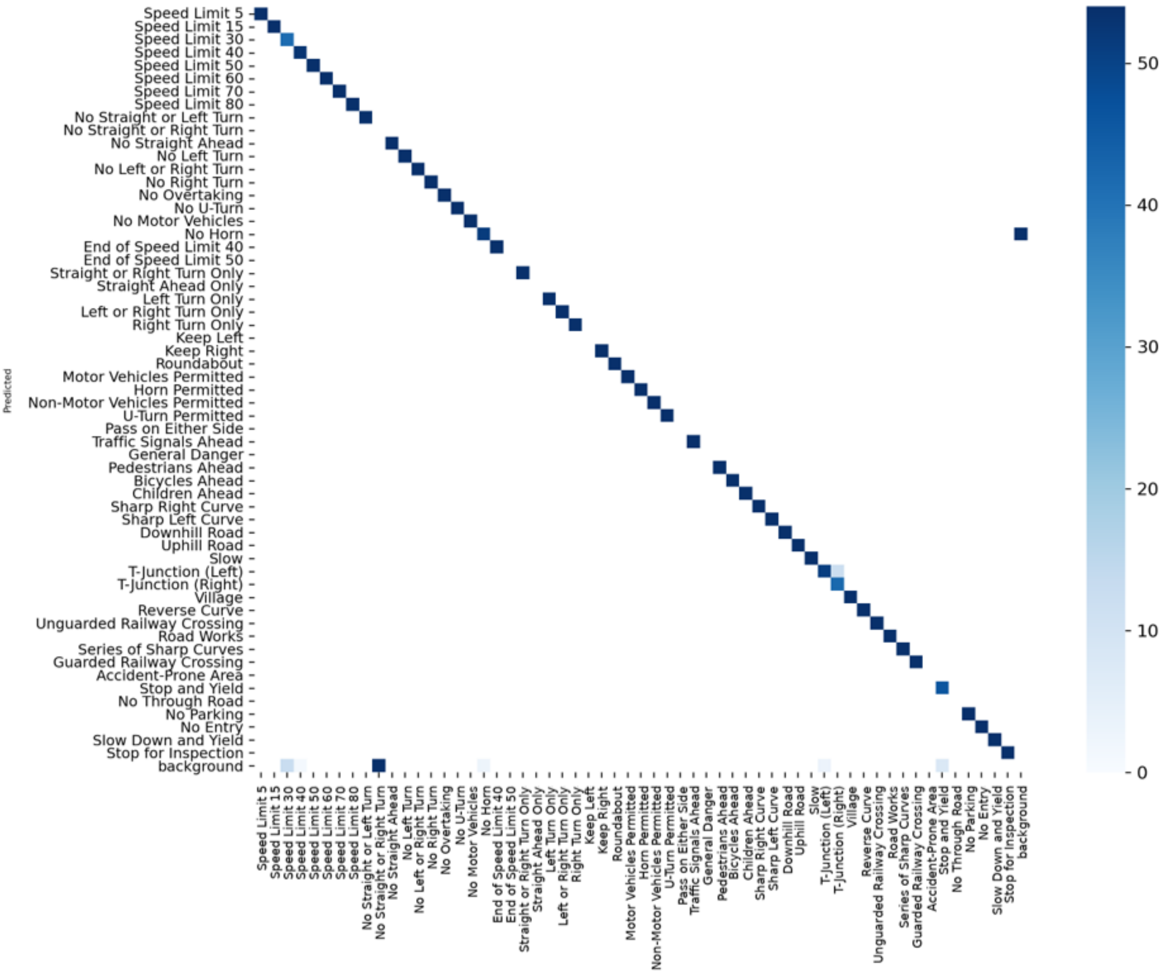

Figure 3. Normalized confusion matrix

3.4. Computational Complexity, Inference Speed, and Practical Applicability of TS-YOLOv11s

In practical intelligent transportation systems, detection accuracy must be balanced with model size, computational complexity, and inference speed to satisfy real-time and deployment requirements. Therefore, the computational efficiency of TS-YOLOv11s is analyzed in terms of parameter count, computational complexity, and inference latency.

Since TS-YOLOv11s does not modify the network architecture of the baseline YOLOv11s and only adjusts training strategies, the model retains the same parameter count of 9.44M and computational complexity of 21.4 GFLOPs, without introducing additional memory or computational overhead. The total number of model parameters can be expressed as:

$$N_p = \sum_{i=1}^{L}(w_i + b_i) \tag{4}$$

where L denotes the number of network layers, and w_i and b_i represent the weights and biases of the i-th layer, respectively.

Experimental evaluation on an NVIDIA GeForce RTX 4060 Laptop GPU shows that TS-YOLOv11s achieves an average inference time of 9.02 ms per image, corresponding to an inference speed of approximately 110 FPS. This performance satisfies the real-time requirements of traffic sign detection in practical driving scenarios, highlighting the suitability of TS-YOLOv11s for deployment in real-time and resource-constrained intelligent transportation systems.

When combined with the detection accuracy results, TS-YOLOv11s achieves an mAP@0.5 of 0.967 without increasing model size or computational complexity, demonstrating a favorable trade-off between performance and computational cost. Overall, the proposed method maintains model lightweight characteristics and high inference efficiency while delivering reliable detection performance, highlighting its practical deployment potential in real-world intelligent transportation systems.

4. Conclusions

This paper presented a training-strategy-driven traffic sign detection method, termed TS-YOLOv11s, to address the challenges of fine-grained category discrimination, small-object detection, and real-time deployment in complex traffic environments. Unlike many existing approaches that rely on architectural modifications or additional modules, the proposed method focuses on optimizing the training process while preserving the original lightweight network structure.

By adjusting loss function weights, introducing label smoothing, and applying data augmentation strategies tailored for small targets and complex scenes, TS-YOLOv11s effectively improves detection performance without increasing model complexity or computational cost. Experimental results on the Chinese Traffic Sign Dataset demonstrate that the proposed method achieves high detection accuracy and robust performance while maintaining real-time inference efficiency, indicating a favorable balance between accuracy and efficiency.

Overall, the results highlight that training strategy optimization can serve as an effective and practical alternative to structural model modification for traffic sign detection tasks. This approach is particularly suitable for real-world intelligent transportation systems with strict requirements on computational resources and response speed. Future work will explore extending the proposed training strategy to other object detection tasks and datasets, as well as further enhancing robustness under more challenging environmental conditions.

References

[1] Fu Rong, Lu Yang, Peng Miao. Improvement of the YOLOv5s Model and Its Application in Traffic Sign Detection [J]. Remote Sensing Information, 2024, 39(06): 87-93. DOI:10.20091/j.cnki.1000-3177.2024.06.011.

[2] Zhao Limin. Research and Application of Urban Road Traffic Sign Detection Based on YOLOv8 [D]. North China Electric Power University, 2024. DOI: 10.27139/d.cnki.ghbdu.2024.000716.

[3] Guo Junqiang, Yang Xiaoxia. An Improved Chinese Text Detection Algorithm for Natural Scene Images Based on YOLOv11s [J/OL]. Intelligent Computer and Applications, 1-8 [2026-02-02]. https://doi.org/10.20169/j.issn.2095-2163.25120701.

[4] Guo Jialin, Cao Yunfeng. Research on an Improved YOLOv11s Method for Detecting Small Aerial Targets [J/OL]. Computer Engineering and Applications, 1-18 [2026-02-02]. https://link.cnki.net/urlid/11.2127.tp.20251213.1324.004.

[5] Qian Wei, Yang Xiao, Liu Quanyi, et al. YOLOv8 Flame Object Detection Algorithm with Attention Mechanism Integration [J]. Journal of Safety and Environment, 2025, 25(01): 75-84. DOI: 10.13637/j.issn.1009-6094.2024.0533.

[6] Wang Meixia. Research on AI Translation Based on Syntax Awareness and Adaptive Label Smoothing [J]. Automation and Instrumentation, 2025, (04): 155-158+163. DOI: 10.14016/j.cnki.1001-9227.2025.04.155.