Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

LRP-RTDETR: A Real-Time High-Precision Traffic Sign Detection Method

Ruodeng Zhao, Shujun Shen, Jie Zhao

How to cite: Zhao R, Shen S, Zhao J. LRP-RTDETR: A Real-Time High-Precision Traffic Sign Detection Method. Textile & Leather Review. 2026; 9:2484-2493. https://doi.org/10.31881/TLR.2026.2484

How to link: https://doi.org/10.31881/TLR.2026.2484

Published: 25 April 2026

LRP-RTDETR: A Real-Time High-Precision Traffic Sign Detection Method

Ruodeng Zhao*, Shujun Shen, Jie Zhao
Faculty of Intelligence Technology, Shanghai Institute of Technology, Shanghai 201418, China
*zhaord2023@163.com

Article
https://doi.org/10.31881/TLR.2026.2484
Published 25 April 2026

ABSTRACT

To address the challenges of small-scale traffic sign detection and constrained computational resources, we propose LRP-RTDETR, an edge-friendly efficient high-precision framework based on RT-DETR. We integrate Large Separable Kernel Attention (LSKA) to broaden the receptive field for fine-grained feature capture. The backbone and neck are re-architected using a GELAN-based framework with RepNCSPELAN4 and ADown modules, reducing structural redundancy while enhancing feature aggregation. Furthermore, an Inner-PIoU v2 loss function, utilizing a non-monotonic attention mechanism, is introduced to mitigate gradient stagnation and improve localization precision. Experimental results on the TT100K dataset show that LRP-RTDETR outperforms the RT-DETR-R18 baseline with a 2.6% increase in mAP@0.5, a 53.4% (30.6 GFLOPs) reduction in computational complexity, and a 12 FPS improvement in inference speed, demonstrating a favorable accuracy–efficiency trade-off for resource-constrained intelligent transportation systems.

KEYWORDS
autonomous driving, traffic sign, LRP-RTDETR, lightweight model

INTRODUCTION

Traffic sign detection (TSD) [1] plays a vital role in autonomous driving perception by providing essential cues for safe navigation and decision-making. In resource-constrained intelligent transportation systems, vehicles and infrastructure are interconnected through V2X communication, forming distributed perception networks where onboard vision modules must operate in real time.

Unlike cloud-based solutions, autonomous vehicles primarily rely on embedded edge computing platforms to process high-resolution video streams locally. Given constraints in computation, memory, and power consumption, detection models must balance accuracy and efficiency. Therefore, lightweight yet high-precision TSD frameworks are critical for stable real-time perception in vehicle–road collaborative environments.

While one-stage detectors, particularly the YOLO series [2], have dominated object detection owing to their streamlined architectures, maintaining both high accuracy and real-time performance under edge constraints remains challenging. Recent enhancements incorporating attention mechanisms [3,4] and lightweight operators [5] have improved feature representation; however, CNN-based models often struggle with multi-scale feature extraction or incur excessive computational overhead. Furthermore, the reliance on Non-Maximum Suppression (NMS) introduces additional post-processing latency, which is undesirable in latency-critical autonomous driving scenarios.

Recently, the Real-Time Detection Transformer (RT-DETR) [6] has emerged as an end-to-end alternative that eliminates NMS and achieves superior efficiency–accuracy trade-offs. Nevertheless, robust traffic sign detection in autonomous driving environments remains challenging due to environmental variability, small-scale targets, and extreme scale differences [7]. To address these limitations, this work targets the performance degradation of real-time traffic sign detection models, particularly under extreme scale variations. Accordingly, we propose LRP-RTDETR, an edge-friendly framework optimized for real-time traffic sign recognition through structural streamlining and refined localization design. The proposed framework incorporates three core architectural innovations:

LSKA-driven Feature Refinement: Utilizing LSKA [8] to broaden the receptive field and enhance the capture of fine-grained sign features in complex backgrounds.

Efficiency-oriented Backbone: Integrating a GELAN-based structure with RepNCSPELAN4 blocks and ADown modules to optimize gradient flow and reduce computational redundancy [9].

Inner-PIoU v2 Localization [10,11]: Implementing an enhanced loss function with a non-monotonic attention mechanism to improve the regression precision of small-scale targets, see Figure 1.

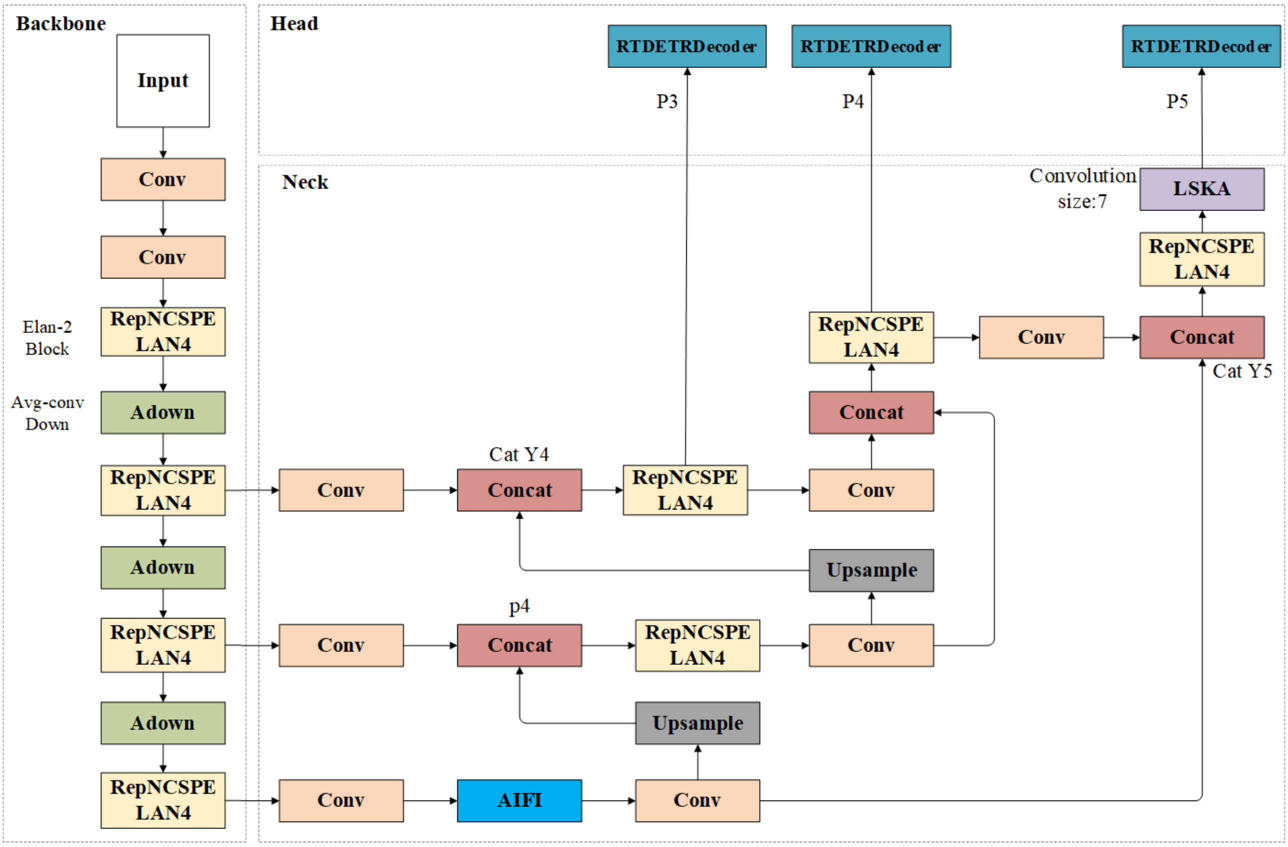

Figure 1. Network Architecture Diagram of LRP-RTDETR

METHODS

The overall architecture of LRP-RTDETR is illustrated in Fig. 1. The framework follows the end-to-end paradigm of RT-DETR and consists of three main components: a feature extraction backbone, a feature aggregation neck, and a transformer-based detection head. To enhance detection performance under real-time constraints, several targeted structural refinements are introduced, including LSKA-driven feature enhancement, a GELAN-based lightweight backbone design, and the Inner-PIoU v2 localization strategy.

LSKA Attention Mechanism

To enhance feature discrimination in cluttered traffic scenes, we incorporate the Large Separable Kernel Attention (LSKA) module with convolution kernel size K=7. By modeling long-range spatial dependencies with a large receptive field, LSKA strengthens contextual feature representation, which is particularly beneficial for small-scale traffic sign perception in autonomous driving scenarios.

Unlike conventional Large Kernel Attention (LKA), which approximates large receptive fields through depth-wise convolution and dilation, LSKA decomposes two-dimensional depthwise and dilated kernels into cascaded one-dimensional horizontal and vertical convolutions. This separable design reduces computational complexity from quadratic to linear with respect to the kernel size K, thereby improving efficiency without compromising contextual modeling capability.

Ultimately, LSKA strikes a deliberate balance: it prioritizes expanded contextual modeling for accuracy-critical detection, while its separable structure provides potential efficiency benefits that mitigate, though not eliminate, the overhead of large-kernel operations.

GELAN

To enhance feature aggregation without the prohibitive overhead of self-attention, we re-architect the backbone using the Generalized Efficient Layer Aggregation Network (GELAN) framework. This design optimizes the gradient path while maintaining a compact parameter footprint. This lightweight aggregation strategy reduces redundant computation and shortens gradient propagation paths, which is particularly beneficial for real-time detection with limited computational budgets.

RepNCSPELAN4

The RepNCSPELAN4 block is adopted as the core computational unit of the backbone to achieve a balance between representational capacity and computational efficiency. Through structural reparameterization, the module maintains multi-branch feature diversity during training and is reconfigured into a compact single-branch structure during inference. This design enhances feature expressiveness while reducing inference complexity, supporting efficient real-time detection in autonomous driving scenarios.

ADown Downsampling

Conventional downsampling operations may degrade fine-grained spatial information, limiting performance on small-scale traffic signs. To address this issue without increasing computational burden, the standard downsampling layers are replaced with the ADown module. ADown employs a dual-branch structure, where one branch performs spatial reduction via average pooling and the other applies shared-weight convolutions for enhanced feature encoding. This parallel design reduces feature aliasing and parameter redundancy, thereby improving small-object representation under real-time processing constraints.

Inner-PioU v2

Conventional IoU-based losses often suffer from gradient stagnation when handling traffic signs with extreme scale variations. When dealing with small objects, the spatial alignment between predictions and reference boxes is often minimal, resulting in weak gradient signals and ineffective bounding box regression. Such scale-sensitive instability becomes particularly critical in autonomous driving perception, where detection modules must maintain reliable localization performance across diverse object scales. To alleviate this issue, we develop Inner-PIoU v2, formulated with two cooperative mechanisms.

Scale-Adaptive Regression

By introducing an auxiliary bounding box governed by a scale factor $ratio \in [0.5, 1.5]$, Inner-IoU enables IoU computation within a transformed spatial region. For small-scale signs, setting $ratio < 1$ (as shown in (1)-(2)) constructs a compact inner overlap region. This strategy amplifies gradient responses when the original IoU approaches zero, alleviating early-stage regression stagnation and improving sensitivity to spatially distant targets.

$$w^{aux} = w \bullet ratio, h^{aux} = h \bullet ratio \tag{1}$$

$$IoU^{inner} = \frac{inter^{aux}}{union^{aux}} \tag{2}$$

Dynamic Gradient Modulation

To stabilize optimization, PIoU v2 introduces a non-monotonic attention mechanism to adaptively reweights gradient contributions according to anchor quality. The non-monotonic attention function $u(\lambda q)$, formulated in (3), leverages the hyperparameter λ to adaptively modulate boundary alignment penalty intensity based on anchor quality. Unlike monotonic strategies, this design emphasizes medium-quality anchors—where localization refinement is most effective—while suppressing low-quality outliers. This mitigates excessive penalties on near-optimal predictions and ensures stable convergence.

$$L_{PIoU\_v2} = u(\lambda q) \cdot PIoU = 3 \bullet (\lambda q) \bullet e^{-(\lambda q)^2} \bullet L_{PIoU} \tag{3}$$

Overall, Inner-PIoU v2 enhances localization robustness against occlusion and scale variations without introducing additional inference complexity. This property supports efficient real-time detection. The final loss is formulated as:

$$L_{Inner-PIoU\_v2} = L_{PIoU\_v2} + IoU - IoU^{inner} \tag{4}$$

The subtraction operation, defined as IoU − IoU_{inner} , is intentionally adopted instead of addition or division for the following reasons. When the predicted box and the ground-truth box exhibit a high degree of overlap, the inner IoU (IoU_{inner}), computed from scaled inner boxes (ratio < 1), is generally smaller than the standard IoU due to the reduced intersection area caused by box shrinkage. This characteristic allows the metric to provide a more sensitive measure for fine-grained localization. By subtracting IoU_{inner} from IoU, an additional positive penalty is introduced, which produces stronger gradients in the high-IoU regime and encourages more precise bounding box regression. In contrast, directly adding IoU_{inner} may lead to excessive penalization even for well-aligned boxes, potentially affecting training stability. Meanwhile, a division-based formulation lacks clear geometric interpretability and may result in unstable gradient behavior during optimization.

EXPERIMENTAL RESULTS AND ANALYSIS

Dataset and Settings

Experiments are conducted on the TT100K dataset, a widely used benchmark for Chinese traffic sign detection. All experiments are implemented in PyTorch with CUDA acceleration. The models are trained with a batch size of 16 for 500 epochs using the AdamW optimizer, and early stopping is applied based on validation performance. During evaluation, the confidence and IoU thresholds are set to 0.25 and 0.35, respectively.

Comparative Results and Analysis

According to Table 1, LRP-RTDETR outperforms RT-DETR-R18, attaining mAP@0.5 and mAP@0.5–0.95 gains of 2.6 and 1.9 percentage points, respectively. Crucially, these accuracy gains are accompanied by a 30.6 GFLOPs (53.4%) reduction in computational complexity and a 12 FPS improvement in inference speed. Compared with mainstream CNN-based and Transformer-based detectors, LRP-RTDETR establishes a superior accuracy-efficiency trade-off. While prior efficient detectors often compromise accuracy for speed, suitable for real-time applications with limited computational budgets.

Table 1. Benchmarking Results on the TT100K Dataset

model	P(%)	R(%)	mAP@0.5	mAP@0.5-0.95	GFLOPs(G)	FPS
RT-DETR-R18	86.9	77.6	81.4	62.5	57.3	73
YOLOv5s	85.7	77.5	85.0	65.3	23.9	65
YOLOv7	77.9	65.5	68.2	52.0	103.9	74
YOLOv8s	86.9	75.1	84.7	64.9	28.5	63
YOLOv9s	87.7	75.6	84.9	65.6	26.8	27
Zhang et al. [12]	88.0	82.5	84.2	64.5	53.4	65
LRP-RTDETR	88.1	82.3	84.0	64.4	26.7	89

Fig. 2 presents qualitative detection results on the test set. Compared with RT-DETR-R18, the proposed LRP-RTDETR exhibits fewer false positives and missed detections when multiple traffic signs appear in the same scene, and demonstrates improved reliability in small-object detection. These qualitative observations are consistent with the quantitative improvements reported above, see Figure 2.

(a) (b) (c)

Figure 2. Detection results of different models on traffic signs: Original image; (b) detection results of the baseline RT-DETR-R18; (c
)

detection results of the proposed LRP-RTDETR

Ablation Studies

Table 2 reports a component-wise evaluation of the proposed model. Incorporating the of LSKA yields consistent accuracy improvements of 2.1% in mAP@0.5 and 1.6% in mAP@0.5-0.95, with only a marginal decrease in inference speed (1 FPS). This reflects a deliberate trade-off between accuracy and efficiency to enhance the model's feature expressiveness for complex traffic scenes. This reflects a deliberate accuracy-efficiency trade-off for enhanced feature expressiveness. The GELAN architecture primarily contributes to computational efficiency without degrading detection performance, establishing the foundation of our efficiency gains. Furthermore, adopting Inner-PIoU v2 provides additional accuracy improvements of 1.0% in mAP@0.5 and 0.8% in mAP@0.5-0.95 without without increasing model complexity, confirming its suitability for small-scale traffic sign localization. Collectively, these components achieve a favorable balance: LSKA and Inner-PIoU v2 prioritize precision, while GELAN ensures real-time feasibility.

Table 2. Ablation study of the proposed components on the TT100K dataset.

LSKA	ADown	RepN	Inner-PIoU v2	mAP@0.5	mAP@0.5-0.95	GFLOPs(G)	FPS
				81.4	62.5	57.3	77
√				83.5	64.1	57.2	78
	√			82.0	63.2	55.6	78
		√		81.6	62.6	50.3	82
	√	√		82.3	62.9	26.8	90
			√	82.4	63.3	57.3	77
√	√	√	√	84.0	64.4	26.7	89

CONCLUSION

This paper presents LRP-RTDETR, an edge-friendly framework designed to mitigate the performance degradation of real-time traffic sign detection models under resource-constrained and extreme scale variations. By systematically optimizing the backbone and localization strategy, the proposed model achieves substantial performance improvements over the standard RT-DETR across multiple metrics. Experimental results demonstrate that our architectural refinements not only enhance the perception of small-scale targets but also streamline the model structure, making it suitable for real-time processing with limited computational budgets. Future research will explore the model's robustness under adverse environmental conditions to further support autonomous driving safety.

Author Contributions

Ruodeng Zhao: Methodology, Conceptualization, Writing – original draft & review & editing, Formal analysis , Validation, Project administration, Software

Shujun Shen: Investigation, Resources, Validation, Visualization, Writing – original draft

Jie Zhao: Data curation, Software, Writing – review & editing

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Funding

This research received no external funding.

Acknowledgements

Not applicable.

REFERENCES

[1] Flores-Calero M, Astudillo CA, Guevara D, Maza J, Lita BS, Defaz B, Ante JS, Zabala-Blanco D, Armingol Moreno JM. Traffic sign detection and recognition using YOLO object detection algorithm: A systematic review. Mathematics. 2024; 12(2):297. doi: 10.3390/math12020297

[2] Han X, Chang J, Wang K. You only look once: Unified, real-time object detection. Procedia Computer Science. 2021; 183(1): 61-72. doi: 10.1016/j.procs.2021.02.031

[3] Guo F, Zhang Y, Tang J, Li W. YOLOv3-A: A traffic sign detection network based on attention mechanism. Journal on Communications. 2021; 42(1):87-99. doi: 10.11959/j.issn.1000-436x.2021031

[4] Yin S, Tan F. YOLOv4-A: Research on Traffic Sign Detection Based on Hybrid Attention Mechanism. Journal of Computers. 2022; 33(6): 181-192. doi: 10.53106/199115992022123306015

[5] Zhang R, Zheng K, Shi P, Mei Y, Li H, Qiu T. Traffic Sign Detection Based on the Improved YOLOv5. Applied Sciences. 2023; 13(17):9748. doi: 10.3390/app13179748

[6] Guo JY, Han K, He W, Liu CJ, Nie Y, Wang CC, Wang YH. Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism. Advances in Neural Information Processing Systems 36. Neural Information Processing Systems Foundation, Inc. (NeurIPS), New Orleans, Louisiana, USA. 2024: 51094-51112. doi: 10.52202/075280-2224

[7] Joselin J, Selvaganapathy S, Barathraj S, Akilan S. AI Applications in Autonomous Vehicles: Challenges in Computer Vision. AI Applications. 2025; 12(2). Available from: https://ijarasem.com/admin/img/13_AI Applications.pdf

[8] Guo MH, Lu CZ, Liu ZN, Cheng MM, Hu SM. Visual attention network. Computational Visual Media. 2023; 9(4):733-752. doi: 10.1007/s41095-023-0364-2

[9] Wang CY, Yeh IH, Liao HYM. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. Computer Vision – ECCV 2024. 2024; 15089:1-20. doi: 10.1007/978-3-031-72751-1_1

[10] Liu C, Wang K, Li Q, Zhao F, Zhao K, Ma H. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism. Neural Networks. 2024; 170:276-284. doi: 10.1016/j.neunet.2023.11.041

[11] Zhang H, Xu C, Zhang S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box. arXiv preprint. 2023; arXiv:2311.02877. Available from: https://arxiv.org/pdf/2311.02877

[12] Qin LM, Zhang YQ, Cui HY, Bian HQ, Wang X. Traffic sign detection method under extreme weather based on improved RT-DETR. Electronic Measurement Technology. 2025; 48(9):56-64. Available from: http://kns.cnki.net/kcms/detail/11.2175.tn.20250326.0931.002.html