Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.AST26571

Technological Evolution of Traffic Sign Recognition

Xiaorui Chen

Institute of Mechanical Engineering, Southwest Jiaotong University, Chengdu, China
yrbcxcn@163.com

Abstract. This paper aims to introduce deep learning approaches developed to overcome challenges in traffic sign recognition, which include YOLOv8-TS, STD-YOLOv5s, Semantic Scene Understanding & Structured Localization, and Cascade R-CNN with Multi-Scale Attention. The paper also delves into critical system architecture considerations, analyzing the roles of various sensors and computing platforms in enabling real-time and efficient TSR. Performance evaluation is discussed using essential metrics such as Precision, Recall, F1-score, and mean Average Precision (mAP), highlighting their importance in quantifying detection and classification accuracy. Optimization strategies are emphasized as crucial for practical deployment, relying heavily on continuous algorithmic advances (like novel network architectures and attention modules) and sophisticated hardware-software co-design to balance computational efficiency with recognition performance. Looking ahead, TSR research is poised to advance significantly in key directions, including multi-modal data fusion for enhanced environmental perception, extreme lightweighting and real-time optimization for edge deployment, achieving cross-regional and cross-language generality, developing robust adversarial defense and safety mechanisms, improving model interpretability, and fostering deeper integration with driving decision systems. These advancements are vital for enhancing the reliability and safety of autonomous vehicles operating in complex, real-world driving environments.

Keywords: deep learning, traffic sign recognition, real-time detection

1. Introduction

Traffic sign recognition, as a key component of intelligent transportation systems and autonomous driving technology, is increasingly gaining attention from researchers and the industrial community. In recent years, China has positioned intelligent driving and smart transportation as core development priorities at the national strategic level, promoting technological innovation and industrial implementation through a systematic policy framework. The year 2025 is becoming a critical milestone for China's intelligent driving development, with central and local policies forming a multi-level support system for project development. The core objective of this project is to achieve efficient and accurate recognition of road signs, providing crucial information for vehicle navigation, driving decisions, and safety assurance [1]. However, with the increasing complexity of the traffic environment and the continuous improvement of vehicle automation, the technical requirements for traffic sign recognition are constantly rising. Currently, under the dual drivers of

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.AST26571

policy and technology, China's intelligent driving in 2025 is undergoing a key transition from technical testing to commercial operation [2].Developing traffic sign recognition technology plays a positive role in enhancing road driving safety and reducing traffic accident rates, among other aspects.

2. Traffic sign recognition challenges

Traffic sign recognition systems face multiple challenges in practical applications. The first is the complexity of the natural environment. Different climatic conditions, such as snow and ice covering signs, overexposure caused by intense lighting, or low contrast under dim light, can severely distort or obscure key visual information of signs [3].Secondly, dynamic road scenarios present significant difficulties. Image motion blur caused by high-speed vehicle movement and bumps, as well as perspective distortion and partial occlusion due to changing viewpoints, not only require strong robustness from the recognition algorithm but also place extremely high demands on the system's real-time processing capabilities. Thirdly, the variability in the sign's own state cannot be ignored. Fading, wear, damage, or coverage by foreign objects due to long-term outdoor exposure, coupled with the appearance of new signs or the obsolescence and modification of old signs triggered by traffic rule updates, requires the system to possess continuous learning and updating capabilities [4].Finally, complex road environments themselves introduce additional interference. Densely distributed or even overlapping signs can easily cause confusion or omission. All these require algorithms to possess stronger contextual understanding and anti-interference characteristics.

3. Technical evolution of traffic sign recognition

3.1. YOLOv8-TS traffic sign detection integrating attention mechanism

Road traffic sign detection is critical in autonomous driving and vehicle networking systems, demanding high real-time performance and accuracy. Huang Zhiyuan et al. proposed a lightweight YOLOv8-TS traffic sign detection network based on YOLOv8s, improving detection performance by integrating attention mechanisms and other technologies.

The specific method is shown in Figure 1. The authors performed an overall lightweight design on the original YOLOv8s, removing redundant detection heads and streamlining the backbone network. They proposed Conv-G7S and CSP-G7S large-kernel convolution modules to reduce parameter count; designed a CSP-SwinTransformer block, introducing window self-attention into the network to enhance context modeling capability; integrated the CBAM (Convolutional Block Attention Module) attention module into the neck network to strengthen the weighting learning of channel and spatial features; and simultaneously improved the loss function, adopting a dynamic non-monotonic bounding box loss (WIoU) to enhance regression performance.

DOI: 10.54254/2755-2721/2025.AST26571

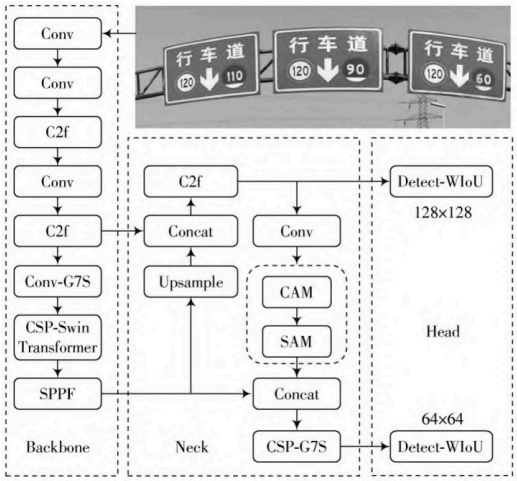

Figure 1. YOLOv8-TS network schematic

The research used YOLOv8s as the baseline model and improved it through an overall lightweight strategy: first, the Detect3 detection head was removed, and the backbone network was streamlined to form a lightweight baseline. Experiments showed that retaining only the Detect1 head could significantly reduce model volume and parameter count while improving detection speed, while basically maintaining accuracy.

Building upon this, to enhance the recognition capability for small targets like traffic signs, novel large-kernel convolution modules Conv-G7S and CSP-G7S were introduced. Conv-G7S expands the receptive field by increasing the convolution kernel size to capture more details; CSP-G7S integrates large-kernel convolution within a cross-stage structure. Comparative experiments showed that adding large-kernel convolution improved mAP@0.5 by 1.6%, validating the effectiveness of the large receptive field. Furthermore, the CBAM attention module was integrated into the neck network. CBAM adaptively weights features through sequential channel and spatial attention mechanisms. Its lightweight design, after integration, increased the model's average precision by approximately 1.2%, effectively strengthening the importance of key channels and spatial regions. Finally, the bounding box loss function was replaced with the dynamic non-monotonic WIoU loss. This loss specifically increases the loss for low-quality prediction boxes and reduces it for high-quality boxes through the product of an exponential term and a modulation term, thereby optimizing the regression effect for small target boxes.

As shown in Figure 2, the authors streamlined YOLOv8s by removing the redundant Detect3 head, reducing parameters to 2.736M with only a 1.3% mAP@0.5 drop and increasing speed to 147.06 fps. Subsequent enhancements through Conv-G7S, CSP-G7S, CSP-SwinTransformer, and CBAM attention yielded the final YOLOv8-TS model. Compared to YOLOv8s, it achieved 96.5% mAP@0.5 with 75.4% fewer parameters at 136.99 fps, while Precision and Recall increased by 6.9% and 3.7% respectively. Comparative experiments confirmed YOLOv8-TS leads in Recall and mAP@0.5 among lightweight models, balancing speed and accuracy effectively.

DOI: 10.54254/2755-2721/2025.AST26571

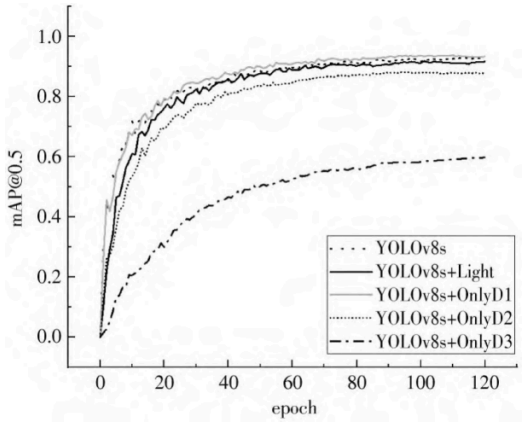

Figure 2. Lightweighting experiment comparison curve [5]

3.2. Application of improved YOLOv5s small target detection algorithm based on multi-attention mechanism in traffic sign recognition

Literature proposed a detection method based on Histogram of Oriented Gradients (HOG), enhancing feature vectors by incorporating color information in CIELab and YCbCr color spaces, and using template images and Support Vector Machines trained with HOG for classification.

The authors designed and improved the STD-YOLOv5s model. Small targets in traffic signs occupy minimal pixel areas in images. The STD-YOLOv5s model made a key enhancement to the YOLOv5s Neck network layer by adding an upsampling operation to capture richer and finer-grained feature information. Increasing the number of feature layers and prediction output layers enables the network to acquire more abundant positional information, which is crucial for small target detection. This solves the problem of insufficient information when the original network processes small targets, significantly enhancing the model's global understanding of images and detection capability for tiny targets.

The STD-YOLOv5s model employs a composite of two attention mechanisms. First, a Coordinate Attention (CA) mechanism is added after each C3 module in the Neck network. The CA mechanism captures precise positional information by extracting features from the input image along the vertical and horizontal directions separately. Subsequently, the feature maps from these two directions are concatenated and normalized to fuse the spatial information. Then, 1x1 convolution operations generate attention weights in the horizontal and vertical directions, respectively, and these weights are applied to the attention calculation of the original feature map. To overcome the limitations of CIoU, the STD-YOLOv5s model replaces the original CIoU loss function with SIoU (Scale-Invariant IoU). It redefines the distance loss and simultaneously considers angle loss, enabling a more comprehensive and precise evaluation of the object's position, size, and orientation.

DOI: 10.54254/2755-2721/2025.AST26571

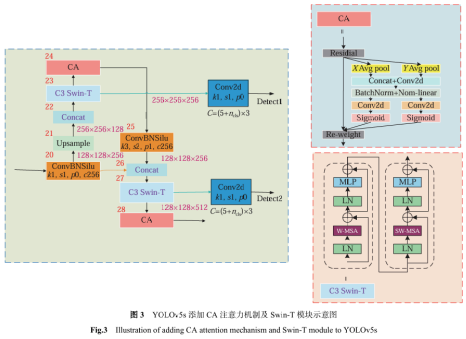

图 3 YOLOv5s 添加 CA 注意力机制及 Swin-T 模块示意图

Fig.3 Illustration of adding CA attention mechanism and Swin-T module to YOLOv5s

Figure 3. Illustration of adding CA attention mechanism and Swin-T module to YOLOv5s

Ablation experiments showed that these improvements effectively enhanced model performance, achieving significant progress particularly in recall rate and average precision. Overall, the STD-YOLOv5s model improved the precision, recall rate, and mAP by 0.77%, 3.21%, and 2.28% respectively, on core metrics, outperforming the original model. It enhanced the detection and localization accuracy for small traffic sign targets and effectively reduced false detections and missed detections. Therefore, this model significantly enhances the recognition capability for small traffic sign targets in complex scenarios such as long distances, occlusion, and poor lighting, possessing deployment potential.

Looking ahead, further research on traffic sign recognition algorithms will focus on extreme model lightweighting, higher accuracy, and stronger portability to fully meet the requirements [6].

3.3. Traffic sign recognition based on semantic scene understanding and structured traffic sign localization

This paper proposed a novel TSR method that introduces a deep learning model to establish credible search regions. This model effectively reduces false positives by leveraging contextual scene information, significantly enhancing the reliability and accuracy of intelligent vehicle TSR systems.

Traditional traffic sign recognition methods face limitations: color-based approaches struggle with lighting variations, while shape-based methods fail with background clutter or deformations. Although CNN-based methods advanced the field, they remain challenged by small-target insensitivity and false-target confusion in complex scenes. This research prioritizes practical robustness over theoretical accuracy by employing an improved Light-weight RefineNet for efficient semantic understanding in resource-constrained scenarios.

The paper first performs initial extraction of road boundaries, involving image binarization, contour detection, interference removal, coordinate definition, and re-fitting. Occlusion handling is also performed. Then, a traffic sign scene structure model is constructed, moving from image-based detection to perception, incorporating spatial reasoning. The article applies MDCOD, a novel network structure that combines the regression idea from SSD (Single Shot MultiBox Detector) and the anchor box mechanism to improve small target detection accuracy. Finally, the improved K-means++ algorithm segments the dataset into small, medium, and large objects before clustering and then optimizing anchor boxes, which enabled faster convergence and higher accuracy.

DOI: 10.54254/2755-2721/2025.AST26571

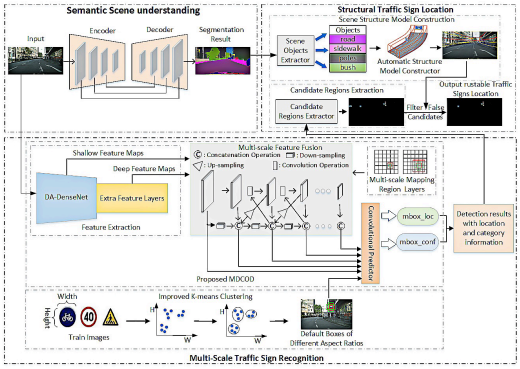

Figure 4. Traffic sign detection method framework composed of semantic scene understanding, structured traffic sign localization, and multi-scale traffic sign recognition

In practical testing, the MDCOD network initially detected 4253 signs, of which 312 similar objects were mistakenly identified as traffic signs. After filtering by the scene structure model, only 26 similar objects remained, while real traffic signs were not filtered out. Combining spatial positional relationships and semantic scene understanding can significantly enhance the robustness of TSR, surpassing what purely feature-based detectors can achieve.

In summary, the research successfully proposes a novel traffic sign recognition method that significantly improves accuracy and robustness in complex real-world driving environments, effectively addressing the key challenges of false positives and small target recognition. It's innovative components [7].

3.4. Cascade R-CNN traffic sign detection based on multi-scale attention and imbalanced samples

The paper proposes a Cascade R-CNN architecture, significantly enhancing robustness through multi-scale attention mechanisms and leveraging hard negative sample mining and data augmentation.

The core idea of Cascade R-CNN is that each stage jointly trains on the bounding box outputs of the previous stage, allowing the model to progressively improve localization accuracy, highlighting the trend towards multi-stage refinement in high-precision tasks. Multi-scale feature fusion integrates high-level semantic information with low-level spatial information, addressing the challenge of small target detection.

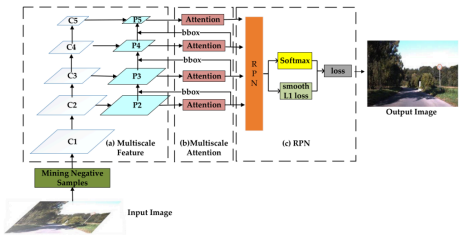

Figure 5. Overview of the method: cascade R-CNN with multi-scale attention and imbalanced samples

The multi-scale attention mechanism is specifically designed to improve detection accuracy in complex environments and enhance the ability to distinguish hard negative samples that are similar

DOI: 10.54254/2755-2721/2025.AST26571

to real traffic signs. Hard negative sample mining is crucial for addressing the imbalance problem in training data, particularly concerning hard-to-detect negative samples, which significantly impact model accuracy and cause false detections. The proposed method proactively mines and increases the number of hard negative samples, forcing the model to learn more robust discriminative features. Without these samples, the model might become overconfident in its classification or biased towards certain types of false positives. Therefore, this method implicitly aims to reduce model bias and improve its generalization to ambiguous scenarios in the real world, a key aspect of trustworthy AI systems.

Next is the data augmentation strategy. A comprehensive data augmentation strategy was adopted to further enhance the model's robustness. Deep learning networks are sensitive to various image corruptions and disturbances prevalent in real-world traffic scenes. Data augmentation aims to simulate these conditions during training, thereby making the model more resilient and robust. This strategy involves generating corrupted images using 15 different algorithms, categorized into three main corruption types: noise corruption, blur corruption, and weather corruption, enabling the model to cope with more complex situations.

On the GTSDB dataset, the model achieved 98.7% accuracy and 90.5% recall, with data augmentation boosting recall from 85.0%. On the Lisa dataset, components improved accuracy by 1.8% and recall by 0.9% without augmentation. Hard negative mining and feature enhancements notably reduced false detections against visually similar signs [8].

4. System architecture and hardware implementation

The performance optimization of traffic sign recognition systems relies on the co-design of hardware and software. In autonomous driving systems, the choice of hardware architecture directly affects the real-time performance, computational efficiency, and robustness of recognition. Typically, such a system includes image sensors, computational units, and auxiliary sensors to provide precise sign detection and classification capabilities. Firstly, traffic sign recognition primarily relies on cameras to acquire road sign information. High-resolution image sensors or High Dynamic Range cameras can provide more stable image input under complex lighting conditions and enhance visibility at night or in harsh weather [9].Secondly, multi-modal perception systems combining LiDAR or millimeter-wave radar help improve recognition capability under occlusion. For example, Tesla's autonomous driving system primarily relies on a camera vision solution [10], while companies like Huawei adopt LiDAR and vision fusion technology [11] to improve recognition accuracy and robustness, showing significant improvement in target recognition under extreme environments.

The collected information needs processing to realize its value. Therefore, the computational unit for traffic sign recognition needs to ensure high accuracy while also considering real-time performance [12]. Currently, mainstream computing platforms include GPUs, TPUs, FPGAs, and dedicated automotive AI chips. GPUs, due to their powerful parallel computing capabilities, are often used for training and inference of deep learning models, but have higher power consumption; TPUs and FPGAs offer better energy efficiency and are widely used in embedded systems. So hardware selection needs to balance cost and processing capability.

5. Performance evaluation and optimization

Evaluating the effectiveness of traffic sign recognition algorithms requires a comprehensive set of metrics. These metrics must measure both the accuracy of detection and classification and the

DOI: 10.54254/2755-2721/2025.AST26571

operational efficiency and robustness needed for practical deployment. They provide a quantitative basis for comparing different methods and understanding their strengths and limitations. The core detection accuracy metrics are: Precision, Recall, F1-score, and mAP [13,14]

Precision (P): Quantifies the proportion of instances detected as traffic signs by the model that are correctly identified. Its formula is True Positives (TP) divided by the sum of True Positives and False Positives (TP / (TP + FP)). High precision indicates the model minimizes false alarms or misdetections [13,14]

Recall (R): Measures the proportion of traffic signs actually present in the image that the model successfully identifies. Its formula is True Positives (TP) divided by the sum of True Positives and False Negatives (TP / (TP + FN)). High recall indicates the model finds all relevant instances, minimizing missed detections [13].

F1-score (F1): The harmonic mean of Precision and Recall, providing a balanced single metric, especially useful when class distribution is imbalanced or both false positives and false negatives carry significant weight. Its formula is 2 * (Precision * Recall) / (Precision + Recall) [13].

Mean Average Precision (mAP): A widely adopted comprehensive metric in object detection, representing the average of the Average Precision (AP) values across all object classes. It is calculated as the area under the Precision-Recall curve for a given class.

Optimizing traffic sign recognition performance involves a multifaceted approach. This can be divided into the following points:

Advanced Network Architectures: Cascade R-CNN: It addresses the challenges of noise interference and imprecise object detection by employing a series of detectors trained with progressively stricter IoU thresholds [15]. YOLOv8-TS: As an improved version of the YOLOv8s network, YOLOv8-TS achieves an overall lightweight design through strategic architectural modifications [16].

Attention Mechanisms: Attention mechanisms enable neural networks to selectively focus on the most relevant parts of the input data, enhancing feature representation and improving detection accuracy by emphasizing salient information and suppressing noise. CBAM: Integrated into the neck network of YOLOv8-TS as a lightweight and effective attention module. This dual attention mechanism enhances the model's ability to learn and emphasize critical information in both channel and spatial dimensions [17,18]

Based on these developing technologies, in the future, traffic sign recognition technology will develop towards multi-modal data fusion, lightweighting, and real-time optimization, cross-regional and cross-language generality, adversarial defense, and interpretability-driven directions [19,20]

6. Summary

Traffic sign recognition is crucial for vehicle navigation, decision-making, and safety. As the traffic environment becomes increasingly complex and automation levels rise, our requirements for traffic sign recognition technology also increase, prompting research dedicated to developing more advanced and robust recognition methods. This study systematically addressed the four major challenges in traffic sign recognition, including complex and variable natural environments, dynamic road scenarios, aging and diversity of traffic signs themselves, and interference from complex road environments. Researchers developed and optimized multiple deep learning models through novel architectural designs, sophisticated attention mechanisms, robust training methods, and context-aware perception, and they advanced theoretical progress in the recognition field. In practice, these achievements significantly improved autonomous driving safety, optimized

DOI: 10.54254/2755-2721/2025.AST26571

intelligent traffic management, enhanced feasibility for edge deployment, and helped reduce development costs.

Future research will focus on multi-modal data fusion, extreme lightweighting, and real-time optimization, cross-regional and cross-language generality, adversarial robustness and safety, interpretability, and deep integration with advanced driving decision-making, to address the complexities of the real world and ultimately achieve safer and smarter autonomous driving systems.

References

[1] Liu, Y. (2025, March 27). Policy, technology, and market demand triple drivers: Future development trends of the smart transportation industry in 2025. China Industry Research Network. Retrieved July 19, 2025, from https: //www.chinairn.com/hyzx/20250327/16413023.shtml.

[2] Tan, R., & Yu, L. (2025, July 9). From "technical verification" to "scenario implementation": Trillion-level market for autonomous driving opens. Securities Times. Retrieved July 19, 2025, from https: //www.stcn.com/article/detail/2473853.html

[3] Liu, S. W., Cai, T. B., Tang, X. F., Zhang, Y. Y., & Wang, C. G. (2022).Visual Recognition of Traffic Signs in Natural Scenes Based on Improved RetinaNet.Entropy, 24(1), 112-112. doi: 10.3390/E24010112.

[4] Wang, C., & Fu, Z. A. (2018).Traffic sign detection based on YOLO v2 model. Journal of Computer Applications(S2), 276-278. doi: CNKI: SUN: JSJY.0.2018-S2-059.

[5] Huang, Z. Y., Fang, Q., & Guo, X. H. (2025).YOLOv8-TS traffic sign detection network integrating attention mechanism. Modern Electronics Technique(01), 179-186. doi: 10.16652/j.issn.1004-373x.2025.01.030.

[6] Ma, G., Li, H. W., Yan, Z. W., Liu, Z. J., & Zhao, Z. J. (2024).Improved small target detection algorithm based on multiattention and YOLOv5s for traffic sign recognition. Chinese Journal of Engineering(9), 1647-1658. doi: CNKI: SUN: BJKD.0.2024-09-014.

[7] Min, W., Liu, R., He, D., Han, Q., Wei, Q., & Wang, Q. (2022). Traffic sign recognition based on semantic scene understanding and structural traffic sign location. IEEE Transactions on Intelligent Transportation Systems, 23(9), 15794–15807. doi: 10.1109/TITS.2022.3145467.

[8] Zhang, J. M., Xie, Z. P., Sun, J., Zou, X., & Wang, J. (2020).A Cascaded R-CNN With Multiscale Attention and Imbalanced Samples for Traffic Sign Detection. IEEE Access. doi: 10.1109/access.2020.2972338.

[9] Yan, H., Virupakshappa, K., Pinto, E. V. S., & Oruklu, E. (2015).Hardware/Software Co-Design of a Traffic Sign Recognition System Using Zynq FPGAs. Electronics(4). doi: 10.3390/electronics4041062.

[10] Gent, E. (2021, July 16). *Tesla places big bet on vision-only self-driving: Full Self-Driving beta software v9.0 shows many improvements, but the road to full autonomy still looks rocky*. IEEE Spectrum. Retrieved July 19, 2025, from https: //spectrum.ieee.org/tesla-places-big-bet-vision-only-self-driving

[11] Blizzardxx. (2022, March 24). *Multimodal fusion 2022 | TransFusion: Robust LiDAR-Camera fusion for 3D object detection with transformers*. CSDN Blog. Retrieved July 19, 2025, from https: //blog.csdn.net/rolandxxx/article/details/123689094

[12] Xie, B. Q., & Weng, X. X. (2019).Real-Time Embedded Traffic Sign Recognition Using Efficient Convolutional Neural Network.. IEEE Access.

[13] Muhe DING. (2024, November 18). Accuracy, precision, recall, F-measure, mAP. CSDN Blog. Retrieved July 19, 2025, from https: //blog.csdn.net/ding_programmer/article/details/89740668

[14] Qingqian. (2025, January 5). Explanation of evaluation metrics for machine learning effectiveness. Zhihu. Retrieved July 19, 2025, from https: //zhuanlan.zhihu.com/p/16545421926

[15] Fan, B. B., & Yang, H. (2021).Multi-scale traffic sign detection model with attention. Proceedings of the Institution of Mechanical Engineers(2-3). doi: 10.1177/0954407020950054.

[16] Zeng, H. F. (2024).Research on Traffic Sign Detection and Recognition Algorithm Based on Improved YOLOv8. Computer Knowledge and Technology(30), 13-16. doi: 10.14004/j.cnki.ckt.2024.1589.

[17] Wei, T. C., Chen, X. F., & Yin, Y. L. (2021).Research on traffic sign recognition method based on multi-scale convolution neural network. Journal of Northwestern Polytechnical University.(4), 891-900. doi: CNKI: SUN: XBGD.0.2021-04-024.

[18] Xu, H., & Srivastava, G. (2020).Automatic recognition algorithm of traffic signs based on convolution neural network. Multimedia Tools and Applications(prepublish). doi: 10.1007/s11042-019-08239-z.

DOI: 10.54254/2755-2721/2025.AST26571

[19] Zhou, X. M., Hu, Y. G., Liu, W. J., & Sun, R. J. (2021).Research on Urban Function Recognition Based on Multi-modal and Multi-level Data Fusion Method. Computer Science(9), 50-58. doi: CNKI: SUN: JSJA.0.2021-09-008.
[20] Shao, Y. X., Meng, W., Kong, D. Z., Han, L. X., & Liu, Y. (2020).Cross-modal Retrieval Method for Special Vehicles Based on Deep Learning. Computer Science(12), 205-209.