Ukrainian Journal of Physical Optics
ISSUE 4
2025

DEEP LEARNING-BASED TRAFFIC SIGN OPTICAL RECOGNITION FOR INTELLIGENT TRANSPORTATION

LING XU [1], JIAAO WANG [2], XIAOLING CHENG [3], WEIPING ZHU [3], YIGUO WAN [1], JINJU TANG [3], XIAOKUN YANG [3], XIANGQING WANG [3], DONGFEI WANG [2,3*]

[1] Jiangxi Vocational and Technical College of Communications, Nanchang, 330013, Jiangxi Province, China
[2] School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, 102627, China
[3] School of Electronics and Information, Nanchang institute of technology, 330044, Jiangxi Province, China
*Corresponding author: wdfchina@126.com

Received: 23.06.2025

Abstract. This paper aims to develop a high-efficiency, high-accuracy driver assistance system by integrating deep learning with an optimized Kalman filter approach. The system is designed to recognize traffic signs in complex road environments, enabling the rapid and accurate identification of critical signage to assist drivers in making correct decisions. This paper addresses key challenges in the field of intelligent transportation: in dynamic traffic environments, conventional object detection algorithms struggle to capture deformation features of traffic signs caused by viewpoint variations, resulting in high miss rates for small and deformed targets; existing tracking systems suffer frequent identity switches in densely populated vehicle scenes due to occlusion, compromising tracking continuity; and complex road conditions significantly degrade recognition robustness. To overcome these limitations, the proposed system integrates an enhanced YOLOv11 object detection framework with a Kalman filter-based multi-object tracking algorithm, forming a real-time, end-to-end processing pipeline. Compared to existing technologies, the proposed approach incorporates a deformable convolutional network to enhance spatial feature deformation modeling. The optimized algorithm combines motion trajectory prediction with appearance feature fusion to reduce the frequency of identity switches and mitigate target loss. The mean average precision value has increased; specifically, with 42 categories, the mean average precision of 50 has improved to 0.9222, and with a mean average precision of 50-95, it can also reach 0.7649.

Keywords: optical recognition, traffic signs, deep learning, YOLOv11 model, Kalman filter
UDC: 535.8
DOI: 10.3116/16091833/Ukr.J.Phys.Opt.2025.04013
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

1. Introduction

Amid rapid technological progress, cars, as the most common form of transportation, have seen a significant rise in individual ownership. This change has fundamentally altered urban mobility patterns. The rapid increase in vehicle ownership and the resulting road congestion have put substantial strain on traffic safety. In recent years, the number of traffic accidents, along with casualties and property damage, has risen sharply, making traffic safety a major public concern.

The restructuring plan for the road transportation industry has shifted the focus of the current transportation system toward the next generation of intelligent vehicles. The biggest difference between smart vehicles and traditional cars is the integration of advanced driver-assistance systems, which allow multi-dimensional interpretation of road conditions to help

drivers make accurate decisions. Survey data show that most drivers believe autonomous vehicles can effectively reduce traffic accidents, and the general public has a positive view of their adoption. In the context of next-generation transportation systems, the performance of driving assistance features has become a key indicator of vehicle safety. As intelligent driving systems are increasingly used in real-world applications, public expectations for their safety and intelligence have risen significantly. As shown in Fig. 1.

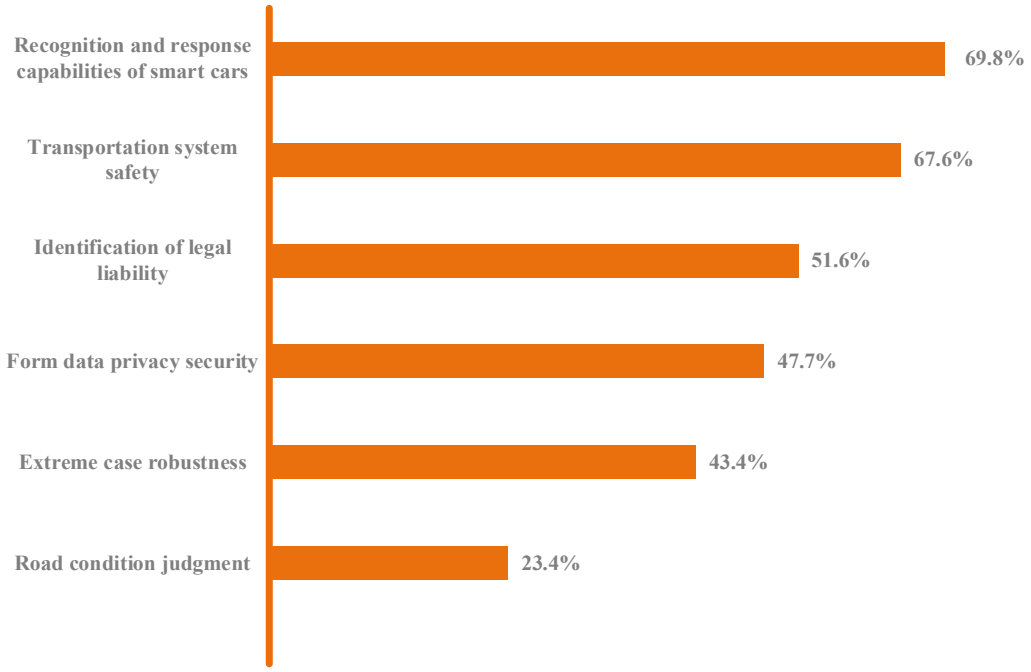

Fig. 1. The level of public expectations for the issues that urgently need to be addressed by intelligent transportation.

The core technology for solving today's critical issues lies in the fast and accurate detection of specific road condition information. The data from upstream detection modules are sent to the central control unit, which then makes precise decisions across multiple aspects. Among these, traffic sign recognition is a vital function of intelligent transportation systems, serving as a foundation for the stable and safe operation of other modules. By capturing images of the road ahead with onboard cameras while driving, traffic signs are detected and accurately identified. The recognition results are then sent to the vehicle's onboard computer or cloud server, helping drivers make informed decisions and ensuring safe, compliant, and stable vehicle operation. Therefore, developing a traffic sign detection system with high speed and accuracy is crucial for the future of the automotive industry. Compared to manual recognition, intelligent onboard cameras are compact and offer several benefits, including lower cost, higher efficiency, and better resilience to environmental conditions, which greatly improves traffic sign detection efficiency.

Currently, the field of object detection includes a wide variety of algorithms, and detecting road traffic information is one of its most important application areas, playing a vital role in developing modern intelligent transportation systems. Current research on object detection in road traffic mainly falls into the following categories.

(1) Manually engineered features.

This category of detection algorithms relies on manually engineered features to enhance the detection capability of classifiers, thereby reducing computational resource consumption to some extent. The detection process based on handcrafted features typically involves three stages: object selection, feature analysis, and classification. Representative methods include edge detection algorithms, the Histogram of Oriented Gradients (HOG), and the line segment detector. The typical pipeline involves applying a sliding window across the target image to select candidate regions, followed by feature extraction using the Viola-Jones detector, and final classification and regression using a support vector machine (SVM) to produce detection outputs.

A major breakthrough in this field was the Viola-Jones detector, introduced by Viola and Jones in their seminal works, "*Rapid Object Detection using a Boosted Cascade of Simple Features*" and "*Robust Real-Time Face Detection*" [1]. This approach combines grayscale imaging techniques, uses functions automatically mapped over grayscale images, and improves recognition performance through the integration of the AdaBoost algorithm and feature selection. However, its robustness is limited when dealing with non-precise or noisy images. Later, N. Dalal [2] proposed using HOG as a supporting method for object recognition in the paper HOG for Human Detection. The gradient-based vector features were input into an SVM to guide the learning process, forming the traditional HOG+SVM framework.

However, in practical applications, the computational process of HOG is relatively complex, leading to longer data processing times. The target selection strategy and the generalization ability of handcrafted features do not meet practical expectations. As a result, the overall performance of the detection algorithm stalls, and detection accuracy is limited by factors such as the number of detection windows and time efficiency. The underlying principle is illustrated in Fig. 2 below.

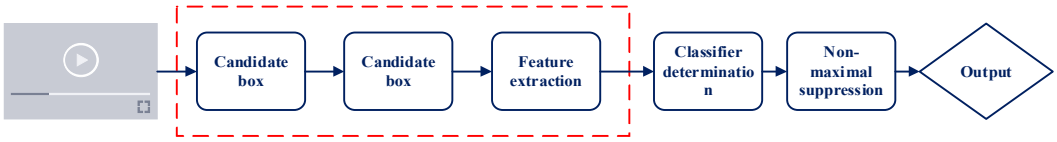

Fig. 2. The process of traditional object detection algorithms.

These algorithms have notable limitations in both recognition accuracy and robustness in complex environments, and they lack the flexibility to develop target-specific optimization strategies. In practice, detection performance depends on factors such as the number, orientation, size, and type of objects, as well as real-time environmental conditions. Achieving accurate recognition amidst these varying conditions remains a primary focus and challenge in the theoretical study of object detection.

(2) Deep learning-based object detection algorithms.

Traditional object detection algorithms, limited by narrow target selection strategies and handcrafted feature processing, failed to achieve breakthroughs after the widespread adoption of HOG+deformable part models. To some extent, the performance of these algorithms stagnated: the former was constrained by the number of detection windows and speed, while the latter was hindered by insufficient accuracy, limiting its practical use. Influenced by machine learning and deep learning approaches, deep learning-based object

detection algorithms have gradually overcome these limitations. By focusing on convolutional neural networks (CNNs) and leveraging their high efficiency and accuracy, real-time object detection has become an increasingly common industry goal.

There are two main approaches to real-time object detection using deep learning. The first approach, known as two-stage detection, creates region proposals with convolutional neural networks and refines target localization using a regression classifier, specifically the region-CNN (R-CNN) [3]. This method, which replaced traditional sliding window techniques, marked a significant advancement in detection and recognition. Additional improvements included spatial pyramid pooling to speed up convolutional sharing, and Ren S. introduced the concept of anchors in Faster R-CNN [4], greatly boosting prediction accuracy. Although two-stage algorithms improve detection accuracy, their complex computations can decrease efficiency.

The second approach, known as one-stage detection, skips the region proposal step and directly predicts bounding box coordinates and class probabilities through regression. The YOLO algorithm demonstrates this method by performing regression across all grid cells to generate detections in a single pass. This approach makes prediction simpler and increases testing speed but usually reduces accuracy compared to the two-stage method.

Different algorithm types optimize various components based on their functional implementations, but each inevitably has certain drawbacks. With ongoing algorithmic improvements and advances in hardware capabilities, object detection has shown strong performance across various complex domains.

Traffic sign detection is essential for traffic sign classification. It also narrows down the search area, lowers the computational demands of subsequent feature extraction algorithms, and can improve recognition accuracy. Traditional traffic sign detection methods rely on differences in shapes, colors, and content features of the signs. In color feature detection research, scholars worldwide have proposed effective solutions such as threshold segmentation and RGB color enhancement to reduce detection errors caused by environmental color variations. Benallal M. et al. [5] studied RGB data variations under different environments and applied threshold segmentation to color models, using difference values as the basis for target recognition. However, this method performs poorly in complex real-world traffic environments with diverse color patterns. A. de la Escalera [6] proposed shape detection techniques based on the Huffman transform and color segmentation; Piccioli G. [7] matched shape detection tasks with predictive models, developing detection and recognition systems with strong performance.

To further enhance detection performance in complex real-world environments, researchers worldwide have proposed fusion detection methods that combine color and shape features of various targets. This efficient, deep learning-based detection approach has gained widespread favor. Natarajan S. et al. [8] applied weighted optimization to the output layer of the convolutional neural network training method.

Object detection systems based on various methods still have limitations in both accuracy and speed [9-10]. Researchers have explored both one-stage and two-stage algorithms to enhance detection accuracy [11-17]; however, these approaches are prone to missed and false detections in complex scenarios involving small or occluded objects. To support intelligent transportation systems, some studies have adopted the YOLOv5 algorithm integrated with

convolutional attention modules [18], which reduces model parameters compared to earlier methods. However, there is still room for improvement in detection accuracy. The fusion of transformer mechanisms with convolutional neural networks enriches feature representation and suppresses redundant feature interference. However, this also increases parameter count and memory consumption [19]. Subsequent studies modified the number of convolutional layers and kernels in the model to improve detection speed, though the overall accuracy remained limited [20]. Some researchers have employed coordinate attention mechanisms [21] to enhance sensitivity to small objects and optimized loss functions using EIoU (EIoU is an optimized loss function that improves upon IoU and CIoU, addressing the issue of class imbalance) [22]. Although these approaches offer certain improvements, they still suffer from limited detection accuracy, high computational complexity, and poor adaptability to highly complex or unusual environments. In response to the current research status and practical needs, YOLOv11 is chosen as the base detection framework, optimized using DCNv2, and corrected with Kalman filtering (KF), resulting in a high-accuracy algorithm for traffic sign detection.

2. Method

This study focuses on traffic sign detection, an important task in driver assistance systems, and suggests improvements to existing object recognition and tracking frameworks. To overcome limitations such as occlusion caused by redundant information, poor feature capture of sign deformation from different viewpoints in traditional detection algorithms, frequent ID switching due to occlusions in busy traffic scenes, and low detection accuracy for small objects and Chinese traffic signs, we propose an optimized hybrid algorithm that combines YOLOv11 and KF. The overall algorithm workflow is shown in Fig. 3.

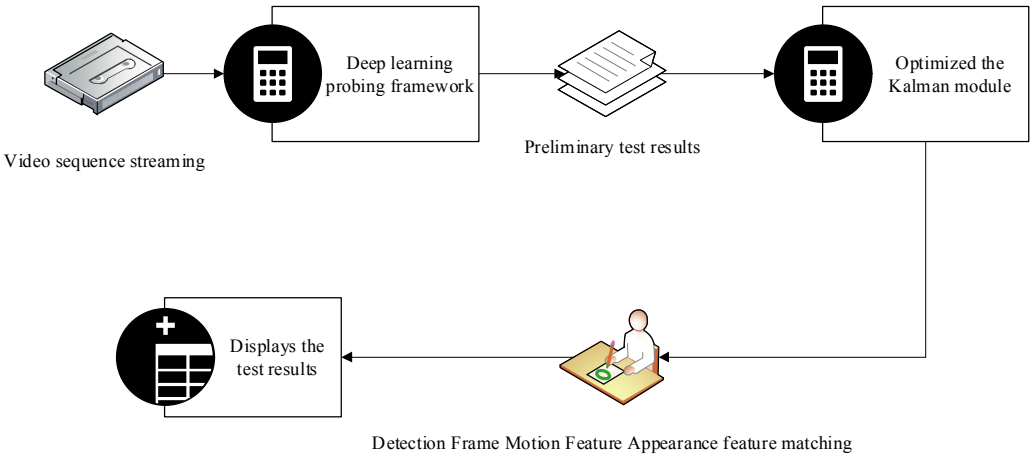

Fig. 3. Algorithm principle and video sequence stream.

The upstream detection module is based on the YOLOv11 framework, which introduces architectural innovations and parameter optimizations to improve practicality in object detection tasks. Compared to other models, YOLOv11 includes advanced feature extraction techniques that enable better detail capture, while also increasing processing speed for real-time performance. The original cloud service provider bottleneck with 2 convolutions block is replaced by the cross-stage partial kernel×2 (C3k2) module, which balances accuracy and

efficiency. Additionally, a new cross-stage partial with pyramid squeeze attention (C2PSA) module is added to enhance detection accuracy for objects of different sizes and positions. Two depthwise convolution modules are incorporated into the decoupled output head to further reduce the parameter count and computational complexity. The architecture has three parts: Backbone, Neck, and Head. The Backbone, strengthened with the C3k2 module, handles feature extraction. The Neck fuses and enhances the extracted features, and the Head produces the final detection outputs.

This study uses a multi-stage processing architecture. First, traffic signs are roughly localized using the YOLOv11 network with an embedded deformable convolutional network v2 (DCNv2) module. Next, an optimized KF algorithm is employed for cross-frame object tracking, and the final output—a warning signal with motion trajectory—is displayed on the in-vehicle head-up display (HUD) system. During detection, the YOLOv11 backbone is extensively modified with deformable convolutional modules to enhance computation. Traditional convolution operations rely on fixed geometric sampling grids (e.g., 3×3 or 5×5), which are ineffective for capturing features of deformed objects like tilted or curved traffic signs. The rigid sampling structure cannot align with actual feature positions, leading to inaccurate feature extraction. The DCNv2 module solves this issue through its dual dynamic mechanism. The principle is shown in Fig. 4.

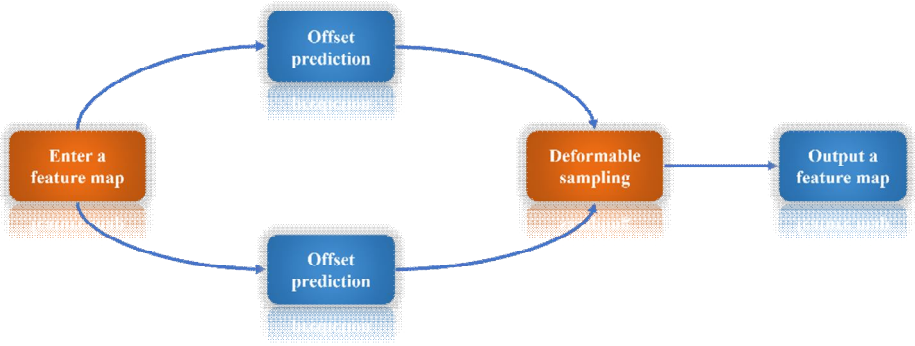

Fig. 4. Principles of DCNv2.

The core formula is: $y(p)=\sum_{k=1}^{K} w_k x(p+p_k+\Delta p_k)\Delta m_k$, where w_k represents the weight, x represents the input feature map, p represents the coordinates of the convolution kernel center, and p_k represents the preset sampling points of the regular convolution kernel.

Here, Δp_k represents the learnable offset, which is predicted by an auxiliary convolutional layer to output spatial dimension information. Δm_k denotes the modulation scalar, predicted via a separate convolutional layer and dynamically adjusted by a Sigmoid activation function to modulate the contribution weight of each sampling point. Furthermore, differentiable sampling is achieved using bilinear interpolation. The corresponding formula is:

$$\frac{\partial y(p)}{\partial \Delta p_k}=\sum_q \frac{\partial y(p)}{\partial x(q)}\cdot\frac{\partial x(q)}{\partial \Delta p_k}. \tag{1}$$

In this context, q iterates over all interpolation points to ensure the offset remains learnable. The dynamic deformable convolution addresses challenges such as image deformation and small-scale target detection in real-world driving scenarios. Its working mechanism is illustrated in Fig. 5.

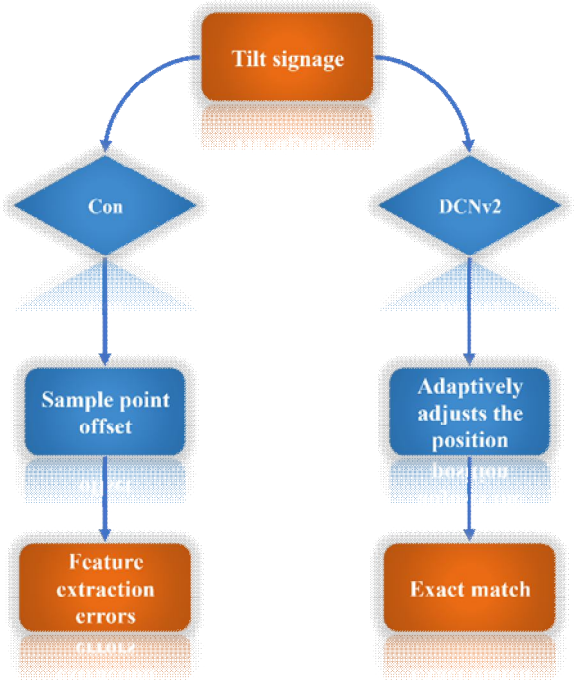

Fig. 5. The application principle of deformation adaptation in traffic sign detection.

The YOLOv11 backbone network has three output layers, corresponding to resolutions of 1/8, 1/16, and 1/32 of the original input image. For an input size of 640 × 640 pixels, the feature map sizes are 80 × 80, 40 × 40, and 20 × 20, respectively. During downsampling, deeper feature maps gain a larger receptive field and capture more global semantic information, while shallower layers focus on local details. Compared to the original YOLOv11 model, the improved network boosts detection accuracy for small objects and deformed targets. Its optimization mechanism within the overall network is shown in Fig. 6.

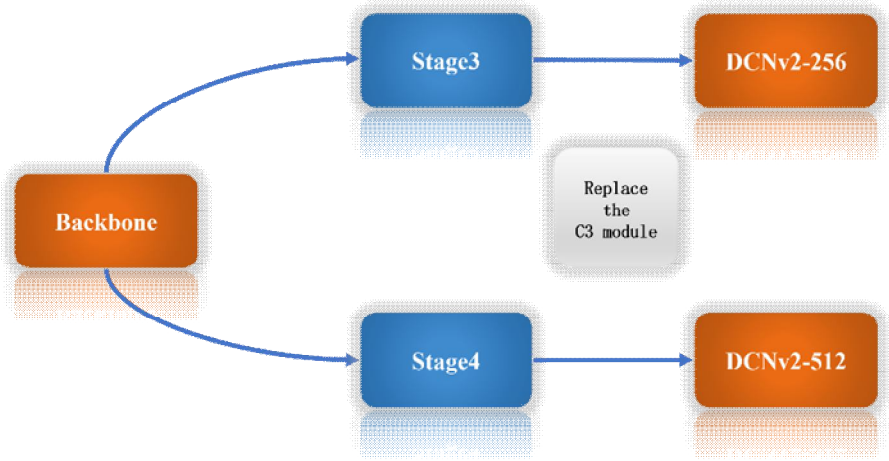

Fig. 6. The optimal placement for the optimization module in YOLO.

Optimization is performed on the middle-to-high-level feature layers to handle regions with significant deformation effectively. In contrast, shallow layers are preserved to avoid excessive displacement that could compromise fundamental features. The optimization results are illustrated in Fig. 7.

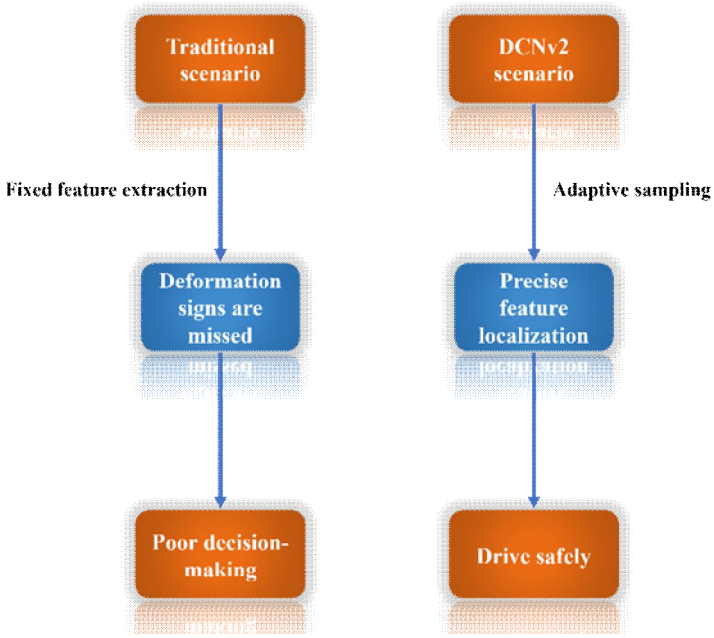

Fig. 7. Comparison of DCNv2 with traditional methods in real-world road conditions.

Leveraging the dual-dynamic mechanism of DCNv2 enables significant improvements in traffic sign detection, especially in deformation adaptation, small object recognition, and occlusion robustness. Its adaptive sampling offset mechanism is well-suited to handle variations in the viewing angles of traffic signs, providing crucial support for autonomous driving perception systems. Additionally, to address the issue of object loss caused by the high speed of vehicles – more pronounced than in other detection scenarios – this work incorporates an optimized tracking framework by introducing the KF for optimal state estimation of the system. The Kalman algorithm estimates the state of a dynamic system from a series of noisy measurements and is computationally efficient to implement. It predicts object bounding boxes based on previous frames and links them with detections in subsequent frames using the Hungarian algorithm.

The linear KF assumes that both the state transition and observation models are linear functions, with all variables following a normal distribution and the noise being Gaussian and uncorrelated. Based on these assumptions, it performs recursive state estimation of stochastic processes. By combining prior state estimates with their associated uncertainties, the algorithm computes an optimal solution that balances multiple error sources and reference quantities, producing results that closely approximate the true state. The KF assumes that variables follow a Gaussian distribution, and it calculates a realistic Gaussian curve using the following four parameters and the Kalman gain.

(1) Prediction: Computed from the observed value at the previous time step.

(2) Observation: In this system, the observed values are provided by the detection framework.

(3) Observation noise: Consists of various factors such as environmental noise, channel interference, and detection inaccuracies.

(4) Prediction error: Comes from the deviation of the predicted value from the previous timestep.

Taking a two-dimensional Gaussian distribution as an example, the state prediction formula is as follows:

$$X_k^P = Ax_{k-1}^t + Bu_{k-1} + w_{k1}. \tag{2}$$

Here, A represents the state transition matrix, B is the control matrix, and w_{k1} denotes the process noise u_{k-1},x_{k-1}^t,X_k^P. Here, u_{k-1} represents the control input at time k-1, x_{k-1}^t represents the true value at time k-1, and X_k^P represents the true value at time k, which is the value to be estimated. At time step k, the measurement Z_k corresponding to the true state x_k satisfies the following equation:

$$Z_k = Hx_k + v, \tag{3}$$

where H represents the observation matrix, and v denotes the observation noise. It is assumed that the random variables w_k and V follow multivariate normal distributions with zero mean and covariance matrices Q and R, respectively; that is, $w_{k1}\sim N(0, Q)$,$V\sim N(0, R)$, where N represents that it follows a normal distribution. The overall process of the KF algorithm consists of two main stages: the prediction stage and the state update stage. During the state update stage, the Kalman gain is used to weigh the relative importance of the observation and the prediction. Essentially, it is computed through operations on covariance matrices to derive an appropriate weighting that leads to an optimal solution. The Kalman gain determines the specific update strategy, yielding a new Gaussian noise distribution. Here, P_t^- - is the predicted estimate covariance matrix, P_t is the updated estimate covariance matrix, and K_t is the optimal Kalman gain. The best estimate of the state $\hat{X}_t^-$ is computed prior to the observation Z_t, and the updated best estimate $\hat{X}_t$ is calculated during the update stage based on the observed Z_t. The principle of the linear KF is illustrated in Fig. 8.

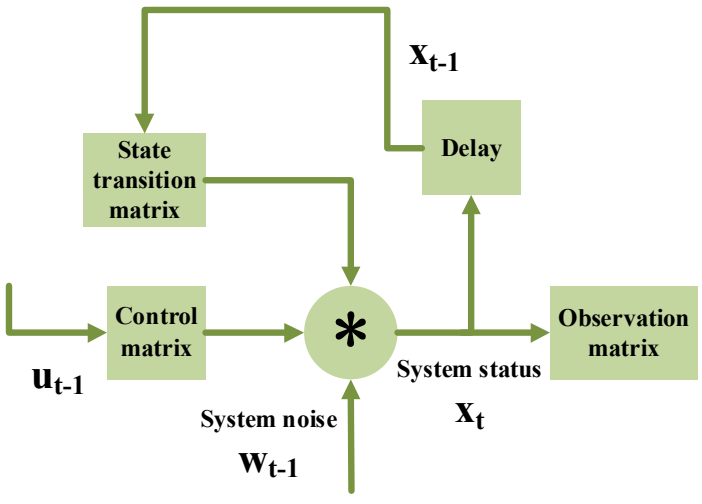

Fig. 8. KF principal scheme. Or uk-1 and k represent a certain moment in time and the previous moment.

The core of the overall process involves calculating the Kalman gain and updating the prior state. The computational flow of the algorithm is illustrated in Fig. 9.

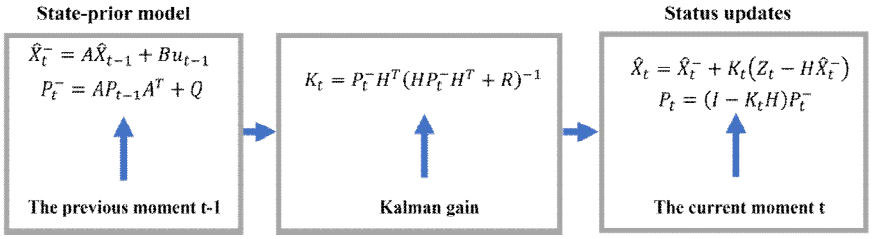

Fig. 9. The scheme of the Kalman calculation process.

KF plays a crucial role in temporal association and state stabilization within traffic sign recognition systems. Its core value lies in addressing the inherent spatiotemporal discontinuities present in single-frame detection models. By utilizing the traffic sign's central position, size, and velocity, combined with real-time vehicle speed, the process noise covariance matrix is dynamically adjusted to ensure accurate motion prediction in high-speed scenarios. At the temporal processing level, the algorithm first predicts the target position using the state transition equation. Subsequently, it associates the predicted bounding boxes with the detections from the current frame via the Hungarian algorithm, integrating a cost matrix that combines both geometric and appearance features. In complex traffic environments, the algorithm continuously predicts the positions of occluded signs using a kinematic model. When the vehicle approaches the sign, the recognition region is dynamically corrected based on a size variation model. Serving as the core of temporal perception, the KF optimizes the YOLOv11 detection model. The filter compensates for the limitations of static detection in dynamic object tracking, while the detection model provides high-precision observation data for the filtering process. Together, they provide the vehicle control module with a continuous and stable foundation for environmental perception. The operational principle is illustrated in Fig. 10.

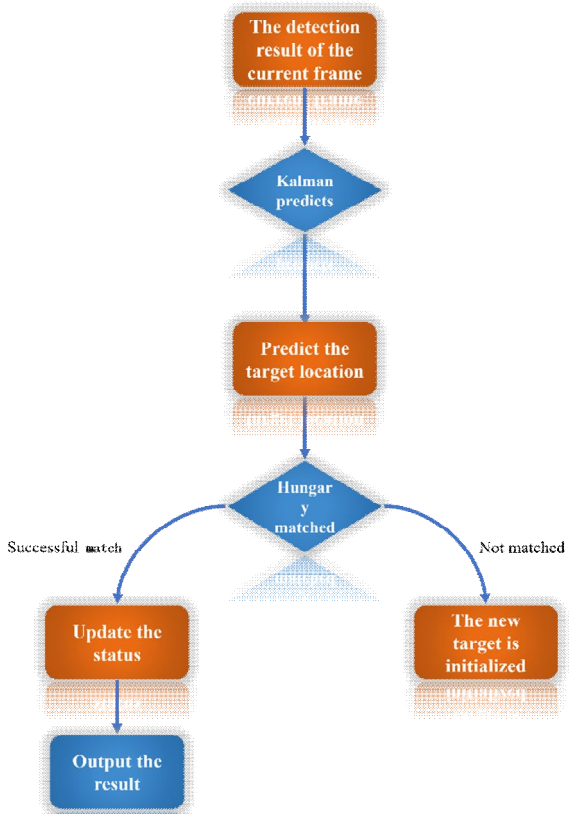

Fig. 10. The principle of the KF algorithm used in traffic sign recognition.

The system adopts an improved YOLOv11 as the foundational detection framework, reconstructing the feature extraction mechanism by integrating a deformable convolution module. Utilizing its dual adaptive mechanism – dynamic offsets and

modulation factors – it precisely addresses perspective deformation and scale variations of traffic signs, enabling the network to dynamically adjust sampling locations based on the actual deformation state of the signs. The modulation factors, output via a Sigmoid-activated independent convolutional layer, generate spatial weight distributions that automatically reduce noise interference in occluded areas while enhancing key information in effective feature regions. To meet temporal continuity requirements, the system integrates a KF-based algorithm that constructs a seven-dimensional state vector describing the spatial position, size ratio, and motion characteristics of the traffic signs. For industrial-grade deployment, a multidimensional lightweight design is employed by embedding the DCNv2 module in the middle layers of the backbone network to replace standard convolutions, and using grouped convolutions to reduce computational load. A "detection-tracking-decision" framework is constructed, where the spatial adaptability of DCNv2 provides high-precision observational input for tracking, and the temporal prediction capability of the KF inversely guides the attention allocation of the detection module. The optimized network architecture is illustrated in Fig. 11.

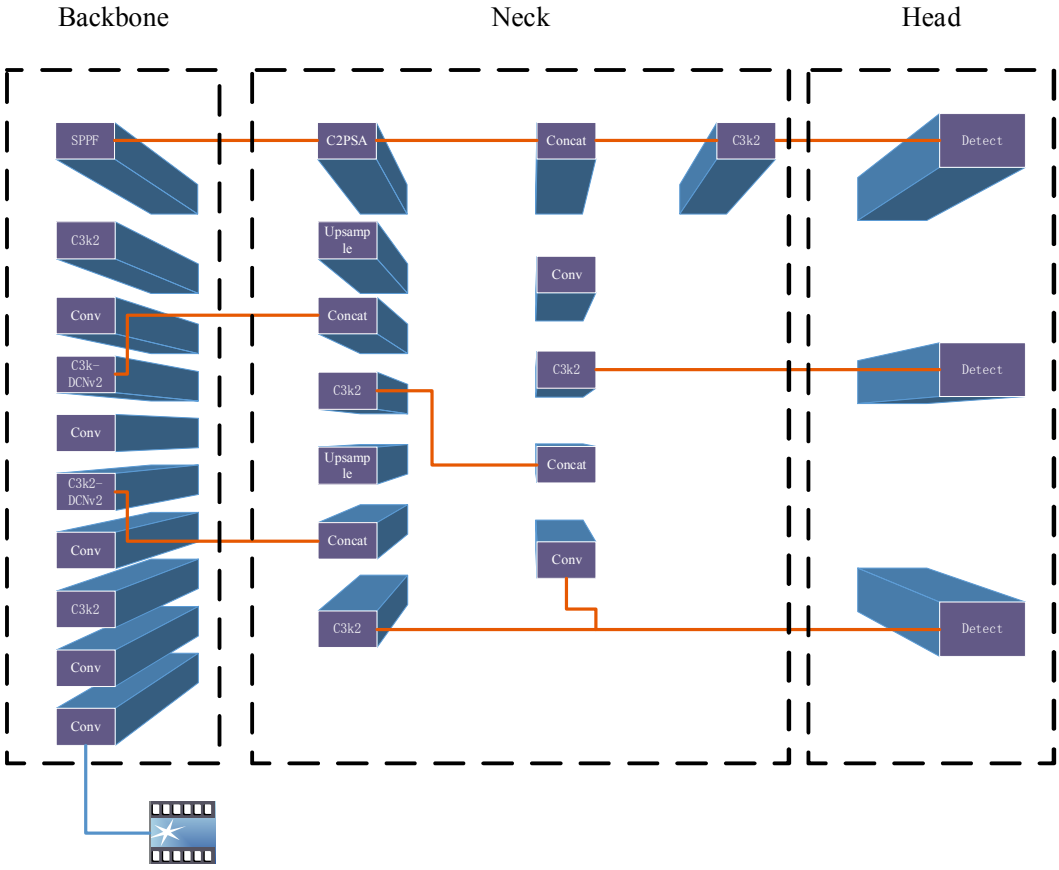

Fig. 11. An optimized traffic sign detection network framework based on YOLOv11.

3. Experimental results

To verify the feasibility of this invention, computer simulations were carried out to validate the authenticity and effectiveness of the aforementioned theory, with training data generated in a virtual environment set up using Miniconda. Chinese road traffic signs typically feature a white

background with black text and black borders, usually in rectangular shapes. Warning signs are yellow and black, shaped as upward-pointing triangles. Regulatory signs have a white background with red borders and black patterns, commonly circular or triangular. Guide signs have a blue background with white patterns, shaped as circles or rectangles. Direction signs feature blue backgrounds with white text (green backgrounds with white text on highways), generally rectangular. Tourist area signs have a brown background with white characters and are rectangular in shape. In practical applications, the influencing factors can be broadly categorized into five types.

(1) Natural environment: Variations in seasons and weather conditions, as well as differences in lighting levels at dawn and dusk—especially under extreme conditions such as heavy rain, snow, and strong winds—result in different recognition performances for the same signs under similar settings.
(2) Redundant interference: In actual road traffic environments, objects with shapes and colors similar to traffic signs exist, and differences among sign categories can also affect detection performance.
(3) Pixel blur: During high-speed vehicle motion, road surface conditions and imaging issues cause distortions due to rapid movement, negatively impacting the images fed into the detection module.
(4) Similar shapes: High vehicle speed causes the captured image regions to be small and difficult to recognize.
(5) Target loss: In real traffic environments, sign locations are often hard to detect, and targets tend to be small in scale. Some signs may be partially damaged or occluded by greenery when fed into the detection module.

To improve detection accuracy for specific application scenarios, the dataset was selected from street scenes captured by onboard cameras under actual Chinese road conditions. By combining the application context and network operational principles, the algorithm ensures its effectiveness across various traffic scenarios. Based on the TT100k dataset, manual annotation and data optimization enhancements were applied to better align with real Chinese road traffic conditions. The dataset consists of training, validation, and test sets containing 6,598, 1,889, and 970 images, respectively. The main traffic signs in China were categorized and clearly labeled, covering 42 types of signs. Most images in the dataset are small in size, reflecting the data stream actually received by the detection module in real driving environments. Based on this data, data augmentation was performed on selected traffic sign images to further improve training efficiency and enhance model robustness. During the accuracy evaluation of the detection algorithm, the following metrics were used: precision, recall, F1 score, precision-recall curve (P-R curve), average precision (AP), and mean average precision (mAP). Finally, a confusion matrix was employed to assess detection accuracy, as shown in Fig. 12.

Real results	Predict the outcome	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Fig. 12. Confusion matrix results.

Recall and precision are defined by the formulas: $P=\frac{TP}{TP+FP}$, $R=\frac{TP}{TP+FN}$, Based on these, accuracy and the F1 score are further introduced to evaluate the classification performance of the network model. The F1 score is the harmonic mean of precision and recall, calculated as:

$$Accuracy=\frac{TP+TN}{TP+TN+FP+FN},\ F1score=\frac{2\times P\times R}{R+P}. \tag{4}$$

The AP value measures the performance of the trained model on each category, while mAP evaluates overall performance across multiple categories. Geometrically, mAP approximates the area under the precision-recall (P-R) curve, effectively summarizing it into a single AP value. The experimental results for the P-R curve and F1 score are shown in Fig. 13 below. The PR curve (precision-recall curve) illustrates the trade-off between the model's accuracy and recall. The horizontal axis represents the recall rate, indicating whether the model can identify the target. The vertical axis shows accuracy, indicating the proportion

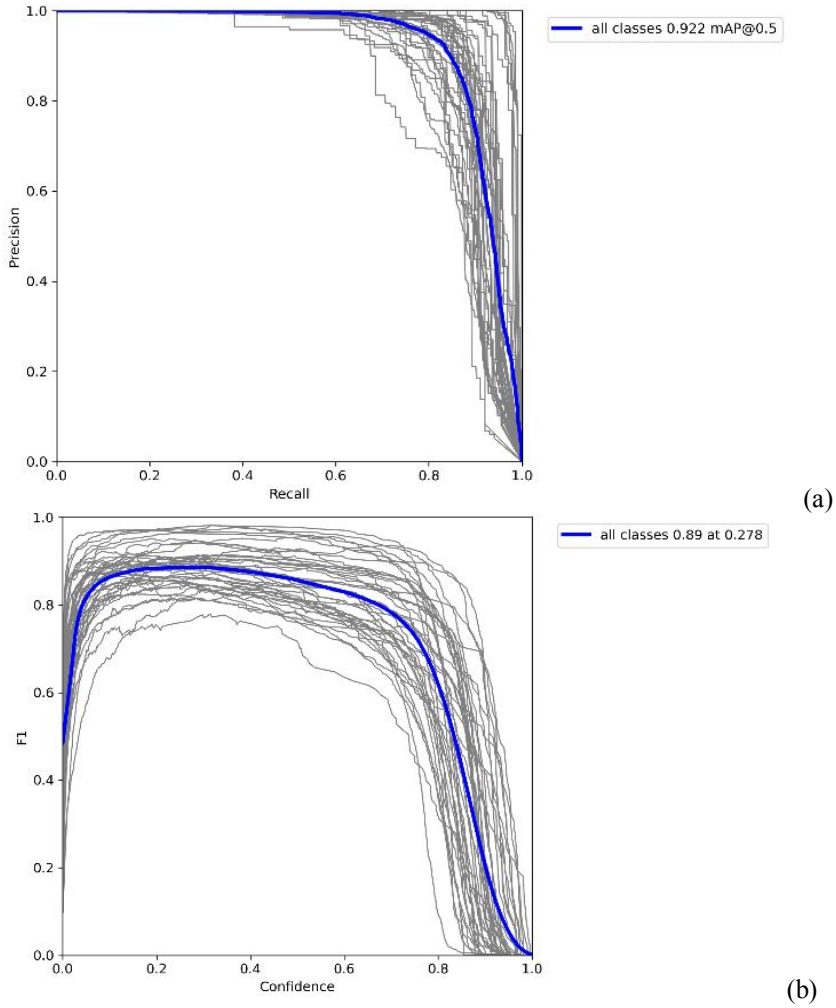

Fig. 13. Experimental performance indicators: (a) dependence of the precision on the recall rate, (b) dependence of the F1 score on the confidence threshold.

of correct identifications. The closer the curve is to the upper right corner, the higher the precision and recall, suggesting that the model is both accurate and comprehensive in identifying traffic signs. The F1 value's variation at different confidence thresholds is dynamically displayed. The horizontal axis indicates the confidence threshold, the minimum confidence required for the model to classify a case as positive. The vertical axis shows the F1 score, the harmonic mean of precision and recall at that threshold. This score balances accuracy and recall, representing the overall performance of both. Only when both are high can the F1 score be high. This figure shows that the model performs stably in most cases and demonstrates strong robustness.

The normalized confusion matrix results are shown in Fig. 14 below. The vertical axis represents the true label, which is the actual category of the data sample. The horizontal axis represents the predicted label, which is the model's predicted classification result for the sample. This figure primarily highlights the model's recognition ability across different categories. The dark cells on the diagonal indicate that the model correctly recognized the vast majority of samples in that category, and the overall classification performance of the model is excellent.

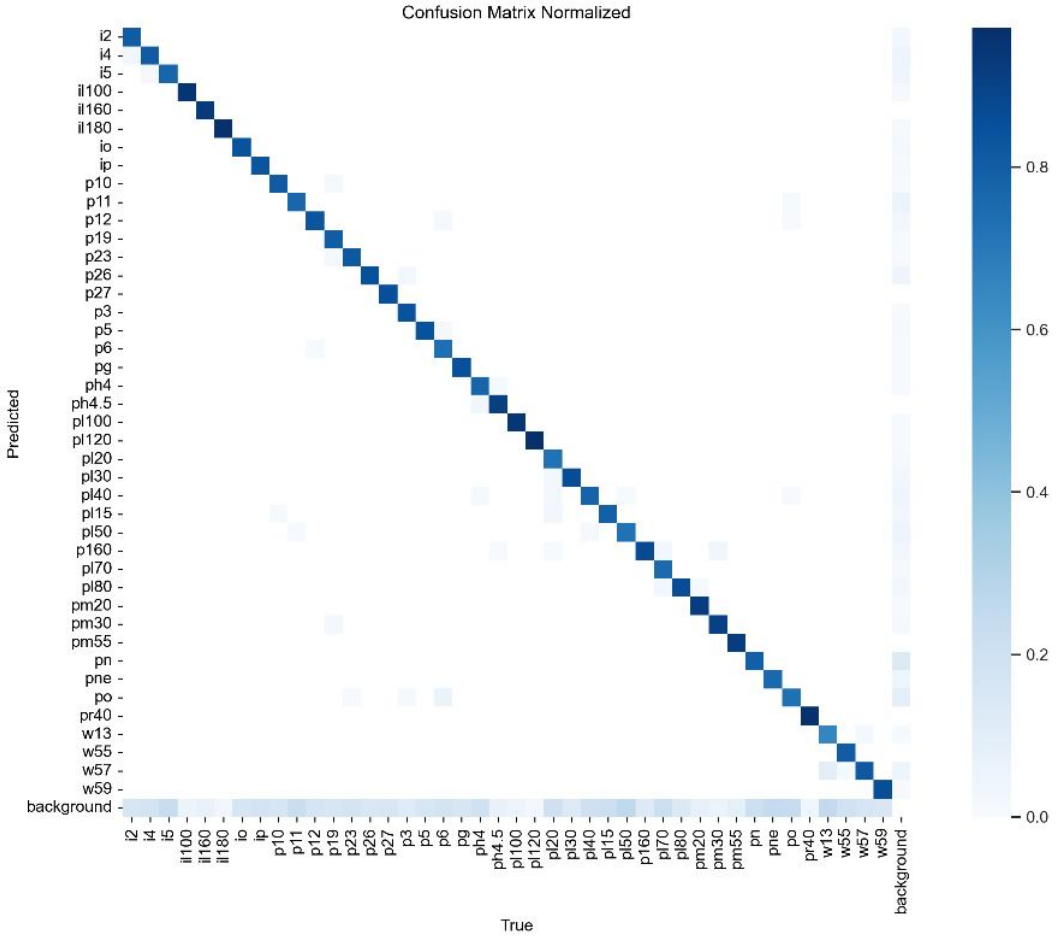

Fig. 14. The normalized confusion matrix results. The vertical axis represents the true label, while the horizontal axis represents the predicted label. (The darker the color of the block, the higher the proportion of correct classifications).

The performance metrics are presented in Table 1 below.

Table 1. Accuracy metrics of the optimized network in traffic sign recognition.

Name	F1	P	R	mAP50	mAP75	mAP50-95
DCNv2	0.8866	0.9477	0.8355	0.9222	0.8745	0.7649

The mAP50 and mAP50-95 values are 92.2% and 76.4%, respectively. By optimizing deformable convolution, the model achieves higher accuracy compared to the baseline and other modules, such as lightweight adaptive weight modules, for this type of task. Additionally, the backbone optimization balances computational performance and parameter stability, making it more suitable for in-vehicle systems. This paper addresses accuracy loss due to viewpoint deformation, partial occlusion, and environmental interference in dynamic traffic scenarios. The technical solutions include dual-core innovations that lead to breakthroughs. Experimental data demonstrate that improvements in detection models and tracking algorithms effectively handle deformation, occlusion, real-time performance, and environmental adaptability. The approach also considers lightweight deployment by incorporating dynamic sampling mechanisms, multi-modal matching strategies, and adaptive noise control mechanisms. The actual detection results are shown in Fig. 15.

Fig. 15 (Part I) .This image displays the recognition results for real road condition images or video data, with 'pl19' being one of the labels and 0.9 representing the probability of it being that label.

Fig. 15 (Part II).

4. Conclusion

This paper proposes an intelligent traffic assistance system based on deep learning to address key challenges in traffic sign detection in dynamic traffic environments. Traditional algorithms often experience high miss rates for small targets due to perspective distortion, frequent ID switches during tracking caused by dense occlusions, and limited robustness under complex road conditions. By integrating an improved object detection framework with an optimized tracking algorithm, an end-to-end real-time processing solution is developed. On the detection side, the YOLOv11 backbone network is innovatively redesigned by incorporating deformable convolution modules into selected mid-to-high-level feature layers. These modules use dynamic offsets to allow sampling points to adapt to the geometric structure of deformed objects.

Additionally, modulation scalars dynamically reweight feature contributions, improving the network's ability to extract traffic sign features affected by perspective distortions such as tilting and bending. The network architecture is further optimized by integrating the C3k2

module to balance detection performance and speed, and by adding the C2PSA module to improve multi-scale detection accuracy while reducing computational load, ensuring efficient deployment on in-vehicle hardware systems. On the tracking side, an optimized multi-object tracking framework based on KF is used. A seven-dimensional state vector models the spatial location, size, and velocity of traffic signs. The process noise covariance is dynamically adjusted based on the vehicle's real-time speed to enhance prediction accuracy in high-speed scenarios. A multidimensional matching strategy combining motion trajectory prediction and appearance feature similarity is employed, utilizing the Hungarian algorithm for inter-frame association to ensure tracking continuity. The detection and tracking modules form a closed-loop system. DCNv2 supplies high-precision observation inputs, while the temporal prediction capability of KF feeds back to guide attention allocation in the detection module. To adapt to real-world Chinese traffic conditions, the dataset and training strategies are refined. Based on the TT100k dataset, 42 types of typical Chinese traffic signs are manually annotated, covering challenging scenarios such as changes in illumination, rain and snow interference, motion blur, small targets, and occlusions. Targeted data augmentation techniques are used to boost model robustness. Experimental results show that the proposed improvements significantly enhance adaptability to deformations and robustness against occlusions, outperforming the baseline model. The mean AP value has increased. With 42 categories, the mAP50 has improved to 0.9222, and mAP50-95 can also reach 0.7649. The final system can output real-time warning signals with motion trajectories, providing drivers with continuous and stable decision-making support for critical traffic information.

Funding and acknowledgment. This work was supported by the first batch of general science and technology projects of Jiangxi Provincial Department of Transportation in 2024 (2024YB031), the Hubei Provincial Natural Science Foundation of China (2023AFB474), the General Project of the 14th Five-Year Plan of Beijing Education Science (CDDB24253), the Beijing Digital Education Project (BDEC2024ZX042), and the Jiangxi Provincial Natural Science Foundation (20232BAB212006). It also received support from the Scientific Research Project of the 14th Five-Year Plan of China Life Sciences Association (K1102024061014) and the Education Reform and Development Project of "Integration of Industry and Education, School Enterprise Cooperation" of China Electronic Labor Society in 2024 (Cea12024136).

Conflicts of interest. The authors declare no conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability. The data can be obtained upon request from the authors. The dataset used is the TT100k dataset, and training and detection were performed using the proposed algorithm on an NVIDIA GeForce RTX 3090.

Reference

1. Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001* (Vol. 1, pp. I-I). IEEE.
2. Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). IEEE.

3. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
4. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, *28*.
5. Benallal, M., & Meunier, J. (2003, May). Real-time color segmentation of road signs. In *CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No. 03CH37436)* (Vol. 3, pp. 1823-1826). IEEE.
6. De La Escalera, A., Moreno, L. E., Salichs, M. A., & Armingol, J. M. (1997). Road traffic sign detection and classification. *IEEE transactions on industrial electronics*, *44*(6), 848-859.
7. Piccioli, G., De Micheli, E., & Campani, M. (1994, May). A robust method for road sign detection and recognition. In *European Conference on Computer Vision* (pp. 493-500). Berlin, Heidelberg: Springer Berlin Heidelberg.
8. Natarajan, S., Annamraju, A. K., & Baradkar, C. S. (2018). Traffic sign recognition using weighted multi-convolutional neural network. *IET Intelligent Transport Systems*, *12*(10), 1396-1405.
9. Yang, B., & Zhang, H. (2022). A CFAR algorithm based on Monte Carlo method for millimeter-wave radar road traffic target detection. *Remote Sensing*, *14*(8), 1779.
10. Zhang, J., Wang, R., Liu, R., Guo, D., Li, B., & Chen, S. (2022). DSP-based traffic target detection for intelligent transportation. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 13180-13191.
11. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, September). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Cham: Springer International Publishing.
12. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988).
13. Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790).
14. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
15. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
16. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence*, *38*(1), 142-158.
17. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, *37*(9), 1904-1916.
18. Zou, W., Li, W., Li, G., Zhou, Q., & Liu, Q. (2023, August). An improved small target traffic sign detection algorithm for YOLOv5s. In *2023 5th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT)* (pp. 223-229). IEEE.
19. Yu, J., Ye, X., & Tu, Q. (2022). Traffic sign detection and recognition in multiimages using a fusion model with YOLO and VGG network. *IEEE Transactions on Intelligent Transportation Systems*, *23*(9), 16632-16642.
20. Wang, Q., Song, W. L., Zhang, Y. Z., Chen, J. H., & Jiang, D. P. (2021). Study on hyperspectral conifer species classification based on improved VGG16 network. *Forest Engineering*, 2021, *37*(3), 79-87.
21. Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13713-13722).
22. Zhang, Y. F., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, *506*, 146-157.

Ling Xu, Jiaao Wang, Xiaoling Cheng, Weiping Zhu, Yiguo Wan, Jinju Tang, Xiaokun Yang, Xiangqing Wang, Dongfei Wang. (2025). Deep Learning-Based Traffic Sign Optical Recognition for Intelligent Transportation. *Ukrainian Journal of Physical Optics*, *26*(4), 04013 – 04031. doi: 10.3116/16091833/Ukr.J.Phys.Opt.2025.04013

Анотація. *У статті представлено розробку високоефективної та високоточної системи допомоги водієві шляхом інтеграції глибокого навчання з оптимізованим підходом фільтра Калмана. Система призначена для розпізнавання дорожніх знаків у складних дорожніх умовах, забезпечуючи швидке та точне виявлення критично важливих знаків для допомоги*

водієві у прийнятті правильних рішень. У роботі розглядаються ключові проблеми в галузі інтелектуального транспорту, зокрема: у динамічних дорожніх середовищах традиційні алгоритми виявлення об'єктів не здатні ефективно фіксувати особливості деформації дорожніх знаків, спричинені змінами кута огляду, що призводить до високого рівня пропущених об'єктів, особливо малих і деформованих; наявні системи відстеження часто мають збої ідентифікації в щільних транспортних потоках через перекриття об'єктів, що погіршує безперервність відстеження; складні дорожні умови значно знижують надійність розпізнавання. Для подолання цих обмежень, запропонована система поєднує вдосконалену структуру виявлення об'єктів YOLOv11 з алгоритмом мультиоб'єктного відстеження на основі фільтра Калмана, формуючи конвеєр обробки в режимі реального часу. На відміну від наявних технологій, запропонований підхід використовує змінювану згорткову мережу для покращення моделювання деформації просторових ознак. Оптимізований алгоритм поєднує прогнозування траєкторії руху з об'єднанням ознак зовнішнього вигляду, що дозволяє зменшити частоту зміни ідентифікаторів та знизити ймовірність втрати цілі.

***Ключові слова**: оптичне розпізнавання, дорожні знаки, глибоке навчання, модель YOLOv11, фільтр Калмана*