

Received 19 September 2023, accepted 28 October 2023, date of publication 1 November 2023, date of current version 7 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3329068

RESEARCH ARTICLE

Detection and Recognition of Obscured Traffic Signs During Vehicle Movement

SHI LUO, CHENGHANG WU¹, AND LINGEN LI

School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China

Corresponding author: Chenghang Wu (wuchenghang1@gmail.com)

This work was supported in part by the National Key Research and Development Program of China under Contract 2019YFB1600500.

ABSTRACT The present study proposes an algorithm to extract obscured traffic sign information from road driving images and fuse it with the vehicle movement process by integrating the vehicle speed. Given the low recognition rate of obscured traffic signs, the proposed algorithm utilizes a traditional color-shape recognition approach to extract potential traffic sign regions. The fusion of traffic sign information from multiple consecutive frames into a single frame enhances the completeness and accuracy of the traffic sign information, bringing it closer to the original characteristics of the traffic sign. The experimental results demonstrate that fusing traffic sign information from the first and third frames improves the template matching similarity by 15.2%. And successfully identifies traffic signs that cannot be recognized by the YOLOV4 and YOLOV8 convolutional neural network when driving at vehicle speeds of 18 (km/h), 36 (km/h), and 54 (km/h). The findings highlight that the proposed algorithm can effectively fuse obscured traffic sign information during vehicle movement to obtain traffic signs with more similar geometric features to the original traffic signs.

INDEX TERMS Traffic sign recognition, image recognition, image fusion.

I. INTRODUCTION

The recognition of traffic signs is an essential aspect of automated and assisted driving, playing a pivotal role in the theoretical and practical underpinnings of automated driving [1], [2], [3]. Target detection usually consists of two steps, first searching for the target in the image and then locating the target using the enclosing frame [4]. The objective of traffic sign recognition is to accurately discern the position and category of traffic signs within complex, real-world settings.

With the explosive development of deep learning, traffic sign recognition is currently divided into two major categories, one is the traditional traffic sign recognition algorithm based on color, shape or a combination of both [5], [6], and the other is the traffic sign recognition algorithm using convolutional neural network. Convolutional neural networks in target detection are divided into single-stage and two-stage procedures, where the single-stage process is exemplified by

the YOLO series of algorithms [7]. And the two-stage process is exemplified by the R-CNN algorithm [8], [9]. In traffic sign recognition, the YOLO series algorithm is mainly used because the inference speed of YOLO algorithm is faster than the two-stage algorithm.

Video object detection has become an increasingly popular research topic due to its specific applications in various scenarios such as unmanned technology, intelligent video surveillance, and robot navigation. While image detection algorithms have been widely used, they primarily focus on static images, which limit their application scope. In contrast, video object detection is more comprehensive and reflects the needs of real-life situations better. However, video object detection poses a more complex task than image object detection. The temporal context of the video contains a significant amount of redundant information, making it difficult to extract motion and temporal information reasonably. Moreover, video object detection also faces obstacles such as occlusion, morphological diversity, lighting changes, and motion blur that are absent in image detection [10]. Consequently, relying solely on image detection algorithms

The associate editor coordinating the review of this manuscript and approving it for publication was Yong Yang¹.

for video object detection may not generate reliable and realistic recognition results. Therefore, developing robust video object detection algorithms that can address these challenges is imperative for achieving accurate and efficient detection performance.

In the context of video target detection, target tracking plays a crucial role. Video object detection and tracking are two closely related tasks, with tracking being regarded as a special kind of detection to some extent. Therefore, combining video object tracking with video object detection is possible. The video target tracking module contains spatio-temporal information concerning the target object, which is essential for the video target detection algorithm. It is worth noting that, unlike target detection, target tracking does not require target identification. Instead, it relies solely on inter-frame information for localization purposes. The output of tracking provides not only the target position and the size of the first frame in the sequence but also predicts the position of the object and the size of subsequent frames. Therefore, the integration of video object tracking and video object detection can enhance the effectiveness and accuracy of target detection [11], as tracking contributes relevant spatio-temporal data and prediction of future states.

However, the majority of the mainstream algorithms currently employed in traffic sign recognition fail to consider the motion of traffic signs with regularity and continuity during vehicle motion. Only a few articles have addressed this issue, such as Wu et al. [12], which mainly focuses on the problem of blurring and defocusing during motion by utilizing multiple frames of images. In the meantime, in the field of traffic sign recognition, particularly in the case of obscured traffic signs in China, no researcher has produced a dataset specifically for recognizing obscured traffic signs. The lack of a targeted dataset has hindered the ability of trained neural networks to accurately recognize obscured traffic signs. Furthermore, the textures of obscured traffic signs differ significantly from those of ordinary traffic signs, and there is no corresponding dataset for targeted training. This lack of data has resulted in trained neural networks being less accurate in detecting and recognizing obscured traffic signs.

Therefore, the algorithm employed in this study builds upon the conventional color-shape traffic sign recognition algorithm [13], aiming to enhance its performance. An additional stage has been introduced, which involves the fusion of information pertaining to obscured traffic signs and subsequent re-identification. Through the incorporation of this novel element, the algorithm demonstrates improved capabilities in traffic sign recognition, making it a valuable contribution to the field.

Our main contributions are as follows:

First, the present study proposes a vehicle driving process-based algorithm for detecting and recognizing traffic sign targets to address the issue of obscured traffic signs that remain undetectable. This approach aims to provide

a solution to the problem of identifying obscured traffic signs in order to enhance road safety and prevent accidents.

Second, to address the problem of traffic sign alignment across multiple frames, a straightforward and efficient image fusion method is proposed in this study. This method leverages vehicle speed and a specific frame to locate traffic signs in the preceding and following frames. By doing so, the proposed approach aims to improve the accuracy of traffic sign detection and recognition, especially in situations where signs are partially obscured.

Third, the algorithm presented in this study is a general and user-friendly approach that has been evaluated in both the CCTSDB(2021) dataset [14], [15] and real-world road scenes in Zhenjiang City. The results demonstrate the effectiveness and practicality of the proposed algorithm in enhancing traffic sign detection and recognition.

II. RELATED WORK

We're going to discuss some of the preparations that have been made, which include, among other things, how others are using frame-to-frame information, and the standards and detection methods for traffic signs.

According to "Road Traffic Signs and Markings" (GB+5768.2-2022), which was published in March 2022 and came into effect on October 1, the size of traffic signs is divided by the speed of vehicles. As the speed of the vehicle increases, the traffic signs become larger accordingly. Take the warning sign as an example, the size is 130×60 when the vehicle speed is $100 \sim 120$, and 90×40 when the vehicle speed is $40 \sim 70$. Meanwhile, different types of traffic signs have different color-shape correspondence. The specific color and shape correspondence of different types of traffic signs are as follows.

- 1) For prohibition signs and restriction signs, red corresponds to the square octagon and inverted triangle. Red plus blue or red plus white corresponds to a circle.
- 2) For directional signs and general road signs, blue corresponds to round and rectangular shapes.
- 3) For warning signs, yellow corresponds to square triangles, rectangles. Meanwhile, green corresponds to the square triangle.
- 4) For freeway and expressway fingerposts, green corresponds to rectangle.
- 5) For tourist signs, the brown color corresponds to the rectangle.

The focus of our study pertains to the presence of traffic signs on urban roads, particularly the first three types among those previously mentioned. These specific types of traffic signs dominate a significant portion of the city's roadways. Concurrently, Chinese city roads are adorned with abundant greenery to enhance the aesthetic appeal of the surroundings. However, this vegetation, particularly trees located alongside the roads, has the potential to obstruct and conceal the visibility of traffic signs situated nearby. Hence, this research

aims to address the issue of detecting and identifying these three types of traffic signs after they have been obscured by other elements.

Next, we will discuss the paper on interframe information. Masmoudi et al. [16] designed an end-to-end self-driving car tracking framework including a You Look Once version 3 (YOLOv3) object detector for recognizing leading vehicles and other obstacles, and a reinforcement learning (RL) algorithm. This framework navigates self-driving vehicles through video frames. Shengyu Lu et al. [17] proposed a real-time target detection algorithm for video based on YOLO network. By preprocessing the image to eliminate the influence of the image background, and then train the Fast YOLO model for target detection, to obtain the target information, using a small convolutional operation instead of the original convolutional operation, which reduces the number of parameters and greatly shortens the time of target detection. Zhang et al. [18] proposed a simple, effective, and generalized correlation method to track by correlating almost all detection frames, instead of only those with high scores. For detection frames with low scores, we utilize their similarity to the tracklet to recover the true target and filter out background detections. Deng et al. [19] propose to address motion blur or occlusion by aggregating neighboring frames to enhance per-frame features. Specifically, we propose Single Shot Video Object Detector (SSVD), a new architecture that integrates feature aggregation into a single-stage detector for video object detection.

Inspired by the principles of human vision, Li et al. [20] proposed an attention mechanism-based approach for adversarial sample defense of traffic signs. In this process, key regions in the image are extracted by the attention mechanism and the pixels are filtered by interpolation. This model simulates the daily human behavior and can improve the recognition success rate of traffic signs with added noise. Chung et al. [21] proposed an attention-based convolutional pooling neural network (ACPN) that applies the attention mechanism to feature maps to acquire key features. the ACPN is robust to external noise and can still classify images with added noise. Zhu et al. [22] created the TT100K dataset and proposed a CNN-based model. Li et al. [23] came out with a cross-layer fusion multi-target detection and recognition algorithm based on Faster R-CNN, which utilizes the five-layer structure of VGG16 to obtain more feature information to solve the problem of small object recognition. In addition, it also includes various levels of occlusion and truncation that can challenge the robustness and adaptability of the computer vision algorithms. Shobha and Deepu [24] present a comprehensive survey of video-based vehicle detection techniques using traditional computer image processing methods. The authors examine various approaches for recognizing, classifying, and tracking vehicles in motion. These approaches include appearance-based models, motion-based models, as well as region-based and feature-based tracking.

Additionally, the article discusses several tracking algorithms used to track vehicles in videos. Liu et al. [25] have introduced an end-to-end deep learning framework known as motion-aid feature calibration network (MFCN) to address the problem of video target detection. The fundamental concept underlying this approach is to leverage the temporal coherence of video features, while also considering their motion patterns detected via the optical stream. By aggregating calibrated features at both the pixel and instance level across frames, the proposed framework aims to enhance the accuracy of detection and improve robustness against changes in appearance. Lou et al. [26] In order to solve this problem of small-size target recognition, they proposed an improved small-size target detection algorithm for special scenes based on YOLOV8. The YOLOV4 algorithm by Bochkovskiy et al. [27] has good performance in traffic sign recognition and detection after training.

We use the YOLOV4 and YOLOV8 algorithms for comparison with our algorithm.

III. METHODS

The algorithm employed in this study comprises two distinct stages for detecting and recognizing obscured traffic signs. The first stage involves detecting potential traffic signs, while the second stage involves fusing images of regions of interest (ROIs) from multiple frames to obtain the final recognition outcome.

The initial stage of the algorithm focuses on identifying potential traffic signs through a series of image processing techniques. The second stage involves merging the ROIs from multiple frames to enhance the recognition accuracy of the obscured traffic signs. The fusion process is performed using advanced image processing techniques that enable the algorithm to isolate the relevant information from each frame and combine it to obtain a more accurate recognition outcome.

A. VEHICLE TRAVEL IMAGE PRE-PROCESSING

During the process of vehicle driving, the location of traffic signs is situated within a complex environment with changing lighting conditions. Road scenes captured by cameras may exhibit variations in angles and perspectives, leading to disparities between the images of traffic signs and their standard counterparts.

To address this issue, the HSV color space is utilized for color gamut segmentation of the input image. This approach can enhance the robustness of the system against external environmental factors. The RGB color space is transformed into HSV space using a conversion formula. In the HSV space, the H, S, and V components are normalized within the range of (0, 255), and the upper and lower limits of each component are constrained to segment the color gamut of red, yellow, and blue in the input image [28]. Notably, traffic signs exhibit higher color saturation, corresponding to a higher

S-component of color in the HSV space, than the standard HSV space color gamut segmentation. The upper and lower limits of each component of red, yellow and blue colors are as follows:

$$\begin{pmatrix} h_{\min} & s_{\min} & v_{\min} \\ h_{\max} & s_{\max} & v_{\max} \end{pmatrix} = \begin{pmatrix} 0 & 43 & 46 \\ 10 & 255 & 255 \end{pmatrix} \cup \begin{pmatrix} 156 & 43 & 46 \\ 180 & 255 & 255 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} h_{\min} & s_{\min} & v_{\min} \\ h_{\max} & s_{\max} & v_{\max} \end{pmatrix} = \begin{pmatrix} 11 & 120 & 46 \\ 25 & 255 & 255 \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} h_{\min} & s_{\min} & v_{\min} \\ h_{\max} & s_{\max} & v_{\max} \end{pmatrix} = \begin{pmatrix} 100 & 115 & 46 \\ 124 & 255 & 255 \end{pmatrix} \quad (3)$$

After performing the Canny edge detection algorithm on the image after the hsv space transformation, the binarized edge detection image is obtained. After the outer contour of the image is obtained by Douglas-Peucker algorithm, image denoising is performed to remove the fine noises.

Since various indications inside the traffic sign can be ignored in the potential traffic sign detection and recognition. Therefore, the edge detection image is turned into a road driving binarized image that can be detected and recognized by using a closed operation to expand the inner and outer contours of the image after hole filling, which facilitates the subsequent traffic sign detection and recognition, as shown in Fig. 1.

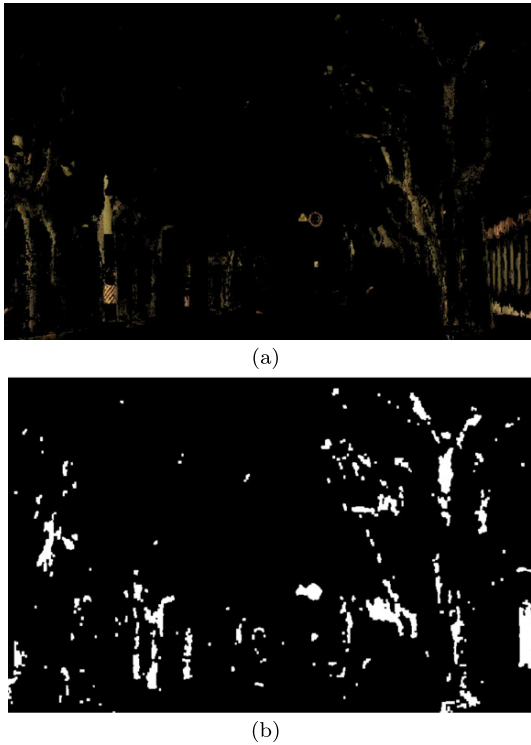


FIGURE 1. Image after color gamut segmentation and morphology processing. (a) Image after yellow color gamut segmentation. (b) Images for detection and recognition.

B. DETECTION AND IDENTIFICATION OF POTENTIAL TRAFFIC SIGNS

In the first stage of the algorithm, the connectivity domain area of the traffic signs that are obscured during vehicle travel is small, while the geometric features are not obvious compared to the standard traffic signs. To avoid removing the obscured traffic signs, the detection and recognition in the first stage contains some images of non-traffic signs. Compared with the general traffic sign recognition algorithm, the algorithm in this paper has more non-traffic sign areas in the first stage detection.

This paper divides the detection of traffic signs in one stage into three categories based on the obviousness of geometric features after being obscured, which are the detection and recognition of circular traffic signs, rectangular traffic signs and general traffic signs.

The detection of circular traffic signs in one stage is obtained by performing the Hough circle transformation through eq.(4) and eq.(5) to obtain the center and radius of the circle, and then the region of interest(ROI) of the current frame is obtained, and the result is shown in Fig. 2(a)

$$(x - a)^2 + (x - b)^2 = \rho^2 \quad (4)$$

$$\begin{cases} x = a + \rho \cos \theta \\ y = b + \rho \sin \theta \end{cases} \quad (5)$$

Rectangular traffic signs have a distinct geometric feature: right angles. Even the rectangular traffic signs that are partially obscured still have the geometric feature of right angle. The detection of rectangular traffic signs for one stage is obtained by eq.(6). After that, the ROI is selected according to the line segment forming the angle, as shown in Fig. 2(b).

$$\begin{cases} k_a = -\frac{y_{a+1} - y_a}{x_{a+1} - x_a} \\ k_{a+1} = -\frac{y_{a+2} - y_{a+1}}{x_{a+2} - x_{a+1}} \\ \theta = 57.3 \cdot |\arctan k_{a+1} - \arctan k_a| \end{cases} \quad (6)$$

Except for circular traffic signs and rectangular traffic signs, the rest of the shapes have no obvious geometric features after being obscured, and the binarized images are used to obtain the ROI by template matching through eq.(7). In eq.(7), $I(x, y)$ is the original image matrix, $T(x, y)$ is the template image matrix, and $R(x, y)$ is the image result matrix. For example, the ROI is obtained by setting a low threshold of template matching similarity, as shown in Fig. 2(c).

$$\begin{cases} T'(x, y) = \frac{T(x, y) - \frac{1}{w \cdot h} \sum_{x', y'} T(x', y')}{\sqrt{\sum_{x', y'} T(x', y')^2}} \\ I'(x, y) = \frac{I(x, y) - \frac{1}{w \cdot h} \sum_{x', y'} I(x', y')}{\sqrt{\sum_{x', y'} I(x', y')^2}} \\ R(x, y) = \sum_{x', y'} (T(x', y') \cdot I'(x + x', y + y')) \end{cases} \quad (7)$$

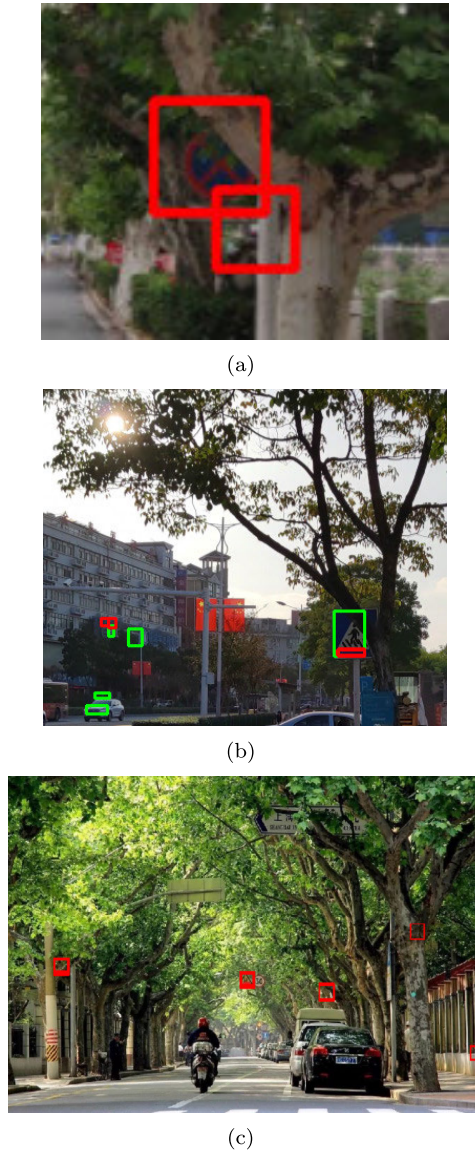


FIGURE 2. One-stage traffic sign detection results. (a) is One stage of circular traffic sign detection results. (b) is One-stage rectangular traffic sign detection results. (c) is One stage of other traffic sign detection results with warning signs as an example.

C. EXTERNAL RECTANGULAR COORDINATE MATRIX OF ROI

The process of traffic sign detection and recognition of obscured traffic signs requires image fusion of multiple consecutive frames of the vehicle in motion. Image alignment is generally required in the process of multi-frame image fusion. However, since the geometry of obscured traffic signs varies greatly from frame to frame, general image alignment algorithms, such as the SIFT algorithm or the ORB algorithm, are unable to achieve image alignment. Meanwhile, the computational overhead of one-stage detection for each frame of multiple consecutive images is too large. In order to solve the problem of how to obtain the ROI of multi-frame images in image fusion, we propose to detect traffic signs

by combining the timing information during the vehicle movement. The ROI of a certain frame and the current vehicle speed are used to calculate the ROI of the adjacent frames.

Moreover, unlike general video target detection that involves detecting moving objects such as a cat or a person, the current study deals with detecting objects within the context of a moving vehicle. This provides an advantage since the state of the vehicle is known, including the vehicle speed, acceleration in each direction, and other relevant parameters, rendering it possible to directly obtain the required data for analysis. Secondly, the study also focuses on detecting obscured traffic signs that are stationary relative to the ground, meaning that the vehicle state is known. Furthermore, the approximate predictable trend of changes in the position of the traffic signs in the image is also available. By leveraging these features, the proposed methodology aims to overcome the limitations that are characteristic of traditional traffic sign detection techniques and provide more accurate results.

The coordinate matrix of the outer rectangle of the ROI is derived as follows: Assume that the coordinates of the smallest point of the external rectangle of the traffic sign in the actual scene are (X_0, Y_0) , the width and height of the external rectangle are a_0 and b_0 , respectively, and the distance between the camera and the traffic sign is Z_0 . The pixel coordinates of the initial maximum point $(u_{0\max}, v_{0\max})$ and the pixel coordinates of the initial minimum point $(u_{0\min}, v_{0\min})$ are obtained by substituting the parameters according to [29] to obtain eq.(8) and eq.(9), respectively.

$$\begin{cases} u_{0\max} = \frac{fw}{m} \cdot \frac{X_0 + a_0}{Z_0} + \frac{w}{2} \\ v_{0\max} = \frac{fh}{n} \cdot \frac{Y_0 + b_0}{Z_0} + \frac{h}{2} \end{cases} \quad (8)$$

$$\begin{cases} u_{0\min} = \frac{fw}{m} \cdot \frac{X_0}{Z_0} + \frac{w}{2} \\ v_{0\min} = \frac{fh}{n} \cdot \frac{Y_0}{Z_0} + \frac{h}{2} \end{cases} \quad (9)$$

the parameters in eq.(8) and eq.(9) are shown as follows: f_x, f_y denote the number of pixels in the x-axis or y-axis direction to quantify the physical focal length, respectively. c_x, c_y denote the number of horizontal and vertical pixels that differ between the pixel coordinates of the center of the image and the pixel coordinates of the image origin, respectively. Z is expressed as the distance between the target and the camera in the world coordinate system. m, n are the width and height of the sensor size, respectively. w, h are the width and height of the actual image, respectively.

Suppose the speed of the vehicle at this time is u_a (km/h), and the frame rate of the video recording is K (fps). Then the pixel coordinates of the maximum point of the pixel coordinate matrix at frame i can be introduced by combining

eq.(8), eq.(9) and deviation Θ , as shown in eq.(10).

$$\begin{cases} u_{i \max} = \frac{fw}{m} \cdot \frac{X_0 + a_0}{Z_0 - \frac{u_a(i-1)}{3.6K}} + \frac{w}{2} + \delta \\ v_{i \max} = \frac{fh}{n} \cdot \frac{Y_0 + b_0}{Z_0 - \frac{u_a(i-1)}{3.6K}} + \frac{h}{2} + \delta \end{cases} \quad (10)$$

The difference between the current frame and the initial frame coordinates can be found by making a difference between the eq.(10) and the eq.(8). The difference is obtained by subtracting the initial minimum pixel coordinates from the difference to obtain the current minimum pixel coordinates. The result is shown in eq.(11). Meanwhile, the ROI coordinate matrix of the i frame is shown in eq.(12):

$$\begin{cases} u_{i \min} = \frac{fw}{m} \cdot \left(\frac{2X_0 + a}{Z} - \frac{X_0 + a_0}{Z_0 - \frac{u_a(i-1)}{3.6K}} \right) + \frac{w}{2} - \delta \\ v_{i \min} = \frac{fh}{n} \cdot \left(\frac{2Y_0 + a}{Z} - \frac{Y_0 + b_0}{Z_0 - \frac{u_a(i-1)}{3.6K}} \right) + \frac{h}{2} - \delta \end{cases} \quad (11)$$

$$ROI_i = \begin{bmatrix} (u_{i \min}, v_{i \min}) & (u_{i \max}, v_{i \max}) \\ (u_{i \max}, v_{i \min}) & (u_{i \max}, v_{i \max}) \end{bmatrix} \quad (12)$$

D. IMAGE FUSION ALGORITHM PROCESS

The process of the algorithm for fusion of traffic sign information from multiple frames is shown in the following Fig. 3: At a vehicle speed of 36 (km/h), two images with an

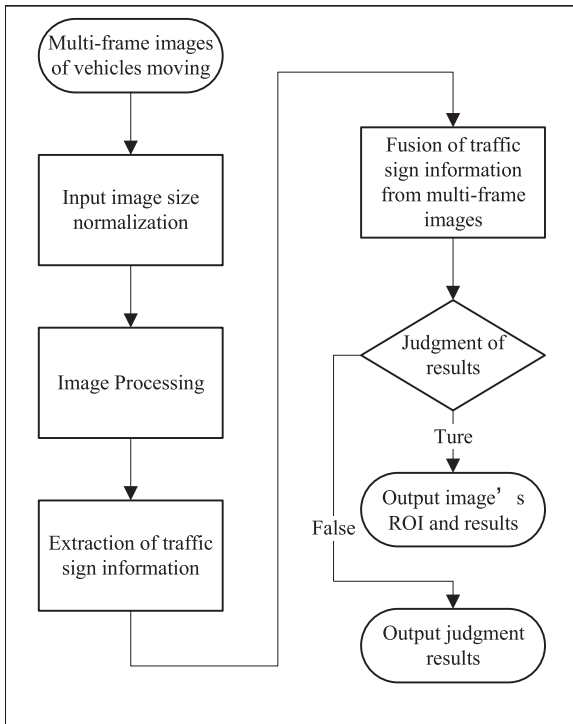


FIGURE 3. Flowchart of the algorithm for traffic sign information fusion.

interval of 4 frames are selected. Taking these two frames as an example, the algorithm flow can be known according to Fig. 3. First, the obtained before and after two frames

are normalized by eq.(13), so that the traffic signs in the before and after two frames can be fused with traffic sign information. As shown in Fig. 4, (a) is a 36×36 image and (b) is a 48×48 image. Both images are 36×36 after normalization.

$$f(x, y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} \begin{bmatrix} x_2 - x & x - x_1 \\ f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{bmatrix} \begin{bmatrix} y_2 - y_1 \\ y - y_1 \end{bmatrix} \quad (13)$$

Next, the partially obscured prohibited traffic signs are extracted, as shown in Fig. 4(c) and (d). The traffic sign images of Fig. 4(c) and (d) are grayed out and binarized to obtain the images of the obscured traffic signs after binarization, as shown in Fig. 4(e) and (f).

TABLE 1. Results of matching circular traffic sign templates for prohibited signs and non-traffic sign areas.

	Initial frame image	The second frame image	Image after fusion
Circular traffic signs	0.525	0.619	0.842
Non-traffic signs	0.151	0.142	0.160

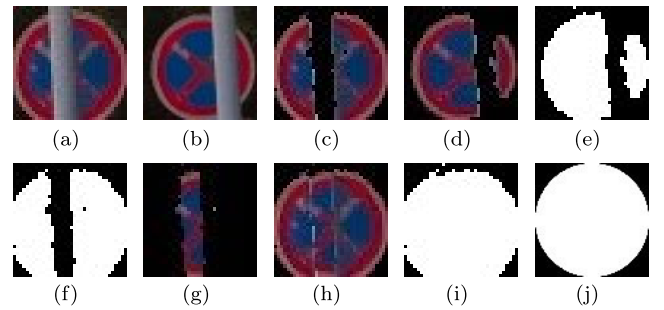


FIGURE 4. Process image of traffic sign information fusion algorithm. (a) Initial frame image of 36×36 . (b) Second frame image of 48×48 . (c) Initial frame color gamut separation. (d) Second frame color gamut separation and normalization. (e) Initial frame binarized image. (f) Second frame binarized image. (g) Fused traffic sign information in the second frame. (h) Fused traffic sign information result image. (i) Fused result binarized image. (j) Circular template image.

After that, the inverse image of Fig. 4(f) is used as a mask, and the traffic sign information is extracted from Fig. 4(b) by logical and operation of image matrix to obtain the information of the missing part of the obscured traffic sign in Fig. 4(a), as shown in Fig. 4(g). After performing image bit operation on Fig. 4(g) and Fig. 4(a), the traffic sign image that fuses the traffic sign information of these two frames is obtained, as shown in Fig. 4(h). From Table 1, we know that when the ROI is not the traffic sign area in the actual scene, there is no information missing from the ROI in the two frames. So the similarity of the fused images after template matching does not improve to a great extent relative to the images before fusion.

The degree of increase in similarity after template matching indicates the degree to which the missing information of the traffic sign is completed, which means the higher the

possibility that the fused image is a traffic sign, and thus the real traffic sign is extracted.

Alternatively, the traffic sign image information contained in the initial frame of the road driving image is relatively complete, and the similarity result after template matching is higher than the set value, and the traffic sign in the image can also be identified.

IV. EXPERIMENTS

A. EXPERIMENTAL ENVIRONMENT AND CONFIGURATION

The experimental environment is Python3.8, and the practical device is a laptop with AMD Ryzen 7 5800H CPU.

B. ANALYSIS OF GENERAL TRAFFIC SIGN RESULTS

Our algorithm mainly performs traffic sign information fusion and complementary recognition for obscured traffic signs in two types of scenarios. The first scenario is when the vehicle is stationary and the object blocking the traffic sign is in motion. As the object moves, the fused traffic sign becomes complete. For example, when waiting for a red light, pedestrians are blocking the traffic sign. The second type of scenario is when the vehicle is in motion, the object blocking the traffic sign is prohibited from moving, and as the vehicle keeps moving, the blocking object keeps moving relative to the vehicle, and the traffic sign information of multiple frames is fused to get a more complete traffic sign. For example, the vehicle driving process of being a pole obscured by the no stop sign. In the actual vehicle driving process, the second situation is more common than the first one.

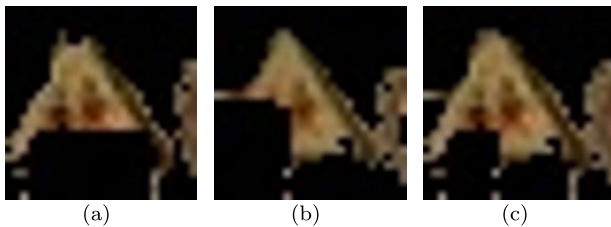


FIGURE 5. Take warning signs as an example, when the vehicle is stationary, the obscuring object is in motion with multi-frame image fusion. (a) initial frame image. (b) the second frame image. (c) image after fusion.

For the above two cases, this paper uses the road driving images in the actual scenario for the study, and the image results as in Fig. 5, Fig. 6 are obtained by the traffic sign information fusion algorithm in this paper, and the data results in Table 2 and Table 3 are obtained by comparing with YOLOV4 algorithm and YOLOV8 algorithm. It can be seen that the recognition result of traffic signs in YOLOV4 neural network in Fig. 5 is 0.641, and the traffic signs in Fig. 6 cannot be detected and recognized by YOLOV4 neural network. And after the traffic sign information fusion by our algorithm, the geometric features of the traffic signs are significantly complemented. The similarity between Fig. 6(c) and the triangle template after template matching reaches 85.8%.

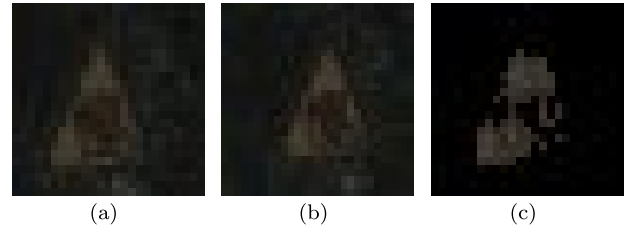


FIGURE 6. Take warning signs as an example, multi-frame image fusion while the vehicle is in motion when the obscured object is stationary. (a) initial frame image. (b) the second frame image. (c) image after fusion.

TABLE 2. Taking warning signs as an example, the recognition results of multi-frame image fusion of obscurers in motion when the vehicle is stationary.

Method	Initial frame image	The second frame image	Image after fusion
YOLOV8	0.559	0.692	—
YOLOV4	0.670	0.641	—
ours	0.639	0.587	0.737

TABLE 3. Taking the warning sign as an example, when the obscured object is stationary and the vehicle is moving, the recognition result of multi-frame image fusion.

Method	Initial frame image	The second frame image	Image after fusion
YOLOV8	0.617	0.618	—
YOLOV4	—	—	—
ours	0.593	0.547	0.858

C. ANALYSIS OF THE RESULTS OF INDICATING TRAFFIC SIGNS

The recognition of indication traffic signs poses a significant challenge due to their non-uniform shape and size, as well as the considerable difference between the geometric shape of these signs and that of ordinary traffic signs when obscured by objects. Additionally, the lack of labeled data for domestic traffic sign datasets further hampers the neural network's recognition ability for indication signs. As a result, the neural network can only recognize certain specific indication signs at close distances, as depicted in Fig. 7. Upon analysis of the CCTDSB dataset, it was discovered that the majority of trained neural networks for the recognition of traffic indication signs were designed to identify circular signs, rather than rectangular ones. As a result, our team developed an algorithm specifically tailored to detect and recognize rectangular indication signs within the realm of traffic signs.

To verify the efficacy of our algorithm, we conducted tests using continuous images from the CCTSDB dataset, as well as road driving images captured from actual road scenes in Zhenjiang City. During these tests, images were captured at an interval of one frame while the vehicle was traveling at a speed of 36 (km/h). The results of these tests are presented in Fig. 8.

Our algorithm was able to accurately identify and recognize rectangular traffic indication signs within the tested datasets. This demonstrates its potential for use in real-world

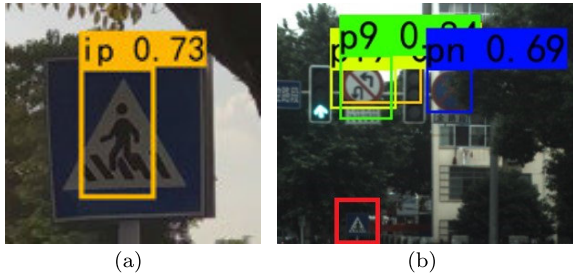


FIGURE 7. Recognition results of crosswalk signs by YOLOV4. (a) is Crosswalk signs successfully recognized by YOLOV4 at close range. (b) is Crosswalk signs that are not recognized by YOLOV4 at a certain distance.

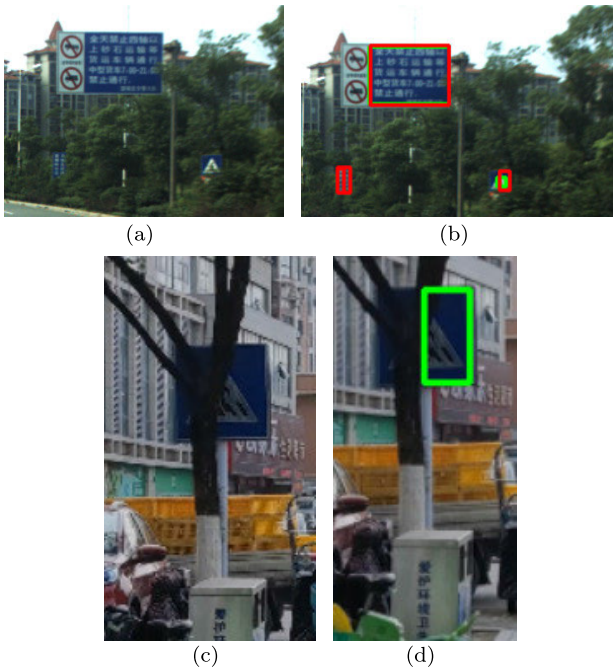


FIGURE 8. Take warning signs as an example, multi-frame image fusion while the vehicle is in motion when the obscured object is stationary. (a) initial frame image. (b) the second frame image. (c) Original image. (d) image after fusion.

scenarios, where the ability to accurately identify signs is crucial for ensuring road safety.

In urban areas, identifying crosswalk signs can significantly enhance vehicle safety during navigation. However, pedestrian crossing signs are prone to being obscured by a range of factors. Given the frequency of pedestrian crossing traffic signs, their susceptibility to obscurity, and their crucial role in ensuring road safety, a template matching method for pedestrian crossing traffic sign templates is employed to identify whether the resulting traffic sign obtained after information fusion is indeed a crosswalk sign. This approach is aimed at improving the accuracy and reliability of crosswalk sign identification. It can be seen from Fig. 9 and Table 4 that pedestrian crossing traffic signs are usually obscured with a large area. Using the combination of the template matching similarity of the initial image and the

increment of the template matching similarity of the fused image as the judgment condition, the obscured pedestrian crossing traffic signs can be recognized.

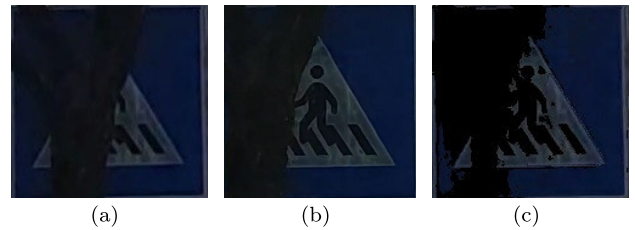


FIGURE 9. Information fusion of crosswalk signs. (a) initial frame image. (b) the second frame image. (c) image after fusion.

TABLE 4. Similarity of matching templates of obscured crosswalk traffic signs and confidence of YOLOV4 and YOLOV8 neural network.

Method	Initial frame image	The second frame image	Image after fusion
YOLOV8	—	—	—
YOLOV4	—	—	—
ours	0.413	0.548	0.786

D. ANALYSIS OF IMAGE FUSION RESULTS AT DIFFERENT VEHICLE SPEEDS AND DIFFERENT FRAME INTERVALS

In this study, the algorithm’s effectiveness and authenticity were ensured by selecting an obscured traffic sign in a section of road in Zhenjiang city as the detection target. To ensure the validity and reliability of the experiment, three constant variables were selected: speeds of 18 (km/h), 36 (km/h), and 54 (km/h), and the same distance was maintained with the obscured traffic sign. A total of 80, 40 and 20 frames were acquired at different speeds, respectively. The part of algorithm in this paper was then employed to conduct one stage of traffic sign detection and tracking. The results of this stage of image processing are presented in Fig. 10.

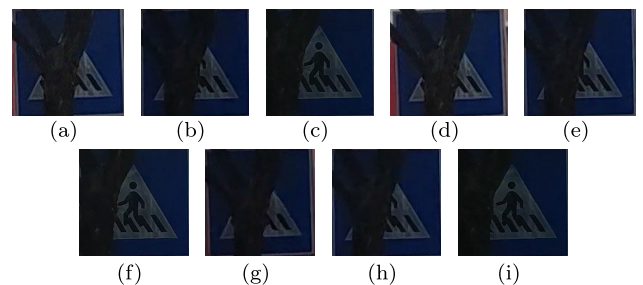


FIGURE 10. Image processing results of one stage. (a)-(c) Three out of eighty frames at 18 (km/h). (d)-(f) Three out of forty frames at 36 (km/h). (g)-(i) Three out of twenty frames at 54 (km/h).

To further evaluate the proposed algorithm, the images were processed with information on obscured traffic signs at intervals of 0 to 3 frames, respectively. Subsequently, image fusion was performed on these processed images. The similarity results are presented in Table 5. Meanwhile, the results after the fusion of images with intervals of 2 frames

are depicted in Fig. 11. Through the above experiments, the algorithm's ability to detect and track obscured traffic signs under different speeds is demonstrated.

TABLE 5. Results of the similarity of traffic sign fusion result maps for different frame intervals and different vehicle speeds.

frame interval	18 (km/h)	36 (km/h)	54 (km/h)
0	0.738	0.728	0.693
1	0.705	0.723	0.676
2	0.706	0.719	0.662
3	0.695	0.701	0.613

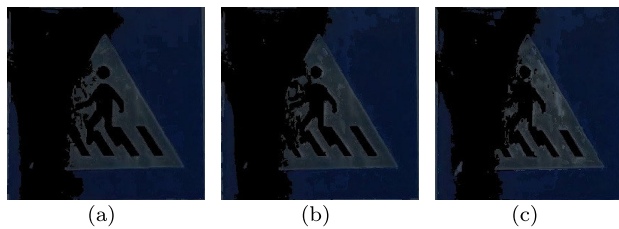


FIGURE 11. The fusion of obscured traffic sign information at different vehicle speeds with a frame interval of two frames. (a) 18 (km/h). (b) 36 (km/h). (c) 54 (km/h).

Based on the findings presented in Table 5, it is evident that the degree of similarity between the traffic sign information fusion map and the template is inversely proportional to the interval between frames, while holding vehicle speed constant. Additionally, it was observed that as vehicle speed increases, this decrease in similarity becomes more pronounced. This phenomenon can be attributed to the widening gap that arises between images undergoing fusion with larger intervals between frames. The increase of vehicle speed further exacerbates this issue, resulting in greater loss of geometric information pertaining to traffic signs. Consequently, the similarity between the traffic sign information map and the template after image fusion decreases significantly. It can be deduced that when navigating through urban roads, a three-frame interval for image fusion may be preferred when traffic speed is low in order to minimize computational overhead. However, when traffic speed is medium, an interval of one or two frames might be ideal. Finally, for relatively high traffic speed in urban roads, no interval or an interval of one frame would be recommended for image fusion.

V. CONCLUSION

Firstly, this paper presents a method for detecting and recognizing traffic signs that are partially obscured during vehicle motion. To achieve this, the image bit operation method is used to extract traffic signs from multi-frame road driving images. Traffic information fusion is then performed to obtain traffic sign images with more complete geometric features. This method is particularly useful when traffic signs are partially obscured during vehicle motion.

Second, the proposed method involves determining the region of interest (ROI) using a color-shape traffic sign detection and recognition algorithm. Groups of multi-frame road driving images with different intervals at different vehicle speeds are then selected and fused to obtain a complete traffic sign image.

Finally, the effectiveness of the proposed algorithm is verified through experiments involving simulated obstacle movement and obscured traffic signs in actual scenes. The results demonstrate that the proposed method is effective in detecting and recognizing traffic signs that are partially obscured during vehicle motion.

The present research paper aims to investigate the detection and recognition of obscured traffic signs within typical urban environments, with a specific focus on cases that do not involve extreme conditions, such as low light intensity, rainy weather, or foggy days. This study does not encompass less common traffic signs found in tourist areas. Additionally, it is essential to note that the algorithm proposed in this paper may not be adequately optimized, resulting in suboptimal recognition speed. Consequently, future research endeavors should incorporate the algorithm presented herein into the YOLO series algorithm of deep learning to enhance both the robustness and inference speed of detecting and recognizing obscured traffic signs.

REFERENCES

- [1] X. Bangquan and W. X. Xiong, "Real-time embedded traffic sign recognition using efficient convolutional neural network," *IEEE Access*, vol. 7, pp. 53330–53346, 2019.
- [2] A. Ruta, Y. Li, and X. Liu, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recognit.*, vol. 43, no. 1, pp. 416–430, Jan. 2010.
- [3] G. Liang, X. Ning, P. Tiwari, S. Nowaczyk, and N. Kumar, "Semantics-aware dynamic graph convolutional network for traffic flow forecasting," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 7796–7809, Jun. 2023.
- [4] C. Cao, B. Wang, W. Zhang, X. Zeng, X. Yan, Z. Feng, Y. Liu, and Z. Wu, "An improved faster R-CNN for small object detection," *IEEE Access*, vol. 7, pp. 106838–106846, 2019.
- [5] C. Liu, F. Chang, and Z. Chen, "Rapid multiclass traffic sign detection in high-resolution images," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2394–2403, Dec. 2014.
- [6] Z. Malik and I. Siddiqi, "Detection and recognition of traffic signs from road scene images," in *Proc. 12th Int. Conf. Frontiers Inf. Technol.*, Dec. 2014, pp. 330–335.
- [7] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
- [8] Y. Zhu, M. Liao, M. Yang, and W. Liu, "Cascaded segmentation-detection networks for text-based traffic sign detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 209–219, Jan. 2018.
- [9] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with Hinge loss trained convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1991–2000, Oct. 2014.
- [10] L. Jiao, R. Zhang, F. Liu, S. Yang, B. Hou, L. Li, and X. Tang, "New generation deep learning for video object detection: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3195–3215, Aug. 2022.
- [11] R. C. Joshi, M. Joshi, A. G. Singh, and S. Mathur, "Object detection, classification and tracking methods for video surveillance: A review," in *Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA)*, Dec. 2018, pp. 1–7.
- [12] Y. Wu, H. Zhang, Y. Li, Y. Yang, and D. Yuan, "Video object detection guided by object blur evaluation," *IEEE Access*, vol. 8, pp. 208554–208565, 2020.

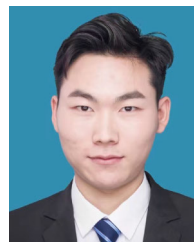
- [13] X. Xu, J. Jin, S. Zhang, L. Zhang, S. Pu, and Z. Chen, "Smart data driven traffic sign detection method based on adaptive color threshold and shape symmetry," *Future Gener. Comput. Syst.*, vol. 94, pp. 381–391, May 2019.
- [14] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Ann. Telecommun.*, vol. 75, nos. 7–8, pp. 369–379, Aug. 2020.
- [15] J. Zhang, X. Zou, L.-D. Kuang, J. Wang, R. S. Sherratt, and X. Yu, "CCTSDb 2021: A more comprehensive traffic sign detection benchmark," *Human-Centric Comput. Inf. Sci.*, vol. 12, pp. 1–19, May 2022.
- [16] M. Masmoudi, H. Friji, H. Ghazzai, and Y. Massoud, "A reinforcement learning framework for video frame-based autonomous car-following," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 111–127, 2021.
- [17] S. Lu, B. Wang, H. Wang, L. Chen, M. Linjian, and X. Zhang, "A real-time object detection algorithm for video," *Comput. Electr. Eng.*, vol. 77, pp. 398–408, Jul. 2019.
- [18] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "ByteTrack: Multi-object tracking by associating every detection box," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 1–21.
- [19] J. Deng, Y. Pan, T. Yao, W. Zhou, H. Li, and T. Mei, "Single shot video object detector," *IEEE Trans. Multimedia*, vol. 23, pp. 846–858, 2021.
- [20] H. Li, B. Zhang, Y. Zhang, X. Dang, Y. Han, L. Wei, Y. Mao, and J. Weng, "A defense method based on attention mechanism against traffic sign adversarial samples," *Inf. Fusion*, vol. 76, pp. 55–65, Dec. 2021.
- [21] J. H. Chung, D. W. Kim, T. K. Kang, and M. T. Lim, "Traffic sign recognition in harsh environment using attention based convolutional pooling neural network," *Neural Process. Lett.*, vol. 51, no. 3, pp. 2551–2573, Jun. 2020.
- [22] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2110–2118.
- [23] C.-J. Li, Z. Qu, S.-Y. Wang, and L. Liu, "A method of cross-layer fusion multi-object detection and recognition based on improved faster R-CNN model in complex traffic environment," *Pattern Recognit. Lett.*, vol. 145, pp. 127–134, May 2021.
- [24] B. S. Shobha and R. Deepu, "A review on video based vehicle detection, recognition and tracking," in *Proc. 3rd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solutions (CSITSS)*, Dec. 2018, pp. 183–186.
- [25] D. Liu, Y. Cui, Y. Chen, J. Zhang, and B. Fan, "Video object detection for autonomous driving: Motion-aid feature calibration," *Neurocomputing*, vol. 409, pp. 1–11, Oct. 2020.
- [26] H. Lou, X. Duan, J. Guo, H. Liu, J. Gu, L. Bi, and H. Chen, "De-YOLOv8: Small-size object detection algorithm based on camera sensor," *Electronics*, vol. 12, no. 10, p. 2323, May 2023.
- [27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [28] Y. Chen, Y. Xie, and Y. Wang, "Detection and recognition of traffic signs based on HSV vision model and shape features," *J. Comput.*, vol. 8, no. 5, pp. 1366–1370, May 2013.
- [29] E. Namazi, R. Mester, C. Lu, and J. Li, "Geolocation estimation of target vehicles using image processing and geometric computation," *Neurocomputing*, vol. 499, pp. 35–46, Aug. 2022.



SHI LUO received the B.S. degree in automatic control from the Shaanxi Mechanical College, in 1992, and the M.S. and Ph.D. degrees in vehicle engineering from Jiangsu University, in 2002 and 2010, respectively. He is currently an Associate Professor with the School of Automotive and Traffic Engineering, Jiangsu University. His research interests include automotive electronics, embedded image recognition and processing, and deep learning in automotive applications.



CHENGHANG WU was born in Zhejiang, China, in 1999. He received the B.S. degree in vehicle engineering from the Ningbo Engineering College, in 2021. He is currently pursuing the M.S. degree with the School of Automotive and Transportation, Jiangsu University. His current research interests include the detection and recognition of traffic signs and the application of deep learning.



LINGENG LI was born in Henan, China, in 1998. He received the B.S. degree in vehicle engineering from the Chung Yuan College of Engineering, in 2020. He is currently pursuing the M.S. degree with the School of Automotive and Transportation, Jiangsu University. His current research interest includes the application of deep learning to the early warning of expected safety features in automobiles.

...