A Review of Occluded Objects Detection in Real Complex Scenarios for Autonomous Driving

Jiageng Ruan[*1], Hanghang Cui[1], Yuhan Huang[2], Tongyang Li[1], Changcheng Wu[1], Kaixuan Zhang[1]

1 Department of Materials and Manufacturing, Beijing University of Technology, Bejing, CHINA

2 Centre for Green Technology, School of Civil and Environmental Engineering, University of Technology Sydney, NSW2007, Australia

Corresponding author address: Beijing University of Technology, 100 Ping Le Yuan, Chao Yang, Beijing, 100123

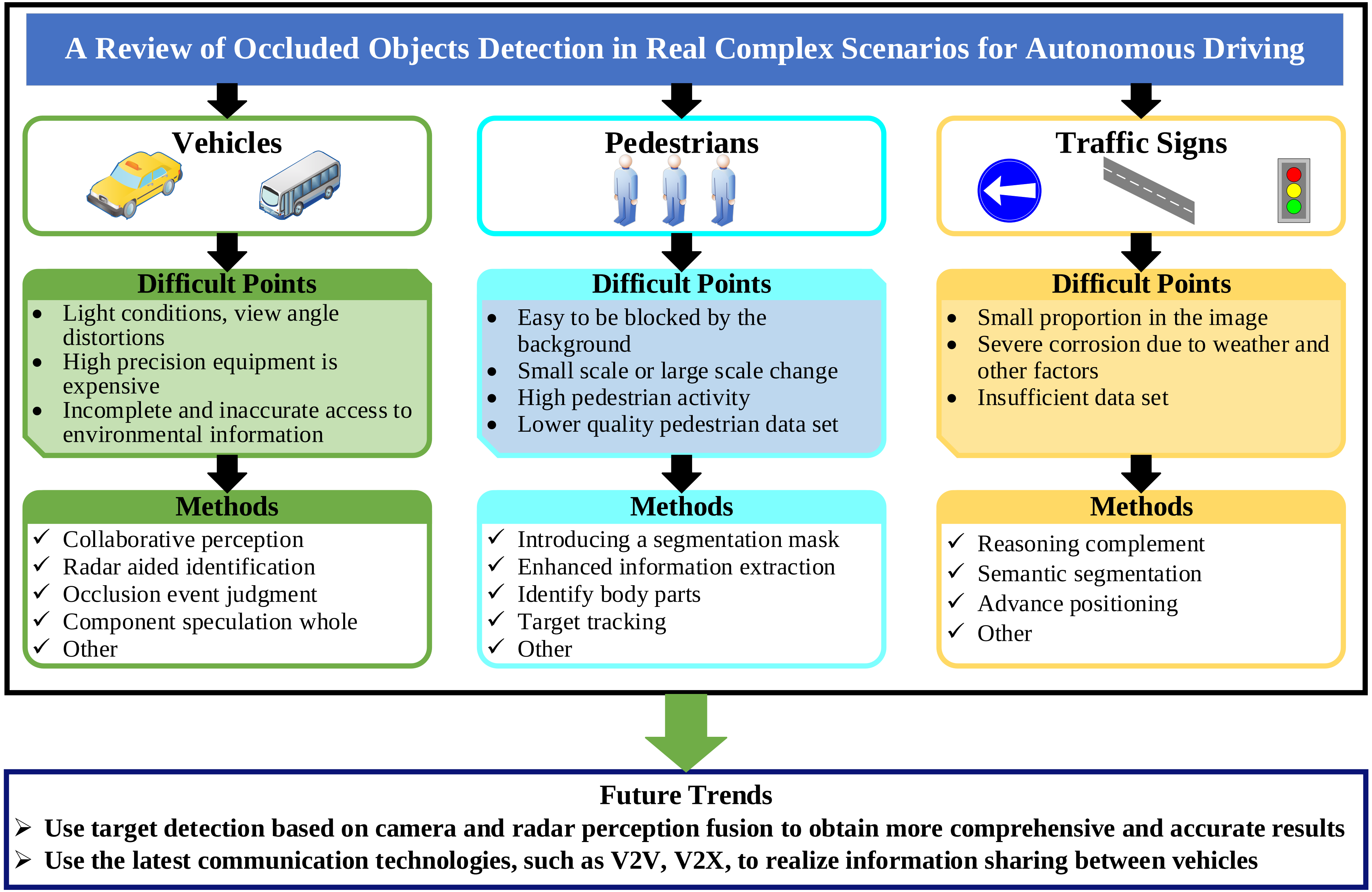
Corresponding author Email: ruanjiageng@bjut.edu.cn

Acknowledgment

This work was supported by the National Key Research and Development Program of China under Grant No. 2022YFE0102700. Dr Yuhan Huang is a recipient of the ARC Discovery Early Career Research Award (DE220100552).

Figure

Graphical Abstract



Highlights:

1. The overlap and differences between various occluded object detection methods are reviewed.
2. Detection methods for occluded vehicles, pedestrians, and traffic signs are studied.
3. The advantages and shortcomings of state-of-art detection methods are compared.
4. Future research trend is predicted based on summarized current solutions.

A Survey of Object Detection for Occluded Objects in Autonomous Driving Scenes

Abstract

Autonomous driving is a promising way to future safe, efficient, and low-carbon transportation. Real-time accurate target detection is an essential precondition for the generation of proper following decision and control signals. However, considering the complex practical scenarios, accurate recognition of occluded targets is a major challenge of target detection for autonomous driving with limited computational capability. To reveal the overlap and difference between various occluded object detection by sharing the same available sensors, this paper presents a review of detection methods for occluded objects in complex real-driving scenarios. Considering the rapid development of autonomous driving technologies, the research analyzed in this study is limited to the recent five years. The study of occluded object detection is divided into three parts, namely occluded vehicles, pedestrians and traffic signs. This paper provided a detailed summary of the target detection methods used in these three parts according to the differences in detection methods and ideas, which is followed by the comparison of advantages and disadvantages of different detection methods for the same object. Finally, the shortcomings and limitations of the existing detection methods are summarized, and the challenges and future development prospects in this field are discussed.

Keywords: Autonomous Driving, Occluded Objects, Object Detection, Vehicles, Pedestrians, Traffic Signs

1. Introduction

As the regulations on carbon emissions are getting tougher around the world, renewable energy-based electric vehicles are treated as the most promising option to address the challenge of climate change. To compete with the traditional internal-combustion-engine vehicle, intelligent driving is taken as an important differentiation competitive advantage by the new energy vehicle. Meanwhile, artificial intelligence technology-based autonomous driving[1] does show a clear blueprint for decarbonized and safer transportation. Functional autonomous driving can be achieved through three basic steps, i.e., environmental perception & recognition[2], planning decision[3], [4], and vehicle control[5]. Fast, accurate, and stable environmental perception & recognition is not only essential information for autonomous vehicles (AV) to complete planning decisions and vehicle control but also the fundamental guarantee for unsupervised safety driving. Object detection, as the primary task in the perception and recognition layer of autonomous driving, has attracted great attention and achieved rapid development in recent years, while is still confronted with great challenges in real scenes with human-vehicle-road-environment complex coupling[6], [7]. The driving environment is complex, dynamic, and unpredictable. The limited and incomplete view of the vehicle, which is obstructed by other vehicles or objects, presents a great challenge to the driving safety of the AV.

Obstacle blocked-view in drivers' sight during driving is a typical impact of complex environments on vehicle self-guiding capabilities, which may lead to a collision in several cases. From partial to complete occlusion, the more parts the object obstructed, the less information vehicle sensors can capture, which makes objects difficult to be recognized.

According to the standard J3016 "Levels of Driving Automation" published by the American Society of Engineers in 2014, autonomous driving capabilities are classified into six levels (0 to 5)[8]. Under Level 4, drivers take the principal responsibility in driving with the help of Advanced Driver Assistance System (ADAS). In other words, vehicles can make decisions based on the information obtained via equipped sensors in Levels 1-3, while the driver is ready to take over the control of the vehicle when necessary. In Level 4, the vehicle is capable to complete the driving task alone, but the driver still can take over the control when he wants. In Level 5, the AV operates independently without driver intervention. Unfortunately, no car maker OEM claims that they have produced a vehicle with Level 4 autonomous driving capability. The biggest challenge is recognizing the objects, which is called perception in autonomous driving, in the surrounding environment accurately[9], [10]. Although the research-oriented and commercialized solutions to recognize the targeted objects in a complex real environment show satisfied accuracy, objects occluded by other items will make the task significantly more difficult.

Zhao et al.[11] and Sharma et al.[12] reviewed the target detection methods, which only focused on the role of deep learning algorithm application. For the autonomous driving scenes, Cui et al.[13] studied the in-depth learning methods used for image and point cloud data fusion processing, which classified the relevant researches according to the fusion level, and identified the gaps and challenges of academic research to the practical application. Tang et al.[14] summarized the studies of target detection and tracking based on radar and vision fusion. Ravindran et al.[15] evaluated the fusion scheme of sensing mode and depth neural network in multi-target detection and tracking. Regarding the specific targets, Badrloo et al., RATEKE et al., and Bouguettaya et al. reviewed the image-based detection of obstacles [16], roads [17], and vehicles [18], respectively. Badrloo et al. compared the monocular and stereo-based obstacle detection algorithm. RATEKE et al. focused on region-based road detection methods for a variety of roads. Bouguetta et al. reviewed the deep learning technology applied in vehicles detection from Unmanned Aerial Vehicles (UAVs) images.

However, the aforementioned studies and reviews mainly focus on the particular methods used in object detection for autonomous driving or particular object detection methods. To the best of the authors' knowledge, very few studies have reviewed, summarized and analyzed the methods for occluded objects, which is one of the great challenges for autonomous driving safety. The scientific contributions of this review are as follows:

- It comprehensively classifies the state-of-the-art studies on occluded objects into three categories according to the targets, namely the detection of occluded vehicles, pedestrians and traffic signs.
- It not only compares various detection methods for the same target, also analyze the same method used in different targets detection cases.
- It provides clear and easy access to useful tools for the researchers and engineers who wants to improve the accuracy of their sensor-based detection system for occluded objects.

The rest of the paper is organized as follows: Section II summarizes the detection methods and applied cases for occluded vehicle detection. Section III outlines the main problems and solutions for occluded pedestrian detection. Section 4 introduces the state-of-the-art methods and related cases in occluded traffic signs detection. Section V concludes the challenges and future research work of occluded object detection in autonomous driving development.

2. Detection of Occluded Vehicles

In real-world driving situations, it is common for vehicles or other traffic participants to block each other due to light conditions, view angle distortions, and road conditions, which pose a huge challenge for object detection in autonomous driving.

2.1 Collaborative Perception-Based Methods

When objects are occluded or too far away, the sensor or lidar in the vehicle usually cannot detect them accurately. However, the development of communication technology, such as Vehicle to Vehicle (V2V)[19], Vehicle to Infrastructure (V2I)[20], and V2X[21], [22], can help vehicles communicate with each other, then understand driving information better. Based on these communication systems, the recognition results of sensors, cameras, and lidars from different vehicles with different angles are aggregated to construct a 3D scene. As shown in Figure 1 below. Then, the 3D scene is fed back to the participating vehicles to improve their perception capabilities, which is known as collaborative perception[23]. Thanks to the V2V communication system, each car can turn the received information into safety warnings after analyzing, encode the warnings into short data packets with numbers, and transmit them to following vehicles through the LED transmitter at the rear of the vehicle. This system based on an LED transmitter and camera can simulate the perspective ability to help the driver perceive the area blocked by the vehicle in front. This non-visual perception system can help the vehicle obtain a virtual see-through ability, helping the vehicle/driver to identify more than 90% of the occlusion. The researchers who proposed the system also added a recommendation generation module to the system. This module can fuse and analyze the driving scene information perceived by the vehicle and the received information, and provide safer and more reliable action suggestions[24]. Although collaborative perception can effectively enhance the perception ability of connected cars, detect occluded targets more accurately, and reduce the possibility of accidents, the lack of an efficient information exchange protocol in the collaborative perception will lead to the same information sent between vehicles repeatedly. The repeated broadcast data pack not only causes data redundancy, and more, also cover or lost key information, which may raise serious safety risks. Aoki et al.[25] proposed a deep reinforcement learning-based collaborative perception scheme. According to the study, the self-car detection field of vision is divided into 15 grids, and the grids state is determined based on the self-car basic information (speed, acceleration, etc.), self-car perception results, received information, and information content. Grids state and network congestion degree are used as state quantities of deep reinforcement learning. The agent performs the action of transmission or abandonment to decide whether to upload the self-sensing information through the V2X communication system. Then, the agent will be rewarded according to the action performed. The reward is determined according to the repeatability, timeliness, and network congestion of information. The lower the repeatability and real-time of the information, the lower the degree of network congestion, and the greater the reward the agent gets. Driven by this reward model, each participating vehicle can obtain rich shared perception information, and at the same time, it also reduces the acquisition of duplicate information. This scheme effectively improves the vehicle detection accuracy and the reliability of the information obtained.

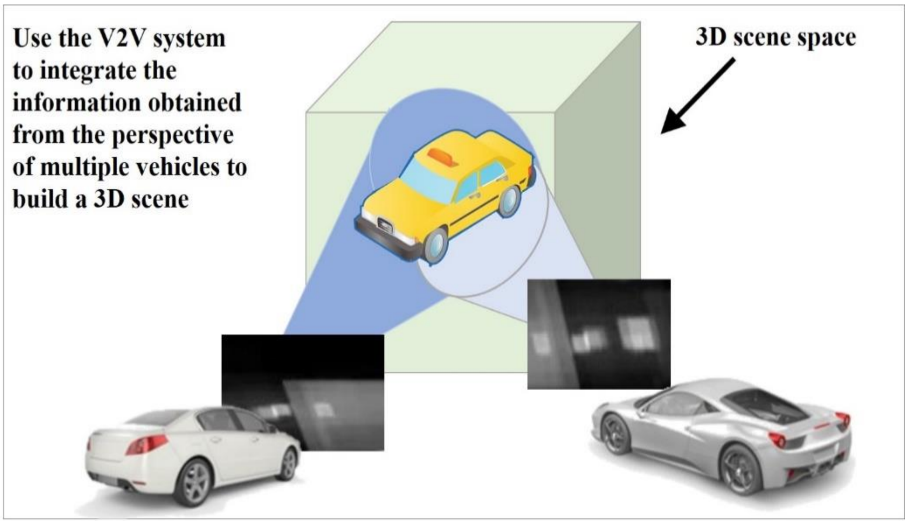


Figure 1. 3D scene fusion

However, it is worth to be noted that the required infrastructure renovation and vehicle device upgrade for V2X communication, such as V2V and V2I, need huge investment and a long time to implement.

2.2 Radar-Based Methods

Considering the initial investment and maintenance costs of the V2X system, a more economical alternative option is to use the available devices and technologies in the current vehicle to detect the occluded vehicles. Radar is a common sensor in the vehicle with a certain degree of penetrating ability, which enables the vehicle to detect and identify obstacles. Compared to other vehicle sensors, radar can capture the speed information directly in a large range and is less affected by weather[26]–[28]. Considering that its penetration ability will degrade when facing distant objects, Pou-Kung et al. use polar sliding window inference to directly apply a short-range trained target detection network to long-range areas of polar space to maintain the radar's penetrating ability[29]. However, due to the limitation of its penetration ability, the current radar information is mostly used to help cars obtain information such as distance and speed of other objects. In specific, when vehicles form a platoon, the obstacle detection area is limited to a relatively small dynamic range, which is based on the distance from the ego car to the lead car, its speed, and acceleration obtained by radar. Then, the vehicle tracking system can generate a safe and reliable driving trajectory by combining the detection results in the dynamic range for the ego vehicle[30]. However, as the length of the vehicle fleet increases, it is dangerous that plan vehicle trajectories only relying on radar-based tracking. To Improve the safety of autonomous vehicle fleets, -Yun Mu et al. [31] proposed a detection method based on YOLOv4 to enhance the detection of the front-vehicle, while the vehicle always maintains the appropriate distance and speed to avoid accidents. The V2X communication system[32] described above also can help vehicles in the fleet to obtain more road information, then reduce the occurrence of collisions. In addition, according to Vivoda. et al.[33], the safety of fleets is related to the road safety support level from their company in a study of 70 companies' fleets. Therefore, in order to enhance the safe driving of the fleet, improving the target detection ability of the front-vehicle, providing efficient communication between vehicles, and strengthening the management of the fleet by the company are feasible solutions.

2.3 Occlusion Judgment-Based Methods

It is common that vehicle controller needs to make decisions based on uncertain and incomplete environmental knowledge[34]. Specifically, the target vehicle could be partially obscured in the sensor’s view due to the relative position, volume, and heading angle to the ego vehicle. could be partially obscured in the sensor’s view due to the relative position, volume, heading angle of the target vehicle, and other factors[35].

Moreover, if driving at high speed, the scales of target vehicles will change dramatically in consecutive photos, which leads to target occlusions with other vehicles in the environment. When vehicle drives at a high speed, it is difficult to identify the size, shape, and location of the obstacle quickly and accurately to avoid potential collision. However, if an occlusion event can be inferred in advance, it can leave more time for the driver and system to drive away from this area safely. Li et al. believe that the lower recognition rate of the target object is, the higher possibility it is blocked[36]. Therefore, a threshold of recognition rate is proposed in their study to determine if the object is obstructed or not by using YOLO V3 for target detection. Similarly, the point cloud obtained by lidar when sampling occluded objects is partial. Taking the lidar data as the original input, the voxelated point cloud data will be converted into a bird's-eye view (BEV) heat map by using statistical coding, which can easily observe the position of objects in the scene. Fully detected objects are represented as elliptical blobs in this map. The occluded object is represented by an incomplete point to indicate potential danger. As shown in Figure 2 below[37]. Following the accurate identification of obstacles, it is necessary to predict the missing part of the cloud, then, supplement the lack of clouds[38]. At the same time, it can also integrate the lidar data and image features to improve the detection accuracy of the obscure object[39], [40].

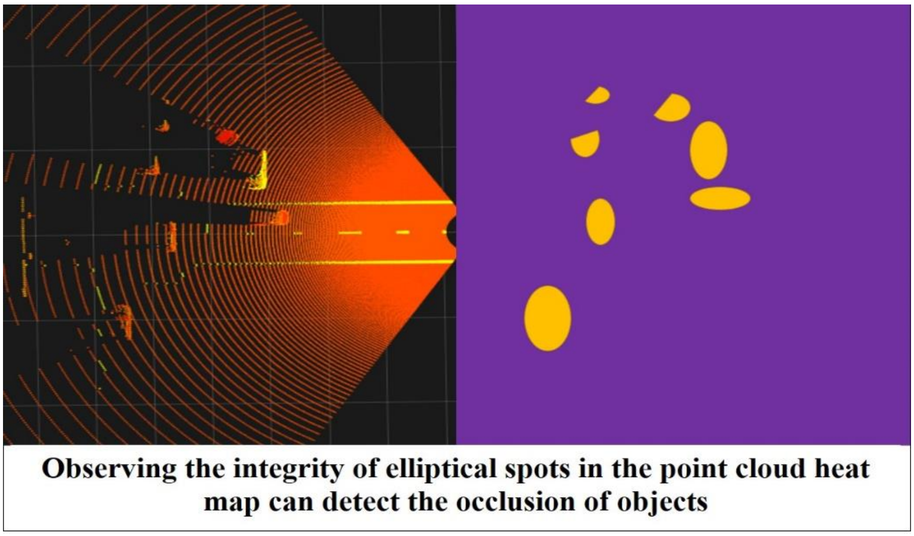


Figure 2. Lidar point cloud(left) and ellipses generated from lidar point cloud (right)

2.4 Component Information-Based Methods

Besides filling the missing data, another feasible way to identify the occluded object accurately is taking the component information as a supplementary feature to the target object[41].

Zhang et al. proposed a method to divide the whole vehicle into independent areas, identifying the parts and the whole target at the same time. Next, coding the parts according to the component affinity field before recombining them into a whole identification frame to improve the recognition accuracy of occluded vehicles. Specifically, they divide the detected object, namely the vehicle, into four semantic parts, i.e., the top, side, front, and back of the vehicle, as shown in Figure 3. Since the characteristics of the left and right sides of the vehicle are generally the same, it is impossible to appear in the front or back of a car at the same time in one image. Therefore, there are only three semantic parts in the test process to encode a combination[42]. Although the above method can achieve satisfactory results in terms of the accuracy of vehicle occlusion detection, the images collected by the camera could be distorted due to the illumination or re-transformation of the image. For example, if a detected object is at the edge of a picture, it will be angularly distorted in the IPM (Inverse Perspective Mapping) domain. The distortion could be restored through the relationship of the vertical edge of the obstacle in the camera image and the corresponding angular distortion slope in the IPM domain is used, the initial shape of the object can be inferred. In consequence, the detection accuracy of the occluded object under different illumination can be improved[43].

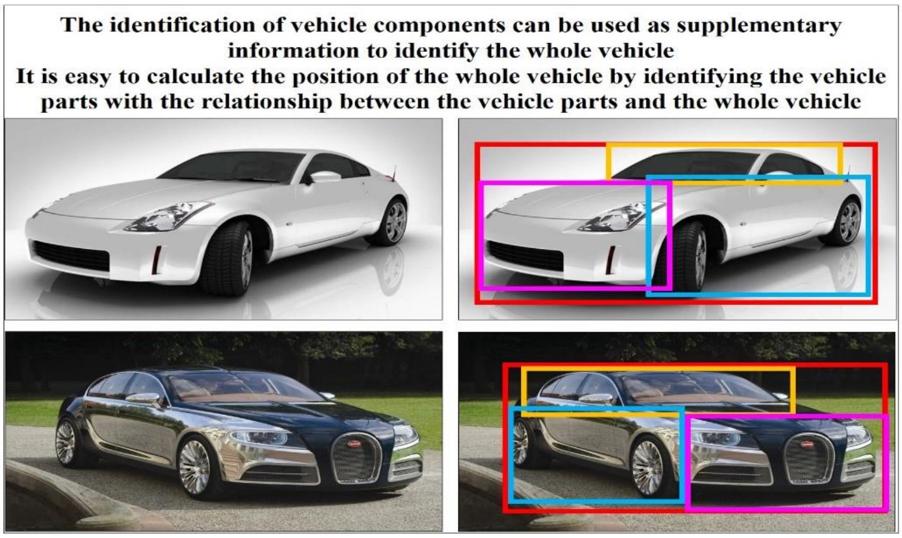


Figure 3. Original image and semantic segmentation of vehicle

2.5 Other Methods and Development Trends

There are some alternative solutions to reduce the impact of object occlusion on detection accuracy. For example, installing a microphone array in the vehicle to identify the occluded vehicles via acoustic detection; Active adjustment of vehicle lateral position to achieve maximum visibility when the view is occluded. However, the first solution is easily affected by environmental noise, while the latter one heavily relies on experienced drivers, The adaptability of both methods is relatively poor[44], [45].

Given the limited detection ability of a single vehicle, especially in the complicated driving scenario, collaborative perception is a promising solution to obtain the information of occluded objects with the development of intelligent connected vehicles.

The information brought by collaborative perception is relatively comprehensive, which can help solve the occlusion problem to some extent. However, the real-time and accuracy of the data are difficult to be guaranteed. In addition, the connected vehicles are vulnerable to the network malicious attacks, especially the unauthorized remote control via the hostile attack of wireless remote interfaces[46]–[48]. Therefore, the object detection methods based on available equipment in the vehicle needs to be further improved before the infrastructure updates for V2X is complete. Aforementioned methods are summarized in Figure 4.

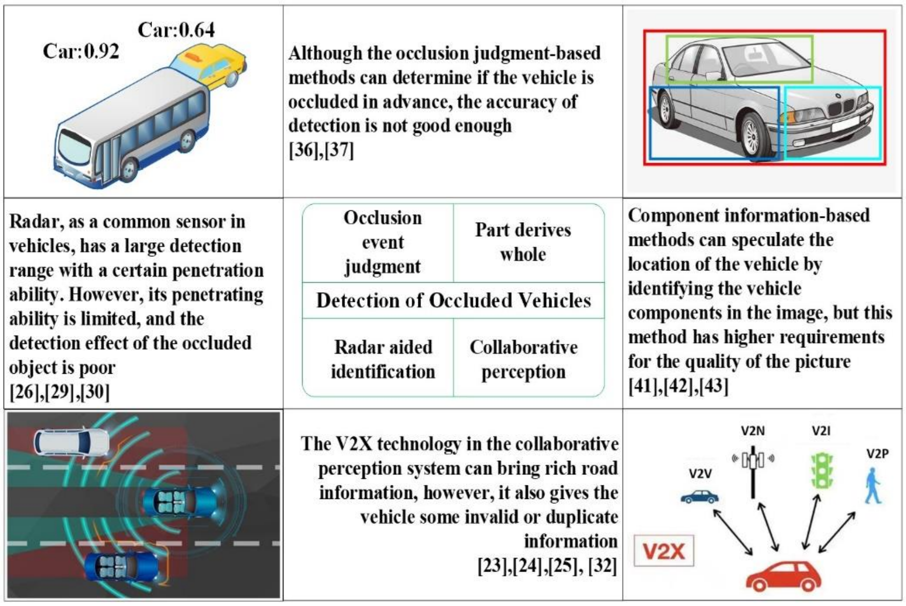


Figure 4. Summary of occluded vehicles detection methods

3. Detection of Occluded Pedestrians

Pedestrians, as the most important participant in the driving scene to be protected, is always the key focus and difficult point of AV detection ability improvement. In practice, there are two major challenges to accurately identifying pedestrians. First, pedestrians are often occluded by other pedestrians or objects since they are small and usually on the moving. Second, the scales of pedestrians are usually small and may vary over a wide range.

3.1 Segmentation Mask-Based Methods

In image processing, segmentation mask can help extract regions of interest while block other regions. Pedestrian occlusion caused by background information can be solved by adding segmentation masks. The segmentation mask is an attention mechanism to make the detector to focus on image regions where potential candidate objects may appear. Integrating the attention mechanism[49] into the target detector will effectively enhance the detection ability of occluded objects via adjusting the network weights in training without learning additional parameters, which will not increase the burden of calculation in the inference process. Meanwhile, the additional segmentation mask channel will also guide the convolution kernel to learn more discriminative features automatically, which makes it easier to distinguish the background from foreground. Another practice method to distinguish the prospect and background in the image is to convert the segmentation mask to a binary cover, in which the cover value uses 1 to indicate that the content of the pixel is the foreground part, and 0 represents the background part. This method could make the target detector to focus on the part with a mask value of 1, and help them learn the characteristics of the detection target, and then identify the occluded object[50].

3.2 Enhanced Information Extraction-Based Methods

Regarding pedestrian detection for AV, the occlusion and the scale change problems often appear simultaneously. Furthermore, the scale varying will increase the difficulty of detecting occluded pedestrians. The input information in different scales has higher requirements for the information extraction ability of the target detector. To meet the challenge of pedestrians' scale change in the detection area, a multi-scale extraction mechanism is proposed to extract more features before achieving satisfactory recognition accuracy[51]. Pyramid network[52], [53], including Image Pyramid Network (IPN) and Feature Pyramid Network (FPN), is a popular extraction network for multi-scale objects. Yin et al. proposed a pedestrian detection method based on an improved FPN, which uses the bottom layer of FPN to predict and detect objects with different scales. Considering the difficulty of obtaining useful information is the major challenge to accurately detect the small-scale and occluded targets, the proposed method adopts densely-connected blocks to enrich the semantic information. The densely-connected block is a combination of the channel attention module and the global attention module. The channel attention module weights each channel according to the importance of the channel to guide the network to emphasize the visible parts of pedestrians, then, solve the occlusion problem. The global attention module can enable the network to obtain the environmental information and spatial relationship of small-scale targets and occluded targets. In consequence, it will help the target detector obtain more feature information of the small-scale targets and occluded targets to improve the recognition accuracy[54]. Compared to FPN, the IPN takes a long time, and the real-time performance is insufficient, so it is difficult to apply to the field of automatic driving, and the feature pyramid network can well make up for the disadvantage of the image pyramid network. Involving the residual network and densely-connected convolutional network to the FPN shows its potential to improve the utilization of input information and the extraction of effective features, which are the keys to effectively increase the detection accuracy of occluded objects[55], [56]. In recent years, some studies use transformer encoder blocks to replace the convolutional blocks and CSP neck blocks in the magnetic head parts of original YOLO V4[57] and V5[58], which is inspired by the visual converter, to generate a transformer predictor head (TPH)[59]. Compared with the original bottleneck block, the transformer encoder block can effectively capture the global information and background information, then, help the detector to identify the occluded and denser objects accurately. The structure of the transformer encoder block is shown in Figure 5.

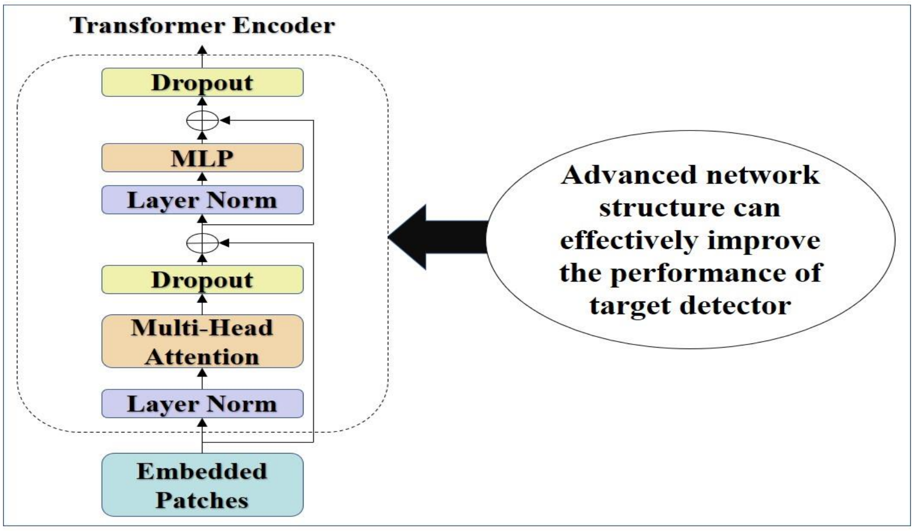


Figure 5. Transformer encoder architecture

3.3 Human Parts Detection-Based Methods

When the pedestrian is in the occluded state, they may appear in a particular part of the body and different motion states. Using the body part information to identify the occluded pedestrians provides a new direction for occluded object detection. In [60], comparing the conventional pedestrian detection neural network with the upper and lower body pedestrian detection neural network, it is found that the two detection methods show similar performance when the mAP (mean Average Precision) of pedestrian completely detection is above 0.75. Therefore, the conventional pedestrian detection model can be replaced by a combination of the upper and lower body detection models without losing accuracy when part of the pedestrian is occluded. Two similar CNN-based occluded pedestrian detection methods are proposed by Islam et al. The first one uses a target detector to identify the parts of the human body. If the confidence of a body part is not less than the predefined threshold for it, it will be marked as an occluded pedestrian. To avoid setting various thresholds for each body part, they use the detected body part with the highest confidence as the basis for occluded pedestrian detection. Based on this assumption, the second method only focuses on the body part with the highest confidence before comparing the confidence score with a fixed threshold. The fixed threshold is set as 0.8, which lacks appropriate justification in this study, to reduce the possible false detection when the threshold is too low[61].

When using body parts to detect occluded pedestrians, it may make misidentification of some similar objects due to the missing interrelated information between human body parts. Based on the relationship between human body parts, Flores-Calero et al. propose a novel classifier for occluded pedestrians. The classifier divides the image containing pedestrians into 12 regions, as shown in Figure 6. In each region, HOG (Histogram of Oriented Gradients) descriptors are used to capture features. Then, SVM (Support Vector Machine) classifiers are to identify body parts. Once body parts are identified, Logic Inference (IL) algorithm is adopted to integrate all regions, in other words, when four regions are classified as true, the detected object is taken as a pedestrian. Testing results show the specialized occluded pedestrian classifier is efficient and stable[62]. Zhang et al. used the structural prior information to divide the human body area into five parts and extract the features through an adaptive pooling unit. Finally, combining the obtained features via the five sub-networks to improve the detection accuracy of occluded pedestrians[63]. Although all the body parts can be used as features to detect the occluded pedestrian, the features of head and shoulders of pedestrians are more stable and not easy to block compared to other body parts. Therefore, LIN et al. improved the architecture of a DetNet. The improved DetNet can generate detection frames for head, shoulders, and bodies simultaneously. To further improve detection accuracy of occluded pedestrian, the prediction weight of different body parts is dynamically adjusted via a built-in mechanism[64]. In future studies, a survey may be necessary to investigate the possibility of occluded body parts in common driving scenarios to reasonably distribute the weighting factors between different parts.

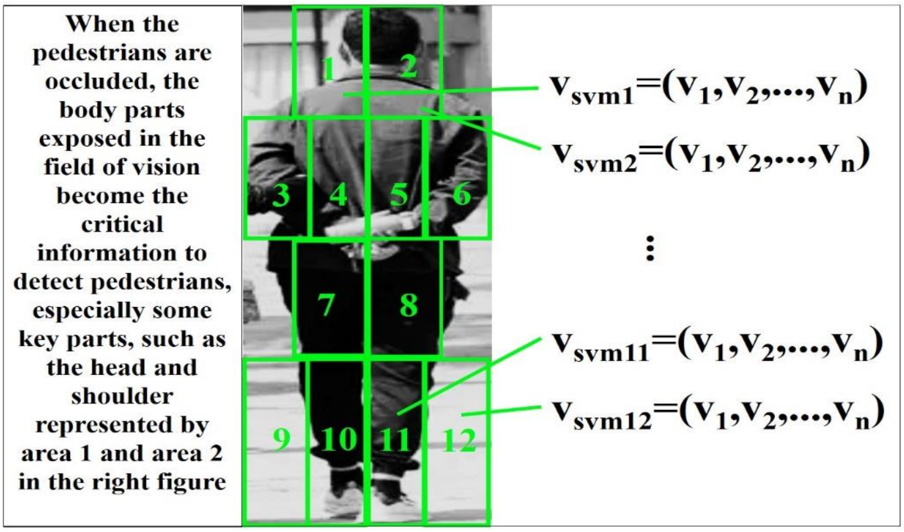


Figure 6. SVM based identifier in divided regions

3.4 Target Tracking-Based Methods

In addition to the partial occlusion problems of pedestrians, their kinetic characteristics affect the detection accuracy. For example, the object moving to a blind spot of the camera will lead to occlusion, however, it will appeal in the view quickly after the pedestrian passes the blind spot. Therefore, it is necessary to predict the trajectory of pedestrians to the existing target tracker has been able to predict target trajectory, although the accuracy can be improved due to the inevitable noise. An effective target tracker can predict occluded pedestrian paths in advance, which will benefit the following path planning[65].

In specific, the pedestrians, who appeared in the view of sensors, can be regarded as a short-term prediction task according to the time series, to calculate the position when it is occluded. And when the occluded pedestrian reappears, the target detection result will be sent to the re-identification network to extract features and compared with the features in the previous pedestrian trajectory to find the original trajectory without switching IDs frequently[66], [67]. Other successful attempt includes the combination of predicted target trajectory with the depth information from monocular depth estimator, which helps detect the occluded pedestrian in 3D scenes[68]. Another promising solution to occluded pedestrian detection is the long-term memory neural networks, which can track the time status of the target through the storage information, which is also an effective means to overcome[69].

Another challenge in occlusion detection is the ‘frame dropping’ that comes from continuous frames including state changes. In terms of AV, it means that the target detection algorithm will ignore the distant target when it just comes into the view, although the target will be recaptured after a very short period, which significantly reduces the response time of drivers and brings great risk. Considering the time interval of consecutive images is short, [70] takes the target state as a continuous variable. They use Kalman filter to predict the target's future state, followed by the Hungarian algorithm used for matching, which constitutes a target-tracking algorithm like human memory. To ensure real-time performance, target tracking can only be performed on images within a few frames. In this way, the occluded object can be detected more accurately. In addition, the combination of optical flow estimation and pixel generation-based prediction methods can also be used to predict occluded objects in high-quality videos with the help of learnable masks[71].

3.5 Other Methods and Development Trends

Another challenge of pedestrian detection is the low quality of pedestrian datasets for algorithm training. The most of current pedestrian datasets are collected by the cameras on the vehicle during driving, which suffers poor background diversity. Particularly, the scenes containing crowded, overlapping, and mutually occluded pedestrians are insufficient, and the density of pedestrians is relatively low in the dataset. For example, the average number of pedestrians per image in the commonly used Caltech dataset and KITTI dataset is less than 1. The performance of pedestrian detectors trained by these datasets is limited. To improve the data quality, Zhang et al. constructs a new dataset for pedestrian detection training. They crawl images from various search engines such as Google, Bing, and Baidu, and use pHash mechanism to remove duplicate images. To increase the pedestrian density in the dataset, they filtered out images with low pedestrian density. The 13,382 images obtained were randomly divided into three subsets: training, validation, and testing with 8000, 1000, and 4382 images in each category, respectively. There are 28.87 persons per image on average. Compared with the current popular datasets, the locations of pedestrians are evenly distributed in the image except at the top. Given the different search strategies of engines, images collected from different search engines show different characteristics in similar scenes, which is a great help to train the object detectors[72].

Although above-mentioned achievements improve the performance of object detectors, too many pedestrians in a single captured view will significantly aggravate the computational burden and reduce the prediction accuracy. A screening module can be included in the detection process to remove the pedestrians who do not affect the route planning and driving actions, which will effectively reduce the computational burden.

In summary, as shown in Figure 7, the aforementioned four pedestrian detection methods, e.g., attention mechanism to distinguish foreground and background, multi-scale extraction network to obtain more useful information, body parts' relationship-based occluded pedestrian detector, and time correlation-based pedestrian route tracking, all achieved outstanding performance. However, it is worth knowing that using body part information and supplementary pedestrian datasets to detect occluded pedestrians needs an extra object detector (network), while other pedestrian detection methods can use the existing detector (network) when a richer dataset containing pedestrians and other targets is available. According to the investigation, optimizing the structure of the detector network, improving the information extraction and anti-interference (background, building, and other object information) ability of the detector, and sharing the detector between pedestrian detection and other object detection is the future development trend of pedestrian detection. In addition, as other occluded object detections, the popularization of V2X is still a promising solution although difficulties in infrastructure and communication is predicted.

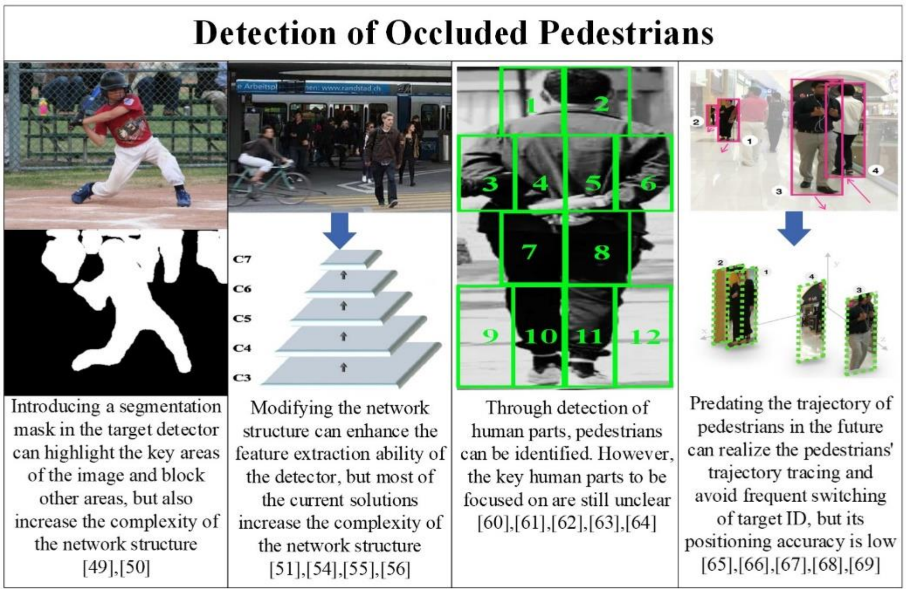


Figure 7. Summary of occluded pedestrian detection methods

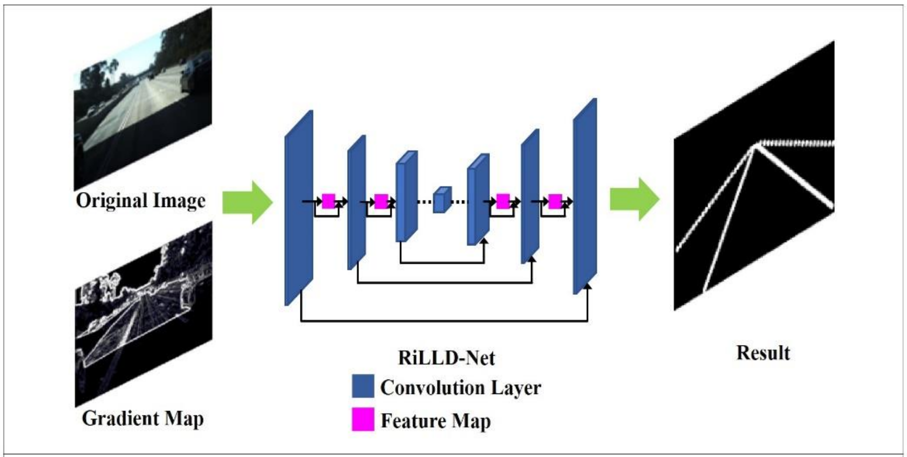
4. Detection of Occluded Traffic Signs

As an important element in AV driving scenarios, traffic signs provide drivers with indispensable road information, which generally includes lane lines, traffic lights, and road signs. Given the traffic sign is usually a small part of the image with relatively fixed positions, it is easy to be occluded by shadows.

4.1 Lane Lines

In general, the purpose of detection for AV is to detect and identify obstacles on the road, rather than all the obstacles in view, especially on the highways[43]. Therefore, it is important to identify the lane lines to improve the accuracy and efficiency of obstacle detection in AV. However, the slender lane lines usually only account for a few pixels in the image, which makes it difficult to detect accurately.

Zhang et al. proposes a novel network-based lane line detector that involves an U-Net encoder-decoder, a residual module with skip connections, and fast connections between the encoder and decoder. Its diagram is shown in Figure 8. The U-Net encoder-decoder is able to position the targets in the background accurately, and the rapid connection between encoders and decoders can help the decoder to lock the lane line details well. Skip connections within residual modules between adjacent feature layers allow the network to process different features in various appropriate ways, enhancing the ability of learning deeper features for lane line detection. The network applies skip connections to all sampling paths to improve learning efficiency and make the most use of features. To better cope with complex and occluded lane lines, Zhang design a new detector by integrating the above-mentioned network, Wasserstein generative adversarial network (WGAN), and multi-objective semantic segmentation together. The proposed network uses the generation ability of WGAN to complete the occluded or missing parts of lines. Then, it divides the collected image into three parts: lane, background environment, and lane line to restore the occluded or incomplete lane lines[73].



The lane line characteristics, neural network, fitting curve, etc. can be used to speculate and generate the lane line ahead. The above figure shows the process of using neural network to generate the lane line

Figure 8. The diagram of the network for lane line detection

Compared to identifying the whole lane lines, it is easier to detect the key points of the lane boundary. A simple and fast lane line detection method proposed by [74], first, converts the camera view to a bird's-eye view via inverse perspective mapping; Then, fits the detected key points of lines by curve fitting method; Finally, it predicts the missing or occluded parts of lane lines based on the characteristics of equidistant and parallel. It performs relatively well when the lane line is simple, but it is difficult to find a suitable fitting curve to predict the missing or occluded lines in complex intersections or areas with many curves. Another promising solution is based on lidar point cloud data, which projects the point cloud data into the bird's-eye view, then, uses a hidden point removal algorithm to classify the data into visible and occluded categories. It is followed by the procession of classified data through a pre-trained neural network to predict road boundaries[75].

Semantic segmentation[76], as a popular method for lane line detection, has achieved convincing results. However, the current CNN-based semantic segmentation network for lane line detection needs to accurately classify each pixel point and use post-processing operations to infer the lane information. The method works well in most cases but shows poor performance when part of the line is occluded where the lines only occupy a small proportion of the image. As a response to the challenge, to reduce the requirements for classification of each pixel to a few points along the lane boundary, take the lane detection and classification problem as the task of CNN regression to predict the occluded lane lines based on the boundary points[77]. Traditional road segmentation uses images as input to mark the roads in the visible areas, while some scholars try to use semantic representation to predict the occluded parts of the road. Unfortunately, the proposed solution can only identify the road boundary in 'ideal' condition and fail to detect the lane lines inside the road area[78].

4.2 Traffic Lights

Unlike the requirements of other traffic markers detection, the identification of traffic lights should be done from a certain distance that ensures an effective warning for AVs, which requires the object detector to be able to identify traffic lights accurately when it appears as a small object in the image. Since the traffic lights usually appear at intersections only, there are many studies taking this advantage to help target detectors narrow down the regions of interest in the images, in consequence, reduce the computational costs and false detection rates. For example, using digital maps to lock traffic lights, marking the shape and position of traffic lights in advance, using vehicle position and road information to find traffic lights, etc.[59]–[62]. However, small-scale targets are easily obscured by relatively large objects in the background that beyond the ability of aforementioned methods. Semantic segmentation can classify the objects at the pixel level with the consideration of background information, and successfully identify the blocked traffic lights at a great distance. Guo et al. proposed a traffic light recognition method by extracting traffic light areas through semantic segmentation and classifying them via a CNN classifier[83]. The recognition process is shown in Figure 9. This proposal improves the detection accuracy of remote traffic lights by 12.8% compared to the vision-based object detector. Furthermore, the classifier shows more than 30% accuracy in color identification than the color threshold-based classifiers.

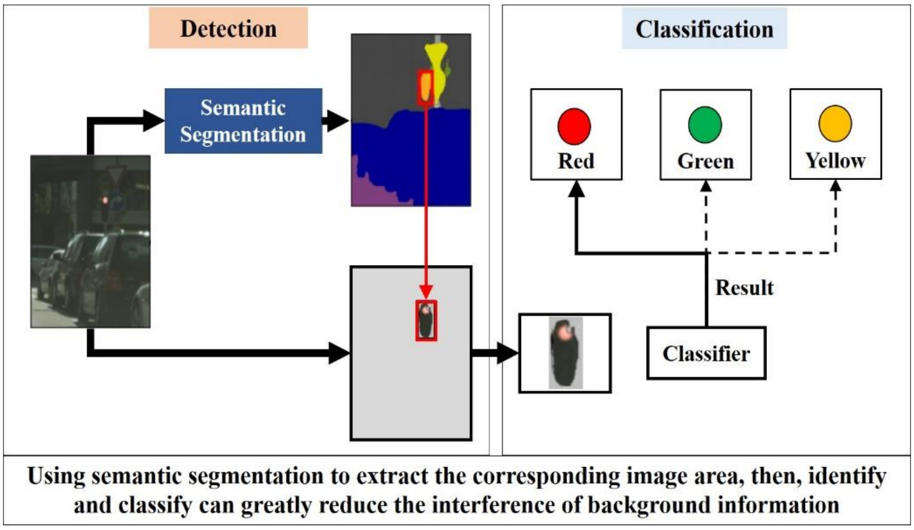


Figure 9. Process flow of traffic light recognition method

4.3 Traffic Signs

There are many types of traffic signs in real driving scenes with significant differences in shape and size. Like the traffic lights, it usually takes a small proportion of images, but is more difficult to detect considering the variable and complex content, potential occlusion, and susceptibility to light[84]. Moreover, occluded road signs can easily convey inaccurate or even wrong traffic information to the AV controller, resulting in traffic accidents. In addition, the integrity of road signs in the wild is considerably difficult to guarantee, which is easily blocked by trees or their shadows. Since the traffic signs are usually coated with a layer of highly reflective material, LiDAR is capable to determine the location in the image and crop the appropriate image for image-based object detection[85].

Similar to pedestrian detection, lacking a sufficient number of images that contains the occluded traffic sign makes it difficult to further improve the performance of existing image-based occluded road signs detector. In addition to creating new appropriate datasets, feature representation transfer learning is another promising solution to the problem of insufficient data. Based on this spot, the Occlusive Symbol Classification Network (OSCN) proposed by Guo et al. shows good performance in occluded traffic sign detection. Compared to the SSD (Single Shot MultiBox Detector) model, the average accuracy of OSCN reaches 96.34%, while t SSD is only 51.26%.

Although the road sign is only a small part of the image, its location is relatively stable in the image, which provides a hot spot region to locate and identify the road signs more quickly and accurately. Researchers can construct a distribution map of detection target locations based on a large number of images containing detection targets, and mark key distribution areas on the distribution map, as shown in Figure 10 below. Rehman et al. combined the above CTF (coarser-to-fine) method with DVL-patches (discriminative vocabulary learned patches) to design an occlusion processing framework. Compared to the method in [86], its mAP on the GTSDB (German Traffic Sign Detection Benchmark dataset) training set is 2.2% higher than the average[87]. Traffic indicators, such as traffic lights and road signs, with relatively fixed positions, where an ambient occlusion model can be constructed using the size and location of nearby objects that may cause occlusion Then, the model can be integrated into the target detector to increase the weight of the corresponding area searching. An optimized searching mechanism allows the detector to achieve higher prior awareness [88]. In this way, the background and obstacle category information can be used more effectively to improve the recognition accuracy of occluded road signs[89].

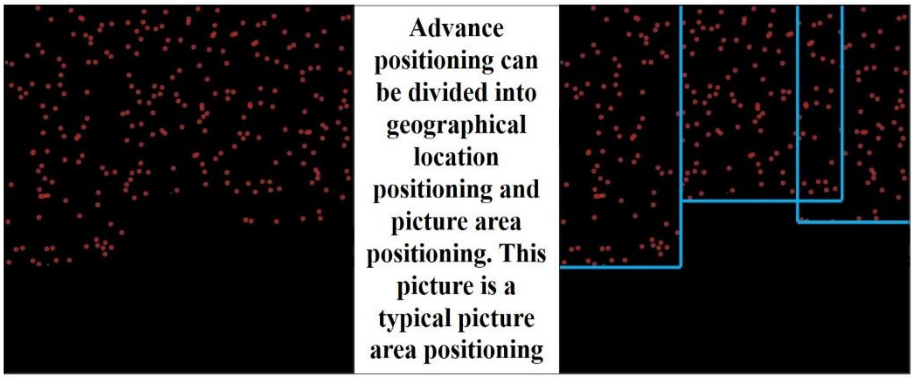


Figure 10. Detection target location distribution map (Left). Key distribution areas (Right)

Some studies try to use multi-level feature extraction methods, deformable convolution, and spatial attention mechanisms to extract richer features to improve the detection accuracy of occluded or small-scale traffic signs[90], [91]. It is worth to know that the optimization of the dataset can help methods of object detection to extract more features[92]. Popular data processing optimization enhancement strategies include Mix Up, Cut Out, Cut Mix, Mosaic, blur, brightness, contrast, hue, and grayscale. we can filter out a more appropriate data set that can be obtained from the optimized and enhanced processed raw data to train the target detector to achieve higher recognition accuracy.

4.4 Other Methods and Development Trends

Although the types of traffic sign are in great number and usually account for a small proportion in the image, which make them hard to detect, they share some similar characteristics. For example, the equidistance and parallelism of lane lines, the high reflective intensity of traffic signs, and the relatively fixed position of traffic lights and signs, etc. As shown in Figure 11, proper use of these features can help intelligent vehicles accurately identify the traffic signs, and even determine the location of traffic signs in advance. With the continuous development of intelligent connected vehicle, the information in the traffic signs obtained in advance via intelligent map, GPS, and V2X systems is more accurate and reliable than that obtained by detection system. It is not difficult to find that the intelligent connected vehicles and intelligent infrastructure are the future trends to solve the detection problem of blocked traffic signs.

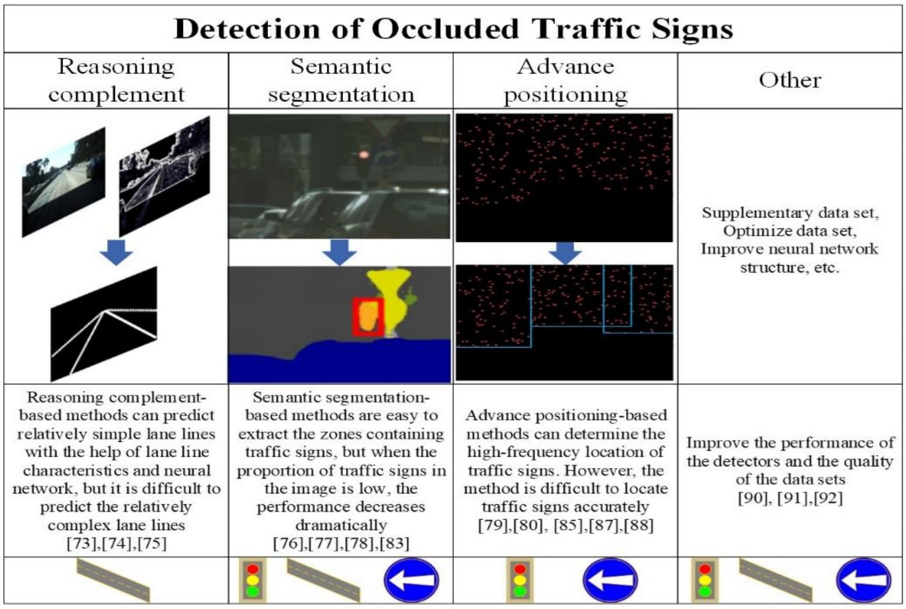


Figure 11. Summary of occluded traffic signs detection methods

5. Summary and Conclusion

This table clearly shows the target detection methods of occluded objects summarized in this paper. In addition, it briefly explains the pros and cons of the reviewed methods from different viewpoints.

Table 1. Summary of target detection methods for occluded objects

Object to be detected	Methods	Advantages	Disadvantages	Related articles
Vehicles	Collaborative perception	Rich information and data interchange	Information redundancy and high cost	[23],[24],[25],[32]
	Radar aided identification	Capable of capturing speed and other information, with a large measurement range	Penetration ability is easy to degrade	[26],[29],[30]
	Occlusion event judgment	Sufficient reaction time is reserved for the vehicle	Unable to accurately identify the occluded object	[36],[37]
	Component speculation whole	Can accurately identify the occluded object	No specific area division standard	[41],[42],[43]
	Other			[44],[45]
Pedestrians	Introducing a segmentation mask	High attention in key areas	Increased network complexity to a certain extent	[49],[50]
	Enhanced information extraction	More information can be obtained from the input	Ditto	[51],[54],[55],[56]
	Identify human parts	Human body parts are strongly correlated	Part division and detection focus is not reasonable enough	[60],[61],[62],[63],[64]

	Target tracking	Avoid id switching, and can estimate the position when it is blocked	Low positioning accuracy	[65],[66],[67] [68],[69]
	Other			[70],[71],[72]
Traffic Signs	Reasoning complement	Low detection difficulty	It is difficult to predict complex road conditions	[73],[74],[75]
	Semantic segmentation	Locate the area of interest	When the target ratio is low, the performance is poor	[76],[77],[78], [83]
	Advance positioning	Identify high-frequency regions in advance	Difficult to pinpoint	[79],[80], [85], [87],[88]
	Other			[90],[91],[92]

At present, the target detection algorithm of occluded objects for AV has achieved considerable improvement. However, there is still a way to the commercialization of L4/L5 autonomous driving. Given the price of Lidar, the current most target detection algorithms still rely on the camera. However, one of the conclusions that can be made according to the review in this study is that thanks to the continuous development of fusion algorithms and the increased use of Lidar on large-scale commercial vehicles, the target detection based on the fusion of camera and radar perception will achieve more comprehensive and accurate results, finally, dominate the detection algorithms in the future. Another conclusion that can be summarized is that, based on the shortcoming analysis of current available theoretical research and technology development, the limited detection and computational ability is a huger barrier to achieving comprehensively driving scenario understanding to avoid any potential accident. Moving the computation-based detection and decision from local (vehicle) to network hub (cloud computation platform) by taking the advantage of latest communication technologies, e.g., V2V, V2X, etc., to realize the information sharing between vehicles, is the most promising solution to the commercialization of L4/L5 AV.

Target detection of occluded objects is one of the most challenging issues faced by the commercialization of autonomous driving. This paper summarizes the state-of-art researches on detection and identification methods for occluded objects in real driving scenarios, including vehicles, pedestrians, and traffic signs, which makes this study different to other object detection researches in AV.

Acknowledgment

This work was supported by the National Key Research and Development Program of China under Grant No. 2022YFE0102700. Dr Yuhan Huang is a recipient of the ARC Discovery Early Career Research Award (DE220100552).

References

[1] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, “A survey of deep learning techniques for autonomous driving,” *J. F. Robot.*, vol. 37, no. 3, pp. 362–386, 2020, doi: 10.1002/rob.21918.

[2] X. Zheng, Y. Li, D. Duan, L. Yang, C. Chen, and X. Cheng, “Multivehicle Multisensor Occupancy Grid Maps (MVMS-OGM) for Autonomous Driving,” *IEEE Internet Things J.*, vol. 9, no. 22, pp. 22944–22957, 2022, doi: 10.1109/JIOT.2022.3187827.

[3] Z. Jian, S. Chen, S. Zhang, Y. Chen, and N. Zheng, “Multi-Model-Based Local Path Planning Methodology for Autonomous Driving: An Integrated Framework,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4187–4200, 2022, doi: 10.1109/TITS.2020.3042603.

[4] H. Li, W. Liu, C. Yang, W. Wang, T. Qie, and C. Xiang, “An Optimization-Based Path Planning Approach for Autonomous Vehicles Using the DynEFWA-Artificial Potential Field,” *IEEE Trans. Intell. Veh.*, vol. 7, no. 2, pp. 263–272, 2022, doi: 10.1109/TIV.2021.3123341.

[5] S. H. Lee and C. C. Chung, “Autonomous-Driving Vehicle Control with Composite Velocity Profile Planning,” *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 5, pp. 2079–2091, 2021, doi: 10.1109/TCST.2020.3029713.

[6] D. Feng, A. Harakeh, S. L. Waslander, and K. Dietmayer, “A Review and Comparative Study on Probabilistic Object Detection in Autonomous Driving,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 9961–9980, 2022, doi: 10.1109/TITS.2021.3096854.

[7] L. H. Wen and K. H. Jo, “Deep learning-based perception systems for autonomous driving: A comprehensive survey,” *Neurocomputing*, vol. 489, pp. 255–270, 2022, doi: 10.1016/j.neucom.2021.08.155.

[8] K. P. Divakarla, A. Emadi, and S. Razavi, “A cognitive advanced driver assistance systems architecture for autonomous-capable electrified vehicles,” *IEEE Trans. Transp. Electrif.*, vol. 5, no. 1, pp. 48–58, 2019, doi: 10.1109/TTE.2018.2870819.

[9] H. H. Jebamikyous and R. Kashef, “Autonomous Vehicles Perception (AVP) Using Deep Learning: Modeling, Assessment, and Challenges,” *IEEE Access*, vol. 10, pp. 10523–10535, 2022, doi: 10.1109/ACCESS.2022.3144407.

[10] Y. Li and J. Ibanez-Guzman, “LiDAR for Autonomous Driving,” *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 50–61, 2020, [Online]. Available: http://www.hesaitech.com/en/autonomous_driving.html.

[11] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, “Object Detection with Deep Learning: A Review,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[12] V. Sharma and R. N. Mir, “A comprehensive and systematic look up into deep

learning based object detection techniques: A review," *Comput. Sci. Rev.*, vol. 38, p. 100301, 2020, doi: 10.1016/j.cosrev.2020.100301.

[13] Y. Cui *et al.*, "Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 722–739, 2022, doi: 10.1109/TITS.2020.3023541.

[14] X. Tang, Z. Zhang, and Y. Qin, "On-Road Object Detection and Tracking Based on Radar and Vision Fusion: A Review," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 5, pp. 103–128, 2022, doi: 10.1109/MITS.2021.3093379.

[15] R. Ravindran, M. J. Santora, and M. M. Jamali, "Multi-Object Detection and Tracking, Based on DNN, for Autonomous Vehicles: A Review," *IEEE Sens. J.*, vol. 21, no. 5, pp. 5668–5677, 2021, doi: 10.1109/JSEN.2020.3041615.

[16] S. Badrloo, M. Varshosaz, S. Pirasteh, and J. Li, "Image-Based Obstacle Detection Methods for the Safe Navigation of Unmanned Vehicles: A Review," *Remote Sens.*, vol. 14, no. 15, pp. 1–26, 2022, doi: 10.3390/rs14153824.

[17] T. Rateke, K. A. Justen, V. F. Chiarella, A. C. Sobieranski, E. Comunello, and A. Von Wangenheim, "Passive vision region-based road detection: A literature review," *ACM Comput. Surv.*, vol. 52, no. 2, 2019, doi: 10.1145/3311951.

[18] A. Bouguettaya, H. Zarzour, A. Kechida, and A. M. Taberkit, "Vehicle Detection From UAV Imagery With Deep Learning: A Review," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 11, pp. 6047–6067, 2022, doi: 10.1109/TNNLS.2021.3080276.

[19] S. Zhang, S. Wang, S. Yu, J. J. Q. Yu, and M. Wen, "Collision Avoidance Predictive Motion Planning Based on Integrated Perception and V2V Communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9640–9653, 2022, doi: 10.1109/TITS.2022.3173674.

[20] X. Duan, H. Jiang, D. Tian, T. Zou, J. Zhou, and Y. Cao, "V2I based environment perception for autonomous vehicles at intersections," *China Commun.*, vol. 18, no. 7, pp. 1–12, 2021, doi: 10.23919/JCC.2021.07.001.

[21] S. Heo, W. Yoo, H. Jang, and J. M. Chung, "H-V2X Mode 4 Adaptive Semipersistent Scheduling Control for Cooperative Internet of Vehicles," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10678–10692, 2021, doi: 10.1109/JIOT.2020.3048993.

[22] Y. Li *et al.*, "V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10914–10921, 2022, doi: 10.1109/LRA.2022.3192802.

[23] R. Yee, E. Chan, B. Cheng, and G. Bansal, "Collaborative Perception for Automated Vehicles Leveraging Vehicle-to-Vehicle Communications," *IEEE Intell. Veh. Symp. Proc.*, vol. 2018-June, no. Iv, pp. 1099–1106, 2018, doi: 10.1109/IVS.2018.8500388.

[24] K. Ashraf, V. Varadarajan, M. R. Rahman, R. Walden, and A. Ashok, "See-Through a Vehicle: Augmenting Road Safety Information Using Visual Perception and Camera Communication in Vehicles," *IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3071–3086, 2021, doi: 10.1109/TVT.2021.3066409.

[25] S. Aoki, T. Higuchi, and O. Altintas, "Cooperative Perception with Deep Reinforcement Learning for Connected Vehicles," *IEEE Intell. Veh. Symp. Proc.*, no. Iv, pp. 328–334, 2020, doi: 10.1109/IV47402.2020.9304570.

[26] M. Meyer, G. Kuschk, and S. Tomforde, "Graph Convolutional Networks for 3D Object Detection on Radar Data," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-Octob, pp. 3053–3062, 2021, doi: 10.1109/ICCVW54120.2021.00340.

[27] J. G. Wang, S. J. Chen, L. B. Zhou, K. W. Wan, and W. Y. Yau, "Vehicle Detection and Width Estimation in Rain by Fusing Radar and Vision," *2018 15th Int. Conf. Control. Autom. Robot. Vision, ICARCV 2018*, pp. 1063–1068, 2018, doi: 10.1109/ICARCV.2018.8581246.

[28] Y. He and L. Li, "A Novel Multi-source Vehicle Detection Algorithm based on Deep Learning," *Int. Conf. Signal Process. Proceedings, ICSP*, vol. 2018-Augus, pp. 979–982, 2019, doi: 10.1109/ICSP.2018.8652388.

[29] P. C. Kung, C. C. Wang, and W. C. Lin, "Radar Occupancy Prediction with Lidar Supervision while Preserving Long-Range Sensing and Penetrating Capabilities," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 2637–2643, 2022, doi: 10.1109/LRA.2022.3144528.

[30] T. W. Kim, W. S. Jang, J. Jang, and J. C. Kim, "Camera and radar-based perception system for truck platooning," *Int. Conf. Control. Autom. Syst.*, vol. 2020-Octob, no. Iccas, pp. 950–955, 2020, doi: 10.23919/ICCAS50221.2020.9268196.

[31] C. Y. Mu, P. Kung, C. F. Chen, and S. C. Chuang, "Enhancing Front-Vehicle Detection in Large Vehicle Fleet Management," *Remote Sens.*, vol. 14, no. 7, pp. 1–18, 2022, doi: 10.3390/rs14071544.

[32] L. Garrote, J. Perdiz, L. A. da Silva Cruz, and U. J. Nunes, "Point Cloud Compression: Impact on Object Detection in Outdoor Contexts," *Sensors*, vol. 22, no. 15, pp. 1–12, 2022, doi: 10.3390/s22155767.

[33] J. M. Vivoda, S. G. Pratt, and S. J. Gillies, "The relationships among roadway safety management practices, collision rates, and injury rates within company fleets," *Saf. Sci.*, vol. 120, no. December 2018, pp. 589–602, 2019, doi: 10.1016/j.ssci.2019.07.033.

[34] P. F. Orzechowski, A. Meyer, and M. Lauer, "Tackling Occlusions Limited Sensor Range with Set-based Safety Verification," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2018-Novem, pp. 1729–1736, 2018, doi: 10.1109/ITSC.2018.8569332.

[35] K. Dai, Y. Wang, J. S. Hu, K. Nam, and C. Yin, "Intertarget Occlusion Handling in Multiextended Target Tracking Based on Labeled Multi-Bernoulli Filter Using Laser Range Finder," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 1719–1728, 2020, doi: 10.1109/TMECH.2020.2994066.

[36] W. Zhou, Y. Zhang, Y. Zheng, P. Li, and M. Xu, "Research on distance measurement method of moving objects in video," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, no. 5, 2020, doi: 10.1088/1757-899X/768/5/052112.

[37] R. Theodosc, D. Denis, C. Blanc, T. Chateau, and P. Checchin, "Vehicle detection based on deep learning heatmap estimation," *IEEE Intell. Veh. Symp. Proc.*, vol. 2019-June, no. Iv, pp. 108–113, 2019, doi: 10.1109/IVS.2019.8814285.

[38] Y. Zhang, D. Huang, and Y. Wang, "PC-RGNN: Point Cloud Completion and Graph Neural Network for 3D Object Detection," *35th AAAI Conf. Artif. Intell. AAAI 2021*, vol. 4B, pp. 3430–3437, 2021, doi: 10.1609/aaai.v35i4.16456.

[39] J. Kim and J. Cho, "A set of single yolo modalities to detect occluded entities via viewpoint conversion," *Appl. Sci.*, vol. 11, no. 13, 2021, doi: 10.3390/app11136016.

[40] M. Csontho, A. Rovid, and Z. Szalay, "Significance of Image Features in Camera-LiDAR Based Object Detection," *IEEE Access*, vol. 10, pp. 61034–61045, 2022, doi: 10.1109/ACCESS.2022.3181137.

[41] Z. Xu *et al.*, "ZoomNet: Part-aware adaptive zooming neural network for 3D object detection," *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 12557–12564, 2020, doi: 10.1609/aaai.v34i07.6945.

[42] W. Zhang, Y. Zheng, Q. Gao, and Z. Mi, "Part-aware region proposal for vehicle detection in high occlusion environment," *IEEE Access*, vol. 7, pp. 100383–100393, 2019, doi: 10.1109/ACCESS.2019.2929432.

[43] C. D. Prakash, F. Akhbari, and L. J. Karam, "Robust obstacle detection for advanced driver assistance systems using distortions of inverse perspective mapping of a monocular camera," *Rob. Auton. Syst.*, vol. 114, pp. 172–186, 2019, doi: 10.1016/j.robot.2018.12.004.

[44] Y. Schulz, A. K. Mattar, T. M. Hehn, and J. F. P. Kooij, "Hearing What You Cannot See: Acoustic Vehicle Detection around Corners," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2587–2594, 2021, doi: 10.1109/LRA.2021.3062254.

[45] P. Narksri, H. Darweesh, E. Takeuchi, Y. Ninomiya, and K. Takeda, "Occlusion-Aware Motion Planning With Visibility Maximization via Active Lateral Position Adjustment," *IEEE Access*, vol. 10, pp. 57759–57782, 2022, doi: 10.1109/ACCESS.2022.3178729.

[46] Q. Luo and J. Liu, "Wireless Telematics Systems in Emerging Intelligent and Connected Vehicles: Threats and Solutions," *IEEE Wirel. Commun.*, vol. 25, no. 6, pp. 113–119, 2018, doi: 10.1109/MWC.2018.1700364.

[47] T. Guan, Y. Han, N. Kang, N. Tang, X. Chen, and S. Wang, “An Overview of Vehicular Cybersecurity for Intelligent Connected Vehicles,” *Sustain.*, vol. 14, no. 9, 2022, doi: 10.3390/su14095211.

[48] L. Zhang, Y. Wang, and H. Zhu, “Theory and Experiment of Cooperative Control at Multi-Intersections in Intelligent Connected Vehicle Environment: Review and Perspectives,” *Sustain.*, vol. 14, no. 3, 2022, doi: 10.3390/su14031542.

[49] R. Khamsehashari and K. Schill, “Improving Deep Multi-modal 3D Object Detection for Autonomous Driving,” *2021 Int. Conf. Autom. Robot. Appl. ICARA 2021*, pp. 263–267, 2021, doi: 10.1109/ICARA51699.2021.9376453.

[50] X. Wang, C. Shen, H. Li, and S. Xu, “Human Detection Aided by Deeply Learned Semantic Masks,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2663–2673, 2020, doi: 10.1109/TCSVT.2019.2924912.

[51] L. Yang, S. Liu, and Y. Zhao, “Deep- Learning Based Algorithm for Detecting Targets in Infrared Images,” *Appl. Sci.*, vol. 12, no. 7, 2022, doi: 10.3390/app12073322.

[52] M. Hong, S. Li, Y. Yang, F. Zhu, Q. Zhao, and L. Lu, “SSPNet: Scale Selection Pyramid Network for Tiny Person Detection From UAV Images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–7, 2022, doi: 10.1109/LGRS.2021.3103069.

[53] J. Luo, H. Fang, F. Shao, C. Hu, and F. Meng, “Vehicle Detection in Congested Traffic Based on Simplified Weighted Dual-Path Feature Pyramid Network with Guided Anchoring,” *IEEE Access*, vol. 9, pp. 53219–53231, 2021, doi: 10.1109/ACCESS.2021.3069216.

[54] R. Yin, R. Zhang, W. Zhao, and F. Jiang, “DA-Net: Pedestrian Detection Using Dense Connected Block and Attention Modules,” *IEEE Access*, vol. 8, pp. 153929–153940, 2020, doi: 10.1109/ACCESS.2020.3018306.

[55] J. Cheng, P. Liu, Q. Zhang, H. Ma, F. Wang, and J. Zhang, “Real-Time and Efficient 6-D Pose Estimation from a Single RGB Image,” *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, doi: 10.1109/TIM.2021.3115564.

[56] Z. Rao, M. He, Y. Dai, Z. Zhu, B. Li, and R. He, “MSDC-Net: Multi-scale dense and contextual networks for stereo matching,” *2019 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA ASC 2019*, vol. c, no. November, pp. 578–583, 2019, doi: 10.1109/APSIPAASC47483.2019.9023237.

[57] S. J. Ji, Q. H. Ling, and F. Han, “An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information,” *Comput. Electr. Eng.*, vol. 105, no. August 2022, p. 108490, 2023, doi: 10.1016/j.compeleceng.2022.108490.

[58] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, “TPH-YOLOv5: Improved YOLOv5

Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2021-Octob, pp. 2778–2788, 2021, doi: 10.1109/ICCVW54120.2021.00312.

[59] C. B. Murthy, M. F. Hashmi, and A. G. Keskar, "EfficientLiteDet: a real-time pedestrian and vehicle detection algorithm," *Mach. Vis. Appl.*, vol. 33, no. 3, pp. 1–15, 2022, doi: 10.1007/s00138-022-01293-y.

[60] J. Henriksson, C. Berger, and S. Ursing, "Understanding the Impact of Edge Cases from Occluded Pedestrians for ML Systems," *Proc. - 2021 47th Euromicro Conf. Softw. Eng. Adv. Appl. SEAA 2021*, pp. 316–325, 2021, doi: 10.1109/SEAA53835.2021.00048.

[61] M. M. Islam, A. Al Redwan Newaz, B. Gokaraju, and A. Karimoddini, "Pedestrian Detection for Autonomous Cars: Occlusion Handling by Classifying Body Parts," *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2020-Octob, pp. 1433–1438, 2020, doi: 10.1109/SMC42975.2020.9282839.

[62] M. J. Flores Calero, M. Aldas, J. Lazaro, A. Gardel, N. Onofa, and B. Quinga, "Pedestrian detection under partial occlusion by using logic inference, HOG and SVM," *IEEE Lat. Am. Trans.*, vol. 17, no. 9, pp. 1552–1559, 2019, doi: 10.1109/TLA.2019.8931190.

[63] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Detecting Pedestrians in a Crowd," *Eur. Conf. Comput. Vis.*, pp. 1–17, 2018, [Online]. Available: http://openaccess.thecvf.com/content_ECCV_2018/html/Shifeng_Zhang_Occlusion-aware_R-CNN_Detecting_ECCV_2018_paper.html.

[64] C. Y. Lin, H. X. Xie, and H. Zheng, "PedJointNet: Joint Head-Shoulder and Full Body Deep Network for Pedestrian Detection," *IEEE Access*, vol. 7, pp. 47687–47697, 2019, doi: 10.1109/ACCESS.2019.2910201.

[65] H. Liang, T. Wu, Q. Zhang, and H. Zhou, "Non-Maximum Suppression Performs Later in Multi-Object Tracking," *Appl. Sci.*, vol. 12, no. 7, 2022, doi: 10.3390/app12073334.

[66] D. Wang, C. Huang, Y. Wang, Y. Deng, and H. Li, "A 3D Multiobject Tracking Algorithm of Point Cloud Based on Deep Learning," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8895696.

[67] J. Chen, F. Wang, C. Li, Y. Zhang, Y. Ai, and W. Zhang, "Online multiple object tracking using a novel discriminative module for autonomous driving," *Electron.*, vol. 10, no. 20, pp. 1–15, 2021, doi: 10.3390/electronics10202479.

[68] T. Khurana, A. Dave, and D. Ramanan, "Detecting Invisible People," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV) 2021*, 2022, no. Iccv, pp. 3154–3164, doi: 10.1109/iccv48922.2021.00316.

[69] N. Engel, S. Hoermann, P. Henzler, and K. DIetmayer, "Deep Object Tracking on Dynamic Occupancy Grid Maps Using RNNs," *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, vol. 2018-Novem, pp. 3852–3858, 2018, doi:

10.1109/ITSC.2018.8569234.

[70] K. Wang and M. Liu, “YOLOv3-MT: A YOLOv3 using multi-target tracking for vehicle visual detection,” *Appl. Intell.*, vol. 52, no. 2, pp. 2070–2091, 2022, doi: 10.1007/s10489-021-02491-3.

[71] W. Lu, J. Cui, Y. Chang, and L. Zhang, “A Video Prediction Method Based on Optical Flow Estimation and Pixel Generation,” *IEEE Access*, vol. 9, pp. 100395–100406, 2021, doi: 10.1109/ACCESS.2021.3096788.

[72] S. Zhang, Y. Xie, J. Wan, H. Xia, S. Z. Li, and G. Guo, “WiderPerson: A Diverse Dataset for Dense Pedestrian Detection in the Wild,” *IEEE Trans. Multimed.*, vol. 22, no. 2, pp. 380–393, 2020, doi: 10.1109/TMM.2019.2929005.

[73] Y. Zhang, Z. Lu, D. Ma, J. H. Xue, and Q. Liao, “Ripple-GAN: Lane Line Detection with Ripple Lane Line Detection Network and Wasserstein GAN,” *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1532–1542, 2021, doi: 10.1109/TITS.2020.2971728.

[74] H. U. Khan, A. Rafaqat Ali, A. Hassan, A. Ali, W. Kazmi, and A. Zaheer, “Lane detection using lane boundary marker network with road geometry constraints,” *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 1823–1832, 2020, doi: 10.1109/WACV45572.2020.9093296.

[75] T. Suleymanov, L. Kunze, and P. Newman, “Online Inference and Detection of Curbs in Partially Occluded Scenes with Sparse LIDAR,” *2019 IEEE Intell. Transp. Syst. Conf. ITSC 2019*, pp. 2693–2700, 2019, doi: 10.1109/ITSC.2019.8917086.

[76] R. Yousri, M. A. Elattar, and M. Saeed Darweesh, “A Deep Learning-Based Benchmarking Framework for Lane Segmentation in the Complex and Dynamic Road Scenes,” *IEEE Access*, vol. 9, pp. 117565–117580, 2021, doi: 10.1109/ACCESS.2021.3106377.

[77] S. Chougule, N. Koznek, A. Ismail, G. Adam, V. Narayan, and M. Schulze, *Reliable multilane detection and classification by utilizing CNN as a regression network*, vol. 11133 LNCS. Springer International Publishing, 2019.

[78] K. Wang, F. Yan, B. Zou, L. Tang, Q. Yuan, and C. Lv, “Occlusion-free road segmentation leveraging semantics for autonomous vehicles,” *Sensors (Switzerland)*, vol. 19, no. 21, pp. 1–15, 2019, doi: 10.3390/s19214711.

[79] K. Yoneda, A. Kuramoto, N. Suganuma, T. Asaka, M. Aldibaja, and R. Yanase, “Robust traffic light and arrow detection using digital map with spatial prior information for automated driving,” *Sensors (Switzerland)*, vol. 20, no. 4, 2020, doi: 10.3390/s20041181.

[80] M. Hirabayashi, A. Sujiwo, A. Monrroy, S. Kato, and M. Edahiro, “Traffic light recognition using high-definition map features,” *Rob. Auton. Syst.*, vol. 111, pp. 62–72, 2019, doi: 10.1016/j.robot.2018.10.004.

[81] H. K. Kim, K. Y. Yoo, J. H. Park, and H. Y. Jung, “Traffic light recognition based on binary semantic segmentation network,” *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–15, 2019, doi: 10.3390/s19071700.

[82] X. Li, H. Ma, X. Wang, and X. Zhang, “Traffic Light Recognition for Complex Scene with Fusion Detections,” *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 199–208, 2018, doi: 10.1109/TITS.2017.2749971.

[83] S. Masaki, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, “Distant traffic light recognition using semantic segmentation,” *Transp. Res. Rec.*, vol. 2675, no. 11, pp. 97–103, 2021, doi: 10.1177/03611981211016467.

[84] S. Santaji, S. Santaji, and S. Hallur, “Detection and Classification of Occluded Traffic Sign Boards,” *3rd Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2018*, no. December, pp. 157–160, 2018, doi: 10.1109/ICEECCOT43722.2018.9001514.

[85] J. Guo, X. Cheng, Q. Chen, and Q. Yang, “Detection of occluded road signs on autonomous driving vehicles,” *Proc. - IEEE Int. Conf. Multimed. Expo*, vol. 2019-July, pp. 856–861, 2019, doi: 10.1109/ICME.2019.00152.

[86] Y. Rehman, I. Riaz, X. Fan, and H. Shin, “D-patches: Effective traffic sign detection with occlusion handling,” *IET Comput. Vis.*, vol. 11, no. 5, pp. 368–377, 2017, doi: 10.1049/iet-cvi.2016.0303.

[87] Y. Rehman, J. A. Khan, and H. Shin, “Efficient coarser-to-fine holistic traffic sign detection for occlusion handling,” *IET Image Process.*, vol. 12, no. 12, pp. 2229–2237, 2018, doi: 10.1049/iet-ipr.2018.5424.

[88] A. Palffy, J. F. P. Kooij, and D. M. Gavrila, “Occlusion aware sensor fusion for early crossing pedestrian detection,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2019-June, no. Iv, pp. 1768–1774, 2019, doi: 10.1109/IVS.2019.8814065.

[89] J. Leng, Y. Liu, D. Du, T. Zhang, and P. Quan, “Robust Obstacle Detection and Recognition for Driver Assistance Systems,” *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1560–1571, 2020, doi: 10.1109/TITS.2019.2909275.

[90] Y. Zhou, S. Wen, D. Wang, J. Mu, and I. Richard, “Object detection in autonomous driving scenarios based on an improved faster-RCNN,” *Appl. Sci.*, vol. 11, no. 24, 2021, doi: 10.3390/app112411630.

[91] A. Mannan, K. Javed, A. U. Rehman, S. K. Noon, and H. A. Babri, “Optimized segmentation and multiscale emphasized feature extraction for traffic sign detection and recognition,” *J. Intell. Fuzzy Syst.*, vol. 36, no. 1, pp. 173–188, 2019, doi: 10.3233/JIFS-181082.

[92] Y. Wang, L. Mo, H. Ma, and J. Yuan, “OccGAN: Semantic image augmentation for driving scenes,” *Pattern Recognit. Lett.*, vol. 136, pp. 257–263, 2020, doi: 10.1016/j.patrec.2020.06.011.