An improved traffic lights recognition algorithm for autonomous driving in complex scenarios

Ziyue Li¹,², Qinghua Zeng¹, Yuchao Liu², Jianye Liu¹ and Lin Li³

Abstract
Image recognition is susceptible to interference from the external environment. It is challenging to accurately and reliably recognize traffic lights in all-time and all-weather conditions. This article proposed an improved vision-based traffic lights recognition algorithm for autonomous driving, integrating deep learning and multi-sensor data fusion assist (MSDA). We introduce a method to obtain the best size of the region of interest (ROI) dynamically, including four aspects. First, based on multi-sensor data (RTK BDS/GPS, IMU, camera, and LiDAR) acquired in a normal environment, we generated a prior map that contained sufficient traffic lights information. And then, by analyzing the relationship between the error of the sensors and the optimal size of ROI, the adaptively dynamic adjustment (ADA) model was built. Furthermore, according to the multi-sensor data fusion positioning and ADA model, the optimal ROI can be obtained to predict the location of traffic lights. Finally, YOLOv4 is employed to extract and identify the image features. We evaluated our algorithm using a public data set and actual city road test at night. The experimental results demonstrate that the proposed algorithm has a relatively high accuracy rate in complex scenarios and can promote the engineering application of autonomous driving technology.

Keywords
Traffic lights recognition, ADA model, autonomous driving, multi-sensor data fusion, deep learning

Introduction
In recent years, the rapid development of autonomous driving has attracted more and more attention worldwide. Simultaneously, with the help of artificial intelligence, significant technological breakthroughs have been made, and industrialization has been deduced to an unprecedented height.¹ The autonomous driving system mainly includes three main functional modules: perception, cognition, and execution. The perception module of the system mainly consists of sensors such as LiDAR, RADAR, camera, IMU, BDS/GPS, and odometer. The difference in measurement principle determines the good complementarity between multi-modal sensing data.² Multimodal data fusion is an effective method to improve the performance of the perception module.³,⁴ According to the multimodal sensors data fusion and its own state, the autonomous driving vehicle realizes the driving trajectory planning.⁵ Furthermore, it uses the control module to finally
obtain control instructions for the steering wheel, accelerator, and brake.6

Traffic light recognition is an essential part of the perception system.7 Although the cooperative vehicle infrastructure system (CVIS) based on V2I communication8 is a reliable and accurate scheme, it still has disadvantages of difficulty in implementation, and the communication easily interferes. Therefore, vision-based traffic lights recognition is an indispensable key technology for the autonomous driving system. However, due to the inherent small object, easily interfered with by the external light sources, and other factors, vision-based traffic lights recognition has always been a technical challenge.9 Especially at night, with extreme weather conditions, intersections, multiple traffic lights, and other complex road conditions, this problem will become more difficult.10 The traffic lights recognition technology has become a bottleneck, restricting the engineering application of autonomous driving.

Essentially, vision-based traffic lights recognition is a problem of image object detection and classification. Based on traditional recognition methods, the technical realization process is generally divided into two stages: object detection and object recognition.11 In the object detection stage, the main task is to obtain the ROI and determine the traffic light’s location in the image. Besides, the main goal of the object recognition stage is to achieve accurate object classification through manual feature extraction and machine learning classification. The typical manual features have good immutability, such as SIFT,12 HOG,13,14 Haar-like,13,14 but the recognition accuracy is significantly reduced when the quality of the image is worse. Simultaneously, the commonly used machine learning algorithms such as AdaBoost,9 SVM10 run fast, while the recognition accuracy is difficult to satisfy the practical requirements of autonomous driving.

With the introduction and application of deep learning15,16 in recent years, the accuracy of traffic lights recognition has been greatly improved, especially since the successful application of AlexNet17 deep convolution neural network in the ILSVRC challenge,18 a variety of excellent improved CNN models such as VGGNet,19 GoogleNet,20 ResNet21 have been proposed one after another. As early as 2011, Sermanet and LeCun22 applied deep learning to traffic sign recognition for the first time, which surpassed the level of human recognition and attracted widespread attention. As Faster-CNN,23 YOLO,24 and other better algorithms were proposed and applied, the accuracy of traffic lights recognition based on deep learning is significantly higher than that of traditional algorithms.25 However, it is slightly insufficient in real-time.26 In addition, the accuracy of the algorithm is still too hard to satisfy the actual requirements in extreme illumination, severe weather, and complex road conditions.

In the actual application scenarios of autonomous driving, vision-based traffic lights recognition is challenging due to the influence of illumination and other light source interference.27 The performance of any single recognition algorithm is difficult to satisfy the needs of the actual application. Therefore, the fusion of multiple technologies based on deep learning has become an important research field.28 On one hand, combining traditional feature extraction and recognition methods with deep learning can make full use of their complementarity to improve system performance.29 On the other hand, integrating real-time positioning and prior map data to obtain ROI and reduce environmental interference is helpful to improve the performance of the traffic lights recognition algorithm.30

Accurate and reliable traffic lights recognition at all-time and all-weather is essential to the safety of autonomous vehicle systems. However, due to the susceptibility to interference from the external environment, the visual-based traffic lights recognition algorithm is complicated. The existing results mainly studied the image recognition of ordinary illumination and weather, which is difficult to meet the actual application requirements of autonomous driving. To solve this question, we proposed an improved traffic lights recognition algorithm based on the multi-sensor data fusion assist (MSDA). The main contributions of this article are summarized as follows.

1. Based on the prior map and multi-sensor data, we propose a method of dynamically adjusting the size of ROI based on different positioning precision (RTK, degraded, outages), achieving the optimal auxiliary effect on the image recognition algorithm.
2. By analyzing the relationship between the sensor error and the optimal size of the ROI, we built an ADA conversion model from sensor parameters to the best ROI.
3. A priori map labeling method based on multi-sensor fusion is proposed, which uses various information such as area size, center point coordinates, semantics, road attributes, sensor working status.

The following content of this article was organized as follows. The second section introduces related research progress of traffic lights recognition from two aspects of deep learning and prior maps and briefly analyzes the shortcomings and research fields of existing achievements. In the third section, the prior map acquisition method was described in detail. At the same time, the principle and technical implementation scheme of our algorithm was introduced. Experiment
and validation of the algorithm were introduced in the fourth section. Compared with the traditional method, the superiority of our algorithm was also analyzed. In the fifth section, the main innovations and achievements of this article were summarized, and further research work was also discussed.

Related work

In our algorithm, multi-sensor data fusion assistance and a prior map were used to obtain the ROI in the image. Furthermore, the ADA conversion model is used to dynamically optimize the ROI online, to improve the accuracy of traffic lights recognition in the complex scenario. At present, deep learning and prior knowledge assistance are two main directions in the field of traffic lights recognition. Therefore, this part mainly introduces the progress of the related research in the above two aspects.

Deep learning–based traffic lights recognition

AlexNet model is a classical deep learning algorithm in the field of image recognition. On that basis, DeepTLR model for traffic lights recognition was proposed, which achieved significantly better performance than manual features without relying on any prior knowledge. With the help of the concise and efficient AlexNet model, Gao and Wang combined deep learning and traditional recognition technology, implemented image segmentation in HSV space, extracted HOG manual features, and merged SVM-based traditional detectors with deep learning detectors. After the iterative development of RCNN to Faster-RCNN, the two-step deep learning algorithm has made significant progress in image recognition accuracy and computational efficiency. Kim et al. applied Faster-RCNN to traffic lights recognition. The latter also analyzed and compared the recognition performance of six color spaces such as RGB, HSV, and three deep learning models composed of Faster-RCNN, R-FCN, Resnet-101, etc. The results showed that in the RGB color space, the recognition accuracy of the deep learning algorithm composed of Faster R-CNN and Inception-Resnet-v2 was better than other combinations.

The one-step algorithms (YOLO, SSD) have more advantages in computational efficiency. They can be used in autonomous vehicle application scenarios with high real-time requirements and limited computing resources compared with the two-step deep learning algorithm. Based on the YOLO algorithm, Behrendt et al. achieved a detection speed of 10 frames per second for a 1280 × 720 resolution image and achieved traffic lights tracking through the combination of a stereo camera and an onboard odometer. At the same time, combining the high-efficiency advantages of YOLO with the high/low exposure dual-channel architecture has apparent advantages in a highly dynamic environment. For the weakness of the one-step deep learning algorithm in small object detection, TL-SSD has improved the SSD algorithm to achieve simultaneous improvement in the recognition precision and recall rate of traffic lights. Ouyang et al. combined the specially designed lightweight convolutional neural network RTTLD (Real-Time Traffic lights Detector) and heuristic ROI detector, which achieved the accuracy rate of 99.3% and 99.7%, respectively, at 10 Hz on an NVIDIA Jetson TX1/TX2 hardware platform. Saini et al. combined convolutional neural network algorithm with traditional HOG feature extraction and SVM classifier to achieve efficient recognition of traffic lights.

Most of the above-mentioned traffic lights recognition algorithms were studied for applications in conventional scenarios (illumination, weather, and traffic conditions). However, in actual application, there are many extreme illuminations (strong lights, backlight, night), severe weather (rain, snow, fog, haze), complex road conditions with various crossroads, and the performance of conventional algorithms are greatly limited. Thus, it is more important to acquire the high-precision map, positioning, attitude, and LiDAR information and fusion with vision information in severe circumstances. The multi-sensor data fusion–assisted improved traffic lights recognition algorithm proposed in this article combines deep learning and prior knowledge assistance, precisely and robustly obtains the ROI through an online adaptive adjustment to improve the recognition accuracy in complex scenarios.

Prior map–based traffic lights recognition

Whether it is traditional or deep learning-based image detection and recognition, with the assistance of real-time positioning and prior maps, the accuracy of traffic signal recognition can be significantly improved. Based on Google’s autonomous driving vehicles, on one hand, Fairfield and Urmson used differential GPS, IMU, and LiDAR to complete a prior map acquisition. On the other hand, with the help of high-precision positioning, vehicle heading and attitude, and prior knowledge, it improved the performance of the traditional traffic lights recognition algorithm. At the same time, through the semantic information of the prior map, the relevant traffic lights were extracted from the image, achieving a recognition accuracy of 99% and a recall rate of 62% under normal lighting conditions. More importantly, there were no false positives of green lights.

Furthermore, Levinson et al. analyzed the error sources in the three stages of prior map matching, positioning, and recognition. With the aid of prior knowledge, the comprehensive recognition accuracy of
traditional image detection and recognition algorithms in intersection scenes can reach 94% in the three-time periods of noon, evening, and night. However, the recognition accuracy rate of green lights and yellow lights was low, and the false positive probability of green lights reaches 1.35%. Besides, the performance of traditional feature extraction and recognition algorithms can be further improved by fully considering the impact of road slope factor on using object location information for auxiliary traffic lights recognition. MSDF-AlexNet integrates the prior map and multi-sensor auxiliary information and effectively improves the image recognition accuracy by acquiring ROI. Those mentioned above prior knowledge-assisted solutions all rely on Real-Time Kinematic GPS (RTK GPS) positioning, and high-precision IMU, while the expensive equipment price brings a huge economic burden. Barnes et al. implemented a prior map-aided scheme based on low-cost GPS positioning, which improved the accuracy of HOG feature extraction and SVM classification recognition algorithm by 40%.

Compared with the traditional algorithm, the prior knowledge-aided algorithm based on deep learning has a better development prospect. Based on the GPS signal-assisted acquisition of the ROI and using the convolutional neural network for feature extraction and classification, John et al. proposed two image recognition schemes suitable for normal light and low light conditions, respectively. Possati et al. Combined the newly proposed deep learning algorithm and prior knowledge assistance and used a prior map to select relevant traffic lights of vehicle driving behavior from the recognition results of YOLOv3.

Those mentioned above prior knowledge-assisted traffic lights recognition algorithms mainly focus on obtaining the ROI through online positioning and assisting traditional, deep learning, or hybrid architecture image recognition solutions. However, they do not consider the impact of sensor errors on the acquisition of ROI.

Different from the existing research achievement of deep learning-based methods such as AlexNet, Faster-RCNN, SSD, and YOLO, this article combines multi-sensor data and prior maps to further improve the algorithm performance by obtaining ROI. At the same time, different from the achievement of a traditional prior map based traffic lights recognition methods, we analyzed the relationship between sensor error and the size of the ROI and built an ADA transition model from performance parameters of the sensors to the optimum ROI. Through adaptive adjusting and matching online in different scenarios, the best ROI was obtained, which effectively improves the environmental adaptability and robustness of the algorithm.

The improved traffic lights recognition method proposed in this paper improves the recognition accuracy by using a prior map and multi-sensor fusion data. In practical applications, if the size of the ROI is too large, it will bring less improvement to the performance of the algorithm. Conversely, if the size of the ROI is too small, it will easily cause the object to missing detection. Therefore, obtaining the optimal size of the ROI is crucial to the improvement of image recognition accuracy. It is known that BDS/GPS receiver is easy to be disturbed by multiple factors such as signal block, electromagnetic interference, multipath effect. Especially in the urban areas, due to environmental factors such as buildings and trees, BDS/GPS signals are often interrupted or the positioning precision degraded, which means that the sensor’s working environment will undoubtedly bring significant influence on acquiring ROI. Different from the existing methods, based on the traditional prior knowledge-assisted traffic lights recognition algorithm, this article’s method can adaptively adjust and match to obtain the best ROI according to the actual performance of the sensors.

Proposal of an improved traffic lights recognition algorithm based on MSDA

Traditional traffic light recognition methods mainly include manual feature extraction and machine learning algorithm classification. The accuracy and environmental adaptability are difficult to satisfy the requirements of autonomous driving. With the development and application of deep learning technology in recent years, feature extraction and classification based on convolutional neural networks have made significant progress in traffic light recognition technology. However, due to the complexity of the road and the external environment, the performance of the algorithm is still far from the demand for autonomous driving in actual scenarios. Any single traffic lights recognition algorithm is difficult to meet the practical requirements. Therefore, various methods combining traditional technologies, deep learning, prior knowledge, and multi-sensor data have been proposed. In this article, an improved traffic lights recognition scheme was proposed to apply autonomous driving in Complex Scenarios. Based on a prior map and multi-sensor data, the best ROI can be obtained through online adaptive matching and adjustment. The implementation scheme mainly includes joint calibration of the sensors, prior map generation, ADA model built, relevant traffic lights selection in the complex crossroad, and feature extraction and recognition based on deep learning.

Overall scheme construction

The vision-based traffic lights recognition algorithm is greatly affected by the external environment. For
example, it is easy to cause the algorithm to miss detection in extreme illumination and severe weather conditions. The interference of similar external light sources (car taillights and neon lights) can also easily cause algorithm detection and recognition mistakenly. Considering the actual application requirements of autonomous driving in Complex Scenarios, an improved traffic lights recognition algorithm was proposed in this article. The architecture mainly includes 4 parts: (1) Sensors’ calibration: joint calibration of the sensors such as camera, LiDAR, IMU, BDS/GPS receiver, unified to the vehicle body coordinate system. (2) Prior map generation: autonomous vehicles integrated with multiple sensors were used to collect and label the traffic lights data (position, bounding box, semantic, attribute, etc.). (3) ADA model built: analyzing the relationship between the sensor’s error and the size of ROI, and building an ADA model from the performance of the sensors to the size of ROI. Based on the real-time sensor data and the prior map, we select and locate relevant traffic lights according to the trajectory planning parameters and control commands of the autonomous driving. (4) ROI acquisition: selecting relevant traffic lights in complex crossroads based on planning decision commands and selecting the best ROI based on the ADA model adaptively. (5) Features extraction and state recognition based on deep learning. The algorithm framework of this article was shown in Figure 1.

As shown in Figure 1, the principle of our proposed traffic signal recognition algorithm is as follows:

1. First, we used the collected multi-sensor data to generate a prior map offline, which contained rich traffic light information (see section “Prior map generation” for details). By analyzing the relationship between the sensor error and the optimal size of the ROI, we built an ADA conversion model from sensor parameters to the best ROI (see section “Adaptively Dynamic Adjustment (ADA) model” for details).

2. Second, the autonomous driving system preprocessed multi-sensor data (LiDAR point cloud, image, attitude, heading, latitude, longitude, height). According to the prior map, we utilized multi-sensor fusion positioning data calculated relative positional relationships of the camera and traffic lights (see section “ROI center point acquisition” for details).

3. Third, according to the decision command of the autonomous driving system, the related traffic lights were obtained (see section “Research on selection rules of relevant traffic lights” for details). And then, we used sensor parameters to calculate the best interest area size online (see section “Adaptively Dynamic Adjustment (ADA) model” for details).

4. Finally, we adopted the state-of-the-art deep learning algorithm YOLOv4 to recognize traffic lights (see section “Feature extraction and recognition algorithm based on deep learning” for details).
Joint calibration of the sensors

Multi-sensor data fusion is an efficient auxiliary tool for image recognition. We used auxiliary data to obtain ROI can significantly improve calculation efficiency and recognition accuracy. The joint calibration of the sensors is an important foundation, which determines the accuracy of the algorithm to a certain extent. It should be emphasized that the calibration work mentioned here refers to the calibration of external installation parameters, which generally does not include the calibration of internal parameters of the sensors. Joint calibration of the sensors refers to the realization of temporal and spatial synchronization of data from different sensors through the conversion of the coordinate system of IMU, camera, BDS/GPS, and LiDAR. In the following content of this article, we emphatically describe the main process of coordinated system transformation of sensors and carry on the derivation and analysis through the mathematical formula.

The joint calibration of the sensors mainly includes rotation and translation. At first, we calibrate the attitude transition matrix between the IMU and the camera. Also, the IMU calculated and output the attitude and heading information of the vehicle in the body coordinate system online. Eventually, the IMU, camera, and LiDAR were all rigidly connected to the vehicle body, and the relative position and attitude relationship were approximately unchanged. Therefore, the attitude and heading information of the camera and LiDAR can be obtained in real time by calibrating the attitude transition matrix between the IMU and the camera, LiDAR. The specific method was to place the vehicle on the horizontal ground and adjust the roll and pitch angle of the camera to keep it horizontal with the help of a leveling instrument. In this case, the output value of the IMU was the difference between these two installation angles. At the same time, we measured their heading difference. And the attitude transition matrix \( C_{\text{IMU}}^{\text{ant}} \) between the IMU and the camera can be obtained by combining the parameters of the attitude and heading. In the same way, we could calibrate the attitude transition matrix \( C_{\text{LiDAR}}^{\text{ant}} \) and \( C_{\text{IMU}}^{\text{LiDAR}} \) between BDS/GPS antenna, LiDAR, and IMU.

Then, we calibrated the position relationship between the BDS/GPS antenna, camera, and LiDAR. The BDS/GPS receiver calculates and outputs the longitude, latitude, and altitude information based on the antenna center online. Similar to the calibration of IMU and camera, we calculated the longitude, latitude, and altitude information of the camera and LiDAR online through the relative position relationship between the antenna and the camera \([x, y, z]\).

In equation (1), we showed the calculation process of the attitude, heading, position, and altitude of the camera. The calculation of LiDAR parameters was the same, so that we will not repeat it here.

\[
\begin{align*}
\begin{bmatrix}
\lambda_C \\
\beta_C \\
H_C \\
\theta_C \\
\gamma_C \\
\varphi_C 
\end{bmatrix} &= \begin{bmatrix}
\Delta CC_{\text{ant}}^{\text{IMU}} & C_{\text{IMU}}^{\text{ant}} & 0_{3 \times 3} & C_{\text{IMU}}^{C} \\
0_{3 \times 3} & C_{\text{IMU}}^{\text{ant}} & 0_{3 \times 3} & C_{\text{IMU}}^{C} \\
\end{bmatrix}
\begin{bmatrix}
\lambda_{\text{ant}} \\
\beta_{\text{ant}} \\
H_{\text{ant}} \\
\theta_{\text{IMU}} \\
\gamma_{\text{IMU}} \\
\varphi_{\text{IMU}} \\
\end{bmatrix}
\end{align*}
\]

where \( \lambda_C, \beta_C, H_C, \theta_C, \gamma_C, \varphi_C \) indicate the camera’s longitude, latitude, altitude, pitch angle, roll angle, and heading angle parameters, respectively. \( C_{\text{ant}}^{C} \) is the transition matrix of the camera from the BDS/GPS receiver antenna body coordinate system to the geographic coordinate system. The receiver antenna coordinate system is defined as the coordinate origin is the geometric center of the receiver antenna. The \( Y \) axis of the coordinate system is the forward direction of the body, and the longitudinal of the body is the \( X \) axis. \( \Delta CC_{\text{ant}}^{\text{IMU}} \) is the transition matrix of the camera from geographic coordinate system to the difference of longitude, latitude, and altitude. \( x_b, y_b, z_b \) are the coordinates of the camera in the antenna body coordinate system. \( \theta_{\text{IMU}}, \gamma_{\text{IMU}}, \varphi_{\text{IMU}} \) are the pitch, roll, and heading output by IMU. \( \lambda_{\text{ant}}, \beta_{\text{ant}}, H_{\text{ant}} \) are the longitude, latitude, and altitude of the antenna output by the BDS/GPS receiver.

Through the attitude transfer matrix, we could get the attitude relationship between the LiDAR, camera, and IMU; through the translation vector, we could get the position of the lidar and camera. Furthermore, we improved the calibration accuracy through cross-calibration between different sensors. On level ground, both the lidar and the camera could detect the columnar objects in front of the vehicle simultaneously. Since we had already calibrated the lidar and the camera, theoretically, the attitude of the columnar objects measured by the two sensors should be the same. However, there are often some calibration errors. We measured the same object through these two sensors, and we can get the difference between the two groups of attitude angles. Based on this, we can get the correction matrix of the attitude transition matrix.

Prior map generation

In the previous studies, the location of traffic lights was generally marked on the map to facilitate the acquisition of ROI. However, there was no detailed description of the marking method. Besides, traffic lights were generally large in size, and merely labeling the position was difficult to achieve the best assistance effect of the prior map. Different from existing achievement, a novel prior map generation method was proposed in this paper and described in detail in the following. We use the center point and bounding box (length and width) to indicate the positioning information of traffic lights,
which was more specific and facilitated the acquisition of the ROI.

Through the sensors’ joint calibration method described in section “Joint calibration of the sensors,” all sensors were unified into the body coordinate system, conducive to data acquisition and labeling. Taking LiDAR as an example, based on accurate attitude, heading, and position parameters, the absolute position of the object can be calculated using point cloud data out from LiDAR in the polar coordinates or Cartesian coordinates system. We generated a prior map through traffic lights data collection, fusion, and manual calibration using multiple sensors, such as the BDS/GPS receiver, IMU, LiDAR, and camera. The prior map mainly contained various parameters of traffic lights such as the center point, bounding box, traffic lights semantics, positioning status of the BDS/GPS receiver, IMU, LiDAR, and camera. The main steps were described as follows:

1. We have manually driven autonomous driving vehicles and adopted multi-sensor to collect urban road images, LiDAR point clouds, position, and other traffic lights information online in various illumination (morning, noon, and night), and weather conditions.

2. We set the road within 200 m of traffic lights as traffic light sections, clustered the LiDAR point clouds, and manually labeled the rectangular bounding box, and calculated the coordinates of the center point (longitude, latitude, and altitude) and the size (length and width) of the bounding box.

3. Due to the occlusion of trees, tall buildings, bridges, and tunnels in urban roads, the BDS/GPS signal was often interrupted. During the BDS/GPS outages, the positioning error of the inertial navigation system would increase rapidly over time. Therefore, it was necessary to label the BDS/GPS receiver (RTK, degraded or outages), road type (such as straight, crossroad, overpass, and roundabout). The acquisition of the ROI based on a priori map mainly contained various parameters of traffic lights such as the center point, bounding box, traffic lights semantics, positioning status of the BDS/GPS receiver, IMU, LiDAR, and camera. The prior map mainly contained various parameters of traffic lights such as the center point, bounding box, traffic lights semantics, positioning status of the BDS/GPS receiver, IMU, LiDAR, and camera. The main steps were described as follows:

(1) We have manually driven autonomous driving vehicles and adopted multi-sensor to collect urban road images, LiDAR point clouds, position, and other traffic lights information online in various illumination (morning, noon, and night), and weather conditions.

(2) We set the road within 200 m of traffic lights as traffic light sections, clustered the LiDAR point clouds, and manually labeled the rectangular bounding box, and calculated the coordinates of the center point (longitude, latitude, and altitude) and the size (length and width) of the bounding box.

(3) Due to the occlusion of trees, tall buildings, bridges, and tunnels in urban roads, the BDS/GPS signal was often interrupted. During the BDS/GPS outages, the positioning error of the inertial navigation system would increase rapidly over time. Therefore, it was necessary to label the BDS/GPS receiver (RTK, degraded, or outages).

(4) By aligning, cleaning, and structured packaging of multi-sensor fusion data, we had generated a priori map, which mainly includes center point, bounding box, traffic lights semantics, positioning status of BDS/GPS, road type.

ROI center point acquisition

The acquisition of the ROI based on a priori map and online positioning is an essential part of the algorithm in this article. In the prior map, the center point and bounding box of the traffic lights area were represented by longitude, latitude, and altitude. At the same time, the real-time positioning data of the camera were also in the CGCS2000 coordinate system. To obtain the ROI of the image, we need to calculate the coordinate parameters of the center point of the traffic lights area in the camera body coordinate system. First, through the center point coordinates, the relative position and angle relationship between the camera and the traffic lights could be calculated. Then, the size and location of the ROI could be calculated by the camera’s attitude and its parameters. The specific process was as follows.

Next, we converted the above two sets of position coordinates to the CGCS2000 earth-centered and earth-fixed (ECEF) coordinate system and the calculation formula as shown in formulas (2) and (3)

$$
\begin{align*}
{x_c}^e &= (R_c + H_c)\cos \lambda_c \cos \beta_c \\
{y_c}^e &= (R_c + H_c)\cos \lambda_c \sin \beta_c \\
{z_c}^e &= (R_c(1 - \varepsilon^2) + H_c)\sin \beta_c
\end{align*}
$$

where $x_c^e$, $y_c^e$, and $z_c^e$ are the coordinates of the camera in the CGCS2000 ECEF coordinate system; $x_s^e$, $y_s^e$, and $z_s^e$ are the coordinates of the traffic lights in the CGCS2000 ECEF coordinate system; $\lambda_c$, $\beta_c$, and $H_c$ are longitude, latitude, and altitude of the camera; $\lambda_s$, $\beta_s$, and $H_s$ are longitude, latitude, and altitude of the center point of the traffic lights bounding box; $\varepsilon$ is the eccentricity, which is the geometric constant of the ellipsoid of the earth in the CGCS2000 coordinate system; $R_c$, $R_s$ are the radius of curvature in prime vertical corresponding to the coordinates of the camera and the center point of the bounding box respectively.

In the CGCS2000 ECEF coordinate system, we solved the vector and converted the attitude matrix according to the coordinates of the camera and the center point of the bounding box and calculated the coordinates of the center point in the geographic coordinate system with the centroid of the camera as the origin. Furthermore, we used the attitude parameters of the camera to calculate the attitude transition matrix from the geographic coordinate system to the body coordinates. Based on this, the coordinate parameters of the center point in the camera coordinate system could be acquired. The calculation formula was shown in formula (4)

$$
\begin{bmatrix}
{x_C}^g \\
{y_C}^g \\
{z_C}^g
\end{bmatrix}
= C_c^g C_c^e
\begin{bmatrix}
{x_c}^e - x_c^e \\
{y_c}^e - y_c^e \\
{z_c}^e - z_c^e
\end{bmatrix}
$$

$x_c^e$, $y_c^e$, and $z_c^e$ are the coordinates of the bounding box center point in the camera body coordinate system; $C_c^g$ and $C_c^e$ are the attitude transition matrix from the geocentric coordinate system to the geographic coordinate system and from the geographic coordinate system to the camera body coordinate system.
According to the center point coordinates in the camera system, we calculated its coordinate parameters in the camera image coordinate system according to the camera’s internal parameters. The conversion process from the camera body coordinate system to the image plane coordinate system was shown in equation (5)

\[
\begin{bmatrix}
u \\
v
\end{bmatrix} = 
\begin{bmatrix}
R_{ex} & 0 & u_0 \\
0 & R_{ey} & v_0
\end{bmatrix}
\begin{bmatrix}
f & 0 & 0 \\
0 & f & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_C \\
y_C \\
z_s
\end{bmatrix}
\]

where \( f \) is the focal length of the camera, \( u_0, v_0 \) are the coordinate of the origin \( O \) of the image coordinate system in the pixel coordinate system, \( R_{ex} \) and \( R_{ey} \) are the resolution of the camera in the horizontal and vertical directions, and \( u \) and \( v \) are the pixel coordinates of the center point in the image coordinate system.

**Adaptively Dynamic Adjustment (ADA) model**

In the process of vision-based traffic lights recognition, if the size of the ROI is smaller, the efficiency of the image recognition algorithm will be higher. However, narrowing the size of ROI means that it is easy to lead to the loss of important information of the object, which will increase the probability of false detection and recognition. Therefore, according to the sensors’ working status, we proposed an online adaptive dynamic adjustment (ADA) method to optimize the size of ROI, taking into account computational efficiency and precision.

In this article, we adopted RTK BDS/GPS, IMU, vision, and LiDAR for autonomous vehicle positioning. The sensors mentioned above are highly complementary. RTK BDS/GPS receivers can output high-precision global positioning information all-time and all-weather, but occlusion and interference can easily cause accuracy degradation or even non-positioning. IMU is not easy to be interfered with by the external environment, but the error accumulates over time. Through lane line recognition, vision can achieve lane-level lateral positioning, but it can only be accurate to the meter level, and it is difficult to obtain global positioning information. LiDAR is also difficult to achieve global positioning. However, it can detect edges and isolation belts on structured roads for relative positioning, especially in tunnels with poor lighting conditions, making up for vision defects. Through multi-sensor fusion, we have achieved accurate and reliable positioning of autonomous vehicles.

Based on the prior map and multi-sensor assistance, the error sources of the ROI mainly include the data labeling error in the process of prior map generation and the measuring error of the sensors in real-time. The data labeling errors of the map consist of joint calibration errors of the sensors, positioning errors, LiDAR measurement errors, point cloud clustering errors, and the asynchronization of the sensors. The calibration errors were generally within the allowable range. In addition, in traffic lights data collection, the vehicle moves very slowly (3–5 km/h). So, we ignored the error caused by an asynchronization of the output data from the multi-sensor.

After the prior map was generated, the error would generally not change on the promise that the traffic lights do not move or malfunction. Therefore, the online dynamic adjustment of the ROI size was mainly related to the error BDS/GPS receiver (positioning error) and IMU (attitude and heading) of the current road section. The real-time positioning in our system mainly relies on the integrated navigation system of BDS/GPS and IMU. The working status of BDS/GPS was easily affected by environmental such as surrounding trees, bridges, and buildings. In order to improve the practical application effect, the working status (RTK, degraded, outages) of the equipment was labeled in the prior map. Next, the relationship between the working status of sensors and the ROI size was analyzed in three conditions: RTK BDS/GPS, degraded, and outages.

(1) **RTK BDS/GPS**

The BDS/GPS receiver equipped with autonomous vehicles has a function of real-time dynamic (RTK) differential, and the positioning precision is 3 cm (root mean square error) under normal working conditions. Compared with the positioning error, the synchronization error of the output data between the BDS/GPS/IMU integrated navigation system, LiDAR, and the camera is more important. According to the update frequency of 25 frames per second, the maximum synchronization error can reach 40 ms, which will result in a distance error of nearly 0.67 meters (assuming that the vehicle travels at a speed of 60 km/h). So, the synchronization error cannot be ignored. Besides, the heading and pitch outputs from the IMU are also the primary sources of error. In the case of a straight-line distance of 100 m between the autonomous vehicle and the traffic signal, an angle error of 0.5° will result in an error of nearly 1 meter on the center point of the traffic light image. The above errors can be expressed in detail in ECEF coordinate system and camera body coordinate system, respectively, as shown in equations (6) and (7)

\[
\delta_{LWD}^C = \begin{bmatrix}
\delta L^e \\
\delta W^e \\
\delta D^e
\end{bmatrix} = \begin{bmatrix}
\lambda_L(\delta_{map} + \delta_{GNSS}^L) \\
\lambda_W(\delta_{map} + \delta_{GNSS}^W) \\
\lambda_D(\delta_{map} + \delta_{GNSS}^D)
\end{bmatrix}
\]
\[
\delta_{LWD}^C = \begin{bmatrix}
\delta L^C_w \\
\delta W^C_L \\
\delta D^C_d
\end{bmatrix} = \begin{bmatrix}
\lambda_L D \tan \vartheta \\
\lambda_H D \tan \vartheta \\
\lambda_D \delta_T
\end{bmatrix}
\] (7)

where \(\delta L^e, \delta W^e, \delta D^e, \delta L^C, \delta W^C, \) and \(\delta D^C\) are the length, width, and vertical errors of the bounding box in the ECEF coordinate system and the camera body coordinate system, respectively; \(\delta_{map}\) is the data collecting and labeling error of the prior map; \(\delta_{t}^{H_{GNS}}\) and \(\delta_{t}^{V_{GNS}}\) are the positioning error of the BDS/GPS receiver in horizontal and height direction; \(\delta_T\) is the synchronization error. Among them, \(\delta_{map}\), \(\delta_{t}^{H_{GNS}}\), and \(\delta_{t}^{V_{GNS}}\) are 3 times of the RMSE (the probability of 99.7%), and \(\delta_T\) is the Max error. \(D, \delta_{t}^L, \) and \(\delta_{t}^H\) are the vertical distance, heading error, and pitch error between the camera and the center point of the traffic lights measured by the IMU along the Z-axis. The road slope is labeled in the prior map, so the error caused by the pitch can be ignored. \(\lambda_L, \lambda_H, \) and \(\lambda_D\) are the reliability coefficient of the sensors used in the calculation of the size of ROI. The larger the value, the lower the confidence in the data output by the sensors and the larger the ROI. Conversely, the smaller the value, the higher the confidence in the data output by the sensors and the smaller the ROI. The value of the parameter is generally larger than 1. In the image coordinate system, the errors in equations (6) and (7) are converted to pixel using equation (8), and the calculation process of the size of ROI is given by equation (9)

\[
\begin{bmatrix}
\delta u \\
\delta v
\end{bmatrix} = C^C_L \begin{bmatrix}
C^C_L & C^C_L \delta_{LWD}^C \\
0_{1 \times 2} & 0
\end{bmatrix} \begin{bmatrix}
\lambda_L D \tan \vartheta \\
\lambda_H D \tan \vartheta \\
\lambda_D \delta_T
\end{bmatrix}
\] (8)

\[
\begin{bmatrix}
u_{ld} & v_{rd}
\end{bmatrix} = C^C_L \begin{bmatrix}
x_{ld}^C & x_{ld}^C \\
y_{ld}^C & y_{ld}^C \\
z_{ld}^C & z_{ld}^C
\end{bmatrix} \begin{bmatrix}
\delta u \\
\delta v
\end{bmatrix}
\] (9)

In equation (8), \(C^C_L\) represents the transition matrix from the camera body coordinate system to the image coordinate system, as detailed in equation (5). \(x_{ld}^C, y_{ld}^C, z_{ld}^C, x_{ru}^C, y_{ru}^C,\) and \(x_{ru}^C\) are the theoretical values of the coordinates of the bottom-left and top-right of the ROI in the camera body coordinates, respectively.

(2) BDS/GPS degraded

When the BDS/GPS receiver RTK differential working mode fails due to obstruction, interference, communication failure, and other factors, the system working status is degraded. The positioning error will reach up to 10 m. It is difficult for the autonomous driving system to obtain an accurate ROI through high-precision positioning and a prior map. However, the system can still determine whether we are currently driving on a traffic light section online and obtain the ROI through the fusion data of the sensors such as IMU, LiDAR, and the camera. In this case, IMU mainly measures the vehicle’s heading and attitude angles. LiDAR realizes the road boundary and separation zone detection, and the camera mainly realizes the lane line detection. We realize the positioning of the autonomous vehicle by multi-sensor data fusion. Through the fusion perception of multi-sensor, the positioning of the vehicle will not deviate from the lane. Therefore, the lateral positioning error will not exceed the lane, and the longitudinal positioning error will be large relatively. Correspondingly, when the BDS/GPS receiver works at degraded mode, the error shown in equation (7) is changed to that shown in equation (10)

\[
\delta_{LWD}^C = \begin{bmatrix}
\delta L^C_w \\
\delta W^C_L \\
\delta D^C_d
\end{bmatrix} = \begin{bmatrix}
\lambda_L (D + \delta_{t}^{H_{GNS}}) \tan \vartheta \\
\lambda_H (D + \delta_{t}^{V_{GNS}}) \tan \vartheta \\
\lambda_D \delta_T
\end{bmatrix}
\] (10)

Although the expressions of the sensors’ error are the same in different working modes, the specific value of each parameter needs to be adjusted according to the actual performance of the device itself.

(3) BDS/GPS outages

During BDS/GPS outages, the lane line and road boundary can be perceived by the camera and LiDAR, and the lateral position error can be controlled within a certain range. Without the assistance of BDS/GPS data, the system will degrade to IMU/odometer integrated navigation mode. Therefore, the RMSE of heading, pitch, and horizontal positioning errors in equation (10) should be adjusted according to the actual accuracy that the integrated navigation system can achieve.

Research on selection rules of relevant traffic lights

When the autonomous vehicle reaches the traffic lights section, it will acquire ROI automatically and then detect and recognize the status of the traffic lights according to the multi-sensor and prior map. However, in complex road environments (such as crossroads, overpass, T-intersection, roundabout.), multiple traffic lights and ROI often appear in the field of vision. It is necessary to select relevant traffic lights of the autonomous vehicle for feature extraction and recognition. Figure 2 shows several typical urban road traffic lights settings in the Chinese national standard “Specifications for road traffic lights setting and installation.” Without the aid of prior knowledge, the semantic recognition of traffic lights based on vision will be a tremendous technical challenge. It can be seen from Figure 2 that when the autonomous vehicle is driving at a complex intersection, there will often be a variety of traffic lights in the field of vision. Under
unfamiliar road conditions, even for human drivers, it will be challenging to recognize the semantics of the traffic lights accurately.

In the previous studies, there was some deficiency in the method of relevant traffic lights selection based on vehicle positioning. In actual applications, the semantics of traffic lights are not exactly consistent with the location of the vehicle. For example, in a complex scene’s driving behavior, although the vehicle is currently in the straight lane, it may switch to the left-turn lane at the next moment. Simultaneously, due to various factors such as signal occlusion and interference, it is not easy to ensure continuous and reliable high-precision positioning. For the positioning-based methods, uncertain positioning errors will significantly increase the probability of incorrect selection of the relevant traffic lights. Therefore, we select the relevant traffic lights according to the autonomous driving trajectory planning and decision commands, effectively reducing the error rate of matching.

In our prior map, the semantics of the traffic lights are labeled in detail. At the same time, the trajectory parameters and control commands can be provided by the controller of the autonomous system in real-time. In the traffic lights section, the autonomous vehicle will match the relevant traffic lights efficiently and accurately. Besides, the recognition accuracy can be effectively improved through the matching of relevant traffic lights. For example, in the left-turn driving behavior, only the traffic lights with left-turn semantics need to be concerned, and other information is classified as less important.

**Feature extraction and recognition algorithm based on deep learning**

In the practical application of traffic lights recognition in complex and open scenarios, the deep learning-based traffic lights recognition method has higher accuracy than traditional algorithms. However, the computational efficiency of which is slightly insufficient. Through comprehensive analysis of various deep learning algorithms such as Faster RCNN, YOLO, and SSD, YOLO has apparent advantages in computing efficiency. Considering that the requirement for real-time performance of traffic lights recognition is very high in autonomous driving. Moreover, the multi-sensor fusion data and prior map assistance can significantly improve the image quality and make up the shortcoming of the one-step deep learning algorithm. So, we adopted the latest proposed YOLOv4 in this article. The YOLOv4 model had been tested on the COCO data set, and the object identification accuracy on the Tesla V100 hardware computing platform reached 65.7% AP50, which was nearly 10% higher than YOLOv3. At the same time, the speed of YOLOv4 had reached 65FPS, which met the performance requirements of real-time in autonomous driving scenarios. The block diagram of our improved traffic lights recognition algorithm based on YOLOv4 was shown in Figure 3.

As shown in Figure 3, the left side is the auxiliary information provided by the real-time fusion data of multi-sensor and prior map for the deep learning algorithm. It mainly includes the center point and bounding box, ADA model, and prior knowledge (attributes of the traffic lights and road). The location of the center point is the first step to obtain the ROI. According to the bounding box, the ROI is preliminarily acquired in the input image. Then, we further optimize the ROI through the ADA model. With the aid of the prior knowledge on the left, the YOLOv4-based traffic lights recognition is realized online in the right part of Figure 3.

![Figure 2. The Setting Methods of Traffic Lights at Several Typical Crossroads on Urban Roads in China.](image)
Experimental verification

The principle of the improved traffic lights recognition algorithm was described in detail in the third section. In this section, experimental verification based on the public dataset and actual road test was conducted, and the results were analyzed in depth. The traffic lights recognition algorithm proposed in this article were based on prior knowledge and multi-sensor data fusion. This article’s primary purpose was to solve the problem of low accuracy of traffic lights recognition in severe weather and extreme illumination. Public datasets such as GTSRB and VIVA were collected and labeled in normal lighting conditions, which were not suitable for verifying our algorithm. Our test was conducted based on the LISA dataset. A large number of sample pictures of the night were collected and labeled in this dataset. We mainly focused on performance testing in severe conditions such as night, rain, and snow. The improvement of our algorithm compared to YOLOv4 was analyzed in detail. Moreover, actual road test at night was carried out in Beijing based on the intelligent networked patrol vehicle developed by the National Engineering Laboratory for Integrated Command and Dispatch Technology.

YOLOv4 deep learning model’s training

The YOLOv4 deep learning model was trained and tested based on the LISA dataset. LISA dataset contains 29,911 pictures in six categories: red lights, yellow lights, green lights, left-turn red lights, left-turn yellow lights, and left-turn green lights. The training data set contains 16,312 daytime pictures (56.23% of the total data) and 12,699 nighttime pictures (43.77% of the total data). The test data set contains 7254 pictures, including 4099 daytime pictures (56.51% of the total data) and 3155 nighttime pictures (43.49% of the total data). Nvidia DGX-1 was used as the training platform of the deep learning model in this paper, which included four pieces of Tesla-V100-16G GPUs, two pieces of Intel Xeon Platinum 3.1 GHz processors, and 30 TB NVMe SSD storage space.

Since the four pieces of GPUs of the training platform had the same performance, the synchronous parallel mode was employed to speed up the training. In the synchronous parallel mode, all GPUs read the value of the parameter at the same time and updated the value of the parameter synchronously when the backpropagation algorithm was completed. The weighted cross-entropy loss function was adopted in the model training, L2 regularization processing was carried out, and a sliding average model was used to control the updating speed of the parameters. The training parameters of the model were set as follows: the batch size was 200, the learning rate and regularization coefficient were both 1e-4, the sliding average coefficient and learning rate attenuation coefficient were both 0.99, and the number of iterations was 30,000.

Based on the NVIDIA Jetson TX2 hardware platform, the YOLOv4 model was tested. With a frame rate of 10FPS, the recognition accuracy of the six types of traffic lights scenarios reached 92.5% for mAP50% and 78.5% for mAP50.

Experimental verification based on LISA dataset

In all, 2094 sample pictures in night were selected from the LISA test dataset, including 681 straight red lights (SRL), 458 straight green lights (SGL), 101 straight yellow lights (SYL), 441 left-turn red lights (LRL), 287 left-turn green lights (LGL), 126 left-turn yellow lights (LYL), as shown in Figure 4.

By manually segmenting the pictures, we obtained the traffic lights ROI of 64 × 64 pixels, 128 × 128 pixels, and 192 × 192 pixels to simulate the auxiliary effects of RTK BDS/GPS, degraded, and outages. YOLOv4 was used to detect and recognize the above three sizes of ROI. The results were detailed in Table 1.
Image segmentation can effectively reduce the influence of interfering light sources such as vehicle tail lights, street lights, and neon lights at night and improve the accuracy of traffic lights detection and recognition. As shown in Table 1, by segmenting 2094 pictures of night scenarios to different degrees, the recognition accuracy can be improved in different ranges. After segmenting the image according to the ROI size of $64 \times 64$ pixels, the recognition accuracy can be increased by 7.9%.

| Test dataset          | Recognition accuracy before segmentation | Segmentation of ROI | Recognition accuracy before segmentation | Increase  |
|-----------------------|-----------------------------------------|---------------------|-----------------------------------------|-----------|
| RTK (2094 photos)     | 78.5% mAP50                             | $64 \times 64$ pixels | 86.4% mAP50                             | 7.9% mAP50 |
| Degrade (2094 photos) | 128% mAP50                              | $128 \times 128$ pixels | 83.9% mAP50                             | 5.4% mAP50 |
| Outages (2094 photos) | 192% mAP50                              | $192 \times 192$ pixels | 82.1% mAP50                             | 3.6% mAP50 |

ROI: region of interest.

**Table 1. YOLOv4 recognition results after image segmentation.**

Image segmentation can effectively reduce the influence of interfering light sources such as vehicle tail lights, street lights, and neon lights at night and improve the accuracy of traffic lights detection and recognition. As shown in Table 1, by segmenting 2094 pictures of night scenarios to different degrees, the recognition accuracy can be improved in different ranges. After segmenting the image according to the ROI size of $64 \times 64$ pixels, the recognition accuracy can be increased by 7.9%.

**Actual road test**

On the urban roads at night, due to low illumination, car tail lights, neon lights, and other interference factors, it was easy to cause traffic lights to be wrongly detected, falsely detected, and missing detected. Based on the NVIDIA Jetson TX2 hardware platform, our algorithm has been engineering realized. At the same time, through the intelligent networked patrol vehicle (Figure 5(a)), an urban road test was carried out at night in Haidian District, Beijing (the test route as shown in Figure 5(b)). Through the indicators such as confusion matrix and accuracy, precision, and recall rate, the performance of the algorithm in actual scenarios was analyzed.

The total length of the test road was about 20.3 km. We started from Houtun Road, the west gate of Dongsheng Science Park, passed through Qinghe Road, Xiaoying West Road, Shangdi East Road, and...
ended at Dongsheng Science Park Road, including T-intersections, Y-intersections, crossroads, overpasses, and other road conditions. There were 49 traffic lights in the testing route, including 39 straight traffic lights, two left-turn traffic lights, eight combinations of straight and left-turn lights, and no right-turn light. In the process of prior map generation, traffic lights data were collected in normal illumination conditions. At night, the actual road test picture was shown in Figure 6. It can be seen from Figure 6 that at night the similar light sources (such as taillights, street lights, neon lights, and other sources) would cause severe interference on traffic lights detection and recognition. In an image, the color, brightness, and shape of the car taillights at a close range were remarkably similar to the red lights in the distance, and the location area of the two in the image had a higher degree of overlap. Especially when the traffic-jammed, the traffic lights were almost covered by a large number of car taillights. The height and shape of the street light were close to that of the traffic light, and it was easy to cause a false detection at a long distance. We used multi-sensor assistance and online dynamic adjustment to minimize the size of the ROI and effectively reduced the interference of the above-mentioned light sources.

In practical application, it is the most harmful and the concerning issue of green lights false positive, such as the red or yellow lights is mistakenly recognized as the straight green lights or the left-turn green lights. On the contrary, it is less harmful than green lights missed detected or mistakenly recognized as red or yellow lights. Therefore, we mainly focused on reducing the false positive probability of green lights by adjusting the threshold parameters.

We tested the performance of the algorithm at night on urban roads. In the process of testing, the results of traffic lights recognition were output online. The advantages of this algorithm were analyzed by comparing it with the recognition results of the YOLOv4 algorithm. The RTK BDS/GPS was working normally in the selected testing road. For a total of 49 traffic lights with left-turn, straight, and a combination of left-turn lights.

Table 2. Analysis table of actual road test result.

| Result | Actual | LS | MSDA | YOLO | LG | MSDA | YOLO | SS | MSDA | YOLO | SG | MSDA | YOLO | Missed recognition | Accuracy (%) | Average (%) |
|--------|--------|----|------|------|----|------|------|----|------|------|----|------|------|-------------------|-------------|-------------|
| LS     | missed recognition | 90.4 | 79.7 | 10.7 |
| LG     | missed recognition | 89.6 | 78.1 | 11.5 |
| SS     | missed recognition | 92.6 | 79.9 | 12.7 |
| SG     | missed recognition | 89.9 | 79.4 | 10.5 |

LS: left-stop; LG: left-go; SS: straight-stop; SG: straight-go; MSDA: multi-sensor data fusion assist.
The shading of table 2 represents the performance achieved by the algorithm in this paper.
and straight in the testing route, 6089 sample pictures were selected, and the result of recognition was analyzed in detail. To make the testing results analysis more intuitive, red and yellow traffic lights were defined as stop, and green lights were defined as pass. The above samples include 1033 left-stop (LS), 579 left-go (LG), 2684 straight-stop (SS), and 1793 straight-go (SG). Sample pictures were shown in Figure 6. The specific results of traffic lights recognition are shown in Table 2 and Figure 7.

It can be seen from Table 2 and Figure 7 that compared with YOLOv4, our algorithm shows significant advantages. The most prominent improvement of these was the recognition accuracy of the straight stop signal, which is improved by 12.7%. Besides, in the night with poor illumination conditions, missed detection and false detection are the main problems in the field of traffic light recognition. Without a prior map and high-precision positioning, it was relatively easy to cause traffic accidents and affect the safety of the system.

In actual application scenarios, the traditional traffic lights recognition algorithm based on YOLOv4, red, or yellow lights mean forbidden to pass, which undoubtedly brings a considerable security risk to autonomous driving. Our algorithm used the real-time positioning and prior map to label the traffic lights section. When the object cannot be detected, the current road section will be forbidden to pass. Therefore, the robustness of the system to the missed detection and mistaken recognition of the traffic lights had been improved. So, the security of the autonomous driving system was relatively enhanced. The following analyzed the test results from the perspective of decision-making safety of autonomous vehicles. First of all, the testing data were defined as follows: the red lights and yellow lights were impassable, the value of which was set to 0; the green lights meant passable, the value of which was set to 1. According to the format of the confusion matrix, the output results of the system had four situations: (1) The traffic lights were correctly recognized (TP, True Positive), including turning left and going straight; (2) The passable traffic lights were recognized as impassable (FN, False Negative); (3) Impassable situation was recognized as passable (FP, False Positive); and (4) Correct recognition of impassable traffic lights (TN, True Negative). According to formula (11), we calculated the accuracy, precision, and recall of the algorithm and YOLOv4, respectively. The results were shown in Table 3

\[
\begin{align*}
\text{Accuracy} & = \frac{(TP + TN)}{(TP + TN + FP + FN)} \\
\text{Precision} & = \frac{TP}{(TP + FP)} \\
\text{Recall} & = \frac{TP}{(TP + FN)}
\end{align*}
\] (11)

In this article, multi-sensor data and a prior map were used to improve the recognition accuracy of traffic lights recognition significantly. Simultaneously, by labeling road sections with traffic lights, the impact of missed objects on the security of the autonomous driving system was reduced considerably. It can be seen from Table 3 and figure that from the perspective of correct control commands for autonomous vehicles, the accuracy and precision of our algorithm have reached 95.4% and 99.4%, respectively. Compared with YOLOv4, the above indicators are increased by 13.9% and 28.2%, respectively, effectively improving the safety of the autonomous driving system in the night environment.

Most of the conclusions of existing research results were obtained in normal environments, which were quite different from our night urban road test. We selected two test results that were relatively close in our

| Actual | Indicator | MSDA | YOLO | YOLOv4 | John et al.9 | [10] |
|--------|----------|------|------|--------|-------------|------|
| 1      | 2131     | 1876 | 269  | 496    | 99.4%       | 71.2%| 28.2% |
| 0      | 13       | 653  | 3669 | 3173   | 88.8%       | 79.1%| 9.7%  |

MSDA: multi-sensor data fusion assist.
The shading of the table 3 represents the performance achieved by the algorithm in this paper.

![Figure 7. Comparison of the accuracy of MSDA and YOLOv4.](image)
test environment and compared them with ours. Compared with John et al.\textsuperscript{9} and Levinson et al.\textsuperscript{10} the performance improvement of our algorithm was relatively insignificant, mainly because of differences in the test environment. Our night test was conducted on urban roads with dense traffic. Illumination, car tail-lights, and street lights are all important factors that affect the recognition of traffic lights. Therefore, we can draw an important conclusion by comparing the results with John et al.\textsuperscript{9} and Levinson et al.\textsuperscript{10} The adaptability of our algorithm to the environment has been significantly improved.

Besides, the definition of precision in our confusion matrix was the probability that the real picture is a green light when the result of a traffic light recognition is passing. We believed that the false positive of green light was the most harmful to traffic safety, and it is the parameter that our algorithm should focus on improving. It can be seen from Table 3 and Figure 8 that the precision of our algorithm reached 99.4\%, which was higher than that of the other three groups. The result also means that our algorithm effectively improved the safety of the autonomous driving system.

Conclusion

Autonomous driving in the urban scenarios, the recognition of traffic lights in various harsh environments such as light source interference, extreme lighting, severe weather, and complex road conditions faces significant technical challenges. However, the problems mentioned above are unavoidable and ubiquitous in practical applications. Although this problem has been optimized and improved through a prior information assistance, multi-technology hybrid architecture, and other methods in the past research results, there is still a big gap between the current technology level and application demand. We proposed an improved method assisted by the multi-sensor fusion data and a prior map in response to this urgent need. The effectiveness and superiority of our method were verified by two ways of a public dataset and actual road test. Through theoretical derivation, experimental verification, and result analysis in this article, the following conclusions can be summarized.

(1) It is challenging to recognize the traffic lights in extreme illumination and complex road environments, and it is difficult for any single method to satisfy the actual application requirements of autonomous driving;

(2) Fusing multi-sensor data such as LiDAR, camera, IMU, BDS/GPS, and prior map to obtain ROI is an effective method to improve the accuracy of traffic lights recognition;

(3) Analyzing the relationship between the error of the sensors and the size of the ROI, the ADA model was built, based on that the best ROI could be acquired, which would effectively improve the accuracy of the deep learning-based traffic lights recognition algorithm.

Our algorithm can effectively improve the security of autonomous vehicles in severe environments. However, there is still a big gap from the actual application, especially the recognition of traffic lights under complex road conditions such as intersections, overpasses, and roundabouts interfered by multiple light sources, facing significant challenges. By integrating traditional image recognition algorithms into our algorithm, further reducing the influence of light source interference will also be an important research direction.

Author Contributions

Q. Z. and Z. L. conceived the idea, wrote the paper, and performed the experiments; L. L. developed the software; J. L. and Y. L. coordinated the research.

Declaration of conflicting interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Funding for Beijing Millions of Talents Project (No. 2020A07), the National Natural Science Foundation of China (Nos. 61533008, 61603181, 61673208, 61873125), the Funding for National Defense Basic Research Program (No. 6142006190201).
References

1. Liu Y and Li Z. Key technology and application of intelligent connected patrol vehicles for security scenario. *Telecomm. Sci.* 2020; 36(4): 53–60.

2. Muresan MP, Giosan I and Nedevschi S. Stabilization and validation of 3D object position using multimodal sensor fusion and semantic segmentation. *Sensors* 2020; 20(4): 1110.

3. Nie J, Yan J, Yin H, et al. A multimodality fusion deep neural network and safety test strategy for intelligent vehicles. In: *IEEE transactions on intelligent vehicles*, https://ieeexplore.ieee.org/document/9207961

4. Jin X-B, Yu X-H, Su T-L, et al. Distributed deep fusion predictor for a multi-sensor system based on causality entropy. *Entropy* 2021; 23: 219.

5. Kim H, Cho J, Kim D, et al. Intervention minimized semi-autonomous control using decoupled model predictive control. In: *2017 IEEE intelligent vehicles symposium (IV)*, Los Angeles, CA, 31 July 2017, pp.618–623. New York: IEEE.

6. Arikan A, Kayaduman A, Polat S, et al. Control method simulation and application for autonomous vehicles. In: 2018 international conference on artificial intelligence and data processing (*IDAP*), Malatya, Turkey, 28–30 September 2018, pp.1–4 New York: IEEE.

7. Diaz M, Cerri P, Pirlo G, et al. A survey on traffic lights detection. In: *International conference on image analysis and processing*, Genoa, Italy, 7–11 September 2015, pp.201–208. New York: Springer.

8. 3GPP. Study on enhancement of 3GPP support for 5G V2X services: TR22.886, v.15.1.0, 2017, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3108

9. John V, Yoneda K, Qi B, et al. Traffic lights recognition in varying illumination using deep learning and saliency map. In: *Proceedings of the IEEE Conference on Intelligent Transport Systems (ITSC)*, Qindao, China, 8–11 October 2014, pp.2286–2291. New York: IEEE.

10. Levinson J, Askeland J, Dolson J, et al. Traffic light mapping, localization, and state detection for autonomous vehicles. In: *2011 IEEE international conference on robotics and automation (ICRA)*, Shanghai, China, 9–13 May 2011, pp.5784–5791. New York: IEEE.

11. Mogelmose A, Trivedi MM and Moeslund TB. Vision-based traffic signs detection and analysis for intelligent driver assistance systems: perspectives and survey. *IEEE Trans Intel Transport Syst* 2012; 13(4): 1484–1497.

12. Ren F, Huang J, Jiang R, et al. General traffic sign recognition by feature matching. In: *2009 24th international conference image and vision computing*, Wellington, New Zealand, 23–25 November 2009, pp.409–414. New York: IEEE.

13. Liang M, Cui X and Song Q. Traffic sign recognition method based on HOG-Gabor feature fusion and Softmax classifier. *J Traffic Transport Eng* 2017; 17(3): 151–158.

14. Lee S, Kim J, Lim Y, et al. Traffic lights detection and recognition based on Haar-like features. In: *International conference on electronics, information, and communication*, Honolulu, HI, 24–27 January 2018, pp.1–4. New York: IEEE.

15. Greenhalgh J and Mirmehdi M. Real-time detection and recognition of road traffic signs. *IEEE Trans Intell Transport Syst* 2012; 13(4): 1498–1506.

16. Ge H and Stethy O. A fast learning algorithm for deep belief nets. *Neural Comput* 2006; 18(7): 1527–1554.

17. Krizhevsky A, Sutskever I and Hinton G. Imagenet classification with deep convolutional neural network. *Adv Neural Inform Process Syst* 2012; 60: 1097–1105.

18. Russakovsky O, Deng J and Su H. ImageNet large scale visual recognition challenge. *Int J Comp Vision* 2015; 113(3): 211–252.

19. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. *Computer Ence*, 2014, https://arxiv.org/abs/1409.1556

20. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: *Proceeding of the IEEE conference on computer vision and pattern recognition*, Boston, MA, 7–12 October 2015, pp.1–9. New York: IEEE.

21. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 2730 June 2016, pp.770778. New York: IEEE.

22. Sermanet P and LeCun Y. Traffic sign recognition with multi-scale convolutional networks. In: *2011 international joint conference on neural networks*, San Jose, CA, 31 July 5 August 2011, pp.28092813. New York: IEEE.

23. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Informat Process Syst* 2015, pp.9199, https://arxiv.org/abs/1506.01497

24. Redmon J and Farhadi A. Yolo3: an incremental improvement. *arXiv*, 2018, https://arxiv.org/abs/1804.02767

25. Jensen MB, Philipsen MP, Moelmose A, et al. Traffic lights detection at night: comparison of a learning-based detector and three model-based detectors. In: *11th symposium on visual computing*, Las Vegas, NV, 1416 December 2015, pp.774783. New York: Springer.

26. Hu Q, Paisitkriangkrai S, Shen C, et al. Fast detection of multiple objects in traffic scenes with a common detection framework. *IEEE Trans Intell Transport Syst* 2015; 17(4): 1002–1014.

27. Siogkas G, Skodras E and Dermatas E. Traffic lights detection in adverse conditions using color, symmetry and spatiotemporal information. In: *International conference on computer vision theory and applications (VISAPP)*, Rome, Italy, 2426 February 2012. New York: Springer.

28. Li Z, Zeng Q, Zhang S, et al. An image recognition algorithm based on multi-sensor data fusion assisted AlexNet model. *J Chinese Inertial Tech* 2020; 28(2): 219225.

29. Wu Y, Geng K, Xue P, et al. Traffic lights detection and recognition algorithm based on multi-feature fusion. In: *2019 IEEE 4th international conference on image, vision and computing (ICIVC)*, Xiamen, China, 57 July 2019, pp.427432. New York: IEEE.
30. Fairfield N and Urmson C. Traffic lights mapping and detection. In: 2011 IEEE international conference on robotics and automation, Shanghai, China, 11-13 May 2011, pp.5421-5426. New York: IEEE.
31. Weber M, Wolf P and Zöllner JM. DeepTLR: a single deep convolutional network for detection and classification of traffic lights. In: 2016 IEEE intelligent vehicles symposium, Gothenburg, 29-32 June 2016, pp.3423-3428. New York: IEEE.
32. Gao F and Wang C. A hybrid strategy for traffic lights detection by combining classical and learning detectors. IET Intel Transp Syst 2020; 14(7): 735-741.
33. Kim J, Cho H, Hwangbo M, et al. Deep traffic lights detection for self-driving cars from a large-scale dataset. In: 2018 21st international conference on intelligent transportation systems, Maui, HI, 4-7 November 2018, pp.2802-2807. New York: IEEE.
34. Wang J and Zhou L. Traffic lights recognition with high dynamic range imaging and deep learning. IEEE Trans Intel Transport Syst 2019; 20(4): 1341-1352.
35. Müller J and Dietmayer K. Detecting traffic lights by single shot detection. In: 2018 21st international conference on intelligent transportation systems (ITSC), Maui, HI, 4-7 November 2018, pp.2662-2673. New York: IEEE.
36. Ouyang Z, Niu J and Liu Y. Deep CNN-Based Real-Time Traffic lights Detector for Self-Driving Vehicles. IEEE Trans Mobile Comp 2020; 19: 3003-313.
37. Saini S, Nikhil S, Konda K.R, et al. An efficient vision-based traffic lights detection and state recognition for autonomous vehicles. In: 2017 IEEE Intelligent Vehicles Symposium, Los Angeles, CA, 11-14 June 2017, pp.606-611. New York: IEEE.
38. Jang C, Cho S, Jeong S, et al. Traffic lights recognition exploiting map and localization at every stage. Expert Syst Appl 2017; 88: 290-304.
39. Barnes D, Maddern W and Posner I. Exploiting 3D semantic scene priors for online traffic lights interpretation. In: 2015 IEEE intelligent vehicles symposium (IV), Seoul, South Korea, 28 June 1 July 2015, pp.573-578. New York: IEEE.
40. Possatti L.C, Guidolini R, Cardoso VB, et al. Traffic lights recognition using deep learning and prior maps for autonomous cars. arXiv, 2019, https://arxiv.org/abs/ 1906.11886#:~:text=Traffic%20Light%20Recognition%20Using%20Deep%20Learning%20and%20Prior%20Maps%20for%20Autonomous%20Cars,-Lucas%20C.&text=Autonomous%20terrestrial%20vehicles%20must%20be,identify%20the%20relevant%20traffic%20lights
41. Meng Q and Hsu L-T. Integrity monitoring for all-source navigation enhanced by kalman filter based solution separation. IEEE Sens J. Epub ahead of print 23 September 2020. DOI: 10.1109/JSEN.2020.3026081
42. Meng Q and Hsu L-T. Integrity for autonomous vehicles and towards a novel alert limit determination method. Proc IMechE, Part D: J Automobile Engineering 2021; 235(4): 996-1006.
43. Meng Q, Liu J, Zeng Q, et al. Improved ARAIM fault modes determination scheme based on feedback structure with probability accumulation. GPS Solutions 2019; 23: 16.
44. Zeng Q, Chen W, Liu J, et al. An improved multi-sensor fusion navigation algorithm based on the factor graph. Sensors 2017; 17(3): 641.
45. AQSIO and SAC. Specifications for road traffic lights setting and installation; GB14886-2016, 2016, https://www.chinesestandard.net/PDF/English.aspx/GB14886-2016
46. Bochkovskiy A, Wang C and Liao HM. Cv4: optimal speed and accuracy of object detection. arXiv, 2020, https://arxiv.org/abs/2004.10934
47. Jensen MB, Philipse M, Mogelmore A, et al. Vision for looking at traffic lights: issues, survey, and perspectives. IEEE Trans Intel Transport Syst 2015; 17: 1800-1815.