

Article

Real-Time Target Detection System for Intelligent Vehicles Based on Multi-Source Data Fusion

Junyi Zou , Hongyi Zheng and Feng Wang *

School of Automotive and Traffic Engineering, Wuhan University of Science and Technology, Wuhan 430065, China

* Correspondence: wangfeng95407@163.com

Abstract: To improve the identification accuracy of target detection for intelligent vehicles, a real-time target detection system based on the multi-source fusion method is proposed. Based on the ROS melodic software development environment and the NVIDIA Xavier hardware development platform, this system integrates sensing devices such as millimeter-wave radar and camera, and it can realize functions such as real-time target detection and tracking. At first, the image data can be processed by the You Only Look Once v5 network, which can increase the speed and accuracy of identification; secondly, the millimeter-wave radar data are processed to provide a more accurate distance and velocity of the targets. Meanwhile, in order to improve the accuracy of the system, the sensor fusion method is used. The radar point cloud is projected onto the image, then through space-time synchronization, region of interest (ROI) identification, and data association, the target-tracking information is presented. At last, field tests of the system are conducted, the results of which indicate that the system has a more accurate recognition effect and scene adaptation ability in complex scenes.

Keywords: machine vision; millimeter-wave radar; multi-source data fusion; YOLOv5 algorithm; target detection



Citation: Zou, J.; Zheng, H.; Wang, F. Real-Time Target Detection System for Intelligent Vehicles Based on Multi-Source Data Fusion. *Sensors* **2023**, *23*, 1823. <https://doi.org/10.3390/s23041823>

Academic Editor: Justyna Tomaszewska

Received: 20 December 2022

Revised: 28 January 2023

Accepted: 1 February 2023

Published: 6 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Current sensors commonly used for environment sensing in smart cars are industrial cameras, LIDAR, RADAR (millimeter-wave radar), inertial measurement unit (IMU), a global positioning system (GPS), etc. [1–5]. These sensing systems not only collect the information to be sensed in the environment, but also process the corresponding signals. For example, artificial intelligence approaches are widely used to process the environmental information extracted from vehicle-mounted cameras. To continuously improve its processing speed and accuracy, new neural networks such as DNN and FRCNN are continuously applied [6,7]. Moreover, deep learning and semantic segmentation methods based on visual data are widely used in various traffic scenarios. For instance, in terms of target detection, the advantages of various road target detection algorithms are compared in Reference [8], and the algorithms' applicable objects and applicable scenarios are explained in detail. A pedestrian detection algorithm for complex scenes is proposed in Reference [9], and an algorithm for the appearance and distance of the measurement to be inspected is proposed in Reference [10]. On the basis of target detection, Reference [11] then introduces a target-tracking algorithm based on convolutional neural networks. After experiments, it is proved that it can complete the track management and trajectory tracking of quantitative targets based on monocular data. In terms of pavement disease detection, a detection method based on a combined semantic segmentation and edge detection model is introduced in Reference [12], which effectively improves the detection accuracy. A perception method based on GPR investigation and improved mask R-CNN is introduced in Reference [13], which achieves accurate detection of vertical joints inside asphalt pavements.

Radar systems are also common in-vehicle sensor sensing systems. Automotive radar systems are responsible for the detection of objects and obstacles, their position, and their

speed relative to the vehicle [14]. However, each sensor of the environmental sensing system can sense environmental information individually, but the single device has high limitations, large errors, and different dimensions of collected information [15], which makes it difficult to achieve the detection of targets in complex scenarios [16]. With the continuous development of intelligent networked vehicle technology, the field of intelligent driving has higher requirements for the precision, accuracy, and real-time performance of sensor-perceived information [17]. To satisfy the above requirements, the theory of data fusion sensing based on two or more sensors is proposed [18,19].

Since different sensors have different properties, fusion schemes between LIDAR, RADAR, and cameras are more frequently proposed for solving vehicle and pedestrian perception problems. A fusion sensing system based on an IR camera and LIDAR was proposed in Reference [20] to solve the problem of sensing under different lighting conditions. In Reference [21], a fused camera and LIDAR sensing system optimized for hardware is proposed. The above fusion perception solutions are based on LIDAR and a camera, but considering that LIDAR is expensive and its recognition accuracy is affected by the lighting conditions in the environment similar to a camera, this paper chooses the fusion perception solution of millimeter-wave radar and a camera.

The fusion of millimeter-wave radar and camera sensing solutions has been initially applied in the field of intelligent driving. In the field of road target perception, the fusion of data collected by in-vehicle cameras and millimeter-wave radar for the perception of vehicles and people in different environments is presented in References [22–27], respectively, in the offline case. Offline here means that the sensors collect the data first and subsequently perform the data fusion analysis work on the computer. A fusion sensing algorithm with radar as the primary sensor and camera as the secondary sensor was proposed in Reference [17] and verified experimentally to prove that the method has a high accuracy rate. In the field of target tracking and track management, a millimeter-wave radar and camera fusion sensing scheme based on a polar coordinate system is proposed in Reference [28], which can effectively improve vehicle tracking accuracy and reduce computational costs. In Reference [29], a fused sensing algorithm for distributed millimeter-wave radar and camera is proposed, which is capable of accomplishing track association between MMW radar and camera, and the fused track state estimation method has excellent performance. A DNN-LSTM-based fusion tracking method is proposed in Reference [30], and its robustness and accuracy are demonstrated. Combining the above two related studies, this paper proposes a real-time fusion sensing system of millimeter-wave radar and camera based on the NVIDIA Xavier hardware system. It is a real-time system and easy to install in the car, and can be directly connected to the car ECU for other secondary development work about the smart driving field subsequently. To further ensure the accuracy of the whole fusion perception system, target tracking is performed by extending the Kalman filter to further improve the accuracy of the perception system and reduce the false detection rate.

In addition, this paper has also made corresponding enhancements in the system operation aspects. Neural network recognition methods are often used for processing visual data and are widely used in the field of vehicular traffic. A neural network based on multi-scale feature fusion was proposed in Reference [31]. This neural network is trained to combine the different features of two scanned images and thus identify the cracks in the pavement. As mentioned above, YOLO, as a neural network, is widely used in the field of intelligent transportation. For target detection, Reference [32] and Reference [33] used, respectively, YOLOv5s and YOLOv3 networks for the recognition of traffic signals and traffic signs. Among them, Reference [32] introduced the AlexNet network to further judge the recognition results of YOLO, which effectively improves the system recognition efficiency. Reference [34], on the other hand, implements the control of driverless vehicles based on the tracking detection results of the YOLOv3 network for pedestrians. In terms of pavement disease detection, a YOLOX-based pavement disease detection method was proposed in Reference [35], which effectively solves the problem of slow identification by traditional methods. To improve the comfort of driverless vehicles, a pavement disease detection

system based on the Scaled-YOLOv4 detection framework was proposed in Reference [36], and the recognition results can be used in the subsequent engineering related to vehicle control. In terms of improving accuracy, a self-supervised learning network based on “SSL YOLOv4” was proposed to solve the problem of manual labeling of data while ensuring the recognition accuracy in Reference [37]. In Reference [38], a detection recognition algorithm based on improved RES-YOLO is proposed, which significantly improves the average recognition accuracy for multi-target vehicles. In this paper, the feature pyramid network of the YOLOv5 network is improved to adapt to the hardware characteristics of the Xavier mobile processing platform, while taking into account the network size and recognition accuracy. The input of this pyramid network is changed from three feature layers to four feature layers and trained with the selected COCO dataset. The network is experimentally proven to have a higher detection rate and a lower miss detection rate.

This paper is conceived as follows, Section 2 describes the system’s framework and working principle. Section 3 presents the development environment and the components. Section 4 introduces the data pre-processing of the different sensors. Section 5 shows the detailed fusion process. Section 6 demonstrates the field tests of the system.

2. System Design

The proposed multi-source data fusion-based vehicle target detection system (referred to as target detection system) takes the approach of fusing millimeter-wave radar data and visual recognition results, and its target detection process is as follows:

- (1) Information acquisition: The jointly calibrated millimeter-wave radar and camera detect the targets in their working area; receive the measured velocity, relative position, azimuth, and other information; and release them in ROS space after pre-processing.
- (2) Temporal synchronization: Because the data format and transmission period of the two sensors are different, without temporal alignment, it is impossible to obtain the data of the two sensors for the same measurement at the same time, and thus it is impossible to start fusion detection. Because the two sensors are installed in different frontal positions, without spatial alignment, it is impossible to describe the spatial indicators of the same measurement in a unified coordinate system (body coordinate system in this paper). Two sensor data frames with the same timestamp are combined into one data frame and published to the ROS space as the temporal synchronization result. Spatial synchronization, on the other hand, projects the target coordinate points detected by the millimeter-wave radar into the camera coordinate system based on the written sensor joint calibration data and publishes them to the ROS space as the spatial synchronization result.
- (3) Fusion decision: Based on the results of spatio-temporal synchronization, the detection area obtained by visual inspection is the main focus, combined with the data detected by millimeter-wave radar to further improve the accuracy of the detection frame and complete the detection process.

The target detection system is mainly composed of vehicle platform, millimeter-wave radar, industrial camera, and intelligent driving domain controller, and its detection process is shown in Figure 1.

Object mode is an operating mode that comes with millimeter-wave radar, and the filtering algorithm built into its controller can filter out some of the stray points. Filter is a pre-processing method proposed in this paper, and its backbone is shown in Filter processing algorithm flow. ID is the number that the ROS environment puts on the data when they are received, and the timestamp is the time when the data are received. The extended Kalman filter (EKF) is used for target tracking and trajectory. The EKF is then used for target tracking and track management.

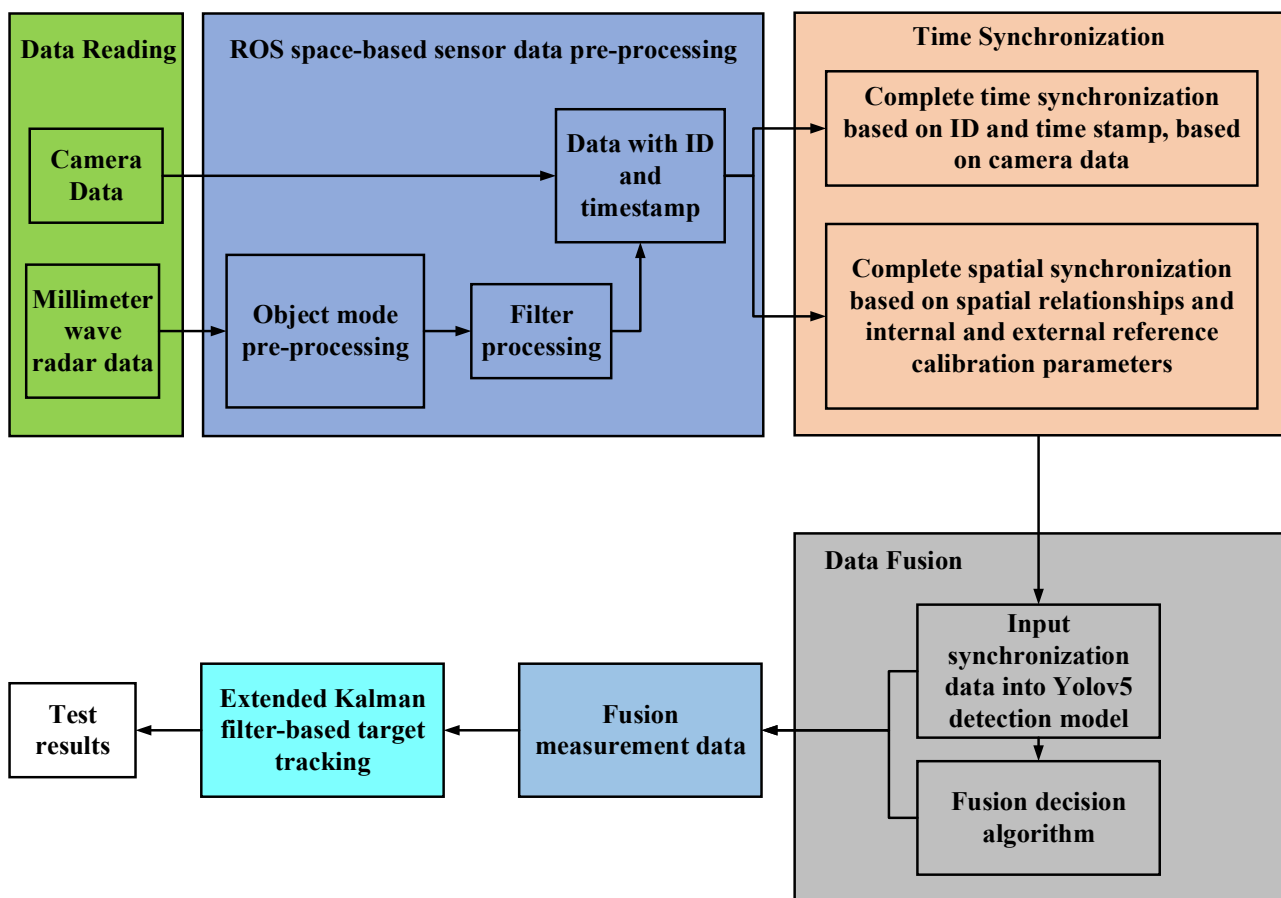


Figure 1. Target detection flow chart. The integration process is described below.

The functions of each part of the target detection system are as follows:

- (1) On-board platform: This section is used to mount the sensors, domain controller, and display. After the sensors are installed, the external parameters of the sensors are jointly calibrated according to their installation position.
- (2) Sensor: It senses the position, size, and azimuth of the target to be detected in the environment and uploads the data to the controller using CAN(Controller Area Network) communication and FPDlink communication.
- (3) Domain controller: Deploy a real-time target detection program to complete sensor data pre-processing, data fusion, and target detection on the domain controller, and publish the results.

3. System Composition

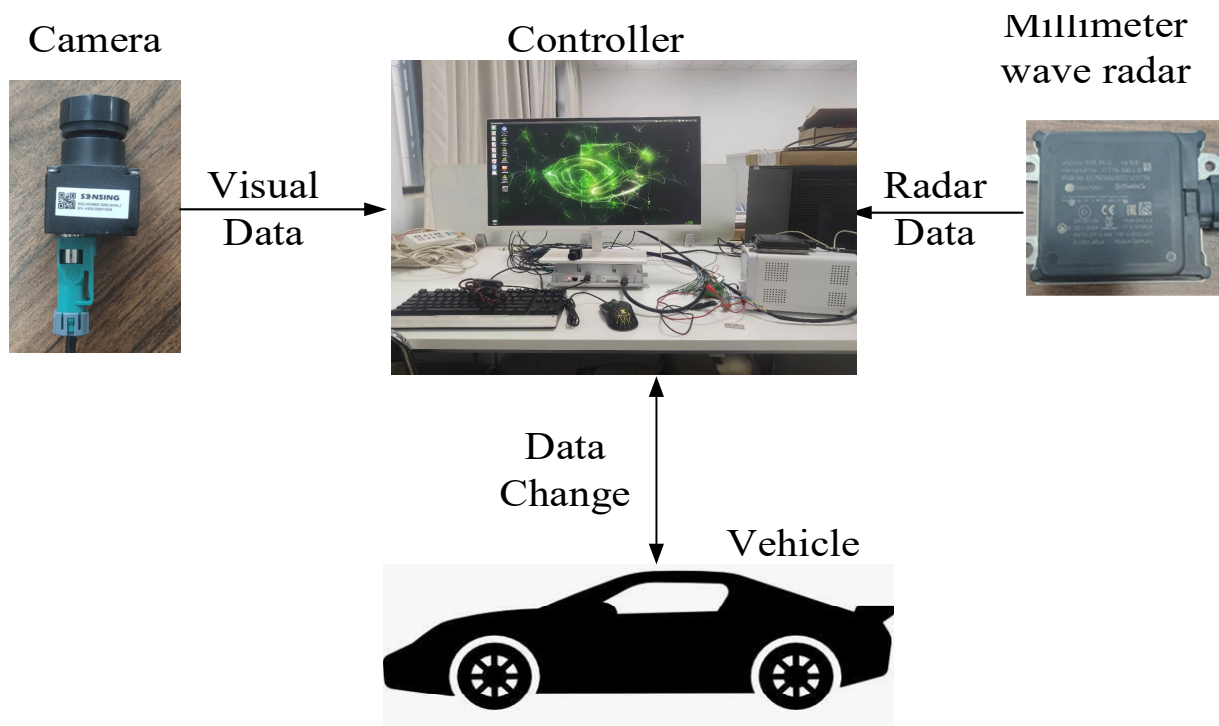
3.1. Hardware System

The hardware components of this target detection system mainly include vehicle platform, controller, sensing equipment, communication module and regulated power supply, etc. The names of the devices and their functions are introduced as shown in Table 1.

Table 1. Target detection system hardware components.

Equipment	Quantity	Function
Senyun Smart Industrial Camera	1	Acquisition of the environment to be measured
Continental ARS-408 mm wave radar	1	Obtain information such as the distance and position of the target to be measured in the environment
Cuckoo Autopilot Domain Controller	1	Deploy programs, receive data, and publish identification results, system power supply
12V DC regulated power supply	1	System power supply
7-inch display	1	Display the recognition effect of the program and related information
Communication Module	1	Data transfer between sensors

A schematic diagram of the hardware platform components and the information flow between the main components is shown in Figure 2.

**Figure 2.** Hardware platform composition diagram.

3.2. Software System

The software environment of this system is mainly based on an Ubuntu 18.04 LTS operating system on the NVIDIA Xavier AGX development board with ROS melodic and common NVIDIA GPU high-performance image processing software. Ubuntu 18.04 is the most stable and widely used open-source operating system, and Jetpack is NVIDIA's software management system specifically for this hardware control platform. The ROS software is a common upper computer software in the field of robot control. Further development is on the basis of these functions to complete space-time synchronization, data fusion, and other functions efficiently and conveniently. Cuda is a computing framework introduced by NVIDIA specifically for high-performance CPU computing, which is conducive to solving high-frequency computing problems. The yolov5 model requires the deployment of PyTorch based on the Cuda framework. PyTorch is for deep learning, and OpenCV is an open-source library that is widely used for processing camera images. The specific software environment is shown in Table 2.

Table 2. Software environment of the Target detection system.

Software Name	Versions
Ubuntu	18.04 LTS
JetPack	4.4.1
Cuda	10.2.89
Opencv	4.1.1
ROS	Melodic
Python	3.6.9
YOLOv5	PyTorch version

For the problem of transferring software data streams, this paper adopts a data communication method based on ROS messages. The system is based on the ROS environment, and the required ROS topics and their contents are defined. The data for the main links of data reading, pre-processing, space-time synchronization, target detection, and tracking can be obtained directly by subscription in the ROS space.

Finally, the UI interface subscribes to the detection results, visualizes the message, and then projects it to the corresponding location to complete all the functions of the detection system.

4. Data Processing

4.1. Visual Data Processing Based on Improved YOLO Algorithm

The vision detection part of this paper uses an improved YOLOv5 detection algorithm. The YOLO algorithm is a machine vision target detection algorithm that balances real-time and accuracy. The original YOLO algorithm was proposed by Divvula et al. in 2016 [39]. The original YOLO network had some drawbacks in terms of detection accuracy and long recognition time [40].

To address the problems of insufficient detection accuracy and long detection time, this paper uses the latest YOLOv5 target detection network for the purpose of improving the real-time detection speed. The network mainly consists of three parts: the Backbone, the feature pyramid network (FPN), and the classifier and regressor (YOLO Head). The Backbone uses the CSPdarknet network to extract features from the input images, resulting in four feature layers of different scales, Parts 1–4. Part 4 is the smallest scale layer for detecting small targets, while Part 1 is the opposite. The data in Parts 2–4 are fed into the FPN layer and continue to extract features, based on which the other feature layers are upsampled and downsampled, and CSP features are processed. After the scales are the same, these data are fused, which in turn output three final YOLO Head detectors for detection.

The system further improves the detection accuracy by changing the input part of the feature pyramid network (FPN) layer from the original four layers to three layers, and its network structure is shown in Figure 3.

In order to verify the recognition effect of the improved network, this paper conducts a comparative test on the detection effect of the improved network based on the COCO dataset. The full name of MS COCO is Microsoft Common Objects in Context, which originated from the Microsoft COCO dataset funded by Microsoft in 2014 for labeling, and is one of the most common datasets currently used in the intelligent driving domain. This dataset contains the training set for the required scenarios, as shown in the table above, with a rich training set and a corresponding test set. The training set and test set composed from this dataset can effectively demonstrate the performance improvement of the improved YOLOv5 network. A total of 4000 images were selected from the COCO dataset to form the test set, and the scenes covered different road environments and different vehicle conditions under various weather conditions. In order to test its training effect, we selected 1000 images in the same structure as a test set to check its recognition effect. The specific division of the data set is shown in Table 3a, and the results are shown in Table 3b.

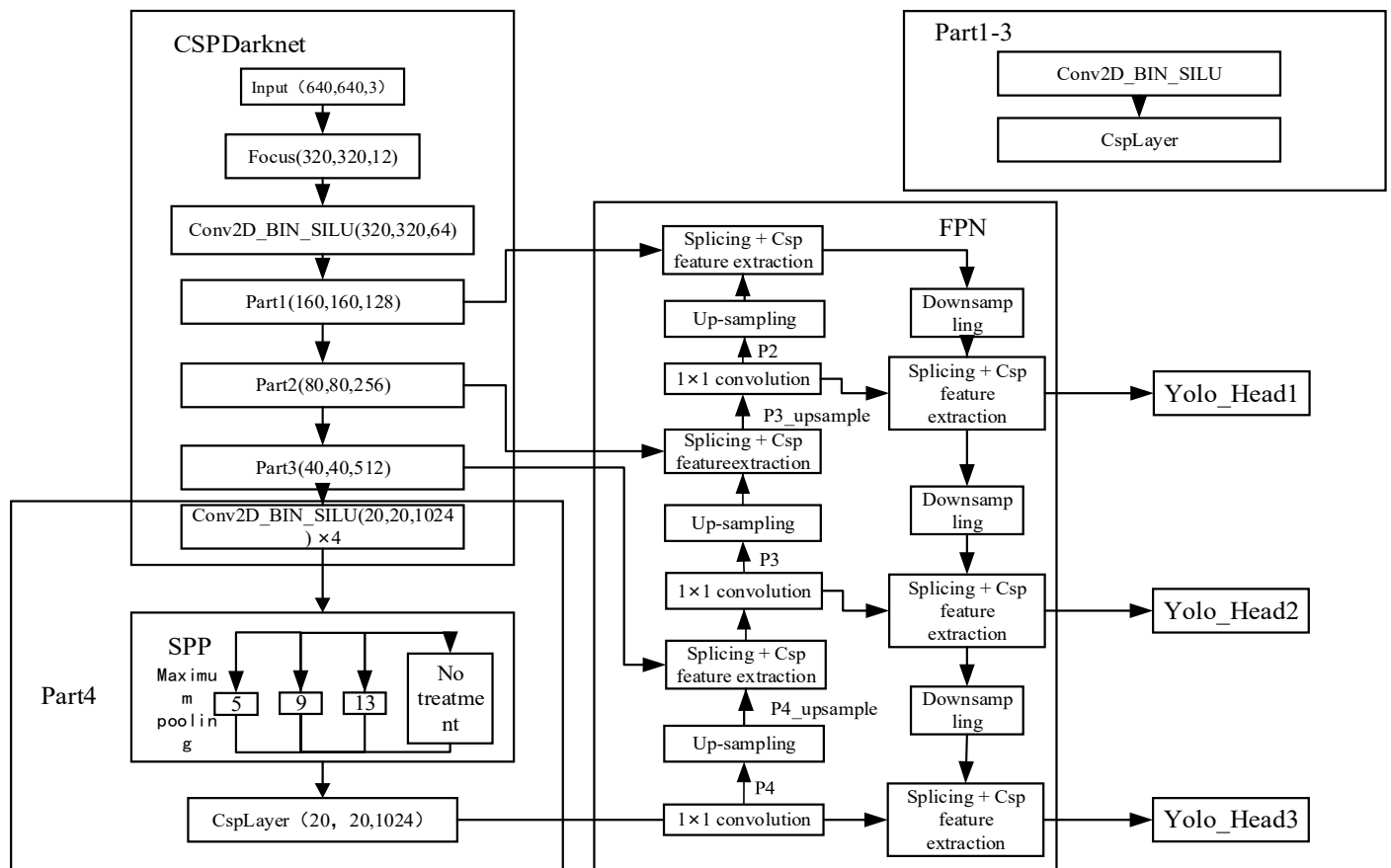


Figure 3. Target detection flow chart.

Table 3. (a) The specific division of the data set. (b) Recognition effect comparison results.

(a)						
Training Set	Sunny	Rain	Cloudy	Test Set	Sunny	Rain
Daytime open space	400	300	300	Daytime open space	100	75
Daytime Road Complex	400	300	300	Daytime Road Complex	100	75
Evening open space	400	300	300	Evening open space	100	75
Evening Road Complex	400	300	300	Evening Road Complex	100	75
(b)						
Name	Model size	Image size	Detection speed (fps)	Average positive inspection rate (%)	Recall	
Improve YOLOv5	168 M	640 × 640	31.9	88.06	80.46	
YOLOv5	90.2 M	640 × 640	32.6	87.21	79.98	
YOLOv3	237 M	640 × 640	16.5	85.54	77.66	

4.2. Millimeter-Wave Radar Data Processing

In this paper, we use Continental ARS408-21 mm wave radar, which has good software and hardware adaptability and can use the CAN network to configure radar parameters and release information related to the detected measurements [41]. This type of radar has Object mode and Cluster mode, and in this paper, Object mode with certain pre-processing functions is chosen. In Object mode, the radar outputs data such as measured distance, velocity, and mass information contained in Cluster mode, but also performs some hierarchical clustering of detected points, which can output data with fewer clutter points including collision avoidance information. A comparison of the point cloud distribution between the two modes is shown in Figure 4.

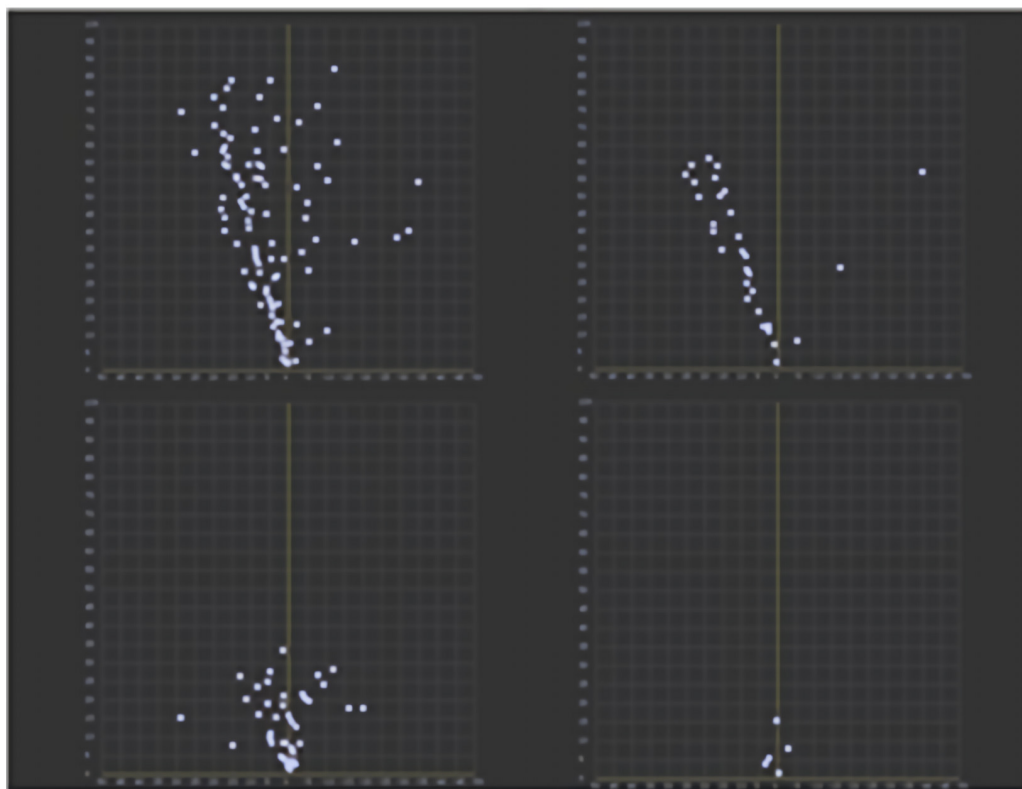


Figure 4. Target detection flow chart.

The top left is the radar point cloud map of a complex environment in Cluster mode, and the top right is the radar point cloud map of a complex environment in Object mode; the bottom left is the radar point cloud map of an open area in Cluster mode, and the bottom right is the radar point cloud map of open area in Object mode. It can be seen that the filtering algorithm has significantly improved the radar point cloud accuracy in different scenarios.

In addition, in this paper, the data output in Object mode is processed for the application environment of the target detection system to further improve the data accuracy. The processing algorithm flow is shown in Figure 5.

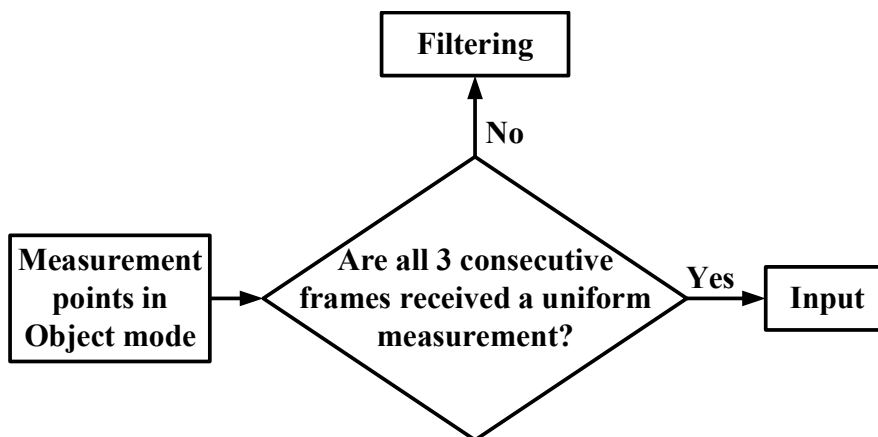


Figure 5. Filter processing algorithm flow.

5. The Key Technologies in Data Fusion

5.1. Time Synchronization

Temporal synchronization of data fusion is when the data collected by two sensors at the same moment are packaged and published [42]. Spatial synchronization of data fusion is divided into soft and hard synchronization. Hard synchronization refers to the homemade multi-sensor time triggers and the use of the triggers to unify the sensor data to be fused in terms of frequency. Soft synchronization, on the other hand, refers to the processing of the received data from different sensors according to the program and system time, thus achieving temporal synchronization of the data at the software level.

The Continental ARS408 mm wave radar and the Senyun industrial camera used in this system are less compatible with hard synchronization, so this system uses a program based on the ROS system to implement millimeter-wave-vision soft synchronization. The two sensor data forms are shown in Figure 6.

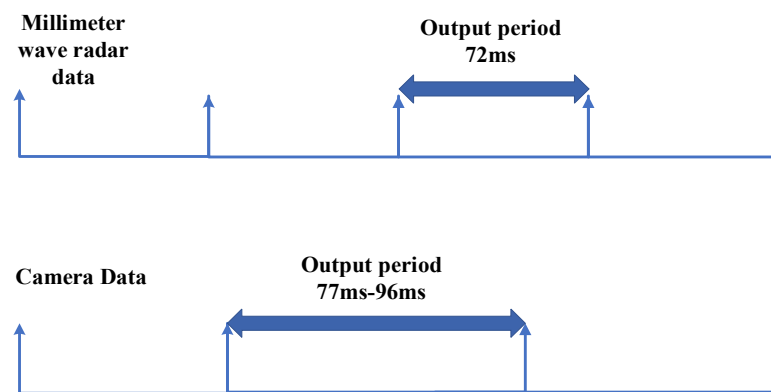


Figure 6. Data form. The axes in Figure 6 are measured in milliseconds.

As can be seen from Figure 6, the millimeter-wave radar and camera send data to the host computer at regular intervals, and this interval is the output period. The output period of the camera data is not fixed, and its time is between 77 ms and 96 ms. The camera data output period is long, and the interval is variable; the time synchronization is based on the camera data, and the nearest millimeter-wave radar data are matched with the camera data. The time synchronization process is shown in Figure 7.

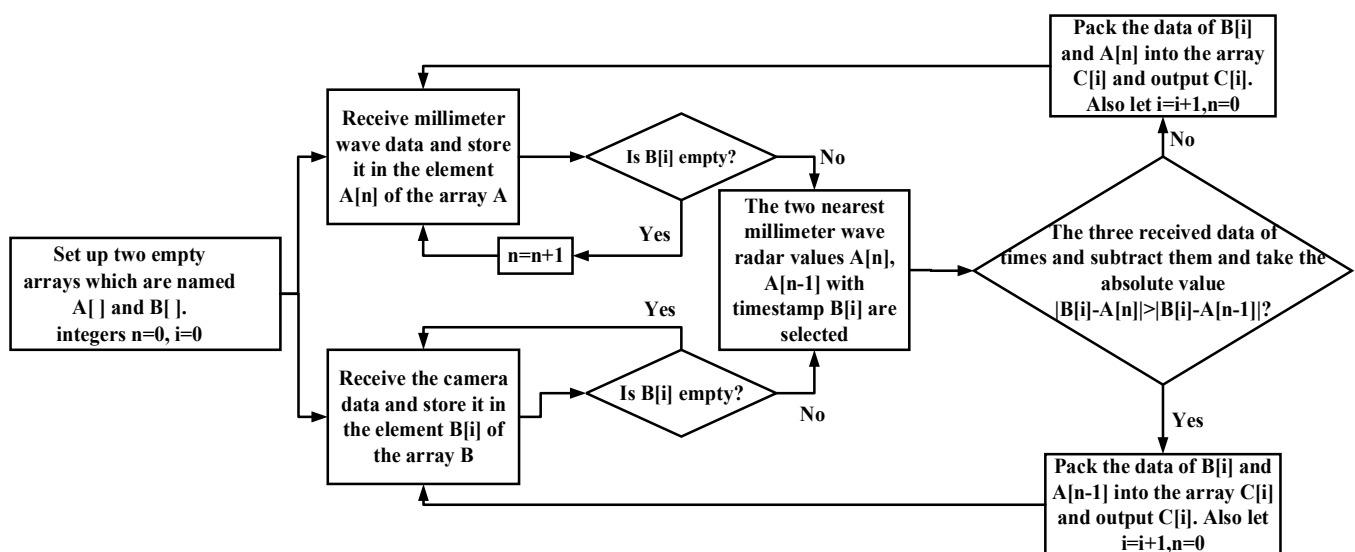


Figure 7. Time synchronization flow chart.

5.2. Space Synchronization

The spatial synchronization of data fusion is the transfer of two sensor data into one coordinate system by matrix transformation [43]. In this system, there are a total of four coordinate systems, such as body coordinate system, camera coordinate system, image coordinate system (ignoring distortion), and millimeter-wave radar coordinate system, whose positions and positive directions are shown in Figure 8.

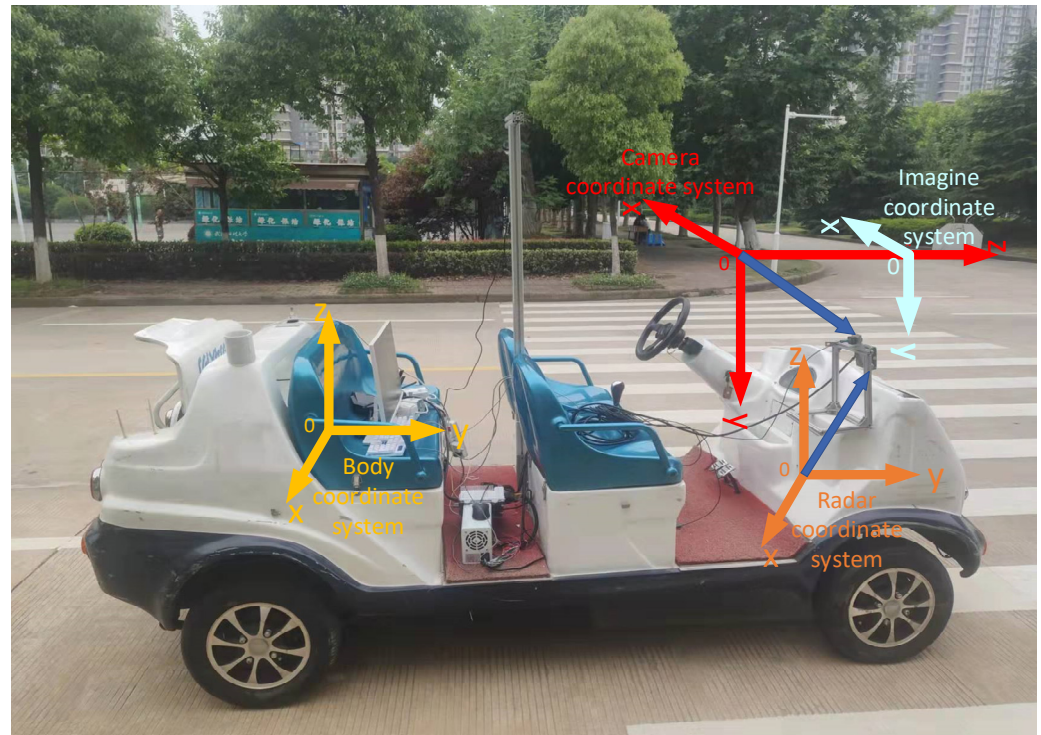


Figure 8. Schematic diagram of the coordinate system of the testing system. The red one is the camera coordinate system, the yellow one is the body coordinate system, the orange one is the millimeter-wave radar coordinate system, and the light-blue one is the image coordinate system.

In the target detection system, if the camera is considered as a coordinate system origin, then it can form a camera coordinate system in the three-dimensional space, as shown in Figure 7, and the points in this coordinate system have three dimensions of coordinates. Because the body coordinate system and the camera coordinate system are in the same three-dimensional space, the points between the two coordinate systems can be obtained by translation, rotation, and other operations. The image coordinate system is the coordinate system that exists for the pictures taken by the camera. It is a two-dimensional plane coordinate system with the center of the image as the origin. The coordinate system is in projection with the camera coordinate system, and the conversion of the points between the two coordinate systems is based on the internal and external reference calibration data of the camera. The conversion formula for the conversion from the measurement points in the body coordinate system to the measurement points in the image coordinate system is shown in Equation (1).

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M_1 M_2 X \quad (1)$$

In Equation (1), x, y are the image coordinate system coordinates; f_x, f_y, c_x, c_y are the camera internal reference matrix parameters, M_1 is the camera internal reference matrix; R, T are the camera external reference matrices, R matrix is the orthogonal matrix,

representing the rotation around the axis, T matrix represents the axial offset. M_2 is the external reference matrix; X , Y , Z are the measurement coordinates in the body coordinate system.

The measurement detected by the millimeter-wave radar is in the radar coordinate system, and in order to perform spatial synchronization, the measurement coordinates need to be converted to the body coordinate system. As shown in Figure 6, the millimeter-wave radar and the body coordinate system are in the same three-dimensional space, so the transformation from radar coordinate system measurement to body coordinate system measurement can be realized by rotation and translation, and the conversion formula is shown in Equation (2).

$$\begin{cases} x = R \times \sin \alpha \\ y = H + R \times \cos \alpha \end{cases} \quad (2)$$

In Equation (2), H is the distance from the projection point of the radar coordinate system origin on the XOY plane of the body coordinate system to the origin of the body coordinate system, R is the distance from the projection point of the obstacle on the XOY plane to the projection point of the radar coordinate system origin on the XOY plane of the body coordinate system, and α is its angle with the y-axis.

Equations (1) and (2) complete the conversion of the image coordinate system, the radar coordinate system to body coordinate system, respectively, and the relationship from the radar coordinate system to the visual coordinate system is obtained by matrix transformation of the two (Equation (3)). So far, the spatial alignment from radar data projection to image data has been completed, and the flow chart is shown in Figure 9.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} R \times \sin \alpha \\ H + R \times \cos \alpha \\ 0 \\ 1 \end{bmatrix} \quad (3)$$

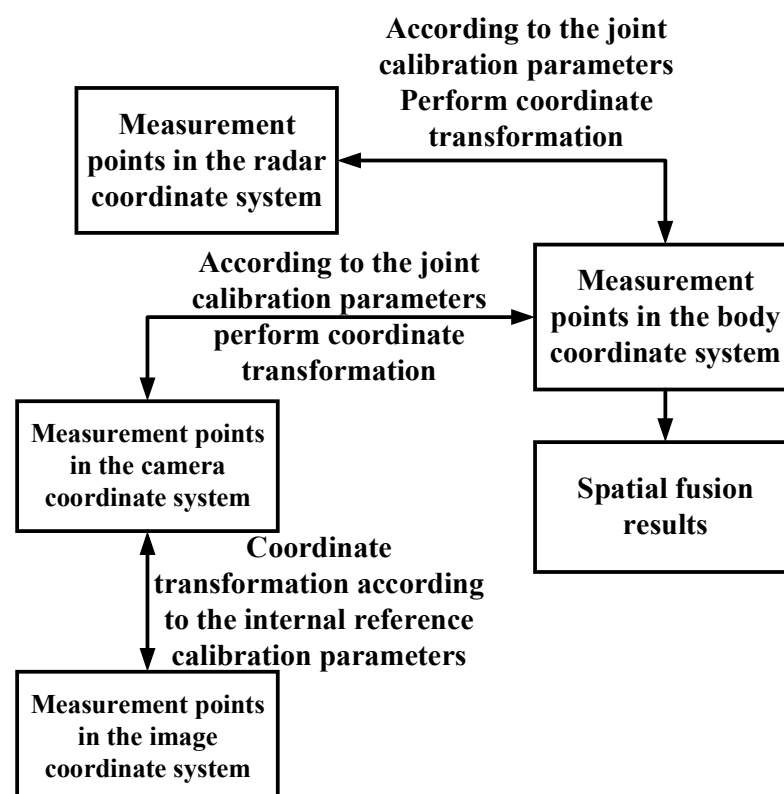


Figure 9. Spatial Alignment Flow Chart.

5.3. Data Fusion

Compared to the region of interest (ROI) generated by millimeter-wave radar [44] used by Yanping Hu et al. in this paper, the improved YOLOv5 network is chosen, and the visual detection frame accuracy has been greatly improved compared to the earlier version of the YOLO network; therefore, the visual detection results are chosen to be directly adopted in this paper.

Due to the extremely metal-sensitive nature of millimeter-wave radar and the fact that the location of the millimeter-wave contact with the detected object is not fixed, it is easy to produce drift of the measurement points [45]. Therefore, when calculating the distance between the measurement and the on-board experimental platform and the detection probability of the millimeter-wave on the measurement, the average value of all millimeter-wave point data in the detection frame is used as the detection result of the millimeter-wave radar.

When both sensors detect the object, the final detection result is output according to Equation (4).

$$\begin{cases} W = P_C \times 0.8 + P_R \times 0.2 \\ Y = P_C \times 0.2 + P_R \times 0.8 \\ X = P_C \times 0.2 + P_R \times 0.8 \end{cases} \quad (4)$$

where W is the measured detection probability, X is the horizontal distance, Y is the vertical distance, and P_C , P_R are the corresponding camera data and radar data, respectively. In Equation (4), we refer to the weighted determination method in Reference [46] on determining the presence or absence of fusion targets, and propose the output method of the measurement metrics based on the weighted data of the two sensors in Equation (4).

The output data mainly include the judgment of the measurement category and the value of the measured horizontal longitudinal distance. Firstly, since there are only two sensors, the coefficients need to add up to 1. Secondly, the two sensors have different hardware sensing principles, and the sensing accuracy of different indicators is relatively high or low. For engineering practice and prior experience, we believe that visual judgment has a higher accuracy rate in target detection, and radar has a unique advantage in distance perception. Therefore, we set the weight of camera data in W to 0.8 and radar data to 0.2, and the weight of camera data in X and Y to 0.2 and radar data to 0.8.

The rest of the cases are output according to the fusion decision shown in Table 4.

Table 4. Integration Strategy.

Camera	Millimeter-Wave Radar	Strategies
Target detected	Target detected	Perform fusion detection
No target detected	Target detected	Output millimeter-wave radar detection data
Target detected	No target detected	Output visual monitoring data
No target detected	No target detected	Judged by the extended Kalman filter

5.4. Target Detection and Tracking

Due to complex scenes or changes in vehicle conditions, the visual detection system and the millimeter-wave radar may lose the target for a short time, and the fused sensory information at this moment will be lost if no relevant processing is performed at this time. To solve this problem, this paper selects the extended Kalman filter (EKF) for fused data points for track management and tracking, which is widely used for tracking and managing target tracks because of its ability to combine sensory information and prediction results to predict the future state of the system [47].

In this paper, the mobile platform deployed by the real-time detection system is the coordinate origin, the forward direction is the y -direction, and the lateral direction is the x -direction, then the state vector of the target measurement is:

$$X = (x, y, v_x, v_y)^T \quad (5)$$

The x, y, v_x, v_y in Equation (5) are the position coordinates and velocity of the target measurement.

The equation of state and the equation of observation for the EKF are, respectively

$$\begin{cases} x_k = f(x_{k-1}, u_k) + w_k \\ z_k = h(x_k) + v_k \end{cases} \tag{6}$$

In Equation (6), x_k, x_{k-1} are the state vectors at the moment of measurement k and $k - 1$, respectively; u_k is the control vector, the system does not use the control strategy in the fusion stage, the item is 0; w_k is the process evolution noise (usually Gaussian noise); z_k is the sensor-perceived measurement observation vector; $f(x), h(x)$ are the system nonlinear state function and measurement function, respectively; v_k is the measurement noise (usually Gaussian noise). The above are the data values at the moment of k .

According to the EKF, the predicted state of the measurement target at the k th moment is

$$X_k = f(x_{k-1}, u_k) \tag{7}$$

X_k in Equation (7) is the predicted value measured at the k th moment.

In addition, the prediction error covariance can be obtained as:

$$P_k = F_k P_{k-1} F_k^T + Q_k \tag{8}$$

In Equation (8), P_k is the prediction error covariance measured at moment k , F is the accordance ratio matrix of $f(x)$, and Q_k is the covariance matrix representing the prediction state.

From Equations (6) to (8), the update equation can be obtained:

$$\begin{cases} K' = P_k H_k^T (H_k P_k H_k^T + R_k)^{-1} \\ P'_k = P_k - K' H_k P_k \end{cases} \tag{9}$$

In Equation (9), K is the Kalman gain, H_k is the transformation matrix that maps the state vector x_k to the space where the measurement value z_k is located; R_k is the covariance array of Gaussian noise of the measurement value, representing the measurement error of the sensor, K', X'_k, P'_k are the updated values.

The final state matrix of the measurement is obtained as

$$X'_k = X_k + K'(z_k - h(X_k)) \tag{10}$$

Equations (7)–(10) combine the observed value at moment $k - 1$ and the predicted value at moment k to calculate the position estimate at moment k . At the same time, the parameters such as the prediction error covariance are corrected in real time by combining the observed value at moment k , and the state matrix of the measurement is updated to realize the cyclic proceeding of the algorithm. The above process achieves the same measurement of past and present state matching.

In summary, the flow of the fusion detection algorithm used in this system is shown in Figure 10.

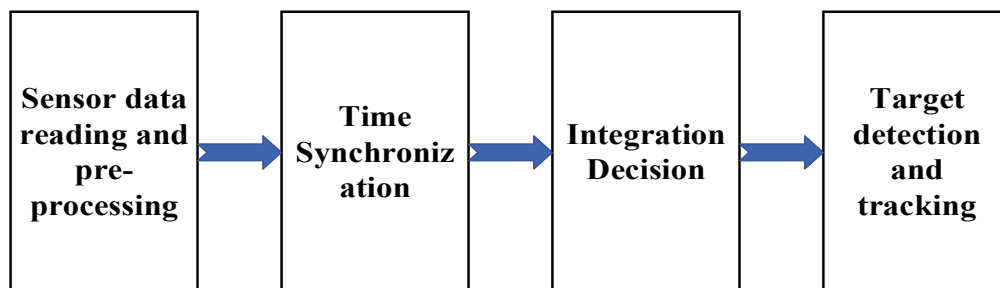


Figure 10. Flow chart of fusion algorithm.

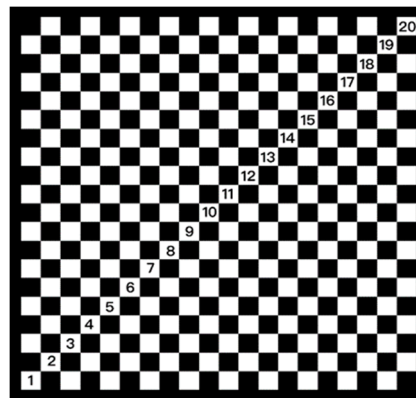
6. Equipment Commissioning and Real Vehicle Test

6.1. Equipment Calibration

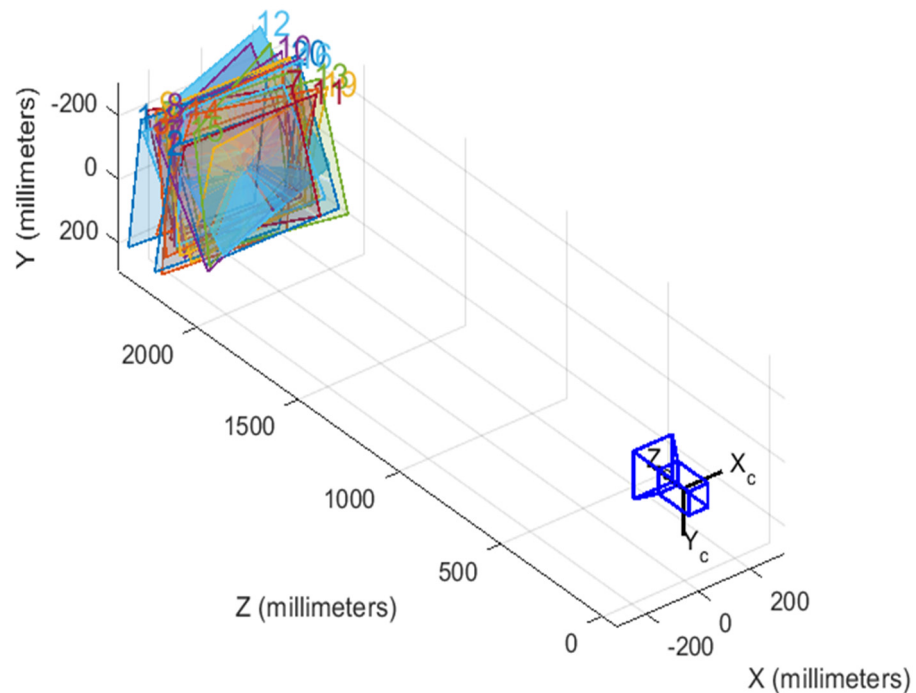
6.1.1. Internal Reference Calibration

The camera internal reference calibration of the target detection system mentioned in this paper is performed using MATLAB's camera calibration toolbox. The calibration plate is shown in Figure 11a, and the spatial position of the 20 calibration plate photos with respect to the camera is shown in Figure 11b, from which the resulting internal reference calibration parameters are given by Equation (11).

$$\begin{pmatrix} 3953.92182439858 & 0 & 1009.14636095190 \\ 0 & 3958.02267342530 & 537.408452046159 \\ 0 & 0 & 1 \end{pmatrix} \quad (11)$$



(a)



(b)

Figure 11. (a) 20 × 20, 25 mm calibration plate; (b) spatial position of the calibration plate with respect to the camera.

6.1.2. External Reference Calibration

In this paper, an open-source offline program is used for the external reference calibration. The time-synchronized millimeter-wave point cloud data and visual image data are

input into this program, and the radar point cloud positions are adjusted to coincide with the measurements on the visual image, and finally the axial offsets of the millimeter-wave point cloud data in X, Y, and Z and the rotational offsets with X, Y, and Z as the rotation axes, respectively, are recorded, and these six parameters represent the external parameters of this system [48]. The final calibration parameters are shown in Equation (12).

$$(0 \ 80 \ 0 \ 0.02 \ 0 \ 0)^T \quad (12)$$

6.2. Fusion Effect

In this system, for the convenience of external parameter calibration and equipment deployment, the millimeter-wave radar and camera are chosen to be deployed on the robot rail experimental stand, installed on the vehicle experimental platform, and then the experimental platform is deployed on the modified electric patrol car, and the final arrangement effect is shown in Figure 12.



Figure 12. Vehicle-mounted experimental platform.

To visualize the fusion effect, the millimeter-wave radar points that have been spatio-temporally synchronized are placed directly on the screen in this paper, and the horizontal coordinates and distances of the measurements are calculated, and the effect is shown in Figure 13. “99.0” is the vertical distance of 99.0 m, “x 11.80” is the horizontal distance of 11.0 m, and the purple point is the millimeter-wave radar measurement point.

To test the effectiveness of the system in detecting vehicles and pedestrians at noon and evening, the system was deployed on a mobile platform, and experiments were conducted on campus scenes, and the vehicle recognition results are shown in Figure 14. The blue box in the figure is the recognition frame of the improved YOLOv5 visual recognition network detection used in this paper, the pink dots are the measurement points detected by the millimeter-wave radar, x is the horizontal distance of the measurement point from the millimeter-wave radar, and y is the longitudinal distance of the measurement point from the millimeter-wave radar. The fusion detection is considered successful when the millimeter-wave radar point and the vision frame appear on the measurement object at the same time.



Figure 13. Fusion display effect.



Figure 14. Vehicle recognition effect map.

In this paper, relevant tests were also conducted in scenes with high pedestrian flow, and the vehicle effect is shown in Figure 15. In the face of complex scenes, the system has a high accuracy of recognition precision and a low rate of missed detection.

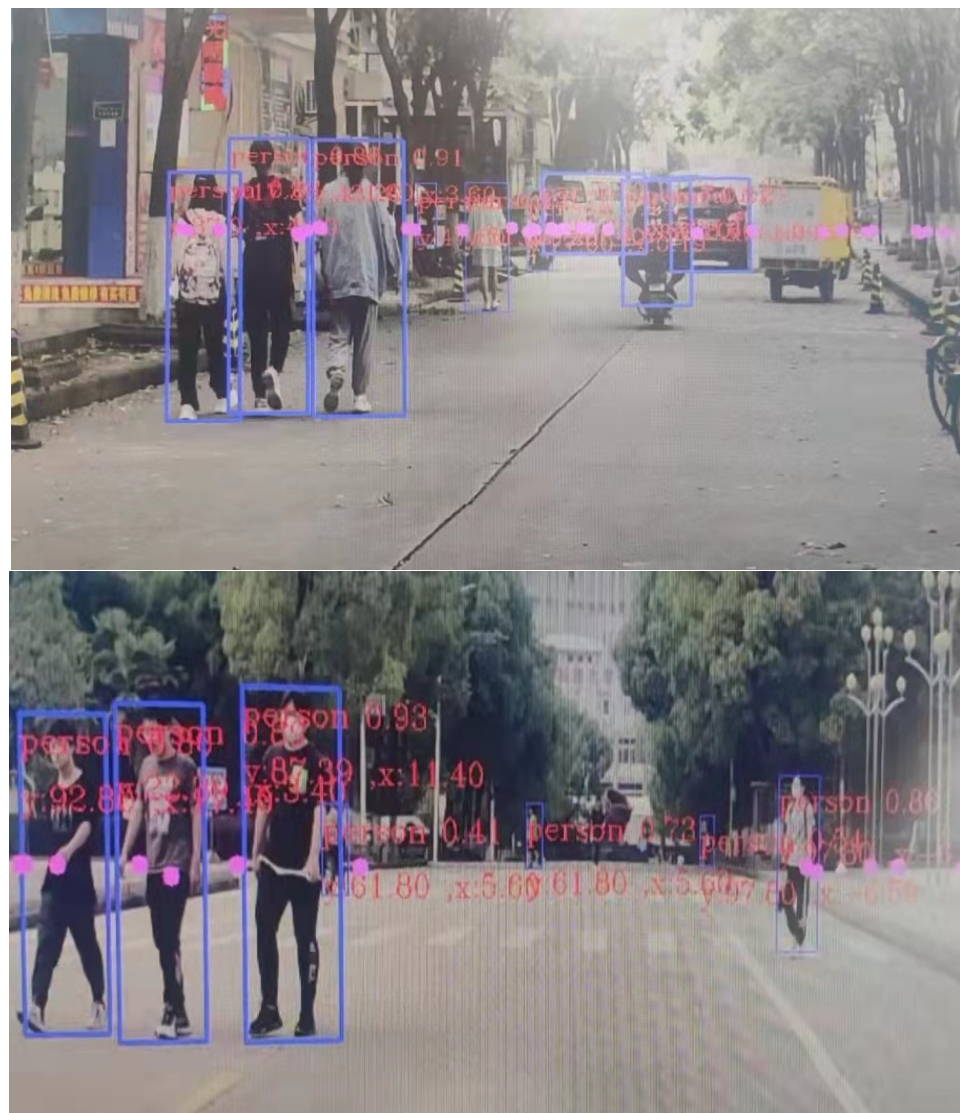


Figure 15. Complex scene recognition effect.

In order to verify the gap between the detection effect of the sensor fusion algorithm and that of a single sensor, 300 sets of data after spatio-temporal alignment were selected and processed with single vision detection, single radar detection, and fusion detection, respectively, and Table 5 shows the processing results of the three methods.

Table 5. Comparison of target detection results. The first data in each cell is for identifying vehicles and the second data is for identifying pedestrians.

Detection Method	Total Number of Targets (Vehicles/People)	Positive Inspection	Misdetction	Missing Inspection	Distance Information Error	Accuracy %	Missing Detection Rate %
Single vision inspection		716/1521	724/1544	116/251	None	86.06 85.84	13.94 14.16
Single radar detection	2604 (832/1772)	631/1323	740/1602	201/449	<2 m	75.84 74.66	24.16 25.34
Fusion Detection		766/1604	74/172	66/168	<2 m	92.07 90.52	7.93 9.48

The results in Table 5 show that the fusion target detection algorithm used in this system has a higher detection rate than the traditional method for road pedestrian and

vehicle identification, and the accuracy rate under fusion detection is higher as seen in Figures 13 and 14. At the same time, this system also combines the advantages of millimeter-wave radar to make up for the shortcomings of visual detection which makes it difficult to output the measured distance information, thus improving the scientific and applicability of the fusion detection algorithm.

7. Conclusions

In this paper, we design and implement an on-board real-time multi-target detection system, which fuses the filtered pre-processed millimeter-wave radar data and the visual data detected by YOLO for target-level data fusion, and then adopts the corresponding fusion strategy for the output of the relevant measurement targets. In addition, in order to reduce the output of pseudo-targets before the fusion results are output, this paper adopts the method of the extended Kalman filter to manage the fused measured trajectories. Compared with the traditional visual recognition detection algorithm, the proposed system in this paper outputs the measurement results with the distance information simultaneously, which improves the applicability of the system.

The experimental results show that the target detection method has a high detection rate and a low false detection rate, and the fusion of visual detection data and millimeter-wave radar data is effectively accomplished in a variety of scenarios, and the target detection effect is better than that of single-sensor detection. While performing the detection, this system can output the distance of the obstacle and the position relative to the vehicle, which improves the richness and accuracy of the detection data.

However, problems such as short delay in the fusion procedure and increased missed detection rate of complex scene target recognition were also found in the process of experimentation. In the subsequent experimental research, the target detection fusion algorithm should be further optimized to shorten the system delay, the YOLO network should be trained for more complex environments, and the algorithm structure should be adjusted by repeatedly conducting multi-scene real vehicle detection tests, so as to further improve the real-time performance and complex scene adaptability of the target detection system.

Author Contributions: J.Z., methodology, writing—original preparation, H.Z., methodology, visualization, resources, investigation, software, F.W., data curation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52202480. This research was funded by the Natural Science Foundation of Hubei Province, grant number 2022CFB732. This research was funded by the Natural Science Foundation of Hubei Province, grant number 2020CFB119. This research was funded by the Hubei Provincial Education Department Scientific Research Program Guidance Project, grant number B2021008.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, H.; Liang, J.; Zhang, Z. Active Fault Tolerant Control of Adaptive Cruise Control System Considering Vehicle-borne Millimeter Wave Radar Sensor Failure. *Transportation-Automobile Safety; New Automobile Safety Study Results from Zhejiang University Described. IEEE Access* **2020**, *8*, 11228–11240. [[CrossRef](#)]
2. Zhang, Q.; Ma, L.; Ren, J.; Zhang, M.; Pan, S. Transmission line distance measurement with millimeter wave radar based on improved robust Kalman filter. *J. Physics Conf. Ser.* **2022**, *2260*, 012029. [[CrossRef](#)]
3. Murata, H. Photonics-Applied High-Frequency Electromagnetic Field Measurement Techniques for 5G/Beyond-5G Mobile Systems and Millimeter-Wave Radars. *IEICE Proceeding Ser.* **2020**. [[CrossRef](#)]
4. Zhang, Y.; Hu, B.; Yuan, X.; Li, Y. Road object recognition method based on improved YOLOv3. *Acad. J. Comput. Inf. Sci.* **2022**, *5*, 1–9. [[CrossRef](#)]
5. Vision St Co. Ltd. Application Titled Road Object Recognition Method And Device Using Stereo Camera. USPTO Patent 20200320314, 2007.
6. Ravindran, R.; Santora, M.J.; Jamali, M.M. Multi-Object Detection and Tracking, Based on DNN, for Autonomous Vehicles: A Review. *IEEE Sens. J.* **2021**, *21*, 5668–5677. [[CrossRef](#)]

7. Singh, C.H.; Mishra, V.; Jain, K.; Shukla, A.K. FRCNN-Based Reinforcement Learning for Real-Time Vehicle Detection, Tracking and Geolocation from UAS. *Drones* **2022**, *6*, 406. [[CrossRef](#)]
8. Liu, Q.; Li, Z.; Yuan, S.; Zhu, Y.; Li, X. Review on Vehicle Detection Technology for Unmanned Ground Vehicles. *Sensors* **2021**, *21*, 1354. [[CrossRef](#)]
9. Cao, J.; Song, C.; Peng, S.; Song, S.; Zhang, X.; Shao, Y.; Xiao, F. Pedestrian Detection Algorithm for Intelligent Vehicles in Complex Scenarios. *Sensors* **2020**, *20*, 3646. [[CrossRef](#)] [[PubMed](#)]
10. Ristić-Durrant, D.; Franke, M.; Michels, K. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways. *Sensors* **2021**, *21*, 3452. [[CrossRef](#)]
11. Dichgans, J.; Kallwies, J.; Wuensche, H.-J. Robust Vehicle Tracking with Monocular Vision using Convolutional Neuronal Networks. In Proceedings of the 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Karlsruhe, Germany, 14–16 September 2020; pp. 297–302. [[CrossRef](#)]
12. Li, P.; Xia, H.; Zhou, B.; Yan, F.; Guo, R. A Method to Improve the Accuracy of Pavement Crack Identification by Combining a Semantic Segmentation and Edge Detection Model. *Appl. Sci.* **2022**, *12*, 4714. [[CrossRef](#)]
13. Liu, Z.; Yeoh, J.K.; Gu, X.; Dong, Q.; Chen, Y.; Wu, W.; Wang, L.; Wang, D. Automatic pixel-level detection of vertical cracks in asphalt pavement based on GPR investigation and improved mask R-CNN. *Autom. Constr.* **2023**, *146*, 104689. [[CrossRef](#)]
14. Patole, S.M.; Torlak, M.; Wang, D.; Ali, M. Automotive radars: A review of signal processing techniques. *IEEE Signal Process. Mag.* **2017**, *34*, 22–35. [[CrossRef](#)]
15. Melendez-Pastor, C.; Ruiz-Gonzalez, R.; Gomez-Gil, J. A data fusion system of GNSS data and on-vehicle sensors data for improving car positioning precision in urban environments. *Expert Syst. Appl.* **2017**, *80*, 28–38. [[CrossRef](#)]
16. Yang, J.; Liu, S.; Su, H.; Tian, Y. Driving assistance system based on data fusion of multisource sensors for autonomous unmanned ground vehicles. *Comput. Networks* **2021**, *192*, 108053. [[CrossRef](#)]
17. Bhatia, S.; Alsuwailam, R.I.; Roy, D.G.; Mashat, A. Improved Multimedia Object Processing for the Internet of Vehicles. *Sensors* **2022**, *22*, 4133. [[CrossRef](#)]
18. Gajda, J.; Sroka, R.; Burnos, P. Sensor Data Fusion in Multi-Sensor Weigh-In-Motion Systems. *Sensors* **2020**, *20*, 3357. [[CrossRef](#)] [[PubMed](#)]
19. Wang, T.-C.; Tong, C.-S.; Xu, B.-L. Recent Findings in Multimedia Described by Researchers from Nanjing University of Aeronautics and Astronautics (AGV navigation analysis based on multi-sensor data fusion). *Multimed. Tools Appl.* **2020**, *79*, 5109–5124. [[CrossRef](#)]
20. Choi, J.D.; Kim, M.Y. A Sensor Fusion System with Thermal Infrared Camera and LiDAR for Autonomous Vehicles: Its Calibration and Application. In Proceedings of the 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Jeju Island, Republic of Korea, 7–20 August 2021; pp. 361–365. [[CrossRef](#)]
21. Mendez, J.; Molina, M.; Rodriguez, N.; Cuellar, M.; Morales, D. Camera-LiDAR Multi-Level Sensor Fusion for Target Detection at the Network Edge. *Sensors* **2021**, *21*, 3992. [[CrossRef](#)] [[PubMed](#)]
22. Garcia, F.; Cerri, P.; Broggi, A.; de la Escalera, A.; Armingol, J.M. Data fusion for overtaking vehicle detection based on radar and optical flow. In Proceedings of the 2012 IEEE Intelligent Vehicles Symposium, Madrid, Spain, 3–7 June 2012. [[CrossRef](#)]
23. Wang, Z.; Miao, X.; Huang, Z.; Luo, H. Research of Target Detection and Classification Techniques Using Millimeter-Wave Radar and Vision Sensors. *Remote. Sens.* **2021**, *13*, 1064. [[CrossRef](#)]
24. Jiang, Q.; Zhang, L.; Meng, D. Target Detection Algorithm Based on MMW Radar and Camera Fusion. In Proceedings of the CONF-CDS 2021: The 2nd International Conference on Computing and Data Science, Stanford, CA, USA, 28–30 January 2021.
25. Kim, K.-E.; Lee, C.-J.; Pae, D.-S.; Lim, M.-T. Sensor fusion for vehicle tracking with camera and radar sensor. In Proceedings of the 2017 17th International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 18–21 October 2017. [[CrossRef](#)]
26. Liu, Z.; Cai, Y.; Wang, H.; Chen, L.; Gao, H.; Jia, Y.; Li, Y. Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions. In Proceedings of the IEEE Transactions on Intelligent Transportation Systems, Indianapolis, IN, USA, 19–22 September 2021; pp. 1–14.
27. Weon, I.-S.; Lee, S.-G. Environment Recognition Based on Multi-sensor Fusion for Autonomous Driving Vehicles. *J. Inst. Control. Robot. Syst.* **2019**, *25*, 125–131. [[CrossRef](#)]
28. Ma, J.; Tian, Z.; Li, Y.; Cen, M. Vehicle Tracking Method in Polar Coordinate System Based on Radar and Monocular Camera. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020.
29. Chen, B.; Pei, X.; Chen, Z. Research on Target Detection Based on Distributed Track Fusion for Intelligent Vehicles. *Sensors* **2019**, *20*, 56. [[CrossRef](#)] [[PubMed](#)]
30. Sengupta, A.; Jin, F.; Cao, S. A DNN-LSTM based Target Tracking Approach using mmWave Radar and Camera Sensor Fusion. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; 688–693. [[CrossRef](#)]
31. Liu, Z.; Gu, X.; Chen, J.; Wang, D.; Chen, Y.; Wang, L. Automatic recognition of pavement cracks from combined GPR B-scan and C-scan images using multiscale feature fusion deep neural networks. *Autom. Constr.* **2023**, *146*, 104698. [[CrossRef](#)]
32. Niu, C.; Li, K. Traffic Light Detection and Recognition Method Based on YOLOv5s and AlexNet. *Appl. Sci.* **2022**, *12*, 10808. [[CrossRef](#)]

33. Mijic, D.; Brisinello, M.; Vranjes, M.; Grbic, R. Traffic Sign Detection Using YOLOv3. In Proceedings of the 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 9–11 November 2020; pp. 1–6. [[CrossRef](#)]
34. Fang, R.; Cai, C. Computer vision based obstacle detection and target tracking for autonomous vehicles. *IMATEC Web Conf.* **2021**, *336*, 07004. [[CrossRef](#)]
35. Wang, L.; Li, J.; Kang, F. Crack Location and Degree Detection Method Based on YOLOX Model. *Appl. Sci.* **2022**, *12*, 12572. [[CrossRef](#)]
36. Fassmeyer, P.; Kortmann, F.; Drews, P.; Funk, B. Towards a Camera-Based Road Damage Assessment and Detection for Autonomous Vehicles: Applying Scaled-YOLO and CVAE-WGAN. In Proceedings of the 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 27–30 September 2021; pp. 1–7. [[CrossRef](#)]
37. Zhang, Y.; Hou, X. Combining Self-Supervised Learning and Yolo v4 Network for Construction Vehicle Detection. *Mob. Inf. Syst.* **2022**, *2022*, 9056415. [[CrossRef](#)]
38. Li, Y.; Wang, J.; Huang, J.; Li, Y. Research on Deep Learning Automatic Vehicle Recognition Algorithm Based on RES-YOLO Model. *Sensors* **2022**, *22*, 3783. [[CrossRef](#)] [[PubMed](#)]
39. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Target Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
40. Gu, J.; Luo, S. An improved YOLO v3 based vehicle detection method. *Agric. Equip. Veh. Eng.* **2021**, *59*, 98–103.
41. Li, X. Research on Modeling and Simulation of Millimeter Wave Radar for Automotive Intelligent Driving. Ph.D. Thesis, Jilin University, Changchun, China, 2020. [[CrossRef](#)]
42. Zhang, J.; Liu, T.; Li, R.; Liu, D.; Zhan, J.; Kan, H. Research on time calibration method for multisensor fusion for autonomous driving. *Automot. Eng.* **2022**, *44*, 215–224. [[CrossRef](#)]
43. Yan, Q. Forward Obstacle Detection Based on Millimeter Wave Radar and Camera Data Fusion. Master's Thesis, Chongqing University of Technology, Chongqing, China, 2021. [[CrossRef](#)]
44. Hu, Y.P.; Liu, F.; Wei, Z.Y.; Zhao, L.F. Millimeter wave radar and vision sensor information fusion for vehicle tracking. *China Mech. Eng.* **2021**, *32*, 2181–2188. [[CrossRef](#)]
45. Niu, C.; Su, O.; Wu, K.; Xu, J.; You, G. Testing and research on sensor ranging accuracy of automobile forward collision warning system. *J. Automot. Eng.* **2019**, *9*, 373–379.
46. Jiang, Q.; Zhang, L.; Meng, D. Target Detection Algorithm Based on MMW Radar and Camera Fusion. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019. [[CrossRef](#)]
47. Tan, G.; Fu, D.; Ding, Y.; Wang, Y. Vehicle driving state estimation based on extended Kalman filter. *J. Guangxi Univ. Sci. Technol.* **2020**, *31*, 18–24. [[CrossRef](#)]
48. Ye, H.; Ruan, H.; Wang, Y. Calibration study of robot-based docking position visual measurement system. *Mechatron. Prod. Dev. Innov.* **2022**, *35*, 9–12.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.