remote sensing

MDPI

Article

A Robust Target Detection Algorithm Based on the Fusion of Frequency-Modulated Continuous Wave Radar and a Monocular Camera

Yanqiu Yang, Xianpeng Wang, Xiaoqin Wu *, Xiang Lan, Ting Su and Yuehao Guo

School of Information and Communication Engineering, Hainan University, Haikou 570228, China; 22220854000131@hainanu.edu.cn (Y.Y.); wxpeng2016@hainanu.edu.cn (X.W.); xlan@hainanu.edu.cn (X.L.); suting4190@hainanu.edu.cn (T.S.); gyhao1996@hainanu.edu.cn (Y.G.)
* Correspondence: xq_wu@hainanu.edu.cn

Abstract: Decision-level information fusion methods using radar and vision usually suffer from low target matching success rates and imprecise multi-target detection accuracy. Therefore, a robust target detection algorithm based on the fusion of frequency-modulated continuous wave (FMCW) radar and a monocular camera is proposed to address these issues in this paper. Firstly, a lane detection algorithm is used to process the image to obtain lane information. Then, two-dimensional fast Fourier transform (2D-FFT), constant false alarm rate (CFAR), and density-based spatial clustering of applications with noise (DBSCAN) are used to process the radar data. Furthermore, the YOLOv5 algorithm is used to process the image. In addition, the lane lines are utilized to filter out the interference targets from outside lanes. Finally, multi-sensor information fusion is performed for targets in the same lane. Experiments show that the balanced score of the proposed algorithm can reach 0.98, which indicates that it has low false and missed detections. Additionally, the balanced score is almost unchanged in different environments, proving that the algorithm is robust.

Keywords: decision-level information fusion; target matching; FMCW radar; lane detection; YOLOv5

check for updates

Citation: Yang, Y.; Wang, X.; Wu, X.; Lan, X.; Su, T.; Guo, Y. A Robust Target Detection Algorithm Based on the Fusion of Frequency-Modulated Continuous Wave Radar and a Monocular Camera. *Remote Sens.* **2024**, *16*, 2225. https://doi.org/10.3390/rs16122225

Academic Editor: Andrzej Stateczny

Received: 13 May 2024
Revised: 15 June 2024
Accepted: 17 June 2024
Published: 19 June 2024

Copyright: © 2024 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

According to surveys, around 1.35 million fatalities and 50 million injuries occur globally each year due to car accidents [1,2]. The quality of vehicle hardware is undoubtedly the foundation of safe driving. However, drivers' inexperience and inattention are the leading causes of car accidents compared to the current mature and reliable technology for producing vehicle hardware [3]. On the contrary, there is no such thing as inattention, fatigued driving, drunk driving, or operating errors with automated driving. Therefore, the development of autonomous driving to reduce collisions has become a focus of discussion in various fields. As one of the core hardware for realizing autonomous driving, vehicle sensors can provide rich environment sensing information like the human sensory system [4]. So, they can be used to reduce traffic accidents caused by drivers. Radar, lidar, and cameras, as several sensors widely used in self-driving cars, are often combined to provide more comprehensive target information and more accurate detection results than a single sensor [5]. The more mainstream approaches to combining sensors include fusing information from radar and cameras, lidar and cameras, or all three [6]. Lidar has two significant drawbacks, namely that it costs more than other sensors and has low reliability in complex environments [7]. Even though the current price of lidar has shown a downward trend, it is still restricted by factors such as lighting conditions, extreme weather, and electromagnetic disturbance [8–10]. Cameras enable object detection and semantic understanding, which are impossible with lidar. For instance, the data obtained through a camera can be utilized to identify traffic signs and gesticulations [11,12]. Radar can accurately obtain information on the range, direction, and velocity of an object [13,14], which are challenging to obtain

with a camera. Therefore, approaches that integrate radar and camera data have become prevalent in both research and commercial applications [15].

In decision-level information fusion, the ranges of interest (ROIs) of different objects usually overlay each other, which may lead to false alarms. In [16], when the radar ROIs of the same target overlap, the authors use the most extensive box as the preferred one and then calculate the average ROI from two ROIs with similar sizes and bottoms. However, this method can only address cases where the ROIs of targets detected by the same sensor overlap. Additionally, radar often makes false detections so that irrelevant targets are considered valid [17]. In order to eliminate spurious detections, the density-based spatial clustering of applications with noise (DBSCAN) method is employed for grouping radar point clouds [18]. Experiments have shown that many radar misdetections can be eliminated by using this additional clustering method.

The visual target detection parts of the above fusion algorithms use a traditional hand-designed feature approach. This approach usually requires a priori knowledge to recognize targets on the road [19]. Nevertheless, the complexity and variability of scenarios will make target detection based on a priori appearance challenging. Unlike traditional manual feature design methods, deep learning methods can automatically learn the features of targets. Meanwhile, deep learning methods unify feature extraction and classifier learning in a single framework and are capable of end-to-end learning. Therefore, they have received extensive attention in recent years [20–22].

The visual detection component in [23] uses Faster R-CNN [24] to detect targets. Although this method can achieve high accuracy without prior knowledge, the two-stage detection algorithm is too slow to use in practical engineering projects. In [25], a technique for detecting objects that combines millimeter wave radar and optical sensors is introduced. In this publication, the YOLOv4 algorithm is utilized for real-time object detection. This algorithm boasts the ability to detect targets promptly. However, the architecture of YOLOv4 is intricate, and it has a substantial model footprint. So, there are still high requirements for computing capability and device storage resources.

Compared to using a single sensor for target detection, all of these fusion methods mentioned above are able to obtain better detection performance. However, these methods still suffer from the disadvantages of low detection accuracy, insufficient system robustness, and high computational cost. In particular, when the ROIs between different targets detected by different sensors overlap, the decision-level information fusion methods will suffer from the problems of a low target matching success rate and imprecise multi-target detection accuracy. In order to address these problems, a robust target detection algorithm based on the fusion of frequency-modulated continuous wave (FMCW) radar and a monocular camera is proposed in this paper.

Firstly, the method employs a reliable lane marker-based detection algorithm [26] to process the image and extract lane information. Then, the two-dimensional fast Fourier transform (2D-FFT), constant false alarm rate (CFAR), Angle-FFT, and DBSCAN algorithms are applied to pre-process the radar measurements. Furthermore, through the exclusion of null targets and the establishment of lane boundary thresholds, the valid targets are selected from the radar detections. Next, YOLOv5 [27] is used to extract the targets from the visual data. Immediately, the targets identified as valid by both vision and radar are sorted into their respective lanes. Ultimately, data from both the radar and camera are integrated within the same lane by aligning their spatial and temporal dimensions. This is achieved through the calibration of the camera, transformation of coordinate systems, and synchronization of sensor sampling instances. The experimental results indicate that this technique not only successfully alleviates the constraints associated with relying on a single sensor for sensing, but also bolsters the precision of information fusion. The primary contributions of the proposed methodology can be encapsulated in the subsequent points:

(1) We introduce a lane detection algorithm to filter out interfering targets outside the lanes, reducing the amount of data processing in the subsequent fusion algorithm;

(2) The positions of the detected lane lines are used as the effective detection ranges of targets instead of the traditional manual setting method, which improves the universality of the algorithm;

(3) We use the YOLOv5s model to detect targets in images and fuse the information of targets with the same lane markings from different sensors. This method can reduce the mutual interference between different targets in neighboring lanes, improving the fusion algorithm's accuracy.

The other parts of this paper are summarized below. Section 2 examines the information fusion framework. We present some critical steps and algorithms for decision-level information fusion in Section 3. Section 4 collects data from realistic environments to validate the algorithm proposed in this paper. Section 5 gives conclusions based on the experimental results.

2. Information Fusion Framework

The framework for information fusion at the decision level [28,29] used in this paper is shown in Figure 1.

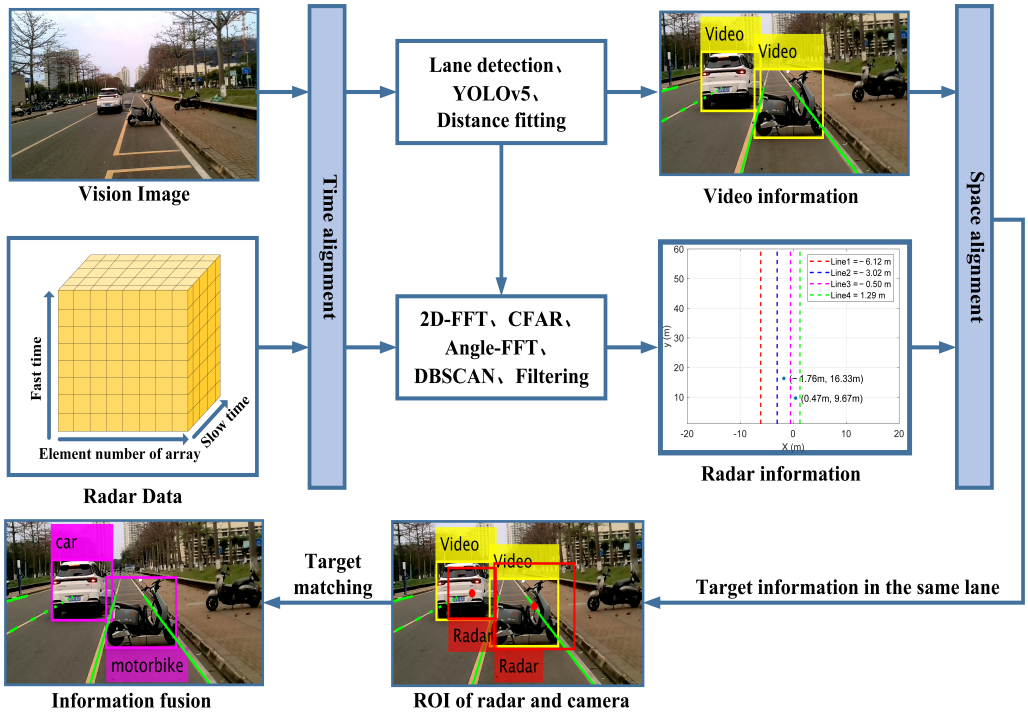


Figure 1. Decision-level fusion framework.

The fusion framework in this paper consists of three parts: visual target detection, radar target detection, and decision-level information fusion. In the visual target detection part, the images are first processed using a lane detection algorithm to obtain lane information. Then, target detection is performed using YOLOv5 and a ranging algorithm based on data fitting [30] to obtain the type and distance of the target. In the radar target detection part, algorithms such as 2D-FFT [31], CFAR [32], and Angle-FFT are first executed to obtain the original point clouds detected by radar. In accordance, the point clouds are clustered by the DBSCAN algorithm [33]. Then, the lanes are utilized to filter out the interfering targets located outside the lanes and group the remaining valid targets. In the information fusion part, targets from the two sensors in the same lane are first spatiotemporally calibrated. Then, the intersection over union (IOU) method is used to match the targets from different sensors. Finally, corresponding fusion strategies are developed based on the matching results.

3. Realization of Fusion

3.1. FMCW Radar Equation

Figure 2 presents the principle of the frequency-modulated continuous wave (FMCW) radar Equation [34]. Its fundamental ideology is to construct periodic frequency slopes, which are proportional to the modulation time. Supposing that the frequency modulation range is B and the duration is T_{chirp}, the expression for slope $f_T(t)$ is derived:

$$f_T(t) = f_0 + \frac{B}{T_{chirp}}t \tag{1}$$

Here, f_0 denotes the initial frequency of the signal, while T_{chirp} takes a value within the range of $[-T_c/2, T_c/2]$. Assuming that the transmitted signal is $S_T(t) = \cos(\psi_T(t))$, the phase $\psi_T(t)$ can be calculated by integrating Equation (1):

$$\begin{aligned}\psi_T(t) &= 2\pi \int_{-T_c/2}^{t} f_T(t)\, dt \\ &= 2\pi \int_{-T_c/2}^{t} (f_0 + \frac{B}{T_{chrip}}t)dt \\ &= 2\pi \left(f_0 t + \frac{1}{2}\frac{B}{T_{chirp}}t^2 \right) - \psi_{T0}\end{aligned} \tag{2}$$

where the value of ψ_{T0} is $\psi_{T0} = \pi(f_0 T_c - \frac{1}{4}\frac{T_c^2}{T_{chirp}}B)$. Phase $\Delta\psi(t) = \psi_T(t) - \psi_T(t-\tau)$ of the down-converted signal is

$$\Delta\psi(t) = 2\pi \left(f_0\tau + \frac{B}{T_{chirp}}t\tau - \frac{1}{2}\frac{B}{T_{chirp}}\tau^2 \right) \tag{3}$$

Here, τ represents the duration required for the emitted signal to contact the target and reflect. The final entry in Equation (3) is negligible since τ/T_{chirp} takes on a value much less than 1. Suppose the target's distance and speed are Y and V, respectively. From this, we can deduce that τ is given by $\tau = 2(Y + Vt)/c$. Bringing τ into Equation (3) would lead to

$$\Delta\psi(t) = 2\pi \left(\frac{2f_0 Y}{c} + \left(\frac{2f_0 V}{c} + \frac{2BY}{cT_{chirp}} \right) t + \frac{2BV}{cT_{chirp}}t^2 \right) \tag{4}$$

In the equation above, c is the velocity of light. The last term in Equation (4), called the distance Doppler coupling [35], is negligible:

$$\Delta\psi(t) = 2\pi \left(\frac{2f_0 Y}{c} + \left(\frac{2f_0 V}{c} + \frac{2BY}{cT_{chirp}} \right) t \right) \tag{5}$$

The down-converted signal, which is also known as the intermediate frequency (IF) signal, is

$$\begin{aligned}S_{IF}(t) &= \cos(\Delta\psi(t)) \\ &= \cos(2\pi(\frac{2f_0 Y}{c} + (\frac{2f_0 V}{c} + \frac{2BY}{cT_{chirp}})t))\end{aligned} \tag{6}$$

The sampling frequency of $S_{IF}(t)$ is set to $f_s = 1/T_s$. Then, the complex form of the IF signal is obtained using the Hilbert Transform [36], which will be used to estimate information about the target:

$$S(n) = e^{i\cdot 2\pi \left(\frac{2f_0 Y}{c} + \left(\frac{2f_0 V}{c} + \frac{2BY}{cT_{chirp}} \right) T_s n \right)} \tag{7}$$

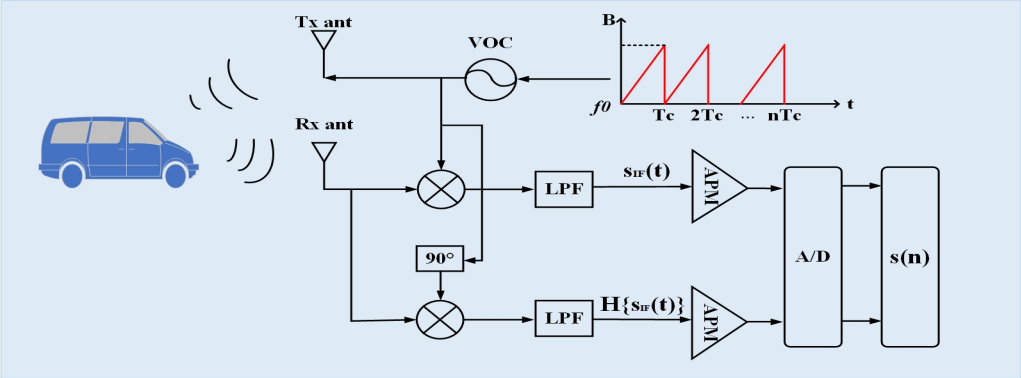


Figure 2. FMCW radar equation schematic block diagram.

3.2. Video Data Pre-Processing

3.2.1. Lane Detection

In this section, an algorithm for detecting lanes that is both robust and efficient is employed, which relies on dependable lane markings. The schematic representation of the algorithm is illustrated in Figure 3.

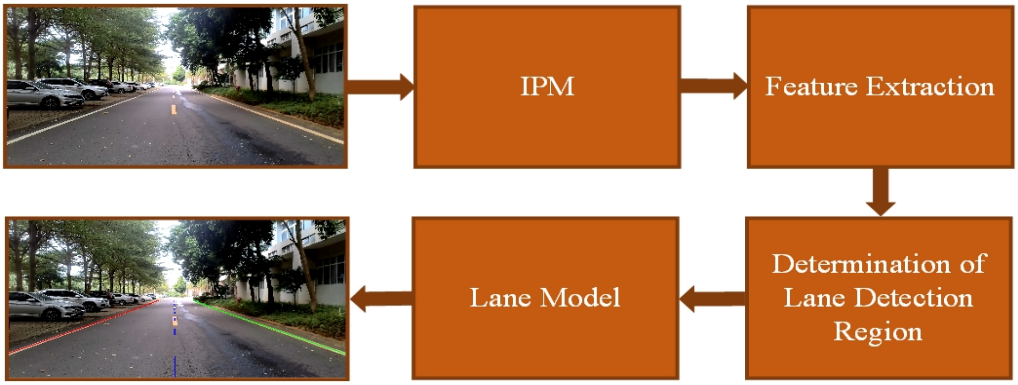


Figure 3. Schematic representation of lane detection algorithm.

The lane detection algorithm is specifically formulated in the following four steps:

(1) To exploit the parallelism of the channels, the raw image is converted into a bird's eye view (BEV) by the inverse perspective mapping (IPM) transform, as shown in Figure 4a.

(2) The RGB image is represented as a mapping, $I_{RGB} : E \times F \times \{R, G, B\} \rightarrow [0, 1]$. $I_{RGB}(i, j, u)$ denotes the pixel value of the RGB image for row i, column j, and color u. Here, E and F are the set of rows and columns, respectively. The greyscale image is represented as the mapping $I_{Grey} : E \times F \rightarrow [0\ 1]$. $I_{Grey}(i, j)$ represents the pixel intensity at the ith row and jth column of the greyscale image. Figure 4b illustrates the conversion of an RGB image to a greyscale image utilizing Equation (8):

$$I_{Grey}(i, j) = 0.299 I_{RGB}(i, j, R) + 0.587 I_{RGB}(i, j, G) + 0.114 I_{RGB}(i, j, B) \quad (8)$$

After that, in order to reduce the amount of image processing and concentrate on lane marking, the ROI of the lane to be detected is selected from the greyscale image, as shown by the yellow bounding box in Figure 4c. Then, the transformation formula of binary image $I_{Bin} : E \times F \rightarrow \{0, 1\}$ is

$$I_{Bin}(i, j) = \begin{cases} 1, \text{if } I_{Grey}(i, j) \geq a \\ 0, \text{if } I_{Grey}(i, j) < a \end{cases} \quad (9)$$

Here, the range of threshold a is $[0, 1]$. Two binary images, I_{Bin1} and I_{Bin2}, are obtained from the greyscale image I_{Grey} by utilizing Equation (9). Subsequently, the binary images are merged using Equation (10):

$$I(i,j) = \begin{cases} 0, \text{if} \left(\sum_{m=i-h}^{i+h} \sum_{n=j-h}^{j+h} I_{Bin1}(m,n) \right) \left(\sum_{m=i-h}^{i+h} \sum_{n=j-h}^{j+h} I_{Bin2}(m,n) \right) = 0 \\ \\ 1, \text{otherwise} \end{cases} \tag{10}$$

Here, h is called the proximity parameter. With these manipulations described above, we obtain the merged binary image I, as shown in Figure 4d. The white parts indicate lane markings, whose element value is 1. Furthermore, all other non-lane portions of the image are black with an element value of 0. At this point, we have completed the feature extraction of the lane lines.

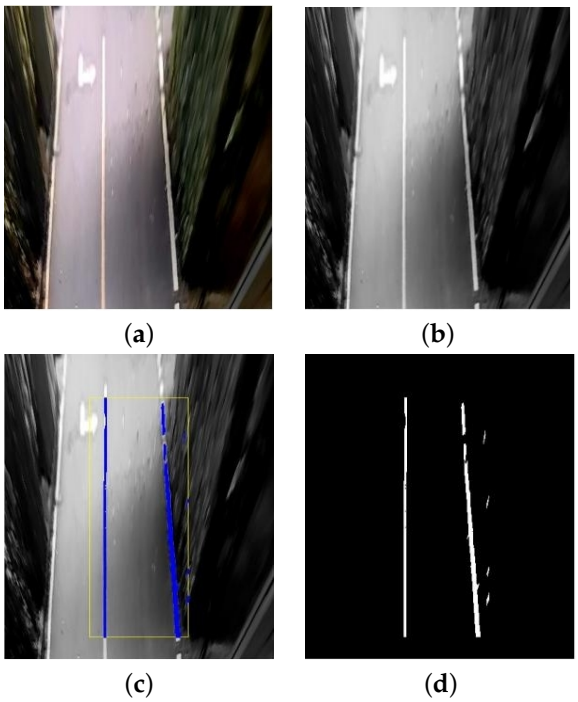


Figure 4. BEV of the lane to be detected. (**a**) RGB format. (**b**) Greyscale format. (**c**) ROI of the lanes to be detected. (**d**) Binary format.

(3) To determine the binary BEV used for lane detection, Figure 5 plots the histograms generated from the binary image I. The histograms are separated into two regions centered on the vehicle. Since the closeness between lane pixel values is crucial for effective lane detection, it is necessary to calculate the standard deviation of histograms.

$$\mu_l = \frac{2}{F} \sum_{j=1}^{F/2} \sum_{i=1}^{E} j \cdot I(i,j); \sigma_l = \sqrt{\frac{2}{F} \sum_{j=1}^{F/2} \sum_{i=1}^{E} I(i,j)(j-\mu_l)^2} \tag{11}$$

$$\mu_r = \frac{2}{F} \sum_{j=F/2+1}^{F} \sum_{i=1}^{E} j \cdot I(i,j); \ \sigma_r = \sqrt{\frac{2}{F} \sum_{j=F/2+1}^{F} \sum_{i=1}^{E} I(i,j)(j-\mu_r)^2} \tag{12}$$

Here, μ_l and μ_r represent the mean values for the left/right areas of the histogram, respectively, while σ_l and σ_r denote the standard deviation.

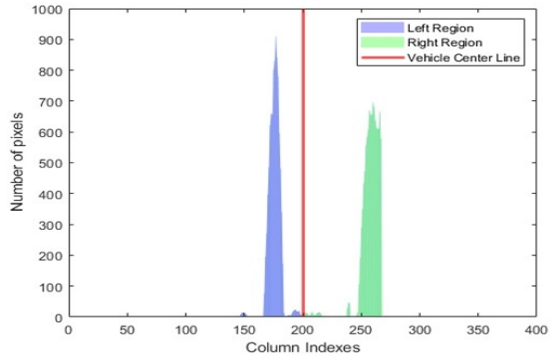


Figure 5. Histograms of lane lines to be detected.

The following criteria are used to determine which area to select for lane detection:

$$\begin{cases} \text{Choose left region, if } (\lambda \mathrm{P_l} - \sigma_l) \geq (\lambda \mathrm{P_r} - \sigma_r) \\ \text{Choose right region, if } (\lambda \mathrm{P_l} - \sigma_l) < (\lambda \mathrm{P_r} - \sigma_r) \end{cases} \tag{13}$$

where $\mathrm{P_l}$ and $\mathrm{P_r}$ denote the maximum value within the histogram's left and right segments. Furthermore, λ is a balancing factor between the maximum value and standard deviation. Eventually, the positions of $\mathrm{P_l}$ and $\mathrm{P_r}$ on the column vector in the histogram are solved, which are the reference positions of carriageway markings used for testing.

(4) To precisely ascertain the shape of the lane pixels, we scan the binary image I line by line along the column indexes of $\mathrm{P_l}$ or $\mathrm{P_r}$. This step is shown in Figure 6. We control the size of the yellow rectangular box to be fixed and scan it from small to large along the Y-axis. Then, a second-order polynomial is used as the lane model:

$$\mathrm{pi}^2 + \mathrm{qi} + \mathrm{c} = \mathrm{j} \tag{14}$$

Considering the F measurements covered by the yellow rectangle as $\{(\mathrm{i_1}, \mathrm{j_1}), (\mathrm{i_2}, \mathrm{j_2}), \ldots, (\mathrm{i_F}, \mathrm{j_F})\}$, the least square estimate (LSE) algorithm is applied to all pixels overlaid by the rectangle. As a result, the parameters p, q, and c of the polynomial in the lane to be detected are calculated:

$$\bar{x} = (B^T B)^{-1} B^T \bar{J} \tag{15}$$

$$\mathrm{B} = \begin{bmatrix} \mathrm{i_1^2} & \mathrm{i_1} & 1 \\ \mathrm{i_2^2} & \mathrm{i_2} & 1 \\ \vdots & \vdots & \vdots \\ \mathrm{i_F^2} & \mathrm{i_F} & 1 \end{bmatrix}, \bar{x} = \begin{bmatrix} \mathrm{p} \\ \mathrm{q} \\ \mathrm{c} \end{bmatrix}, \bar{J} = \begin{bmatrix} \mathrm{j_1} \\ \mathrm{j_2} \\ \vdots \\ \mathrm{j_F} \end{bmatrix} \tag{16}$$

In this way, the square of the error vector is minimized:

$$\mathrm{e} = \|\mathrm{B}\bar{x} - \bar{J}\|_2^2 \tag{17}$$

Subsequently, the lane polynomial to be computed is moved to the peak column position in another area to extract the subsequent lane marker.

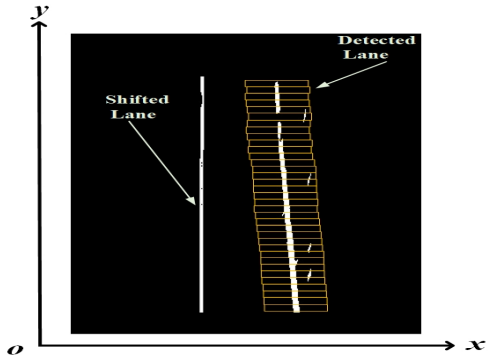


Figure 6. Visual representation of lane detection.

3.2.2. YOLOv5 Target Detection Algorithm

YOLOv5 [37] is a one-stage target detection algorithm, which is classified into five versions: n, s, m, l, and x. Considering the model's detection speed, accuracy, and size, YOLOv5s is chosen in this paper. Its structure is shown in Figure 7. YOLOv5 mainly consists of the backbone, neck, and head. The backbone network mainly consists of the focus structure, the cross-stage partial network (CSPNet), and the spatial pyramid pooling (SPP) structure. The focus structure adopts the slicing operation to reduce the loss of information caused by the downsampling. The CSPNet can improve the learning ability of the model and reduce the amount of data transfer and computation of the network. Furthermore, the SPP uses maximum pooling at different scales to improve the perceptual field of view. The neck network uses the pyramid attention network (PAN) to efficiently fuse semantic features by combining top-down and down-top approaches. The head network is used to detect objects at different scales and match the optimal target bounding box by non-maximal suppression (NMS).

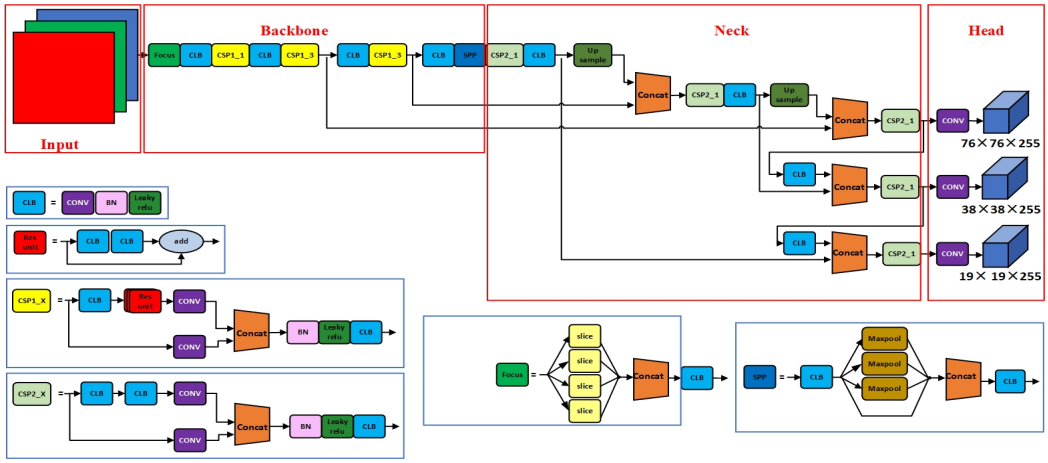


Figure 7. Network structure of YOLOv5s.

After obtaining the bounding box of a target making use of the YOLOv5s algorithm, the pixel occupied by the target in the image will be used to estimate its actual distance. The traditional small-hole imaging method [38] for distance detection requires prior knowledge of the height of a target. However, determining the height of an object using a monocular camera is challenging. There is a data fitting algorithm that can be used to estimate the actual distance to the target. The basic principle of this method is to place the target 3.6 m directly in front of the camera and take a shot every 0.6 m of movement. Figure 8 shows a portion of images used to solve for distance.

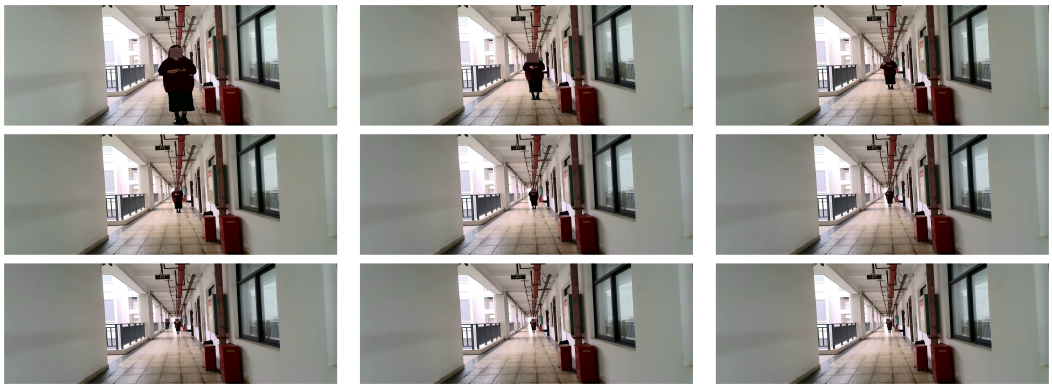


Figure 8. Different locations of targets.

Figure 9 illustrates the result of obtaining a distance fitting curve using a rational function.

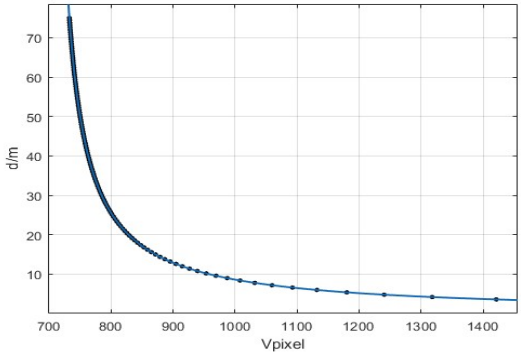


Figure 9. Distance fitting curve.

The vertical coordinate d represents the practical distance of targets. Furthermore, the horizontal coordinate V_{pixel} represents the vertical pixel positions, which are located at the center of the underside of bounding boxes. Eventually, the relationship between V_{pixel} and d is

$$d = \frac{2603}{V_{pixel} - 698.2} \tag{18}$$

3.3. Spatiotemporal Calibration

3.3.1. Temporal Calibration

The millimeter wave radar was set to acquire 10 frames per second, while the camera was at 30 frames per second. Then, according to the relationship between the data acquisition frequencies of the two sensors, the data from the radar were processed frame by frame, and the data from the camera were processed every three frames. By performing this, we can make the data acquisition frequency of both sensors consistent. Finally, assuming that the camera is activated earlier than the radar by $\Delta\tau$ seconds, the corresponding number of frames for early activation is obtained as

$$\text{frameVid} = \Delta\tau \cdot V_{fps} \tag{19}$$

Here, V_{fps} is the data acquisition frame rate of the camera, whose value is 30 fps. Thereby, the first frame of data acquired by the radar is temporally aligned with the frame frameVid of the camera, as shown in Figure 10.

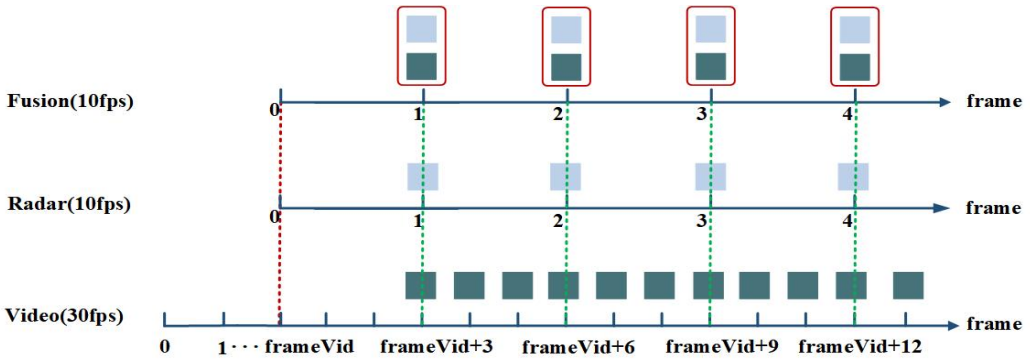


Figure 10. Sampling method of frame.

3.3.2. Spatial Calibration

Since the fusion is performed on a pixel coordinate system, the targets detected by the radar are projected onto the pixel coordinate system, which is called spatial calibration between multiple sensors [39,40]. Suppose Q is the target to be detected, and its coordinates in the world, radar, and camera are (X_w, Y_w, Z_w), (X_r, Y_r, Z_r), and (X_c, Y_c, Z_c), respectively. To simplify the derivation process, it is assumed that the camera and world share the same reference coordinate system. Then, the position of point Q in each coordinate system is shown in Figure 11a. Using R and T to denote the angular rotation and translation information between the two sensors, respectively, the target detected by the radar could be transformed into the camera coordinate system using the following equation:

$$\begin{bmatrix} Xc \\ Yc \\ Zc \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} Xr \\ Yr \\ Zr \\ 1 \end{bmatrix} \tag{20}$$

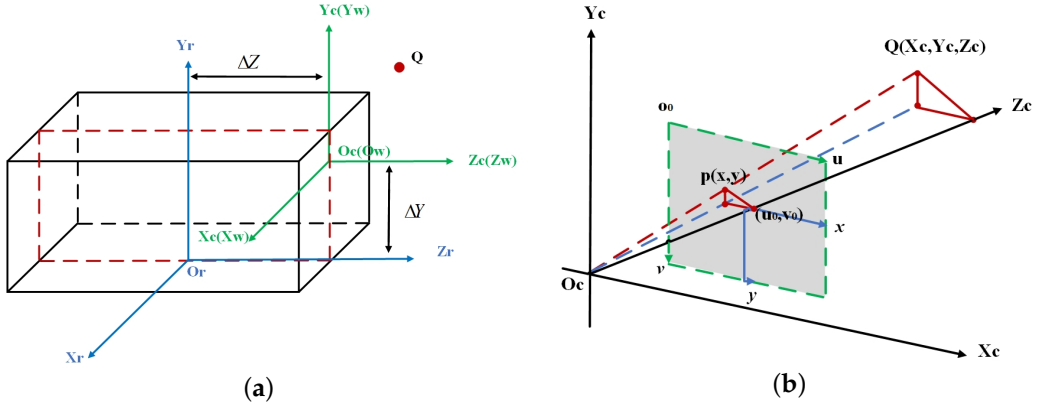


Figure 11. Schematic of coordinate system relationships. (**a**) Position of radar coordinate and camera coordinate. (**b**) Position of camera coordinate, image coordinate, and pixel coordinate.

Using the principle of similar triangles, the point $Q(X_c, Y_c, Z_c)$ in Figure 11b is projected onto $p(x, y)$, where o-xy represents the image coordinate system.

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Xc \\ Yc \\ Zc \\ 1 \end{bmatrix} \tag{21}$$

Here, f denotes the focal length of the optical sensor used in this study. The computer stores image data using a pixel-based coordinate system that takes the point o_0 in the shaded portion of Figure 11b as the starting point. Suppose there is a pixel point (u_0, v_0)

that corresponds to the home point of the image coordinate system. The transformation relationship from image coordinate to pixel is

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{22}$$

In the above equation, dx and dy denote the pixel density along the horizontal and vertical axes in the image coordinate system, respectively. Ultimately, the conversion formula that links the radar coordinate to the pixel is presented as shown in the following mathematical expression:

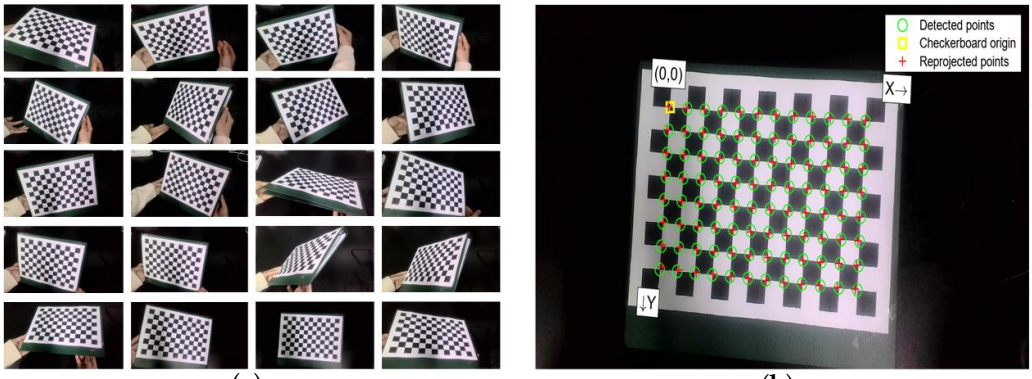
$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_r \\ Y_r \\ Z_r \\ 1 \end{bmatrix} = N_1 N_2 \begin{bmatrix} X_r \\ Y_r \\ Z_r \\ 1 \end{bmatrix} \tag{23}$$

$$N_1 = \begin{bmatrix} f/dx & 0 & u_0 & 0 \\ 0 & f/dy & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{24}$$

$$N_2 = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \tag{25}$$

The elements of matrix N_1 are associated with the camera's intrinsic properties, which are determinable through a calibration technique involving a checkerboard pattern. The elements of matrix N_2 are determined by the mounting position of the camera, which can be obtained by measuring the spatial arrangement relative to the radar.

In this paper, we employ a checkerboard calibration board with 10×8 squares to adjust the intrinsic properties of camera. Here, the side length of each square is 20 mm $\times$ 20 mm. Firstly, 20 pictures of the checkerboard grid are taken from different angles using a camera, as shown in Figure 12a. Then, the corner points in the checkerboard grid pictures are extracted and corrected using the calibration toolbox. The findings are presented in Figure 12b. Where the arrows indicate the direction of increase in the X and Y axes.



(**a**) (**b**)

Figure 12. Camera calibration. (**a**) Camera calibration chessboard graph. (**b**) Corner extraction and correction of checkerboard.

Ultimately, the obtained internal parameter matrix N_1 of camera is

$$N_1 = \begin{bmatrix} 2032.1 & 0 & 1338 & 0 \\ 0 & 2033.2 & 741.8 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{26}$$

The camera's external parameter calibration determines the values of R and T. Figure 13 depicts the placement of the sensors in relation to each other, where the camera is encased in a black box. The relative positions between the two sensors are adjusted so that the pitch and azimuth deflections are approximated to 0^{o}. Therefore, the size of the rotation matrix between the two sensors is R = I. Then, the Z-axis offset between the center of the camera and the radar is measured to be $\Delta Z = 170$ mm, the Y-axis offset is $\Delta Y = 25$ mm, and the X-axis offset is 0 mm. The value of the translation matrix is thus obtained as $T = [0\ -25\ 170]^T$. Based on R and T, the camera's external parameter matrix N_2 is expressed as

$$N_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -25 \\ 0 & 0 & 1 & 170 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{27}$$

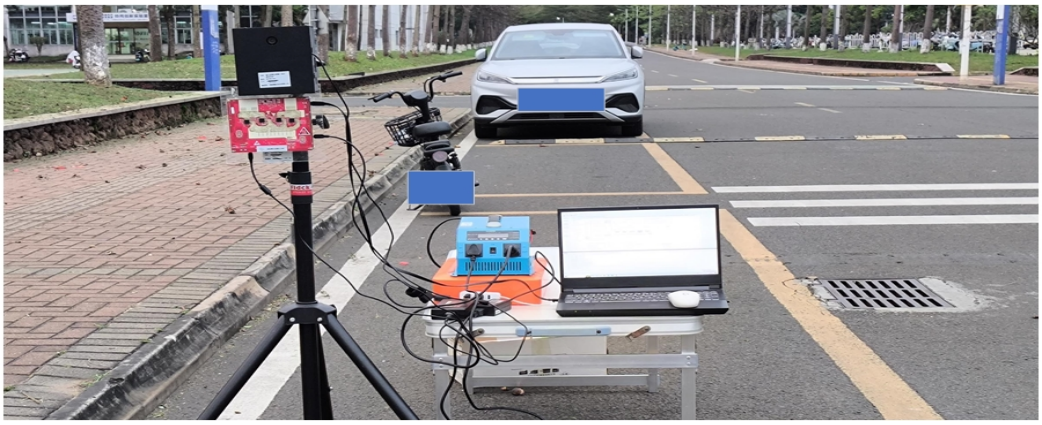


Figure 13. Hardware system for information fusion of radar and monocular camera.

3.4. Target Matching

In order to make consistent judgments on the targets detected by the radar and camera, it is necessary to generate the corresponding vehicle ROI based on each pixel point mapped by the radar in the image [41]. The vehicle's external dimensions specify that the aspect ratio of the vehicle's rear should range from 0.7 to 1.3. In this paper, the width of the vehicle is set to be 2.4 m, and the aspect ratio is taken to be 1.2. The formula to establish the pixel width (W) and height (H) of the target projection point detected by radar in the image is finally obtained as follows:

$$\begin{cases} W = 2.4 \times \frac{\alpha f}{dx \cdot Y_r} \\ \\ H = 2.0 \times \frac{\alpha f}{dx \cdot Y_r} \end{cases} \tag{28}$$

Here, α is the ratio between the physical and pixel size.

In Figure 14, the red rectangle indicates the radar ROI with an area of S_R, and the green rectangle indicates the visual ROI with an area of S_V. In this manner, the IOU can be determined using the following equation:

$$IOU = \frac{S_V \cap S_R}{S_V \cup S_R} \tag{29}$$

Then, target matching is performed based on the value of the IOU, and the corresponding fusion strategies are formulated based on the matching results. The specific fusion schemes are as follows:

If the value of the IOU is in the range $(0, 1]$, it can be judged that the two sensors detect the same target. Considering the respective detection advantages of the two sensors, the distance, speed, and angle detected by the radar, as well as the category, bounding box,

and lane detected by the camera, are output. If IOU = 0, it means that there is no overlap between the radar and vision bounding boxes. In this condition, it is essential to determine whether the values of S_V and S_R are 0. If $S_R = 0$ and $S_V \neq 0$, it means that the radar fails to detect the target. At the same time, if the confidence score of the target detected by the camera is greater than or equal to 0.50, the camera is judged to have detected a valid target. Thus, the information detected by the camera is output. If $S_R \neq 0$ and $S_V = 0$, it means that camera fails to detect the target. Then, target tracking is performed to determine if the radar detects a valid target. If it is a valid target, the information detected by the radar is output. Otherwise, the invalid target is discarded.

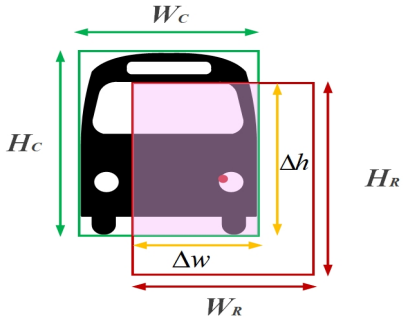


Figure 14. Target matching algorithm of radar and camera.

3.5. Evaluation Indicators

This paper evaluates the algorithm's performance using precision, recall, and F1 score. The F1 score is also known as the balanced score, which takes into account both the precision and recall of the algorithm. The specific formulas are as follows:

$$\text{precision} = \frac{T_P}{T_P + F_P} \tag{30}$$

$$\text{recall} = \frac{T_P}{T_P + F_N} \tag{31}$$

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{32}$$

where T_P denotes the number of targets correctly detected within lanes, F_P denotes the number of false alarms, and F_N denotes the number of missed detections.

4. Experiments and Results

4.1. Experimental Platforms and Environments

Information on the software and devices used in this study is shown in Table 1.

In all experiments, the camera and radar were always placed on a tripod, as shown in Figure 13. The tripod was always placed on the ground, and the distance between the camera's optical center and the ground was 1.28 m. We kept the roads on the campus as the experimental scenes except for special instructions.

Table 1. Information on the software and devices.

Software and Device	Version	Function
Operating system	Windows11	–
CPU	i5-11400	–
GPU	NVIDIA GeForce RTX 3060	–
CUDA	12.3	–
Pytorch	11.3	–
Python	3.8	–
Pycharm	2023	Running the YOLO algorithms
MATLAB	R2022b	Running the lane detection, radar, and fusion algorithms
Camera	Hewlett-Packard (HP) 1080p	–
Radar	AWR2243 [42]	–

4.2. Visual Detection Results

4.2.1. Validation of Lane Detection Algorithm

We placed the camera on four different roads to capture images with normal light, ground icon interference, ground shelter interference, and weak light. The results of the four environments for lane detection are shown in Figure 15. Where the detected lane lines are drawn in red, green and blue.

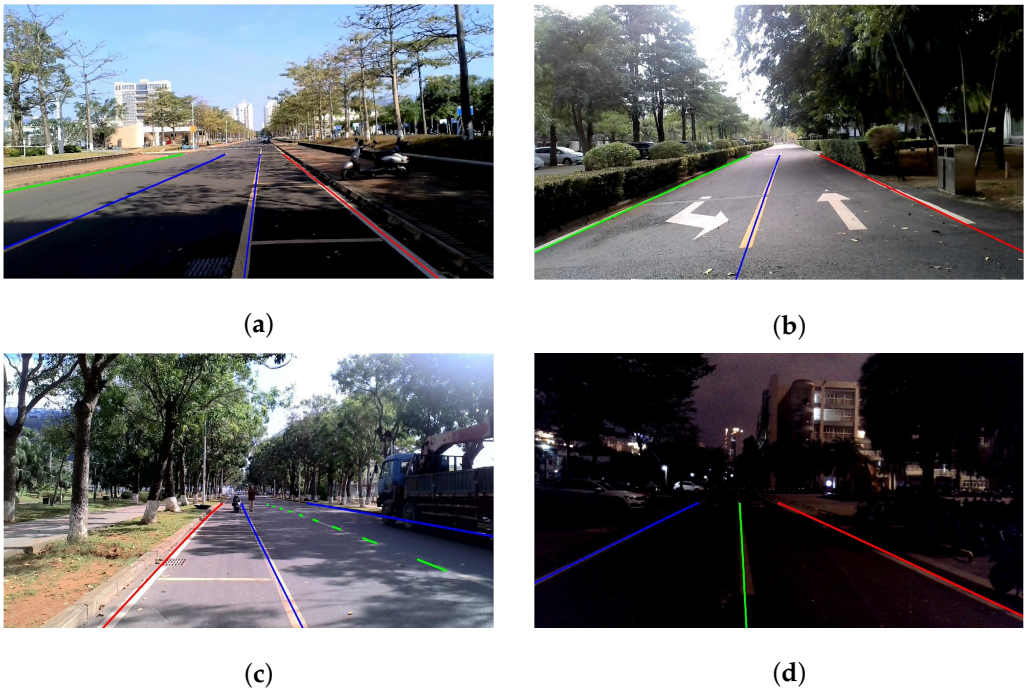


Figure 15. Lane detection results. (**a**) Normal light. (**b**) Ground icon interference. (**c**) Ground shelter interference. (**d**) Weak light.

From the experimental result in Figure 15a, it can be verified that the lane detection algorithm can precisely detect the lane lines under normal lighting conditions. Even if there are some shadows near the lane lines, excellent detection results can be obtained. Moreover, the lane detection algorithm can also detect lane lines in complex situations, such as lines with ground icon interference, areas occupied by other vehicles, and situations

with weak lighting. Examples of these situations and the correct detection results are given in Figure 15b–d.

4.2.2. Validation of YOLOv5 Algorithm

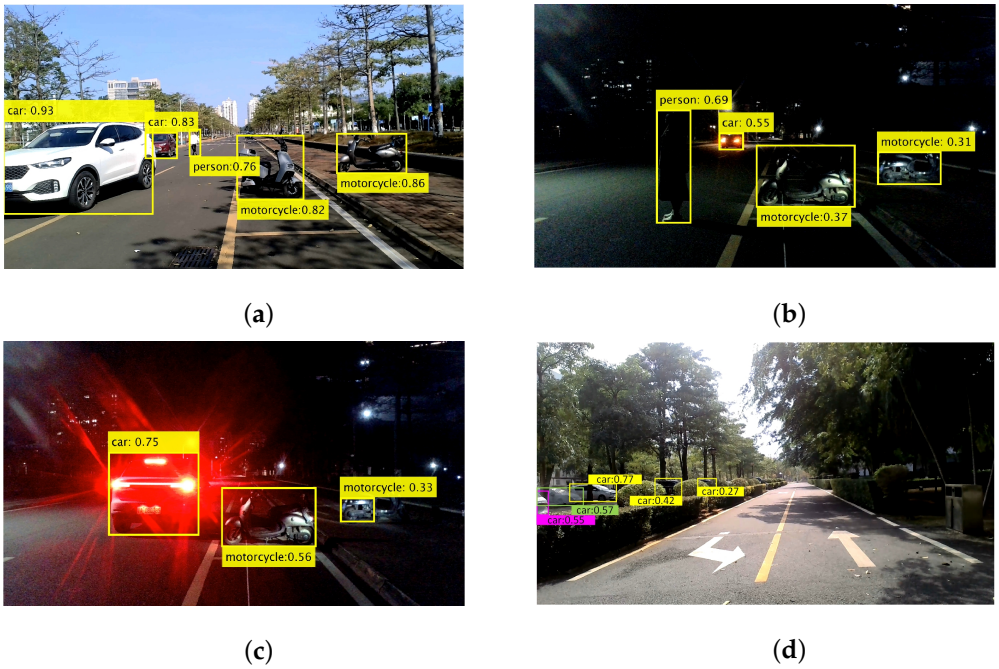
In order to verify the reasonableness of using the YOLOv5s model to detect targets in images, we analyzed its performance in three aspects: time, space, and accuracy. We used the pre-trained models of YOLOv5s, YOLOv7_tiny, YOLOv8s, and YOLOv9s [43] for comparison experiments. Their performance is shown in Table 2.

Table 2. Performance of different models.

Model	Params/M	FLOPs/G	AP_{50}^{val}/%	FPS (bs = 32)
YOLOv5s	7.2	16.5	56.8	106
YOLOv7-tiny	6.2	13.9	51.3	123
YOLOv8s	11.2	28.6	61.8	101
YOLOv9s	7.1	26.4	63.4	–

Where the values of Params, FLOPs, and AP_{50}^{val} are official published results obtained on the COCO 2017 dataset. The FPS was obtained by taking a video with our camera and processing it using the four models mentioned above. Considering the limitations of the hardware, as well as the accuracy and real-time performance of the algorithm, YOLOv5s has apparent advantages. Therefore, the experiments justify using YOLOv5s to detect targets in images.

To verify the ability of the YOLOv5s model to recognize targets in natural environments, we captured images with normal light, weak light, intense light, and targets occluded by trees. The experimental results are shown in Figure 16. As we can see from the figures, YOLOv5s can accurately recognize targets even in harsh environments.



(**a**) (**b**)

(**c**) (**d**)

Figure 16. Detection results of YOLOv5s. (**a**) Normal light. (**b**) Weak light. (**c**) Intense light. (**d**) Targets occluded by trees.

4.3. Validation of Radar Target Detection Algorithm

In order to verify the effectiveness of the radar target detection algorithm and the reliability of using lane lines to screen valid radar point clouds, we set up two targets on the road. Taking the radar center as the coordinate origin, the positions of these two targets are (−0.93 m, 21.62 m) and (0.50 m, 9.40 m), as shown in Figure 17, where line1, line2, line3, and line4, respectively, represent the lateral distance between the four lane lines and the center of the radar.

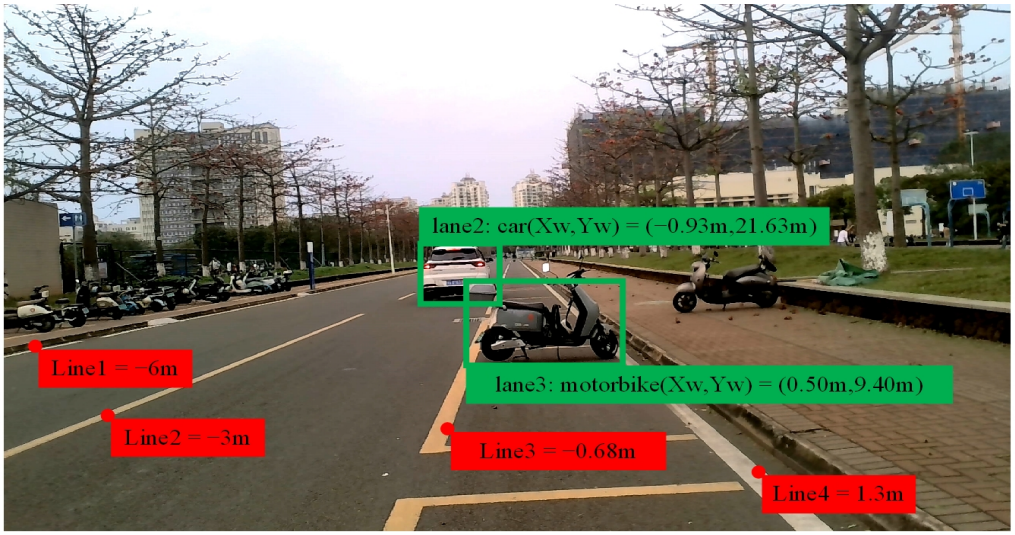
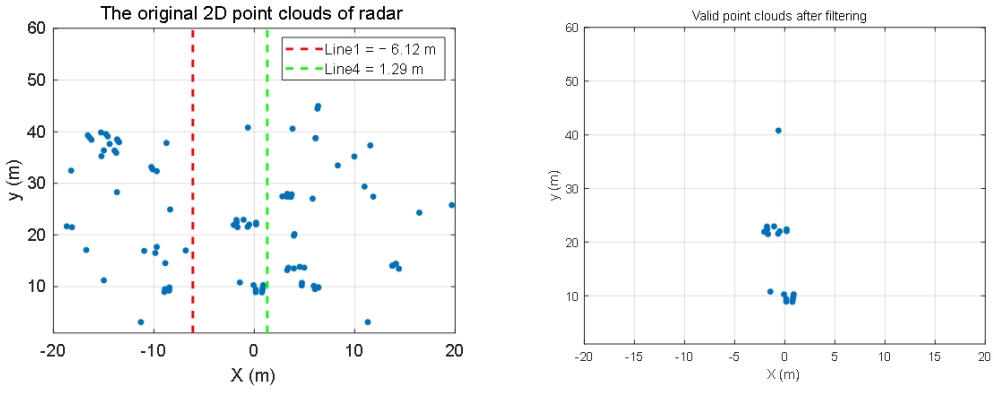


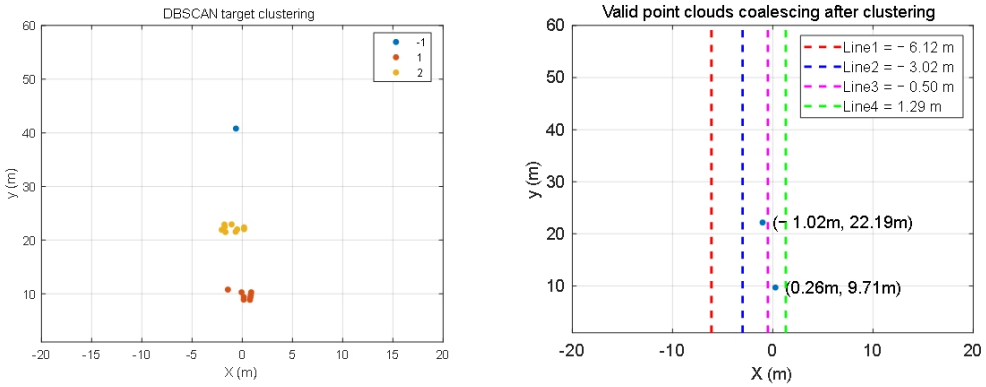
Figure 17. Actual positions of targets.

The original point clouds obtained after employing the 2D-FFT, CFAR, and DOA estimation algorithms are shown in Figure 18a. The two vertical lines in Figure 18a are the locations of line1 and line4, which were obtained based on the lane detection algorithm. Figure 18b shows valid point clouds after filtering the interfering targets outside lanes. The results of clustering point clouds within lanes using the DBSCAN algorithm and coalescing the coordinates of point clouds with the same category labels are, respectively, shown in Figure 18c,d. The experiments demonstrate that the radar sensor correctly detects two valid targets, and all of the interfering targets outside lanes are filtered out.



(**a**) (**b**)

Figure 18. *Cont.*



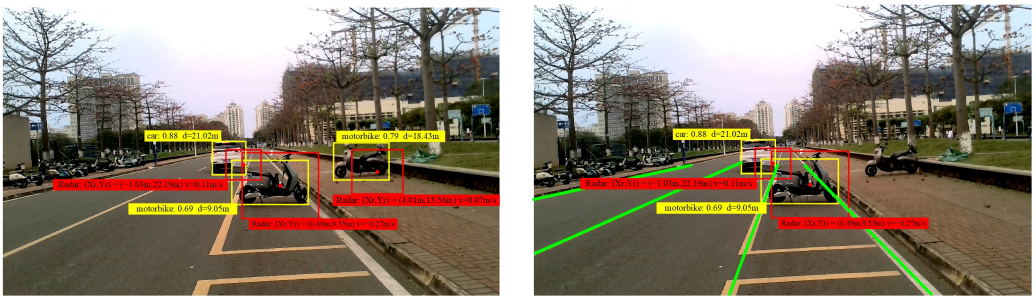
(**c**) (**d**)

Figure 18. Radar detection results. (**a**) Original 2D point clouds of radar. (**b**) Valid point clouds after filtering. (**c**) DBSCAN target clustering. (**d**) Valid point clouds coalescing after clustering.

4.4. Validation of Information Fusion

4.4.1. Comparison with Fusion without Lane Detection

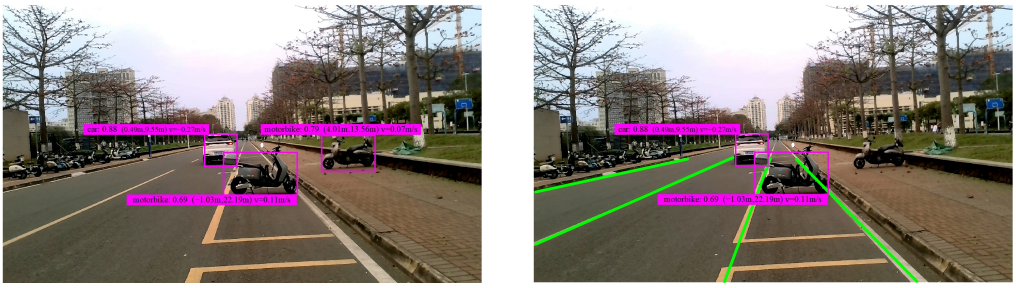
Keeping the positions of the sensors and targets the same as in Section 4.3, we performed comparative experiments between the proposed fusion algorithm with lane detection and the case without lane detection. The difference between the two algorithms is whether they contain lane detection, while the rest of the features are consistent. Figure 19 shows the results of the spatiotemporal calibration of the information detected by multiple sensors. The yellow bounding boxes indicate the targets detected by the camera, and the red bounding boxes indicate the targets detected by the radar. As shown in the experiments without lane detection in Figure 19a, neither of the interfering targets outside the lane detected by the two sensors are filtered out completely. This is because a fixed lane width needs to be assumed at this point, and this method is often inaccurate. In contrast, in Figure 19b, all of these disturbances outside lanes are accurately filtered.



(**a**) (**b**)

Figure 19. The results of multi-sensor spatiotemporal calibration. (**a**) Without lane detection. (**b**) With lane detection.

Figure 20a,b show the fusion results without lane detection and with lane detection, respectively. In Figure 19a, the ROIs of different targets detected by different sensors in lane 2 and lane 3 overlap with each other, which leads to wrong target matching as shown in Figure 20a. On the contrary, in Figure 20b, we perform information fusion on the same lane, which can accurately fuse the information of the same target from different sensors.



(**a**) (**b**)

Figure 20. Information fusion results. (**a**) Without lane detection. (**b**) With lane detection.

4.4.2. Comparison with Other Related Studies

We collected information about the targets with radar and camera in weak, intense, and normal light. Furthermore, 199 frames of data were collected for each environment. The three scenarios are shown in Figures 21–23. In the same experimental scenario, comparison experiments were performed between the results of camera-only detection, the proposed fusion algorithm, and another decision-level fusion method [41]. As the methods of camera-only detection and [41] need to pre-set the lane width to filter out the interfering targets, we set their lateral effective range to $[-6.0\text{ m}, 1.3\text{ m}]$. On the contrary, the proposed method does not require us to set the lane threshold. The performance of different algorithms is shown in Table 3. The F1 score of the proposed algorithm is able to reach 0.98, which indicates that it has low false and missed detections. Moreover, the F1 score is almost unchanged under different lighting environments, which suggests that the performance of our algorithm is consistent under various environments. In addition, the proposed fusion algorithm outperforms [41] in all aspects. This is attributed to the fact that we used the lane detection results to assist the fusion of the radar and camera, whereas [41] had difficulty in solving the problem of mutual coverage of the bounding boxes of different targets on adjacent lanes.

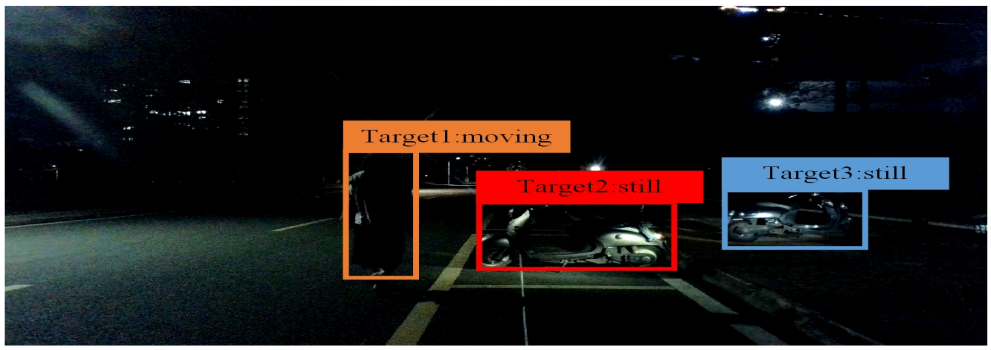


Figure 21. Weak light.



Figure 22. Intense light.

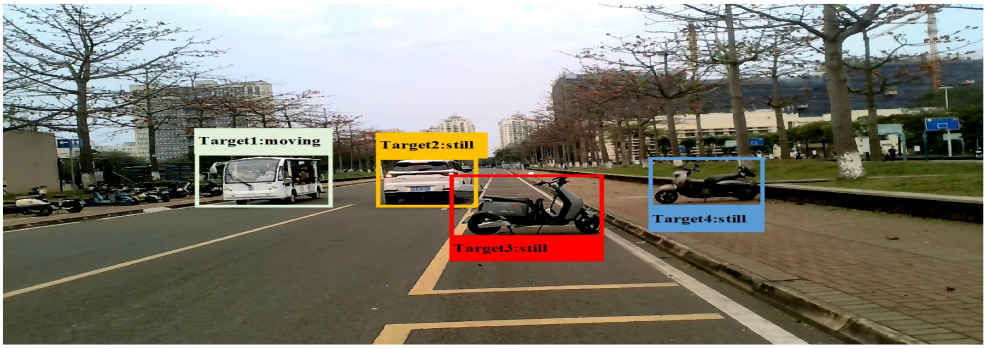


Figure 23. Normal light.

Table 3. Performance in different environments.

Environment	Algorithm	Precision	Recall	F1
Normal light	Camera-only detection	86.91%	88.44%	0.88
	Decision-level fusion	86.81%	90.95%	0.89
	Fusion (ours)	96.83%	99.75%	0.98
Weak light	Camera-only detection	79.18%	86.93%	0.83
	Decision-level fusion	82.50%	82.91%	0.83
	Fusion (ours)	96.32%	98.74%	0.98
Intense light	Camera-only detection	69.10%	97.59%	0.81
	Decision-level fusion	62.03%	97.81%	0.76
	Fusion (ours)	96.15%	98.46%	0.97

Finally, in order to analyze the algorithm's inference time and running memory, we processed 199 frames of data from the scene shown in Figure 23. After temporal-calibrating the radar and camera using the algorithm in Section 3.3.1, each frame of data was sampled at a period of 0.1 s. The time and space required for the algorithms to run are shown in Table 4. As a limitation of the software, the average inference time per frame of both algorithms is much larger than the actual acquisition time, which makes it difficult to achieve real-time processing. However, the inference time of the proposed algorithm increased by only 0.04 s compared to [41]. Therefore, combining Tables 3 and 4, we can conclude that the proposed algorithm is able to achieve superior target detection performance at the expense of a certain time and space.

Table 4. Performance of different algorithms in time and space.

Algorithm	Inference Time/s	Running Memory/GB
Fusion (ours)	0.94	1.26
Decision-level fusion	0.90	0.99

5. Conclusions

For the shortcomings of the low success rate of target matching and the imprecise accuracy of multi-target detection in decision-level information fusion, a robust target detection algorithm based on the fusion of FMCW radar and a monocular camera is proposed in this paper. The method utilizes a lane detection algorithm to process the image so as to obtain the actual distance of the lane lines concerning the sensors. Then, the lane lines are used to assist the radar and camera in filtering interferences outside the lanes and grouping the remaining valid targets. Spatiotemporal calibration of sensors is achieved through camera calibration, coordinate system conversion, and harmonization of sensor sampling moments. Finally, the information is fused using a decision-level fusion method for targets with the same lane markings detected by the radar and camera. Experiments

show that the F1 score of the proposed algorithm can reach 0.98, which indicates that it has low false and missed detections. Moreover, the performance of our algorithm is consistent under various environments. However, the algorithm still presents significant challenges in terms of both time and space. In the future, we will aim to achieve superior detection accuracy while reducing the time and space expenses of the algorithm.

Author Contributions: Research design, Y.Y. and X.W. (Xianpeng Wang);data acquisition, X.W. (Xiaoqin Wu), X.L. and Y.G.; writing—original draft preparation, Y.Y.; writing—review and editing, X.W. (Xianpeng Wang), X.L. and T.S.; supervision, X.W. (Xianpeng Wang) and X.W. (Xiaoqin Wu); funding acquisition, X.W. (Xianpeng Wang) and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Research and Development Project of Hainan Province (ZDYF2023GXJS159), the Natural Science Foundation of Hainan Province (620RC555), the National Natural Science Foundation of China (No. 61961013, 62101088), and the Innovative Research Projects for Graduate Students in Hainan Province (Qhys2023-292).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Hassan, R.; Abdel-Rahim, A.; Hadhoud, R. Study of Road Traffic Accidents Cases admitted to Ain Shams University Hospitals during Years 2017 and 2018. *Ain Shams J. Forensic Med. Clin. Toxicol.* **2022**, *38*, 1–10. [CrossRef]
2. Ahmed, S.K.; Mohammed, M.G.; Abdulqadir, S.O.; El-Kader, R.G.A.; El-Shall, N.A.; Chandran, D.; Rehman, M.E.U.; Dhama, K. Road traffic accidental injuries and deaths: A neglected global health issue. *Health Sci. Rep.* **2023**, *6*, e1240. [CrossRef]
3. Shams, Z.; Naderi, H.; Nassiri, H. Assessing the effect of inattention-related error and anger in driving on road accidents among Iranian heavy vehicle drivers. *IATSS Res.* **2021**, *45*, 210–217. [CrossRef]
4. Lu, J.; Peng, Z.; Yang, S.; Ma, Y.; Wang, R.; Pang, Z.; Feng, X.; Chen, Y.; Cao, Y. A review of sensory interactions between autonomous vehicles and drivers. *J. Syst. Archit.* **2023**, *141*, 102932. [CrossRef]
5. Yeong, D.J.; Velasco-Hernandez, G.; Barry, J.; Walsh, J. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors* **2021**, *21*, 2140. [CrossRef]
6. Wang, Z.; Wu, Y.; Niu, Q. Multi-sensor fusion in automated driving: A survey. *IEEE Access* **2019**, *8*, 2847–2868. [CrossRef]
7. Sushma, R.; Kumar, J.S. Autonomous vehicle: Challenges and implementation. *J. Electr. Eng. Autom.* **2022**, *4*, 100–108. [CrossRef]
8. Kim, J.; Park, B.j.; Kim, J. Empirical Analysis of Autonomous Vehicle's LiDAR Detection Performance Degradation for Actual Road Driving in Rain and Fog. *Sensors* **2023**, *23*, 2972. [CrossRef] [PubMed]
9. Bhupathiraju, S.H.V.; Sheldon, J.; Bauer, L.A.; Bindschaedler, V.; Sugawara, T.; Rampazzi, S. EMI-LiDAR: Uncovering Vulnerabilities of LiDAR Sensors in Autonomous Driving Setting Using Electromagnetic Interference. In Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks, Guildford, UK, 29 May–1 June 2023; pp. 329–340.
10. Giannaros, A.; Karras, A.; Theodorakopoulos, L.; Karras, C.; Kranias, P.; Schizas, N.; Kalogeratos, G.; Tsolis, D. Autonomous vehicles: Sophisticated attacks, safety issues, challenges, open topics, blockchain, and future directions. *J. Cybersecur. Priv.* **2023**, *3*, 493–543. [CrossRef]
11. Gautam, S.; Kumar, A. Image-based automatic traffic lights detection system for autonomous cars: A review. *Multimed. Tools Appl.* **2023**, *82*, 26135–26182. [CrossRef]
12. Sharma, V.K.; Dhiman, P.; Rout, R.K. Improved traffic sign recognition algorithm based on YOLOv4-tiny. *J. Vis. Commun. Image Represent.* **2023**, *91*, 103774. [CrossRef]
13. Guo, Y.; Wang, X.; Lan, X.; Su, T. Traffic target location estimation based on tensor decomposition in intelligent transportation system. *IEEE Trans. Intell. Transp. Syst.* **2022**, *25*, 816–828. [CrossRef]
14. Wang, X.; Guo, Y.; Wen, F.; He, J.; Truong, T.K. EMVS-MIMO radar with sparse Rx geometry: Tensor modeling and 2D direction finding. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 8062–8075. [CrossRef]
15. Tang, X.; Zhang, Z.; Qin, Y. On-road object detection and tracking based on radar and vision fusion: A review. *IEEE Intell. Transp. Syst. Mag.* **2021**, *14*, 103–128. [CrossRef]
16. Bombini, L.; Cerri, P.; Medici, P.; Alessandretti, G. Radar-vision fusion for vehicle detection. In Proceedings of the International Workshop on Intelligent Transportation, Toronto, ON, Canada, 17–20 September 2006; Volume 65, p. 70.
17. Chipengo, U.; Commens, M. A 77 GHz simulation study of roadway infrastructure radar signatures for smart roads. In Proceedings of the 2019 16th European Radar Conference (EuRAD), Paris, France, 2–4 October 2019; pp. 137–140.

18. Hsu, Y.W.; Lai, Y.H.; Zhong, K.Q.; Yin, T.K.; Perng, J.W. Developing an on-road object detection system using monovision and radar fusion. *Energies* **2019**, *13*, 116. [CrossRef]
19. Abbas, A.F.; Sheikh, U.U.; Al-Dhief, F.T.; Mohd, M.N.H. A comprehensive review of vehicle detection using computer vision. *TELKOMNIKA (Telecommun. Comput. Electron. Control)* **2021**, *19*, 838–850. [CrossRef]
20. Xiao, Y.; Tian, Z.; Yu, J.; Zhang, Y.; Liu, S.; Du, S.; Lan, X. A review of object detection based on deep learning. *Multimed. Tools Appl.* **2020**, *79*, 23729–23791. [CrossRef]
21. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
22. Ciaparrone, G.; Sánchez, F.L.; Tabik, S.; Troiano, L.; Tagliaferri, R.; Herrera, F. Deep learning in video multi-object tracking: A survey. *Neurocomputing* **2020**, *381*, 61–88. [CrossRef]
23. Dimitrievski, M.; Jacobs, L.; Veelaert, P.; Philips, W. People tracking by cooperative fusion of radar and camera sensors. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 509–514.
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2016**, *39*, 1137–1149. [CrossRef]
25. Lin, J.J.; Guo, J.I.; Shivanna, V.M.; Chang, S.Y. Deep Learning Derived Object Detection and Tracking Technology Based on Sensor Fusion of Millimeter-Wave Radar/Video and Its Application on Embedded Systems. *Sensors* **2023**, *23*, 2746. [CrossRef] [PubMed]
26. YenIaydin, Y.; Schmidt, K.W. A lane detection algorithm based on reliable lane markings. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4.
27. Sun, F.; Li, Z.; Li, Z. A traffic flow detection system based on YOLOv5. In Proceedings of the 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Shanghai, China, 15–17 October 2021; pp. 458–464.
28. Long, N.; Wang, K.; Cheng, R.; Yang, K.; Bai, J. Fusion of millimeter wave radar and RGB-depth sensors for assisted navigation of the visually impaired. In Proceedings of the Millimetre Wave and Terahertz Sensors and Technology XI, Berlin, Germany, 5 October 2018; Volume 10800, pp. 21–28.
29. Zhong, Z.; Liu, S.; Mathew, M.; Dubey, A. Camera radar fusion for increased reliability in ADAS applications. Electron. *Imaging* **2018**, *2018*, 251–254.
30. Lv, P.; Wang, B.; Cheng, F.; Xue, J. Multi-Objective Association Detection of Farmland Obstacles Based on Information Fusion of Millimeter Wave Radar and Camera. *Sensors* **2022**, *23*, 230. [CrossRef] [PubMed]
31. Song, M.; Lim, J.; Shin, D.J. The velocity and range detection using the 2D-FFT scheme for automotive radars. In Proceedings of the 2014 4th IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 19–21 September 2014; pp. 507–510.
32. Yuan, Y.; Li, W.; Sun, Z.; Zhang, Y.; Xiang, H. Two-dimensional FFT and two-dimensional CA-CFAR based on ZYNQ. *J. Eng.* **2019**, *2019*, 6483–6486. [CrossRef]
33. Lim, S.; Lee, S.; Kim, S.C. Clustering of detected targets using DBSCAN in automotive radar systems. In Proceedings of the 2018 19th International Radar Symposium (IRS), Bonn, Germany, 20–22 June 2018; pp. 1–7.
34. Winkler, V. Range Doppler detection for automotive FMCW radars. In Proceedings of the 2007 European Radar Conference, Munich, Germany, 10–12 October 2007; pp. 166–169.
35. Mukherjee, A.; Sinha, A.; Choudhury, D. A novel architecture of area efficient FFT algorithm for FPGA implementation. *ACM SIGARCH Comput. Archit. News* **2016**, *42*, 1–6. [CrossRef]
36. Barnhart, B.L. The Hilbert-Huang Transform: Theory, Applications, Development. Ph.D. Thesis, The University of Iowa, Iowa, United States, 2011.
37. Xiaoling, Y.; Weixin, J.; Haoran, Y. Traffic sign recognition and detection based on YOLOv5. *Inf. Technol. Informatiz.* **2021**, *4*, 28–30.
38. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
39. Guo, X.p.; Du, J.s.; Gao, J.; Wang, W. Pedestrian detection based on fusion of millimeter wave radar and vision. In Proceedings of the 2018 International Conference on Artificial Intelligence and Pattern Recognition, New York, NY, USA, 18 August 2018; pp. 38–42.
40. Mo, C.; Li, Y.; Zheng, L.; Ren, Y.; Wang, K.; Li, Y.; Xiong, Z. Obstacles detection based on millimetre-wave radar and image fusion techniques. In Proceedings of the IET International Conference on Intelligent and Connected Vehicles (ICV 2016), Chongqing, China, 22–23 September 2016; pp. 1–6.
41. Cai, G.; Wang, X.; Shi, J.; Lan, X.; Su, T.; Guo, Y. Vehicle Detection Based on Information Fusion of mmWave Radar and Monocular Vision. *Electronics* **2023**, *12*, 2840. [CrossRef]
42. Su, Y.; Wang, X.; Lan, X. Co-prime Array Interpolation for DOA Estimation Using Deep Matrix Iterative Network. *IEEE Trans. Instrum. Meas.* **2024**, 1. [CrossRef]
43. Wang, C.; Yeh, I.; Liao, H. YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv* **2024**, arXiv:2402.13616.

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.