Deep Learning based Object Distance Measurement Method for Binocular Stereo Vision Blind Area

Jiaxu Zhang
Automation School
Xi'an University of Technology
Xi'an, China

Shaolin Hu
Automation School
Guangdong University of Petrochemical Technology
Guangdong, China

Haoqiang Shi
Automation School
Xi'an University of Technology
Xi'an, China

***Abstract*—Visual field occlusion is one of the causes of urban traffic accidents in the process of reversing. In order to meet the requirements of vehicle safety and intelligence, a method of target distance measurement based on deep learning and binocular vision is proposed. The method first establishes binocular stereo vision model and calibrates intrinsic extrinsic and extrinsic parameters, uses Faster R-CNN algorithm to identify and locate obstacle objects in the image, then substitutes the obtained matching points into a calibrated binocular stereo model for spatial coordinates of the target object. Finally, the obstacle distance is calculated by the formula. In different positions, take pictures of obstacles from different angles to conduct physical tests. Experimental results show that this method can effectively achieve obstacle object identification and positioning, and improve the adverse effect of visual field blindness on driving safety.**

***Keywords*—deep learning; computer vision; binocular stereo vision; intelligent transportation**

I. Introduction

With the improvement of peoples living standards, the number of vehicles in China has continuously increased. As of the end of 2016, the number of motor vehicles in China has reached 19.4 million. More and more cars not only make life more convenient, but also accompanied with security risks, and 2016 accidents happen 165,200 times a year in road accidents. Some of these car accidents were caused by poor back sight. Usually, during the reversing process, the driver can only judge the rear environment of the vehicle through technologies such as rear-view mirror and parking sensor. These methods inevitably have problems such as single function or blind area, which leads the driver to make wrong judgments. With the growth of automotive safety and intelligence requirements, how to effectively achieve rapid detection and positioning of the reversing environment, and distinguish obstacle types become one of the focuses of research on intelligent traffic safety.

Compared with active measurement sensors such as ultrasonic sensors, laser sensors, and infrared sensors, vision sensors have the advantages of simple installation, rich information, and high efficiency, which makes the application of visual information more and more in the field of traffic intelligence. Binocular vision technology imitates human eye structure and uses two or more cameras to collect images of different orientations of the same target. The 3D information of the target can be accurately calculated by the matching image points between the left and right images under the binocular camera model [1]. In the technology of binocular positioning measurement, how to extract the features in the image and match the left and right images is a difficult and important point in the binocular vision field. The traditional stereo feature matching method is to find corresponding points in the image pair from features such as edges, colors, and textures in the image [2]. However, these methods are computationally intensive, lacking in pertinence, and are prone to mis-matching unnecessary targets, which reduces the performance and accuracy of the algorithm. In fact, the distance detection of the blind area of the vehicle only needs to obtain the coordinates of the target to be detected in a complex traffic scene, and the target detection algorithm can achieve target recognition and segmentation. At present, there are many researches on target detection. Among them, the background difference method [3] and optical flow method [4] can be used to extract moving objects from the background in continuous video frames and apply them to the background of intelligent traffic [5], [6]. The target detection can also be accomplished by using features such as Histogram of Gradient [7] and Scale-invariant feature transform [8] on the image, and a good detection effect is achieved in visual navigation [9] and intelligent video surveillance [10]. In recent years, with the rapid development of deep learning in the field of image and video processing, the target detection algorithm has developed by leaps and bounds. Deep learning overcomes the limitations of traditional hand-designed features and can accomplish goals in complex lighting, angular rotation, and occlusion conditions. A growing number of methods based on convolutional neural networks (CNNs) [11] have been proposed in the target detection direction, such as the YOLO [12] based on the regression idea proposed in 2016, and the R-CNN based on classification thought proposed in 2014 and the later improved fast-R-CNN [13] and faster-R-CNN [14]. Compared with traditional methods, target detection based on deep learning has a great improvement in performance and speed.The fastR-CNN detection method combines RPN(Region Proposal Networks) and convolutional neural network. Compared with the traditional method, the network greatly improves the robustness of the target detection algorithm and has strong adaptability to illumination changes. The accuracy rate has been greatly improved, in addition to the advantages of fewer parameters and easier updating.

In view of the accuracy, intelligence and real-time requirements of distance measurement for traffic obstacles, a new method combining deep learning and binocular vision is proposed in this paper. Using deep learning to detect obstacles in the scene, it avoids the ineffective calculation in traditional

methods. Binocular vision is combined with the deep learning detection algorithm to achieve the distance measurement of obstacle targets. It is verified by experiments that the algorithm can realize the distance measurement task.

II. Principles and Methods

A. Distance Measurement System Overall Structure

In order to achieve intelligent distance measurement of blind obstacles in vehicles, an overall architecture is designed as shown in Fig 1.

The entire system consists of three parts: camera calibration module, deep learning detection module and distance detection module. The deep learning detection module in the system contains two parts: model building and training. The target detection network built in this paper can learn to accurately extract the features of the target to be detected in the continuous gradient descent iteration, and realize the target recognition and segmentation location in the image. The binocular camera calibration module is used to construct the linkage between the binocular image plane coordinates and 3D spatial position in the image pair data. In the calibration module, the data collection platform of the binocular camera is set up first, and the intrinsic and extrinsic parameters of the left and right cameras are obtained by the camera calibration, and the relationship between the image coordinate system and the world coordinates can be obtained under the binocular vision system. Finally, in the distance detection experiment, the binocular camera collects the image pairs of the different angle of the obstacle in the scene. The data input target detection module is used to find the coordinates of the target in the image. The coordinates of the matching points between the left and right cameras and the binocular vision model can be combined to restore the space position of the target. Finally, the intelligent recognition and distance measurement of the obstacles are realized. The system uses the binocular vision principle and deep learning method to quickly and accurately realize the functions of obstacle object identification, detection, and distance measurement.

B. Object Detection Based on Deep Learning

Based on the above system design, this paper first uses the Faster R-CNN [14] model for non-contact obstacle recognition, which can complete the stereo matching task of the image plane during distance measurement. Faster R-CNN is an end to end target detection algorithm, which unify the steps of feature extraction, region proposal, bounding box regression and classification to the framework as shown in Fig 2 This structure avoids the problem that the algorithm needs extra space to cache a large number of region proposal, and reduces the difficulty of training, and has a great improvement in the accuracy and speed of detection and positioning. The test process can be divided into four steps. First, features are extracted from the original image using a convolutional neural network and combined into a more abstract feature map; then, the feature map is input into the RPN (Region Proposal Networks) to extract the candidate region of the target. Then use the ROI pooling layer to pool the target candidate area to a fixed scale connected full-connected layer, and finally, the softmax regression algorithm is used to classify the target, and the target boundary frame is obtained by using the multi task loss function. The output of the network is a 5-dimensional vector containing the target category and location information.

1) Feature Extraction Based on Deep Learning: The difference between deep learning and traditional pattern recognition is that it can automatically learn visual features rather than traditional manual design features. Traditional feature extraction relies on prior knowledge and parameter adjustment experience, and the performance of the system is mainly determined by the quality of the feature design. In the face of rich and diverse mass data, traditional methods cannot take advantage of the number of data. Each feature can only extract a specific feature of a small area, such as texture, color or combination of various features. In order to realize the global feature learning of the image and get pixel-level information of the image, LeCun proposed a convolutional neural network (CNN) [11] based on the traditional neural network, which is composed of the Convolutional layer and the pooling layer alternately. Among them, the convolutional layer and the pooling layer match each other, automatically discovering the hidden features in the image data, and have satisfactory results on computer vision problems. This method not only reduces the complexity of the implementation, but also has a strong scalability for different types of image data with different illumination and occlusion conditions.

The target detection in this paper requires the use of VGG convolutional neural network [15] for feature extraction, and then the extracted feature map is sent to the fully connected layer to complete the task of classification or bounding box regression. The VGG model is the same as a conventional convolutional network and consists of a convolutional layer, a pooled layer, and a fully connected layer. The network structure, as shown in Fig 3, consists of 16 layers of convolution, 5 layers of maximum pooling, 3 layers of full connectivity, and a softmax classifier. The difference compared with the traditional convolutional network is that a 77 convolution kernel is decomposed into multiple convolution kernels with a size of 33 and a stride size of 1, and the number of convolution layers per layer is doubled. Each pooling layer halved the side length of the feature map. This structure allows the network to use less parameters to get a deeper network structure, add more nonlinearity to the network, and the details of the image can be better expressed. Convolution through layers makes it possible to minimize the amount of subsequent calculations while retaining the features of the input image, providing input for region proposal networks.

2) Region Proposal Networks: The Region Proposal Networks (RPN) [16] takes the feature map extracted by the convolutional neural network as input, and outputs bounding boxes of the detection target , and each candidate frame corresponds to the probability score and location information of a target. The RPN network structure used in this paper, as shown in Fig 4, uses a 33 convolution sliding window to traverse the entire feature graph, and maps each sliding window to a 256-dimensional feature vector. An anchor is set for each sliding window. Each anchor corresponds to a total of k region proposals with multiple scales and different aspect ratios(This article uses a scale of 128,256,512, length, width ratio of 1:1, 1:2, 2:1 total of nine anchor). Each low-dimensional feature

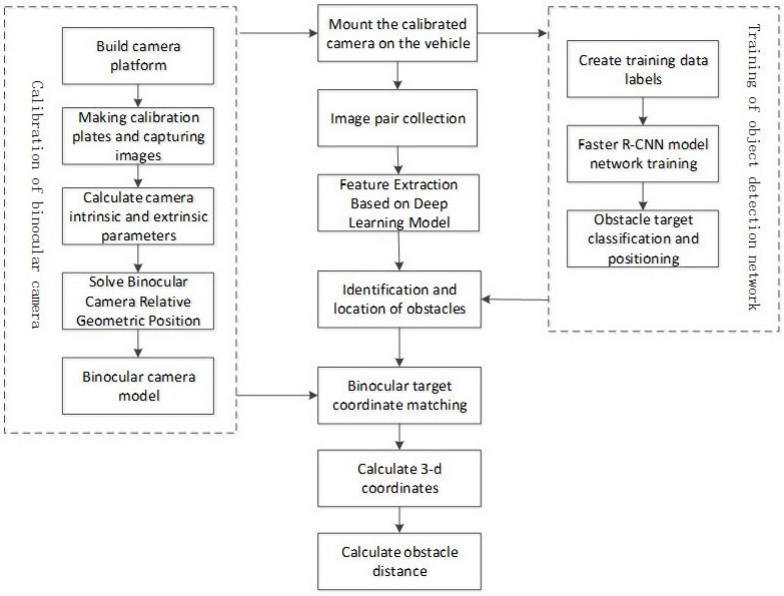

Fig. 1. System architecture diagram

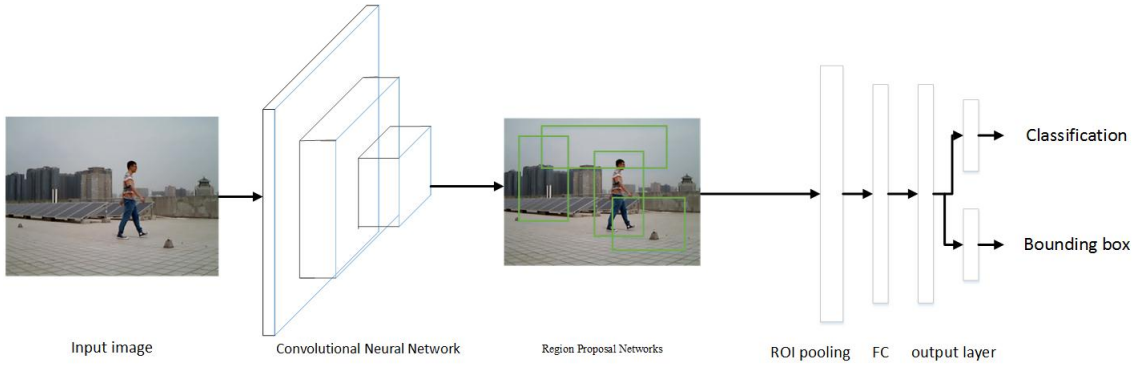

Fig. 2. Object detection model based on deep learning

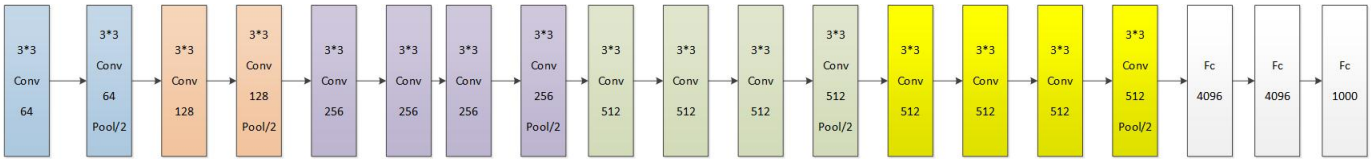

Fig. 3. VGG network structure

vector connects two parallel full-connection layers, which are used to implement bounding box regression and classification, respectively. Then all the candidate regions of the feature map are ranked according to the probability. Finally, the first 300 regions are selected as candidates for the Faster R-CNN network.

3) Obstacle Detection Model Based on Faster R-CNN: After the feature extraction and candidate region proposal have been performed on the input image, the candidate regions with different sizes are connected to the ROI pooling layer, and the feature maps of these candidate regions are converted into a fixed size. Finally, the fixed-size feature map is connected to the full-connection layer , and the full connection layer is divided into two branches, which are the classification layer using Softmax and the regression layer that rectifies the target bounding box. The multitask loss function used by the model is shown in Formula (1).

$$L(p_i, t_i) = \frac{1}{N_{cls}} \sum L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum p_I^* L_{reg}(t_i, t_i^*) \quad (1)$$

$$L_{cls}(p_i, p_i^*) = -\log(p_i^* p_i + (1 - p_i^*)(1 - p_i)) \quad (2)$$

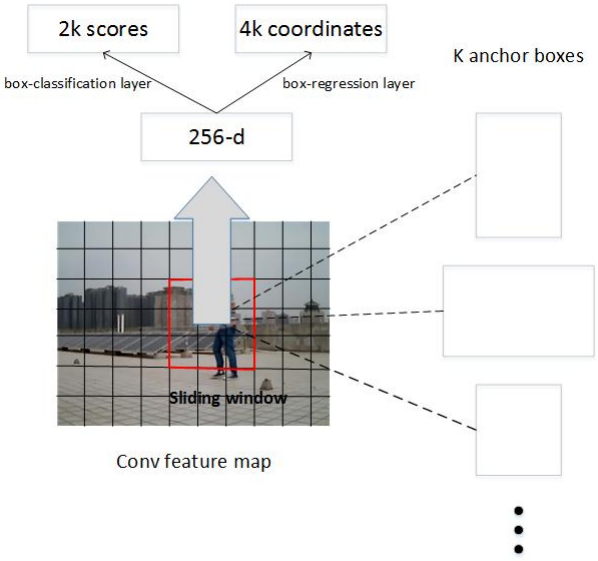

Fig. 4. Region Proposal Networks

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) = smooth_{L1}(t_i, t_i^*) \tag{3}$$

$$smooth_{L1}(t_i, t_i^*) = \begin{cases} 0.5(t - t^*)^2, if|t - t^i| < 1 \\ |t - t^*| - 0.5, otherwise \end{cases} \tag{4}$$

Formula (1): p_i is the target prediction probability, if the region contains the target p_i^* is 1, otherwise it will be 0. t_i is the predicted value of bounding frame coordinates, and t_i^* is the true coordinates of the bounding box. N_{cls} and N_{reg} are the normalized parameters of the classification and regression terms, respectively, and $Lambda$ is the balance weight. L_{cls} is the cross-entropy loss of the classification and $Lreg$ is the regression loss, where R is the robust loss function.

Faster R-CNN implements an end-to-end network structure to achieve target positioning and classification by sharing convolutional features. Save a lot of computing and caching needs, and achieve fast and accurate target detection.

C. Binocular Stereo Distance Measuring Principle

Binocular stereo ranging uses two cameras with different positions to shoot the same target, and then uses the corresponding image points to get the depth information of the target. The distance measuring system is mainly realized by three steps: camera imaging, three-dimensional reconstruction and distance measurement.

1) Camera Model: The image acquisition in the system is completed by two cameras of the same type. The camera projects the three-dimensional scene onto the two-dimensional imaging plane through the optical lens. In order to describe the perspective mapping relationship between the plane image and the space target, this paper uses a pinhole camera model, as shown in Fig 5, to describe the camera imaging process. The image point x is formed by intersecting a straight line connecting the three-dimensional point X and the camera optical center C with the image plane. The dashed line represents

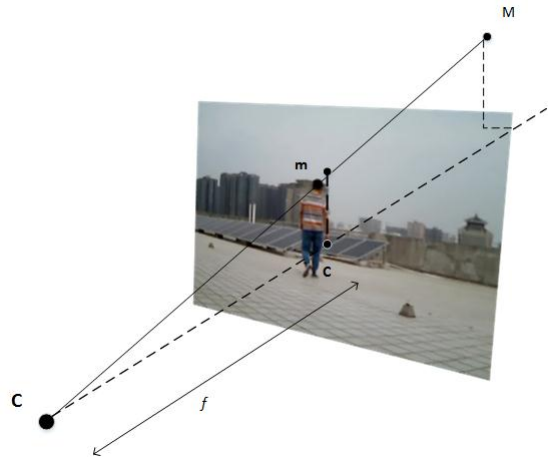

Fig. 5. Pinhole camera model

the camera's optical axis and f represents the camera's focal length.

In the pinhole camera model, the mapping relationship between the three-dimensional space points and the two-dimensional projection point coordinates is as shown in formula (5), and the homogeneous coordinate form of formula (5) is as shown in formula (6).

$$m = PM \tag{5}$$

$$\begin{aligned} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{aligned} \tag{6}$$

m is the coordinate of the spatial point M projected on the imaging plane. Among them, f represents the focal distance, that is, the distance between the image plane and the camera center, and the $f_x = \frac{f}{d_x}$ and $f_y = \frac{f}{d_y}$ and (d_x, d_y as the physical dimensions of a single pixel) represent the normalized focal length on the X and Y axis, respectively. The intersection point of optical axis and image plane is called the origin of image coordinates,$c = [c_x, c_y]$. These parameters are called intrinsic parameters, and they are composed of 3×3 matrix K called the camera's intrinsic parameter matrix to describe the projection relationship of the camera. R is a rotation matrix describing the direction of the camera, t is a three-dimensional translation vector describing the center position of the camera, and a 4×4 matrix consisting of R and t is called an extrinsic parameter matrix.

2) Binocular Vision Principle: The camera can only get two-dimensional images and lose depth information in the process of photography. Therefore, in order to realize the visual distance measurement, we can imitate the human eye with two cameras in different locations to collect the images, and then use the projection relationship between the position of the

image point and the three-dimensional space point to calculate the target space position. In this experiment, the positions of two cameras are arranged as shown in Fig 6. The optical axes of the left and right cameras do not need to be parallel but the visual field overlaps. O_L and O_R are the optical centers of the left and right cameras respectively. The projections of the space point M on the imaging planes of the two cameras are x_1 and x_2, respectively, and Z_L and Z_R are the optical axes of the two cameras.

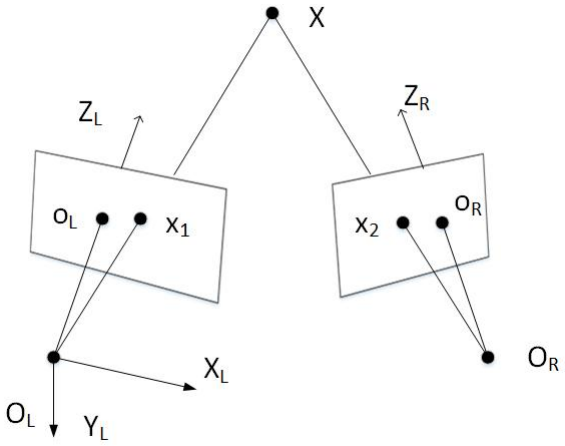

Fig. 6. Binocular 3-dimensional reconstruction model

When analyzing binocular geometric relationships, the relative positional relationship between the two cameras is represented by a homography matrix. Since only rigid body transformation is performed between the left and right cameras, that is, the coordinate system is transformed only by the homography matrix, the geometric structure between the two cameras can be represented by the rotation matrix R and the translation matrix T. For calculation convenience, coordinate axis and origin of coordinates are aligned with the first camera coordinate system. That is, the origin of the space coordinate system is the left camera optical center O_L, the coordinate Z axis coincides with the optical axis Z_L and is perpendicular to the image plane, and the X_L and Y_L axes are parallel to the x axis of the image y axis. Thus, the projection matrix formula of left and right cameras can be obtained (7)(8):

$$P_1 = K_1(I|0) \tag{7}$$

$$P_2 = K_2(R|t) \tag{8}$$

According to the projection model formula (5) of the camera, we can obtain the relationship between the three-dimensional point M and the two-dimensional projection point P_L and P_R as formula (9). The projection models of the two cameras are p_1 and p_2 respectively, and the projection points of 3D point M are x_1 and x_2.

$$\begin{bmatrix} P_1 & -x_1 & 0 \\ P_2 & 0 & -x_2 \end{bmatrix} \begin{bmatrix} M \\ 1 \\ 1 \end{bmatrix} = 0 \tag{9}$$

During distance measurement, the corresponding coordinates of the left and right images are $x_1(u_L, v_L)$ and $x_2(u_R, v_R)$, respectively. The intrinsic and extrinsic parameters in the projection matrices P_1 and P_2 can be obtained by stereo camera stereo calibration. The spatial point $M(x_L, y_L, z_L)$ can be calculated by Equation (9).

3) Distance Measurement Principle: The binocular camera model and the deep learning method are combined to obtain the three-dimensional coordinates (x_L, y_L, z_L) of the target in the left camera coordinate system, and then the distance between the obstacle object and the binocular camera device is calculated. Therefore, the formula (10) of the position coordinate of the right camera optical center $(x_r, y_r, z_r, 1)^T$ in the coordinate system can be obtained according to formula (7)(8).

$$\begin{bmatrix} X_R \\ Y_R \\ Z_R \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & t_{3\times1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix} \tag{10}$$

From the known three coordinates, from Cosine's Theorem (11), the angle between the left camera and the obstacle line v_1, the left camera and the right camera line of the line v_2 can be calculated, and the distance between the target and the binocular camera is obtained. The distance D between the target obstacle and the vehicle is as shown in formula (12).

$$\cos\theta = \frac{v_1 v_2}{\|v_1\|\|v_2\|} \tag{11}$$

$$D = \sin\theta \cdot v_1 \tag{12}$$

III. Test and Result Analysis

During the test process, the data acquisition system consists of two CCD cameras, model GY0814-3MP, a USB 2.0 interface CCD industrial digital camera with the same parameters and an industrial lens with 3 million pixels. Select a 210mm*297mm calibration plate, computer and corresponding software to achieve target distance measurement. The data acquisition platform is shown in Fig 7. The two cameras are fixed on the platform with adjustable spacing, angle and height to meet the needs of different distances and effective ranges. The experimental part includes two parts: The first part is to detect the corresponding image points of the target on the imaging planes of the left and right cameras through deep learning; the second part is to obtain the three-dimensional coordinates and distances by substituting the image points into the calibrated camera model.

Fig. 7. Binocular experimental platform

A. Faster R-CNN Network Training

In the distance measurement system, we use the detection result (object centroid) of the deep learning model in Section 2.2 as the matching point between the left and right image planes. In order to verify the positioning capability of the Faster R-CNN model, video data was collected by a vehicle digital camera to verify the model performance. The data set is divided into six video sets ($Set_{L1} - Set_{L3}, Set_{R1} - Set_{R3}$).Extract 5 frames of images per second and save them as jpg images of size 800*600 for a total of 2983 images. The data set was randomly divided into training set (80%) and test set (20%) to achieve object detection network training. Each picture also has a detailed tag in XML format for recording image information. The tag contains the target category, image width, height depth and bounding box coordinate information. Finally, 580 images and corresponding tag files are obtained for training and testing models.

In the training process, the initial learning rate is 0.004. As the number of iterations increases, the initial learning rate decreases at a decay rate of 0.95. The model first unified the training set image size to 300×300, and then iterated 8,000 times using RMSProp algorithm until the loss function reached a minimum, so that obstacles with different distances and angles can be detected and identified, as shown in Fig 8 . When using the test set to verify model performance, the Precision indicator evaluation model was used. The results are shown in Fig 9. With the increase of the number of trainings, the detection accuracy of deep learning can reach 99%, which indicates that Faster R-CNN can accurately locate the obstacles.

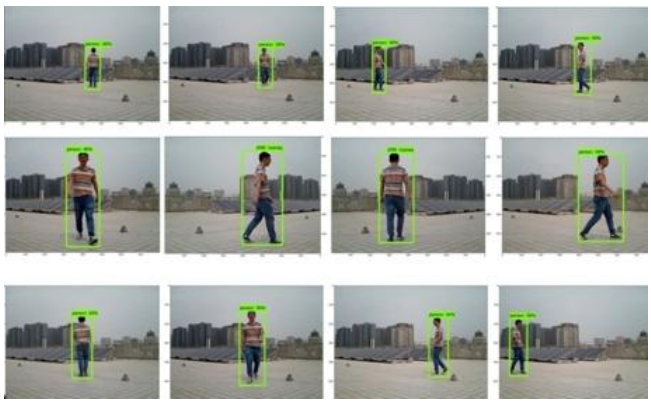

Fig. 8. Test set detection results with different distances and different angles

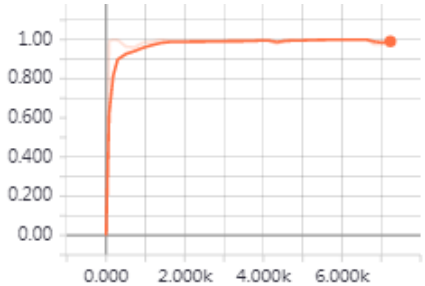

Fig. 9. Performance of object detection network

Fig. 10. Checkerboard calibration board

B. Binocular Camera Calibration and Distance Measurement

Next, in order to achieve the calculation of three-dimensional position through two-dimensional image coordinates. The centroid matching points x_1 and x_2 are substituted into the binocular model (9) to calculate the three-dimensional coordinates. The parameters P_1 and P_2 of the camera in the formula need to be calibrated, and he calibration result has a very important influence on the measurement accuracy. The distance measuring system uses the calibration plate as a reference for binocular vision calibration to solve the camera's intrinsic and extrinsic parameter matrix. In the calibration process, the positions and angles between the binocular cameras are adjusted, and take 28 pictures from different angles on the checkerboard calibration board shown in Fig 10. Then use software Matlab calibration tool stereoCameraCalibrator [17] to extract the corner points of the calibration plate. After obtaining the corner points, the intrinsic parameters of the two cameras are solved using the calibration method [18] as shown in Table 1. Then the mapping relationship between the left and right images is calculated by stereo matching, that is, the extrinsic parameter rotation matrix R and the shift vector T. The result is shown in Table 2.

The binocular camera calibration results are shown in Table 1 and Table 2.

TABLE I. BINOCULAR STEREO VISION SYSTEM INTRINSIC PARAMETER CALIBRATION RESULTS

Intrinsic parameters	f_x	f_y	c_x	c_y
Left camera	1083.08	1086.58	416.18	324.09
Right camera	1092.60	1096.39	421.79	257.48

TABLE II. BINOCULAR STEREO VISION SYSTEM EXTRINSIC PARAMETER CALIBRATION RESULTS

$R =$	0.9976	−0.0122	0.0677
	0.0109	0.9997	0.0198
	−0.0679	−0.0191	0.9975
$t =$	2.1858×10^2	−2.2566	5.7440

In order to verify the method proposed in this paper, physical measurement experiments were conducted. Binocular cameras take pictures of obstacles at different distances and angles. The obtained image data is input into the target detection network that has been trained in Section 3.1. The model can automatically distinguish obstacle types and centroid coordinates x_1 and x_2.Substituting the matching points $x1$, x_2 and the calibrated camera projection matrix P_1, P_2

into equation (9), the three-dimensional coordinate X of the obstacle in space can be obtained. Using distance formula (11) (13), the distance between obstacles and vehicles can be calculated.

The distance measurement experiment was performed on a moving obstacle with a distance of 5-7.5 meters from the binocular camera. As shown in Table 3, the accuracy is analyzed by calculating the error with the actual distance. The experimental results are shown in Table 3 and Fig 11.

TABLE III. BINOCULAR STEREO VISION SYSTEM EXTRINSIC PARAMETER CALIBRATION RESULTS

Real distance/m	Measuring distance/m	Error/m	relative error/%
4.5	4.311	0.189	4.20
4.5	4.452	0.047	1.04
4.5	4.101	0.398	8.84
4.5	4.456	0.044	0.90
4.5	4.374	0.126	2.80
5.5	5.110	0.390	7.09
5.5	5.756	-0.256	4.65
5.5	5.271	0.229	4.16
5.5	5.382	0.118	2.15
5.5	5.264	0.236	4.29
6.5	6.814	-0.314	4.83
6.5	6.577	-0.077	1.71
6.5	6.283	0.217	3.34
6.5	6.329	0.171	2.63
6.5	6.073	0.426	6.55
7.5	7.171	0.329	4.39
7.5	7.804	-0.304	4.05
7.5	7.098	0.402	5.36
7.5	7.747	-0.247	3.29
7.5	7.560	0.06	0.80

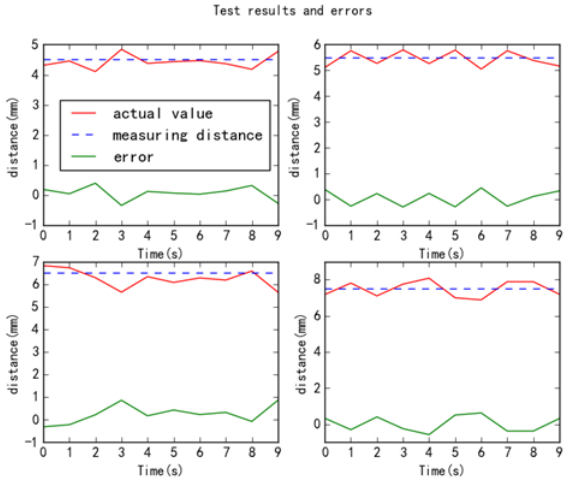

Fig. 11. Test set detection results with different distances and different angles

IV. CONCLUSION

In order to avoid the traffic accident caused by the driver's limited vision during reversing, a system based on binocular vision and deep learning is designed to realize obstacle recognition and distance detection. In the system, the calibrated binocular camera equipment is installed in the rear of the vehicle, the image data is detected by the deep learning corresponding image points. The triangulation method and the binocular camera model are used to calculate the distance information of the target. Tests were performed on targets with different viewing angles and different distances to verify the performance of the system. Compared with traditional methods, deep learning has the advantages of robustness, accuracy, and speed, and it can achieve real-time, high-precision recognition and distance measurement through the combination of binocular stereo vision. The method proposed in this paper can effectively identify the target object and determine the distance, which ensures the safety of the driver's personal and property during driving.

Although this article has achieved some results in ranging, there are still many problems to be solved. Here are the next research plans for this article:

1: In the study of obstacle detection in this paper, only one category of pedestrians is involved, but in actual situations, there are many types of obstacles and the shape is varied. In the future research, it is necessary to select different types of obstacles for research and improve the applicability of the research content.

2: The method used in this paper can detect the distance of the obstacle, and in the future can detect more details of the target, including speed, direction, angle and so on. Achieve more accurate target detection and ensure the safety of vehicle driving.

ACKNOWLEDGMENT

This paper is supported by the National Nature Science Foundation of China (Grant No.61473222 and No. 91646108

REFERENCES

[1] Xiaochao Tian,Zhongke Li. Research on Computer Vision Feature Representation and Learning for Image Classification and Recognition. Electronics Optics & Control, 2015,22(03):54-57+96.

[2] Zhirui Wang,Cailiang Yan. Method of Image Feature Extraction. Journal of Jishou University(Natural Science Edition), 2011,32(05):43-47.

[3] Yu bo,Guili Wang,Yanxu Zhou,Peng. Investigation on Video Moving Target Detection Algorithm in Traffic Monitoring System. Computer Technology and Development, 2017,27(08):156-158+163.

[4] Yijing Hu, Zhengfang Li, Yueming Hu. Theories and Applications of Motion Analysis Based on Optical Flow. Computer Measurement & Control, 2007,15(2):219-221.

[5] Huiyin Zhu, Li Zhang, Yunyan Li. Intelligent Traffic Event Detection System Based on Background Difference. Journal of wut (information & management engineering) , 2011, 33(2):224-227.

[6] Xilai Li, Aihua Li, Xiangfeng Bai. Moving Vehicles Detection in Intelligent Transportation Systems Based on Optical Flow. Electro-Optic Technology Application, 2010, 25(2):75-78.

[7] Pass G, Zabih R. Histogram refinement for content-based image retrieval// IEEE Workshop on Applications of Computer Vision. IEEE Computer Society, 1996:96.

[8] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 2004, 60(2):91-110.

[9] Xiandong Xu, Bingrong Hong, Yiguan, , et al. A somparison of several feature points methods used in mobile robot visual navigation. Journal of Huazhong University of Science and Technology(Natural Science Edition), 2011, 39(s2):200-203.

[10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Engineering & Science, ,2015,37(10):1947-1951.

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[12] Uijlings J R R, Sande K E A V D, Gevers T, et al. Selective Search for Object Recognition. International Journal of Computer Vision, 2013, 104(2):154-171.

[13] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. 2015:770-778.

[14] He K, Zhang X, Ren S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1904-1916.

[15] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science, 2014.

[16] Felzenszwalb P F, Girshick R B, Mcallester D A, et al. Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[17] Wei Zhang, Hong Cheng, SuiWei. Design and Implementation of Camera Calibration System . Computer Engineering, 2007, 33(2):255-256.

[18] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations," Proceedings of the Seventh IEEE International Conference on Computer Vision(ICCV), Corfu, Greece, 1999, pp. 666. doi:10.1109/ICCV.1999.791289