2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) | 978-1-7281-8535-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/IMCEC51613.2021.9482067

Advanced Driver Assistance System based on Machine Vision

Jiajun Zhu, Fanglei Shi, Jiaying Li
College of Internet of Things Engineering, Hohai University, Changzhou, China
1052127739@qq.com, 1939525225@qq.com, 2582427350@qq.com

***Abstract*—In view of the frequent occurrence of traffic problems such as urban traffic congestion and traffic accidents, people pay more and more attention to traffic safety. This paper takes the key technical problems such as front vehicle detection and identification and pre-collision detection in the advanced driving assistance system as the research object, this paper puts forward a computer vision solution based on YOLOv5 algorithm and monocular camera distance calibration, which provides more comprehensive driving environment information for the advanced driving assistance system and improves the active safety of the vehicle. Finally, a case study is given to verify the superiority of the algorithm proposed in this paper.**

Keywords—machine vision; YOLOv5; monocular camera distance calibration ; the advanced driving assistance system

I. INTRODUCTION

With the growth of car ownership and the continuous improvement of highway grades,there are more and more traffic accidents. Therefore, how to reduce the incidence and mortality of traffic accidents has become an urgent problem to be solved. The advanced driving assistance system studied in this paper is a key component of intelligent transportation, which refers to the technical processing of static and dynamic object identification, detection and tracking by using all kinds of sensors installed on the vehicle to collect the environmental data inside and outside the vehicle at the first time. In order to improve the active safety of driving vehicles, this paper studies that in the actual road environment, machine vision is used to detect and track the vehicle in front, and YOLOv5 target detection technology is used to assist drivers to solve the problems of front vehicle detection and front pre-collision detection in complex environment. At the same time, the real-time road condition information of the road ahead is collected by monocular CCD, and the extracted vehicle position information is fed back to the driver, so that the driver can avoid the corresponding risk in time according to the feedback information.

II. METHOD

A. Network structure and working principle of YOLO

Target detection methods based on deep learning are mainly divided into two categories: two-stage target detection algorithm and one-stage target detection algorithm. The former first generates a series of candidate boxes as samples by the algorithm, and then classifies the samples by convolution neural network, which is characterized by slow speed but high accuracy, while the latter does not need to generate candidate boxes. the problem of target frame location is directly transformed into a regression problem, which is characterized by one step, and the algorithm is fast[1]. In the research process of this system, the YOLO algorithm which requires higher prediction speed of the model is selected, and the YOLO algorithm belongs to the one-stage target detection algorithm. It is the most advanced real-time target detection system at present, and it is another landmark target detection algorithm after RCNN, FASTER-RCNN and SSD.

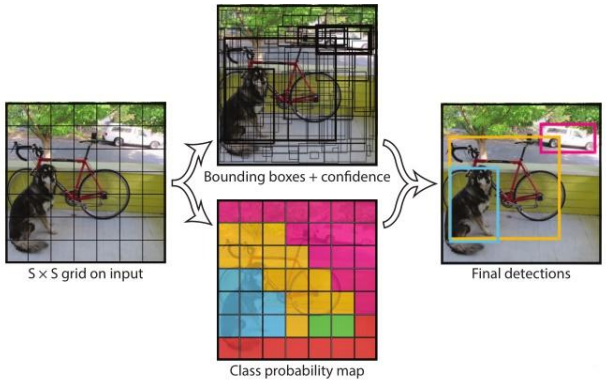

Fig. 1. Schematic diagram of working principle of YOLO algorithm

The core idea of YOLO algorithm is to use the whole graph as the input of the network and directly return the position and category of boundingbox in the output layer. An image is divided into several grids, and if the center of a detected target falls on this grid, the grid is responsible for predicting the target. Each grid predicts eight boundingbox, and each boundingbox not only returns to its own location, but also predicts a confidence value. Confidence is reflected in two aspects: one is the possibility that the bounding box contains the target, and the other is the accuracy of the bounding box. The former is written as Pr (object). If object falls in a gridcell, Pr (object) = 1, otherwise, Pr (object) = 0. The accuracy of the bounding box can be determined by the IOU

(intersectionoverunion) of the prediction box and the actual box (groundtruth)[2]. So confidence can be define as

$$confidence = \Pr(Object) * IOU_{pred}^{truth} \quad (1)$$

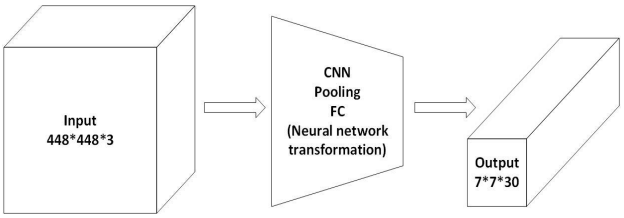

Fig. 2. Model structure of abbreviated version of YOLO algorithm

There are 5 values for each bounding box to predict and confidence, and each grid also predicts a category information, which is recorded as Class C. Then there are three grids, each grid needs to predict B bounding boxes, and C categories should be pre-measured. That is, the output value of all the grids is S*S*(B*5+C). Each grid predicts the probability of C conditional class probability, The coordinates (x, y) represent the relative value of the predicted center of the boungingbox to the grid boundary, and (w, h) is the width and height of the bounding box, which is proportional to the width and height of the whole picture. Confidence is the predicted IOU value of bounding box and ground truth box. As a result, the class, size and position of the tested target are predicted.

B. Yolov5 algorithm

It has been five generations from YOLOv1 to YOLOv5. The advanced assistant driving system based on machine vision studied in this paper adopts the computer vision solution based on YOLOv5 algorithm for target detection. YOLOv5 algorithm is a lightweight and high-performance real-time target detection method recently introduced. This algorithm adds a new improved idea on the basis of YOLOv4, and greatly improves the performance in terms of detection speed and accuracy.

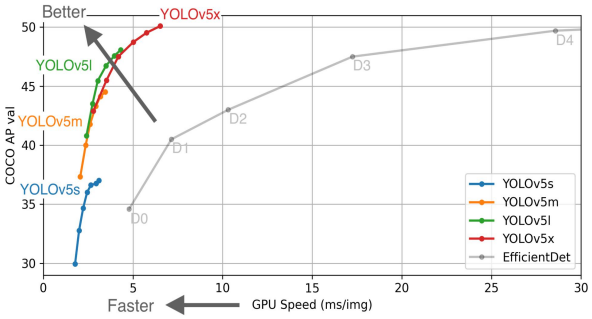

Fig. 3. Algorithm performance Test Chart of YOLOv5

There are four network models in YOLOv5, which are Yolov5s, Yolov5m, Yolov5l and Yolov5x. This paper focuses on the network structure of Yolov5s. Yolov5s can be divided into four parts: input, Backbone, Neck and Prediction.

- Input: represent the input picture, and the network inputs an image with the size of 608 to 608, and the image preprocessing and normalization are carried out in this stage. YOLOv5 uses Mosaic data enhancement operation to improve the training speed of the model and the accuracy of the network, and proposes an adaptive anchor frame calculation and adaptive picture scaling method.

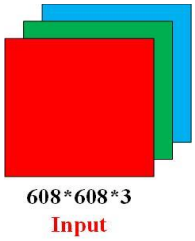

Fig. 4. YOLOv5s network structure Input

- Backbone: it is usually the network of some classifiers with excellent performance, which is used to extract some common characteristic tables. YOLOv5 uses CSPDarknet53 structure and Focus structure.

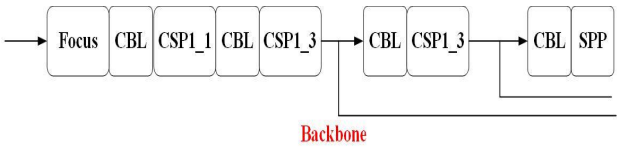

Fig. 5. YOLOv5s network structure Backbone

- Neck: to further improve the diversity and robustness of features, YOLOv5 also uses SPP module, FPN plus PAN module.

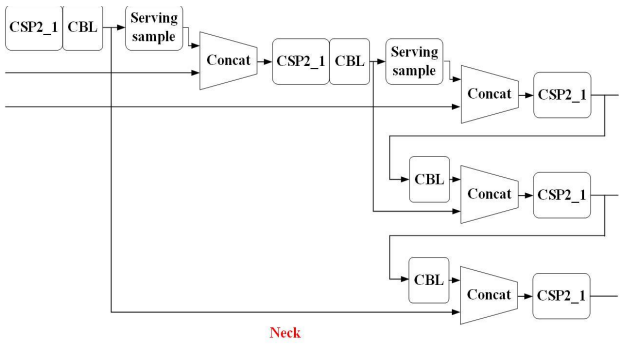

Fig. 6. YOLOv5s network structure Neck

- Prediction: used to complete the output of target detection results. The output branches of different

detection algorithms usually contain a classification branch and a regression branch.

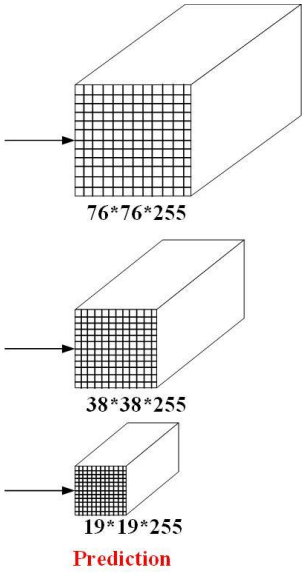

Fig. 7. YOLOv5s network structure Prediction

Compared with the previous YOLO series, the YOLOv5 algorithm considers the neighborhood positive sample anchor matching strategy and adds positive samples; through flexible configuration parameters, we can get models of different complexity; through some built-in super-parameter optimization strategies, we can improve the overall performance; and use mosaic enhancement to improve the performance of small object detection. At the same time, YOLOv5 is small and the weight file is nearly 90% smaller than 4, so that YOLOv5 can be deployed on embedded devices. Compared with YOLOv4, YOLOv5 has higher accuracy and better ability to identify small targets[3].

C. Monocular camera distance calibration

The advanced driving assistance system uses the advantages of YOLOv5 algorithm to calibrate the distance of monocular camera based on machine vision. Monocular vision ranging adopts the projection model of geometric relationship combined with camera calibration and shadow width ranging. The former derives the internal parameters of the camera, which must be obtained by special calibration methods. The latter parameters can be obtained by external measurement [4].

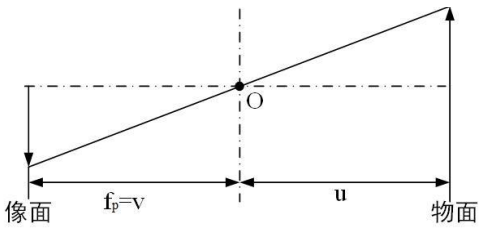

Fig. 8. Pinhole model

The imaging principle of the camera is keyhole imaging. It is assumed that the reflected light on the surface of the object is projected on the plane through a keyhole. The focal length fp of the pinhole model is equal to the distance from the light center O to the imaging plane, and the object distance u is equal to the distance from the light center O to the object surface.

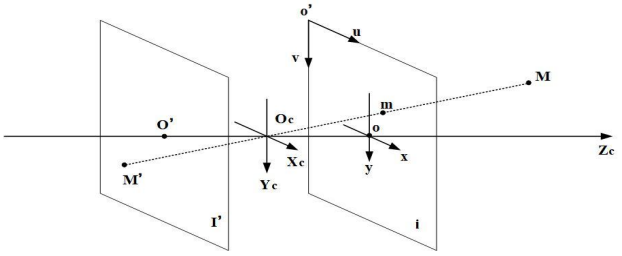

Fig. 9. Imaging schematic diagram of camera pinhole model

As shown in figure 9, the geometric relationship among the object point M, the light center Oc and the image point M'is a collinear equation. OcO'=Oco=f,f is the focal length of the camera. The projection model in which the image plane is located in front of i is usually used to facilitate the conversion between the spatial position of the image point and the corresponding object point, and this process does not affect the results of image analysis.

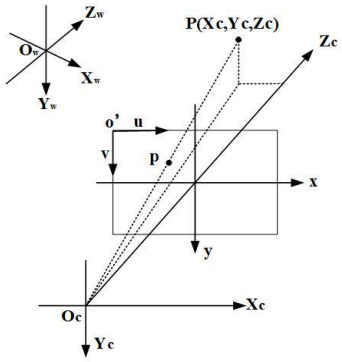

Fig. 10. Four reference coordinate systems

In fact, from the image captured by the camera to the final imaging display in the computer needs to go through the conversion between multiple coordinate systems. The conversion steps are as follows:

Step 1: Convert the 3D spatial coordinates of the target object into camera coordinates after a rotation and translation.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = [R \quad T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2}$$

Where (Xw,Yw,Zw) is the coordinate of the target object in the world coordinate system. (Xc,Yc,Zc) is the coordinate of the camera coordinate system; R is the rotation matrix of 3 ~ 3; T is the translation matrix of 3 ~ 1.

Step 2: According to the principle of keyhole projection transformation, the camera coordinates are transformed into two-dimensional image plane coordinates of the photosensitive screen.

$$Z_c \times \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{3}$$

In the formula, (x, y) is the coordinate of the two-dimensional image plane on the photosensitive screen; f is the effective focal length of the camera.

Step 3: The image plane coordinate system is transformed into the frame storage coordinate system in the computer memory by the internal parameter equation.

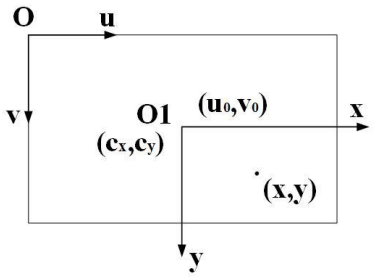

Fig. 11. Schematic diagram of coordinate system transformation

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

For the transformation from image coordinate system to pixel coordinate system, the image has fx pixels per millimeter in x direction and fy pixels per millimeter in y direction. Where cx,cy is the coordinate of the origin of the image coordinate system in the pixel coordinate system.

From equation (2) (3) (4), it can be concluded that the relationship between the world coordinate system and the pixel coordinate system is:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{5}$$

The above formula can be simplified as follows:

$$Z_c \times P = A \times M \times P' \tag{6}$$

In the formula, P is the target pixel coordinate, P' is the target world coordinate, A is the camera internal parameter matrix, M is the camera external parameter matrix, the position of the camera will affect the external parameters.

In reality, due to the influence of lens on light propagation, camera imaging will produce radial distortion and tangential distortion. Zhang Zhengyou calibration method[5] only focuses on the radial distortion, using the optimal solution of maximum likelihood estimation, the distortion parameters are obtained, and the image is dedistorted, and then the dedistorted image is used to determine the internal parameters of the camera. Finally, the external parameter array is estimated based on the internal parameter array. Camera calibration is the process of determining the internal and external parameters of the camera.

After completing the above steps, the computer can manipulate the image.

III. Contrast Experiment

In order to compare the difference in accuracy and speed between YOLOV5 and YOLOV4， In Experiment 1, we compared the Average precision of yolov5s and yolov4-custom based on the COCO2017 data set.

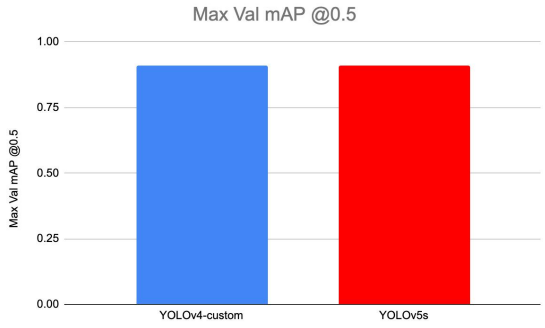

Fig. 12. Comparison of Max Val mAP between yolov4-custom and yolov5s

YOLOv4 maximum validation mAP @0.5 is 0.91, YOLOv5s Tensorboard - maximum val mAP @0.5 is 0.91 .mAP was similar between the two models on our task as both models achieved their max.

In Experiment 1, we compare inference time run under both networks with the environment configurations specified (Tesla P100).

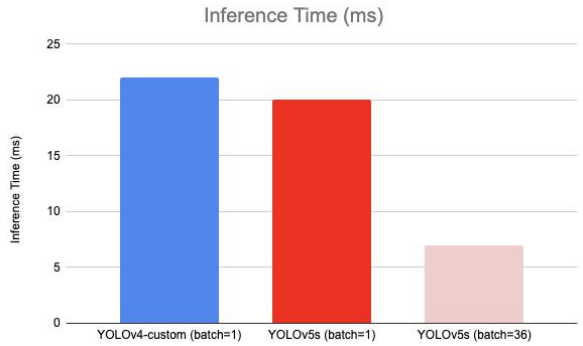

Fig. 13. Comparison of inference time between yolov4-custom and yolov5s

On single images (batch size of 1), YOLOv4 inferences in 22 ms and YOLOv5s inferences in 20ms and YOLOv5s inferences in 7 ms (140 FPS) when you infer in batch.

IV. CONCLUSION

This paper makes an in-depth study of the safety-assisted driving part of modern intelligent vehicles, which can feedback the potential dangers on the traffic highway to the driver in time, reduce traffic safety accidents and realize the safety guarantee of vehicles in the process of driving. it has rich practical significance. This paper mainly uses machine vision to realize the research of advanced driving assistance system, and puts forward the target detection technology based on YOLOv5 algorithm. On the basis of the advantages of YOLOv5 algorithm, combined with the principle of pinhole imaging, single-phase distance calibration is realized.

First of all, the paper briefly explains the advanced driving assistance system and the overall design of the system, and then in the second part introduces the network structure and working principle of YOLO algorithm, and describes the superiority of YOLOv5 algorithm and the calibration method of internal and external parameters of vehicle camera, that is, monocular camera distance calibration. The third part analyzes the experimental phenomena and verifies the superiority of Yolov5 algorithm in end-to-end training and inference, reforming regional suggestion frame target detection framework and real-time target detection. The speed of the algorithm is fast, which improves the performance of the system to a great extent. Finally, as the advanced driving assistance system is a rich research object, there are many challenging tasks that need to be further studied.

REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick,et. You Only Look Once: Unified, Real-Time Object Detection [J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016(6): 27-30.

[2] Maohui Li, Chuanping Wu, Yan Bao,et. On the Application of YOLO algorithm in Machine Vision [J]. Educational Modernization, 2018, v.5 (41): 180-182.

[3] Yifan Liu, BingHang Lu, Jingyu Peng, Zihao Zhang. Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition[J]. World Scientific Research Journal,2020,6(11).

[4] Zhuoyuan Tong. Design of front vehicle detection and ranging system based on machine vision [D]. Harbin Institute of Technology, 2015.

[5] Wei Wei, Laolong Liu, Wei Zhang. Camera Calibration and GUI realization based on Zhang Zhengyou plane Calibration method [J]. 2010.