Research On The Key Frame Selection Method Of YOLOv5-Based Intelligent Damage Determination System

Qianqian Zhu
Automotive Data of China Co.,Ltd., China Automotive Technology and Research Center Co.,Ltd.,
Tianjin, China.
zhuqianqian@catarc.ac.cn

Yingnan Liu
Automotive Data of China Co.,Ltd., China Automotive Technology and Research Center Co.,Ltd.,
Tianjin, China.
liuyingnna@catarc.ac.cn

*Xuening Wu**
Automotive Data of China Co.,Ltd., China Automotive Technology and Research Center Co.,Ltd.,
Tianjin, China.
wuxuening@catarc.ac.cn*

Jiaying Huang
Automotive Data of China Co.,Ltd., China Automotive Technology and Research Center Co.,Ltd.,
Tianjin, China.
huangjiaying@catarc.ac.cn

2024 9th International Conference on Automation, Control and Robotics Engineering (CACRE) | 979-8-3503-5030-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CACRE62362.2024.10635092

***Abstract*—In vehicle damage claims, the intelligent damage determination method based on computer vision instead of manual observation has become the future direction of technology. In the process of wise damage determination, because of the difficulty that the images taken manually cannot be quickly shot with the image quality applicable to the system, this paper proposes a keyframe selection method based on the YOLOv5 model, which solves the problems of convenience and ease of use of the intelligent damage determination system by adding the real-time operation guidance function at the time of the shooting. Firstly, the mobile real-time detection model YOLOv5 detects and counts the vehicle appearance parts. Secondly, the underlying logic determination model guides the user to complete the shooting distance adjustment. Experimenting with the self-built vehicle image data set, the average accuracy of mAP50 reached 94.1%, and YOLOv5 can meet the requirements of real-time and accuracy of keyframe selection in shooting environments such as uneven lighting and complex surrounding background.**

Keywords- YOLOv5；Deep Learning；Computer Vision

I. INTRODUCTION

With the rapid development of the economy, vehicles have been widely used in all areas of family life and industrial production. With the increasing consumption of cars, the dangers caused by vehicles are also increasing, and road safety issues are becoming more and more prominent. Accidents occur, insurance companies must promptly issue insurance and conduct on-site manual damage determination. Damage determination is an integral part of the insurance claims process, and the traditional manual damage determination is inefficient and prone to subjective judgment bias, so it needs to rely on advanced computer vision technology to improve the accuracy and efficiency of damage determination [1]. This paper aims to study the critical frame selection method of a YOLOv5-based intelligent damage determination system to achieve precise damage determination and output of damage-occurring parts.

In a vehicle's intelligent damage determination system, selecting keyframes to capture is critical to damage determination. However, the distance of the image capture is often a critical factor in the intelligent damage determination process. When the distance is too close to the vehicle, the number of vehicle parts captured is small. Because only a partial view is available, it is impossible to determine the vehicle damage's appearance accurately. At the same time, when the distance is too far from the vehicle, the number of vehicle parts captured is significant, but the damage is not clear enough to make an accurate damage type determination. Therefore, we must take explicit and judicious images for intelligent damage analysis at a moderate distance.

In a vehicle's intelligent damage determination system, selecting keyframes to capture is critical to damage determination. However, the distance of the image capture is often a critical factor in the intelligent damage determination process. When the distance is too close to the vehicle, the number of vehicle parts captured is small. Because only a partial view is available, it is impossible to determine the vehicle damage's appearance accurately. At the same time, when the distance is too far from the vehicle, the number of vehicle parts captured is significant, but the damage is not clear enough to make an accurate damage type determination. Therefore, we must take explicit and judicious images for intelligent damage analysis at a moderate distance.

II. SYSTEM ARCHITECTURE

The whole system structure is shown in Fig.1, which can be divided into two parts. The first part is the target detection part, which requires the user to open the app and point it at the vehicle when the YOLOv5 model on the mobile side detects and counts the vehicle appearance parts in the camera in real-time. When the number of parts is within the range of [a,b], the distance is considered appropriate, and the user is allowed to shoot; when the number of parts is less than a, the distance is considered too close to the vehicle and needs to be far away from the vehicle; when the number of parts is more significant than b, the distance is considered too far from the vehicle and needs to be close to the vehicle.

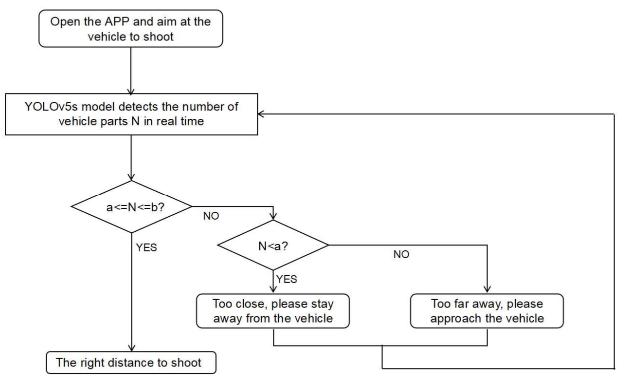


Figure 1: System architecture flowchart

A. Model building

Implementing the critical frame selection function of the intelligent damage determination system mainly relies on two parts of functions. The first one is the model detection function, which realizes the real-time detection and accurate counting of vehicle appearance parts by the YOLOv5 model; the second one is the logical determination function, which outputs a reasonable determination basis by counting the relationship between the number of vehicle appearance parts in the video frames and the shooting distance.

1) *YOLOv5 Target Detection Model:* YOLOv5 is a deep neural network model proposed by Ultralytics LLC. The YOLO series has been continuously upgraded and updated with YOLOv1 [2], YOLOv2 [3], YOLOv3 [4], YOLOv4 [5], YOLOv5, and other algorithms. The YOLOv5 model is based on the YOLOv3 model and is improved, there are four network models, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, and the comparative analysis of the results of their different network deconstruction is shown in Fig.2 below.

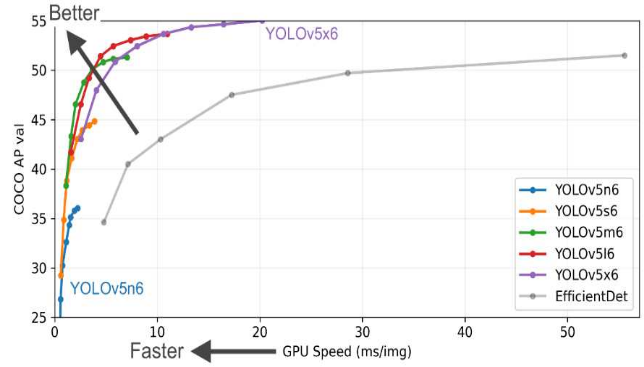


Figure2: Comparison of different network structures of YOLOv5.

YOLOv5 model controls network width and depth by depth_multiple and width_multiple, which has high detection accuracy and real-time performance. In this paper, the YOLOv5s model is applied in an intelligent damage determination system to detect and count exterior vehicle parts accurately. It utilizes its small size, high speed, and high accuracy, thus providing important information for damage determination and location output.

YOLOv5 comprises four parts with special functions: Input, Backbone, Neck, and Head. The Input side usually contains an image pre-processing stage, i.e., scaling the input image to the input size of the network and performing operations such as normalization. Mosaic data enhancement operations are mainly applied to improve the model's training speed and the network's accuracy, and adaptive anchor frame and adaptive image scaling methods are proposed.

2) *Logical Decision Models:* The target detection model obtains the number of vehicle appearance parts, such as wheels, windows, lights, etc. The logical decision model is mainly based on the more near and less far principle, i.e., if the vehicle is close, the number of the parts mentioned above is less. If the vehicle is far, the number of the abovementioned parts is more. Specifically, suppose the number of parts is less than a. In that case, the number of parts is determined to be too close to the vehicle and needs to be moved away from the vehicle. For example, a corresponding message can be displayed to the user: "Please move away from the vehicle." Suppose the number of parts is more significant than b. In that case, it is determined that the vehicle is too far away and needs to be approached, and again, a corresponding message is displayed to the user, e.g., "Please approach the vehicle." If the number of parts is within the range [a, b], the distance is judged to be appropriate, and the user is allowed to take pictures, and the user can be prompted with a message such as: "Distance is appropriate, you can take pictures."

For the specific values of a and b, we counted the number of exterior vehicle parts in each video frame. We analyzed them against their shooting distance to obtain a suitable range. The specific steps are as follows:

- Obtain the number of vehicle appearance parts detected by the YOLOv5 model and record them.
- Obtain the actual shooting distance between the vehicle and the camera, which is calculated using information such as the depth sensor, the camera's viewpoint, and position.
- For each video frame, the number of vehicle appearance parts and the shooting distance are recorded, and the distribution of the number of different parts and the distance is analyzed through a box line diagram.
- Based on the analysis results, the appropriate [a,b] range is determined so that the distance too close or too far can be accurately judged, and the corresponding alert message is eventually displayed to the user.

B. *Dataset construction*

There is no dataset for vehicle appearance parts in the open-source dataset, so we built a set of vehicle appearance parts datasets for this paper [6]. In this paper, part of the original data comes from the vehicle vehicles taken independently. Part of it comes from the authentic accident images kept by insurance companies in the actual damage claim process, covering as many natural scenes as possible, such as blurred vehicle images, clear vehicle images, dark vehicle images, distant vehicle images, and closed vehicle images, as shown in Fig.3. A total of 2112 images were collected as data to be annotated.

Figure3: Image schematic diagram.

1) Data Annotation: As shown in Fig.4, we use libeling tool to annotate the vehicle appearance parts with rectangular boxes and save them directly as txt tag files available in YOLOv5. The labeled data covers the vehicle's appearance under different angles and lighting conditions to simulate the actual damage determination scenario.

Figure4: Schematic diagram of Lalelimg data annotation Photograph.

2) Data Augmentation: In this paper, to address the situation of insufficient data, firstly, standard data enhancement methods such as image flipping, image rotation, and image cropping are applied [7]. In addition, this method of Mosaic data enhancement is used to enrich the data set, which can add many small targets and enhance the robustness of the network. The Mosaic data enhancement algorithm combines multiple images into a single image by random scaling and random arrangement so that the model can recognize targets in a smaller area. The Mosaic data enhancement algorithm refers to the CutMix data enhancement algorithm. The CutMix data enhancement algorithm uses two images for stitching. In contrast, the Mosaic data enhancement algorithm generally uses four for stitching, but the principles of the two algorithms are very similar. The image enhancement example as shown in Fig.5 below.

Figure 5: Data enhancement image flip diagram Photograph

III. EXPERIMENTAL DESIGN AND ANALYSIS OF RESULTS

In this paper, the YOLOv5 model is trained using annotated datasets, and the parameters of the model are adjusted to improve its detection accuracy. After the training, the relationship between the number of exterior vehicle parts and the distance is counted to infer the suitable shooting distance range according to the principle described in the previous paper. Finally, the vehicle images taken using the intelligent damage determination system were verified by the detection results of vehicle appearance parts and the actual shooting distance.

A. *Experimental design*

The hardware devices used in this paper are NVIDIA GeForce GTX 1080 graphics card, Ubuntu 20.04 operating system, and Python 3.7 programming language.

In carrying out model training, from the perspective of real-time and accuracy of vehicle appearance parts detection, this paper applied YOLOv5s, YOLOv5m, YOLOv5L, and

YOLOv5x to conduct experiments, and through repeated training and tuning, models with better performance on vehicle appearance parts detection tasks were obtained. The experimental results are shown in Tab.1 below.

TABLE I. YOLOV5 MODEL PREDICTION RESULTS.

Model	*YOLOv5s*	*YOLOv5m*	*YOLOv5L*	*YOLOv5x*
mAP50	94.1%	96.8%	97.9%	98.2%
mAP	78.5%	85.8%	89.0%	90.6%
Inferred time	98ms	224ms	430ms	766ms

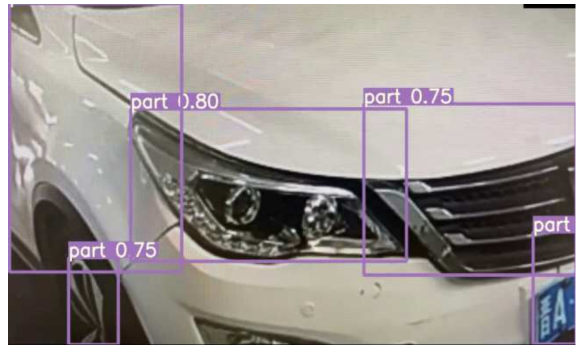
Considering the model accuracy and the real-time detection efficiency of the mobile model, YOLOv5s was finally selected as the critical frame selection model of the intelligent damage-fixing system.

B. *Analysis of results*

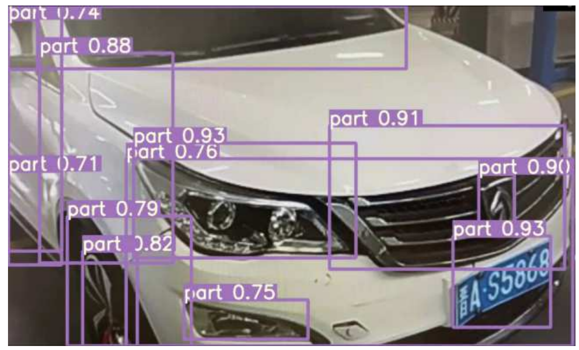
In our experiments, we selected a representative set of vehicle videos and processed these videos with the trained YOLOv5 model. With the model's output results, we counted the number of vehicle appearance parts in each video frame and analyzed them in comparison with their shooting distances. Based on these results, the values of a=8, b=18 were determined as the appropriate range of shooting distances to achieve clear and damage-determining images as crucial frame selection. These keyframes will provide a reliable basis for the subsequent damage determination and position output.

The effect of the vehicle appearance parts detection model is shown in Fig.6 below. In example (a), the number of vehicle appearance parts is 26, the distance is far from the vehicle, and the intelligent damage system can not determine the type of damage; in example (b), the number of vehicle appearance parts is 1, close to the vehicle, the intelligent damage system can not determine the vehicle appearance parts where the damage is located. Finally, in Example (c), when the number of vehicle appearance parts is 14, the distance is suitable, and the intelligent damage system can determine the type of damage and the output of the vehicle appearance parts where the damage occurred.

(a): Too far away, Damage not recognized.



(b): Distance too close, parts not recognized.



(c): Moderate distance, Parts identifiable Damage identifiable.
Figure 6: YOLOv5s detection rendering.

IV. SUMMARY

The experimental results verify the effectiveness of the critical frame selection method based on the YOLOv5 model in the vehicle intelligent damage determination system. By reasonably selecting the shooting distance range, we obtained images with high definition and judgment capability, laying the foundation for accurate determination and location output of vehicle damage. This method has a vast application potential and can provide a more efficient and accurate solution for vehicle damage determination.

REFERENCES

[1] Patil, Kalpesh, et al. Deep learning based car damage classification. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2017. p. 50-54.

[2] Redmon, Joseph, et al. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 779-788.

[3] Redmon, Joseph; Farhadi, Ali. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7263-7271.

[4] Redmon, Joseph; Farhadi, Ali. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[5] Bochrovskiy, Alexey; Wang, Chien-Yao; LIAO, Hong-Yuan Mark. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

[6] Simonyan, Karen; Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[7] Shorten, Connor; Khoshgoftaar, Taghi M. A survey on image data augmentation for deep learning. Journal of big data, 2019, 6.1: 1-48.