Surface Defects Detection of Car Door Seals Based on Improved YOLO V3

Bo Lv\textsuperscript{1,a}, Nanfeng Zhang\textsuperscript{2,b}, Xintao Lin\textsuperscript{3}, Yanxi Zhang\textsuperscript{4}, Tianfen Liang\textsuperscript{5}, Xiangdong Gao\textsuperscript{6}\textsuperscript{*}

\textsuperscript{1}Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology, Guangzhou, China
\textsuperscript{2}Huangpu Customs Technical Center, Dongguan, China
\textsuperscript{3}Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology, Guangzhou, China
\textsuperscript{4}Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology, Guangzhou, China
\textsuperscript{5}Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology, Guangzhou, China
\textsuperscript{6}Guangdong Provincial Welding Engineering Technology Research Center, Guangdong University of Technology, Guangzhou, China

\textsuperscript{a}1418899082@qq.com
\textsuperscript{b}253642845@qq.com
\textsuperscript{*}Corresponding author: gaoxd666@126.com

Abstract—At present, the surface defects of car door seals are generally detected by manual operation, which is inefficient and costly. This paper presents a new detection method to find the surface defects of car door seals with the application of modified YOLO V3 algorithm. Firstly, the K-means clustering algorithm was employed to analyze the number and size of the candidate box of the target detection. Then, the feature pyramid was constructed in the first module of the trunk feature extraction network. Finally, the attention mechanism SE module was added in the output part of darknet53 and two feature gold towers extraction network of the trunk feature. The improved YOLO V3 algorithm was carried out on the defect data set of automobile door seals. The results indicate that compared with the original YOLO V3 algorithm, the improved YOLO V3 mAP reaches 52.71%, increasing by 3.28%.

1. INTRODUCTION

The task of target detection is to find the location of the region of interest from the image and determine the category of the regional target. Since Geoff Hinton et al. proposed the deep learning algorithm [1], deep learning technology has developed rapidly. Scholars at home and abroad have proposed many target detection algorithms based on deep learning [2-4]. Convolutional neural networks can independently learn the characteristics of the target data set and improve their own models. At present, target detection algorithms based on deep learning can be divided into two categories. The first category is a two-step target detection algorithm, represented by Fast R-CNN, Faster R-CNN, Mask R-CNN [5], etc., which is
divided into two stages. The first stage is using the regional candidate frame (RPN) to extract the target information, and the second stage is using the detection network to determine the location and category of the target. The second category is single-step target detection algorithm represented by SSD (single shot multibox detector), YOLO (you only look once) series [6], etc. Compared with the two-step target detection algorithm, the single-step target detection algorithm reduces the steps of regional candidate frame generation. It is an end-to-end target detection algorithm with faster detection speed.

In this paper, the attention mechanism SE module is introduced in the backbone extraction network of the yolov3 model, and the feature map output by the first residual block and the second residual block is used to construct a feature pyramid network (FPN), and the SE module is added before the input result to detect surface defects of the car door seals.

2. YOLO V3 TARGET DETECTION MODEL

2.1 YOLO V3 network

The network structure of YOLO V3 is shown in Figure 1. The network structure includes three parts: the backbone feature extraction network darknet53, FPN, head (used for target recognition and positioning). Darknet53 contains 5 residual blocks. Each residual block is composed of multiple residual units. The residual operation is performed with two DBL units through input. The DBL unit includes convolution, batch normalization and Leaky ReLU activation function. The FPN structure enhances the features extracted by the backbone, and finally outputs the target feature map contained in the last three times of the backbone feature extraction network’s five down-sampling, so as to achieve the detection of targets on three scales.

![Figure 1. YOLO V3 network](image)

2.2 Improved YOLO V3 network

The improved YOLO V3 network structure is shown in Figure 2. The input image is first subjected to convolution operation, and then transmitted to the SE module for 5 down-sampling operations. Among them, the second down-sampling feature map (104x104) undergoes 6 convolutions after the up-sampling and the first down-sampling feature maps are stacked, the stacked feature maps are then passed to the SE module for processing through 6 convolutions. The above operation is to construct the feature pyramid FPN0. The process of the last three down-sampling results is similar to that of the first two down-sampling. After convolution, up-sampling, stacking, convolution, and SE modules, the feature pyramid FPN1 is constructed. The SE module structure is shown in Figure 3, where U represents the input feature map. First, the input feature map U is globally averaged pooled to obtain a 1x1xC feature map, and then two fully connected layers are processed to obtain the weight of each layer, and finally the weight of each layer with each element of the corresponding channel of the original feature map HxWxC is multiplied to obtain a new feature map . The result of SE module processing is passed to the head for identification.
and positioning. The improved yolo v3 network constructs the feature pyramids FPN0 and FPN1, and introduces the attention mechanism SE module before the residual block operation of the backbone network and after the feature pyramid net operation, and realizes the recognition and positioning of targets at four scales of 13×13, 26×26, 52×52, and 208×208. The added 208×208 has a larger shape size than the other three scales of the original network, and contains more target information of the image.

**Figure 2. Improved YOLO V3 network**

### 3. DATA SET OF SURFACE DEFECTS OF AUTOMOBILE DOOR SEALS

The surface defects of automobile door seals can be divided into points, lines and planes in terms of geometry. Combining the geometric characteristics, manifestations and into 7 types of pits, holes, wrinkles, bulges, cracks, scars, and bulges, as shown in Table 1, and Figure 4, diagrams of different types of defects. By analyzing the data, it is found that the similarity of different types of defects on the leather surface is relatively high, the differences are not obvious, the size of different types and the same type of defects are both large, and most of the defects in the data set are small (the length and width of the defect imaging size are smaller than 10% of the image size), it is a small defect detection.

| No. | Defect category | Characteristic description |
|-----|----------------|---------------------------|
| 1   | Pit            | Point-type or planar, irregularly shaped pits |
| 2   | Hole           | Planar, large size, with obvious holes visible to the naked eye |
| 3   | wrinkle        | Planar, large area, complex features |
| 4   | bulge          | Planar or point, generally conical |
| 5   | crackle        | Linear, generally longer in size |
| 6   | scar           | Planar, generally round or polygonal, the surface is dark gray |
| 7   | swelling       | Planar, generally larger in volume than convex, and the image is not obvious |
4. EXPERIMENT AND RESULT ANALYSIS

4.1 Evaluation Index

The result of target detection is generally judged by a confusion matrix. In the confusion matrix, P (Positive) and N (Negative) are used to evaluate the judgment result of the model, that is, positive or negative, and T (True) and F (False) are used to evaluate whether the judgment result of the model is correct. The formulas for the precision and recall rate defined by the confusion matrix are as follows (1)(2):

\[ \text{Precision} = \frac{TP}{TP+FP} \]  
\[ \text{Recall} = \frac{TP}{TP+FN} \]  

Precision indicates that the part that the classifier considers to be a positive class and is indeed a positive class accounts for the proportion of the classifier that is considered to be a positive class, and the proportion of a positive sample that is actually a positive sample; Recall means that the part that the classifier considers to be a positive class and is indeed a positive class accounts for all that is indeed a positive class proportion. Recall rate is a measure of coverage, which measures how many positive samples are divided into positive samples. The precision and recall rate vary with the detection threshold. The PR coordinate system is established with the recall rate as the abscissa and the precision as the ordinate, and the area under the PR curve is defined as AP (Average Precision). The larger the value of AP, the better the detection effect of a single category. mAP (Average Precision) is obtained by averaging...
the AP values of multiple targets. The larger the value of mAP, the better the comprehensive detection effect of the model on all types of targets.

4.2 Result analysis

The experimental hardware is a Dell Inspiron 7500 computer, and the hardware configuration includes Intel-COREi7-9th, NVIDIA GeForce GTX1050GPU, 16G memory. The software used are python3.6, pytorch2.0-gpu deep learning framework. The data set comes from defective products in the production process of a company in Guangzhou, a total of 2565 images are collected (70% for training, 30% for validation), which are labeled using K-means clustering analysis on the labeled candidate boxes, the 12 candidate box sizes are: (20,25), (24,42), (30,30), (37,41), (40,60), (45,101), (52,70), (57,32), (60,53), (71,82), (99,50), (106,122). The experimental results are shown in Figure 5 and Table 2. It can be seen from the table that the AP of the scar and pit defects of the improved yolov3 has been slightly reduced, and the APs of other types of defects have been improved to varying degrees. The mAP of the improved yolov3 model has increased by 3.28% compared with the original model.

5. CONCLUSION

Manual detection of surface defects of automobile door seals is easily affected by factors such as environment and subjective feelings, and has disadvantages such as high cost and low efficiency. In this paper, in order to automatically detect the surface defects of automobile door seals, a certain number of defect images of automobile door seals were collected, then the defects from multiple angles was classified, and an improved method of YOLO V3 was proposed. The experimental results show that the improved yolov3 algorithm has significantly improved the detection effect of most targets, and the model mAP is increased by 3.28%. However, the improved yolov3 algorithm has a large increase in parameters and a decrease in detection speed, which is difficult to meet the practical application of engineering. In the future, simplification the network optimization parameters can be tried to improve the detection speed, and the deformable convolution and dilated

![Figure 5. Detection results. (a) hole (b) wrinkle (c) scar (d) swelling (e) bulge (f) pits (g) crackle](image-url)
| Detection algorithm | scar | pit | bulge | swelling | hole | wrinkle | crackle | mAP/% |
|---------------------|------|-----|-------|----------|------|---------|---------|-------|
| Yolov3              | 90   | 67  | 49    | 47       | 44   | 34      | 15      | 49.43 |
| Improved yolov3     | 88   | 59  | 56    | 54       | 45   | 43      | 24      | 52.71 |

convolution can be increased to expand the receptive field to improve the detection accuracy. In the case of improving accuracy, ensuring real-time detection is still the main direction of future research.

ACKNOWLEDGMENT
This research is financially supported by the Guangzhou Municipal Special Fund Project for Scientific and Technological Innovation and Development (202002020068, 202002030147).

REFERENCES
[1] Ju Moran, Luo Haibo, and Wang Zhongbo, He Miao, Chang Zheng, Hui Bin. Improved yolo v3 algorithm and its application in small target detection. Acta Optica Sinica, 2019,39(7),0715004.
[2] Hua Xia, Wang Xingqing, Wang Dong, Ma Zhaoye, and Shao Faming. Multi-objective detection of traffic scenes based on improved ssd. Acta Optica Sinica,2018,38(12),1215003.
[3] Wang Wenxiu, Fu Yutian, and Dong Feng, Li Feng. Infrared ship target detection method based on deep convolution neural network. Acta Optica Sinica, 2018,38(7),0712006.
[4] Luo Haibo, Xu Lingyun, Hui Bin, and Chang Zheng. Status and prospect of target tracking based on deep learning. Infrared and Laser Engineering,2017, 46(5), 502002.
[5] Lan Zhuo, Hong Yang, and Li Yuan. An improved YOLOv3 method for PCB surface defect detection. 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). IEEE.2021 Shenyang, China,22 to 24 Jan.
[6] Huang Jiahao, Zhang Haiyang, Wang Lin, Zhang Zilong, Zhao Changming. Improved YOLOv3 Model for miniature camera detection. Optics and Laser Technology,2021,142,107133.