Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **8**, Number **2**, May **2023**

An Improved YOLOv3 Algorithm for Small Target Detection

Hui Lv

School of Electronic, Electrical Engineering and Physics
Fujian University of Technology
Fuzhou, 350118, China
1635518661@qq.com

Zhe Chen

School of Electronic, Electrical Engineering and Physics
Fujian University of Technology
Fuzhou, 350118, China
951411819@qq.com

Shi-Hao Huang*

School of Electronic, Electrical Engineering and Physics
Fujian University of Technology
Fuzhou, 350118, China
haoshihuang@fjut.edu.cn

*Corresponding author: Shi-Hao Huang

Received October 2, 2022, revised December 11, 2022, accepted January 29, 2023.

Abstract. *In order to improve the recognition effect of the model on small targets while improving the detection accuracy of large targets, this paper proposes a YOLOv3-SPP-CBAM-104 model for automotive glass defect detection algorithm. Firstly, a channel feature refinement mechanism is introduced into the model to add a small target detection layer 104×104, then a spatial pyramid structure SPP is added before the 13×13 detection branch and a CBAM attention mechanism is added after the spatial pyramid structure, and finally the prior frame is clustered by a K-means clustering algorithm. The experiments were tested on Pascal VOC2007 dataset and self-built dataset. The experimental results showed that the improved YOLOv3 algorithm reached 72.34% mAP on the Pascal VOC2007 dataset, an improvement of 3.82% compared with the original YOLOv3, and 95.19% mAP on the self-built dataset, an improvement of 3.85% compared with the original yolov3 model. The improved model effectively solves the problem of low detection accuracy for small targets and significantly improves the detection effect for large targets.*
Keywords: Deep learning, Glass defects, YOLOv3 algorithm, channel feature refinement mechanism, spatial pyramid structure, attention mechanism

1. **Introduction.** Object Detection is a popular research effort in the field of computer vision [1, 2, 3], and its main task is to identify all objects of interest in an image and determine their category and location [4, 5, 6]. In current research work, small-target detection is challenging. Defects such as inclusions, bubbles, stains and other defects introduced in the glass production process can lead to serious safety hazards during the

use of automobiles. At the same time, the accurate identification and classification of automotive glass defects is of guiding significance for production. Therefore, the detection of automotive glass defects is very important. At the same time, smaller defects in glass are one of the difficulties in detection.Because small targets make up a smaller portion of the image, small targets have less impact on overall loss. Therefore, it is easy to be ignored as the training progresses. At the same time, small targets have higher requirements for noise immunity [7, 8], for example, images obtained in harsh environments; The low-quality images caused by the size, position, shape, and the shooting method of the photographer at the time of the drawing will significantly increase the probability of missing and mischecking of small targets, so improving the effectiveness of feature extraction and aggregation for small target detection is an important issue in the study of object detection [9, 10, 11, 12].

With the rapid development of deep learning [13, 14, 15, 16], breakthroughs have been made in object detection. Object detection algorithms based on Convolutional Neural Networks (CNN) can be roughly divided into two types, namely one-stage object detection algorithm and two-stage object detection algorithm [17]. The detection speed of the first-stage object detection algorithm is faster, but the detection accuracy is lower, and the detection accuracy of the second-stage object detection algorithm is significantly higher than that of the first-stage object detection algorithm, but its detection speed is slower [18]. The classic models of the one-stage object detection algorithm are the YOLO [19] series model and the SSD model [20]. For the YOLOv3 algorithm, the loss of shallow detail information results in low accuracy in small object detection.

Therefore, many scholars have proposed that increasing the depth of the network can be more effective in solving this problem [21, 22]. In fact, increasing the depth of the network is essentially to enhance the characterization capabilities of the object detection model [23]. For existing machine vision, object detection models are evaluated against the image as a whole. Therefore, the detection accuracy of large targets is generally better, and the detection accuracy of small and medium targets is poor. And YOLOv3 must enter a fixed-size image, for large-size images, it will lose some of the relevant information and cause a decrease in detection accuracy. In view of the above problems, this paper proposes to use the object detection algorithm of YOLOv3-SPP-CBAM-104. Its main contributions are summarized as follows:

1. Proposed a target detection algorithm to improve the YOLOv3 model. First, a feature refinement mechanism is introduced to increase the detection rate of small targets by connecting shallow detail information with deep semantic information. Secondly, by adding the SPP layer, the feature map of different sizes can output a vector of a fixed length after passing through the SSP layer. This fixed-length vector is then fed into the fully connected layer for subsequent classification detection tasks. Finally, the CBAM attention mechanism is added, and the focus area is paid more attention to during the training process, thereby improving the detection accuracy of the target

2. Pre-experimental training and testing on the PASCAL VOC2007 dataset. Through the model comparison experiment and ablation experiment, the results show that the object detection model proposed in this paper is 3.82% higher than the original YOLOv3 model on mAP, and the test results are visualized

3. The experiment was tested on a self-built glass defect dataset, and the experimental results showed that the mAP was improved by 3.85% compared with the original model, and finally part of the test results were displayed

The first section of this article mainly introduces the detection of small objects and related work in the field of deep learning. The second and third sections mainly introduce the target detection model proposed in the experiment. Sections 4 and 5 compare the

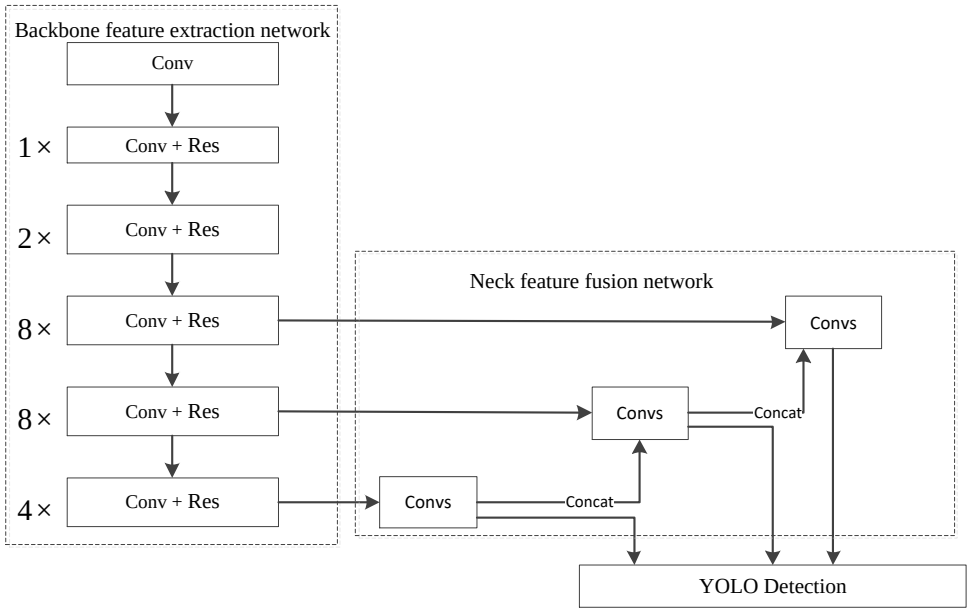


FIGURE 1. YOLOv3 model structure.

proposed algorithm with the current mainstream algorithms on typical datasets, and train and test on self-built datasets and show the results of experiments. Finally, a summary is given.

2. The YOLOv3 algorithm.

2.1. **YOLOv3 model structure.** The results of the YOLOv3 model are shown in Figure 1. The YOLOv3 model consists of three parts: backbone feature extraction network, neck feature fusion network, and head detection network. The backbone feature extraction network is Darknet53, which contains 53 convolutional layers, and downsamples are achieved by changing the size of the convolutional kernel step; In the neck feature fusion network, the feature layer and the neck network in the backbone deep network are fused by upsampling and splicing operations, and three predicted feature branches are obtained.

In the prediction part of the model, the YOLOv3 algorithm first extracts the features of the input $S \times S$ (S takes a value of 13, 26, 52), and each feature map contains three prior boxes to predict large targets, medium targets and small targets respectively. When the center point of the target object falls into the feature map, the object is predicted. After obtaining the confidence score of each feature map, the detection box below the set threshold is excluded by non-maximum suppression, and the final bounding box is obtained, and the prediction result includes predicting the target object category, the target bounding box and the confidence level.

2.2. **Optimized the YOLOv3 algorithm.** The improved model is shown in Figure 2. In the backbone feature network, the image size of 416 × 416 is entered into the Darknet53 network, and after 53 convolutional layers are extracted for feature extraction, and after 5 downsampling, the image size is 208×208, 104×104, 52×52, 26×26, 13×13. In the neck feature fusion network, the feature map of 13×13 is directly output as the prediction feature layer, and after upsampling, it is spliced with the 26×26 feature map in the trunk feature network to obtain the predicted feature layer of 26×26; The feature map of 26×26 is upsampled and spliced with the 52×52 feature map in the trunk feature network to

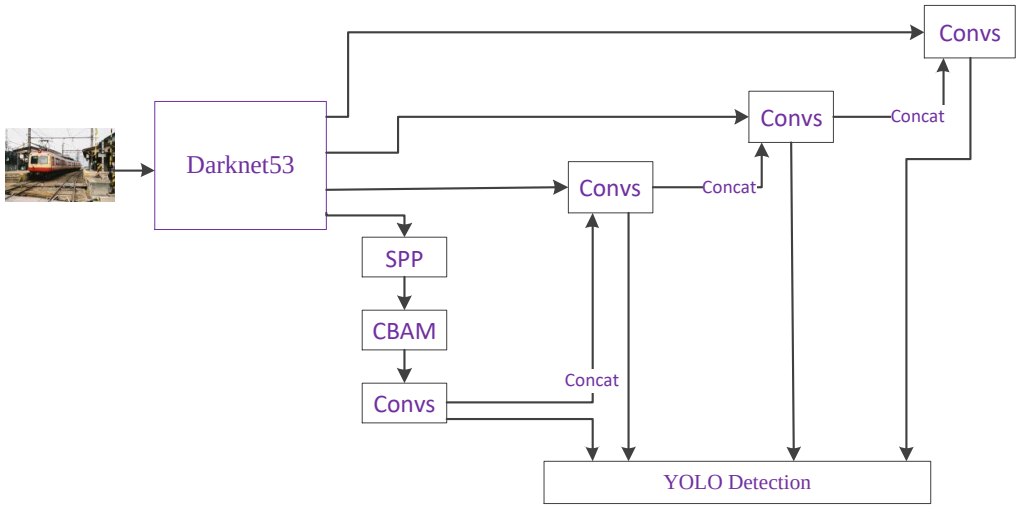


FIGURE 2. Improved model network structure.

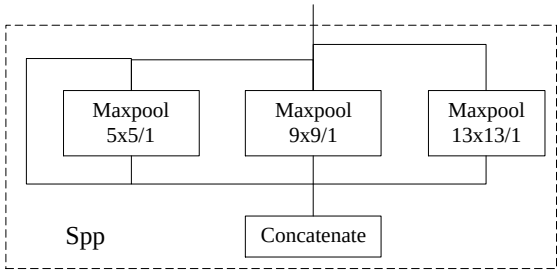


FIGURE 3. SPP structure.

obtain a predicted feature layer of 52×52; The feature map of 52×52 is spliced with the feature map of 104×104 in the trunk feature network by upsampling and splicing the feature map of 104×104 to obtain a predicted feature layer of 104×104. In the improved model, the semantic information of deep features is rich for detecting large targets, and the position information of shallow features is more suitable for the detection of small targets.

Between the trunk feature network and the neck feature fusion network, the SPP (Spatial Pyramid Pooling) module is added, [24, 25] with the SPP structure shown in Figure 3, the SPP structure. In the improved model, an SPP structure is added between the fifth and sixth convolutional layers before the 13×13 prediction feature layer, and the feature map is input to the SPP module to achieve local and global feature fusion in the feature map.

The convolutional attention mechanism (CBAM) [26, 27] is shown in Figure 4. CBAM is composed of channel attention module and spatial attention module, compared to other attention mechanisms, CBAM can be weighted sum of the input picture information in

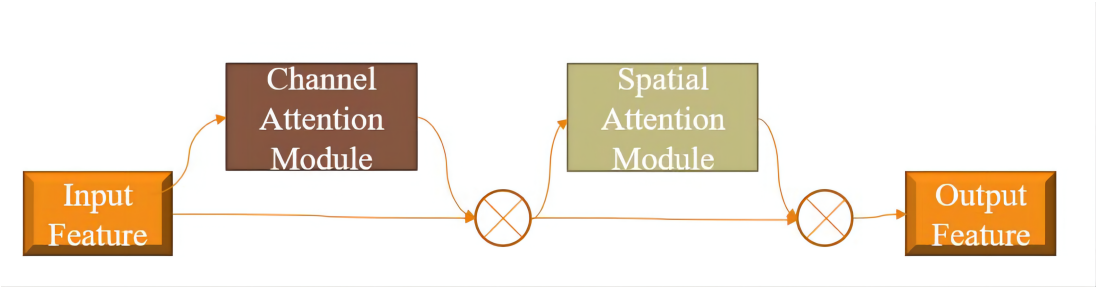


FIGURE 4. CBAM attention mechanism.

two different dimensions, get more dimension information, CBAM is a general module, can be seamlessly connected to the convolutional network.First of all, the input picture is divided into two branches, one enters the channel attention module, gets the channel attention weight, and then performs elementwise multiplication operation with the other branch to get the input of the space attention module, and then divides it into two branches, one into the space attention module, the space attention weight, and the other branch to do multiplication operations to get the final picture feature value.

In the neck feature fusion network, the CBAM (Convolutional Block Attention Module) attention mechanism was progressively embedded in front of the three prediction scales [28], CBAM was added before the detection layer of 13×13 as YOLOv3-a, CBAM was added before the detection layer of 13×13, 26×26 as YOLOv3-b, and CBAM was added before the three detection layers as YOLOv3-c. The experiment used the PASCAL VOC2007 dataset to test the three models, the experimental results are shown in Table 1, adding CBAM to the prediction branch of 13×13 is the best, the mAP reaches 70.18%, which is 1.66% higher than the original YOLOv3, and Recall is also improved than the original version, so the experiment adds CBAM before the first detection layer of YOLOv3.

TABLE 1. CBAM test results at different positions

Model	mAP/%	Recall/%
YOLOv3	68.52	69.97
YOLOv3-a	70.18	70.76
YOLOv3-b	69.37	71.10
YOLOv3-c	68.54	70.06

2.3. **K-means clustering algorithm improvement strategy.** The experiment uses K-means clustering algorithm to cluster the size of the prior box, first divide the labeled target box into k classes, then calculate the center point of each class, and then calculate the European distance between each target box and k center points, by

comparing the distance size, the target box is divided into the category with the smallest distance, through continuous iteration, find the optimal k center points, and finally the classification ends, and the clustering result is shown in Figure 5.

The experiment increased the a priori boxes of 9 scales in the original YOLOv3 to 12, and added the 3 prior boxes of the fourth predictive feature map to detect small targets, and the accuracy rate (ACC) of K-means clustering reached 71% of the prior box size as shown in Table 2.

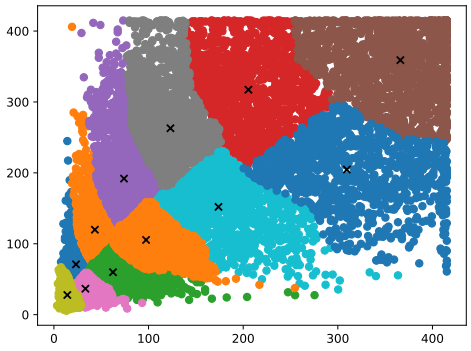


FIGURE 5. K-means clustering results.

TABLE 2. Priori box sizes

Feature diagram	priori box			ACC
13×13	189×316	302×208	359×363	0.71
26×26	58×173	165×268	105×241	
52×52	64×60	38×105	92×109	
104×104	14×26	37×35	21×66	

3. **Experiments.** The software environment used in the experiment is Windows 10 operating system, the development language is phython, the development environment is anaconda, the deep learning framework is tensorflow, and the pycharm software is used for experimental training. The hardware environment is the 11th Gen Intel(R) Core(TM) i7-11700K @ 3.60GHz, 16GB of MEMORY, and the display adapter is the NVIDIA GeForce RTX 3060 Ti.

3.1. **Evaluation indicators.** The YOLOv3 model uses two metrics, accuracy rate (P) and recall rate (R), to measure the stability and quality of the model [29]. The expression for P, R is:

$$\mathrm{P} = TP/(TP + FP) \tag{1}$$

$$\mathrm{R} = TP/(TP + FN) \tag{2}$$

Wherein: TP is the positive class, predicted as the number of positive classes, FP is the negative class, predicted as the number of positive classes, FN is the positive class, predicted as the number of negative classes. The accuracy rate is the proportion of the sample that is judged to be positive and the proportion of the sample that is predicted to be positive, and the recall rate represents the proportion of the sample that is judged to be positive in the sample to the actual positive sample. The accuracy is too large, the recall rate is too small or the accuracy rate is too small, the recall rate is too large, and the test results will have an impact, so the average accuracy mean is used as an indicator to measure the model standard, and the average accuracy AP (area under the PR curve) for each defect is calculated first, and then the average accuracy mean mAP (mean Average Precision) is calculated.

3.2. **Analysis of Pascal VOC2007 experimental results.** The experiment first used the Pascal VOC2007 public dataset to verify the effectiveness of the improved model, including 20 categories, a total of 9963 pictures, including 5011 training sets and 4952 test sets. The number of epochs in the experiment is set to 100, the momentum coefficient

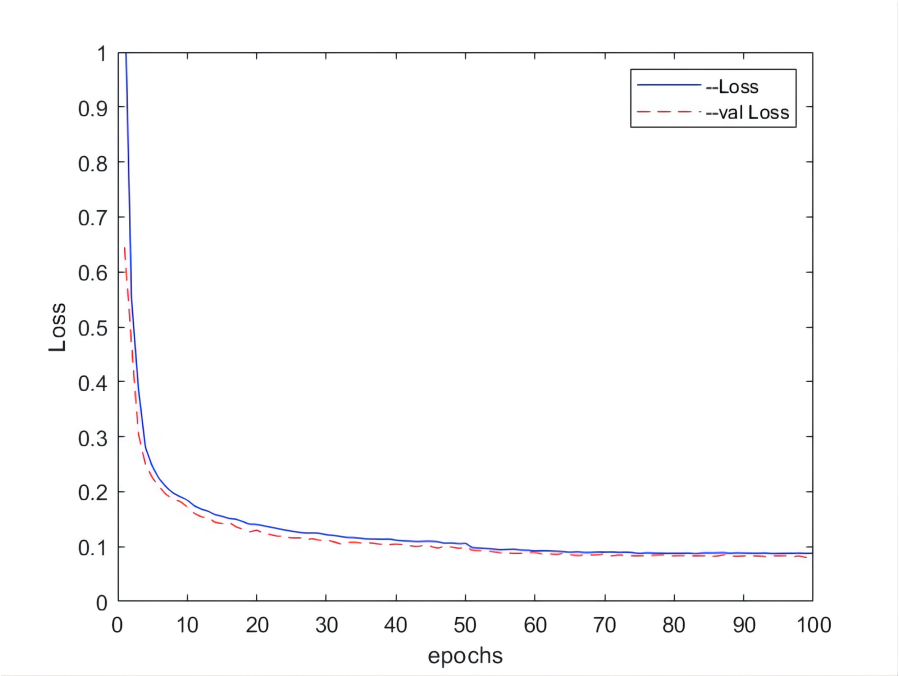


FIGURE 6. Loss map.

is 0.9, the decay value is 0.0005, the initial learning rate is 0.001, the total loss value curve of the optimized model training and the loss curve of the validation set are shown in Figure 6, when the epochs reach about 80 times, the loss value tends to be stable, the total loss drops to about 0.087, and the verification set loss drops to about 0.082.

In order to further verify the effectiveness of the model for small object detection, the YOLOv3-SPP-CBAM-104 (YOLOv3-SC104) model is compared with the SSD, two-stage classic model Faster R-CNN, and the new object detection algorithm YOLOv4 [30] model. Five models were trained on the Pascal VOC2007 dataset, mAP was selected, recall indicators were compared, and the experimental IOU value was set to 0.6, and the test results were shown in Table 3.

TABLE 3. Model comparison results

model	Backbone	Size	mAP/%	Recall/%
YOLOv3	Darknet53	416×416	68.52	69.97
SSD	VGG-16	300×300	65.54	68.89
Faster R-CNN	Resnet50	600×600	71.75	81.03
YOLOv4	CSPDarknet-53	416×416	71.77	69.52
YOLOv3-SC104	**Darknet53**	**416×416**	**72.34**	**75.05**

As can be seen from Table 3, compared with SSD, Faster R-CNN and other models, the improved model has achieved greater advantages in accuracy or recall, compared with the original YOLov3, the mAP value of the improved model has reached 72.34%, an increase of 3.82%, and the recall rate is 75.05%, an increase of 5.08%. Through data comparison, the improved model is more suitable for the detection of small targets. Some images selected from the test results are visualized in the experiment, as shown in Figure 7. The first is the original image, the second is the YOLOv3 recognition result, and the third is the recognition result of the improved model. It can be seen that the detection

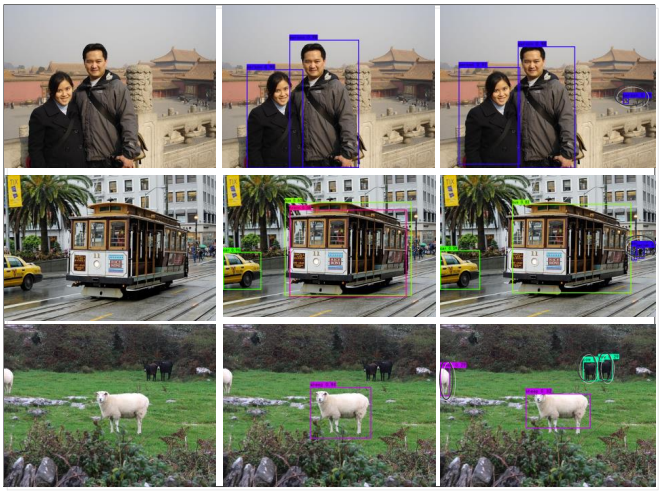


FIGURE 7. Visualization of test results.

ability of the improved model for small targets is significantly improved, and the detection confidence for large targets is also increased.

3.3. **Ablation experiments.** In order to verify the effectiveness of the improved model in detecting small targets, the experiment performed ablation experiment on the public dataset Pascal VOC2007, and introduced three improved methods of spatial pyramid structure SPP, channel feature refinement mechanism and CBAM attention mechanism on the basis of the YOLOv3 model, namely model YOLOv3-1, YOLOv3-3, YOLOv3-2, by comparing the difference in average detection accuracy, recall rate and FPS three indicators. After analyzing the performance of the three improvement points of the optimized model, the YOLOv3-4, YOLOv3-SC104 model is to gradually introduce the improvement method on the basis of YOLOv3, and the ablation experimental results of the model are shown in Table 4.

TABLE 4. Results of ablation experiments

model	SPP	CBAM	Channel feature refinement mechanism	mAP/%	Recall/ %	FPS
YOLOv3				68.52	69.97	16.3
YOLOv3-1	✓			70.71	50.11	15.1
YOLOv3-2		✓		70.18	70.76	15
YOLOv3-3			✓	68.83	61.67	15.9
YOLOv3-4	✓	✓		71.80	71.16	14.56
YOLOv3-SC104	✓	✓	✓	72.34	75.05	14.2

In terms of verifying the effectiveness of SPP, the experiment takes YOLOv3 as the basis model, and adds the SPP module before the 13×13 feature map to enhance the small target feature information in the prediction feature layer, and the mAP value of the model is increased by 2.19%, which proves that after passing the SPP module, the prediction feature map contains more parameter information useful for small target detection.

In order to verify the effect of introducing channel feature refinement mechanism on model detection, based on the original model of YOLOv3, a 104×104 predictive feature layer is added to the neck feature fusion network, and the small target information in the shallow network is fused. Compared with YOLOv3, the mAP of the YOLOv3-104 model increased by 0.31%, which proved that the mechanism of increasing channel feature refinement makes the predicted feature layer after feature fusion contain more features of small targets.

In order to verify the effectiveness of the CBAM attention mechanism on the model, CBAM was added to the 3 neck feature fusion branches of YOLOv3, which increased the mAP by 1.66% compared with the YOLOv3 model, which proved that after joining the CBAM, more information on the space and channel of the target object in the picture was obtained, and more attention was paid to the characteristics of the target.

In order to verify the effect of the introduction of the SPP module and the CBAM attention mechanism in the model, such as the YOLOv3-4 model in Table 5, the accuracy is significantly higher than that of the YOLOv3-1 and YOLOv3-2 models through the comparison of mAP and Recall indicators, which is 3.28% higher than that of the original YOLOv3 and YOLOv3-4 models, which shows that after passing the SPP module, the information of the small target in the feature map is passed into the CBAM attention mechanism. Small target features are further extracted in spatial and channel dimensions, improving accuracy at the expense of detection time.

YOLOv3-SC104 model structure added SPP module, the introduction of channel feature refinement mechanism and the addition of CBAM attention module, compared with the original YOLOv3 model, more useful small target information was extracted, mAP increased by 3.82%, it can be seen that the improved model is more suitable for small object detection. As the improved model structure becomes more complex, the number of images detected per unit of time is reduced.

3.4. **Self-built dataset detection.** The experimental data set comes from the defect detection project of Shenzhen Zhongyanchuang Technology Co., Ltd., and the sample used in the experiment is the windshield of Fuyao Automobile. Stain,Scratches, Ston, inclusions, bubbles, some glass defects as shown in Figure 8.

In the experiment, the image is normalized to 416×416, and labelImg is used to label the data set. Since the target detection technology requires a sufficient amount of data set to achieve a high enough accuracy of the model, the labeling of the data set takes a lot of time. In order to solve this problem, this paper draws on the idea of pseudo-label generation method, uses the weights trained by small samples to automatically label new samples, realizes data expansion, and makes the data set more diverse through data enhancement. The time required for dataset labeling at training time is greatly reduced by using pseudo-label generation. Finally, the model is obtained by training the obtained pseudo-label data set and manually labeled samples as a self-built data set. The process of obtaining pseudo-label samples is shown in Figure 9.

The final experimental data set is 8814, including 7190 training sets, 771 test sets, and 853 validation sets. Table 5 shows the final data set.

Targets whose dimensions are less than one-tenth of the original drawing are considered small targets. Bubbles and spots contain a large number of small targets, few features, difficult to detect, and the two defect features are relatively similar, it is easy to misexamine; The proportion of furuncles and inclusions in the original figure is large and the characteristics are obvious; In the original drawing, the scratches are slender and have a large span, which is clearly distinguished from other defects. After using the Pascal VOC2007 dataset to verify the effectiveness of the improved model, the experiment

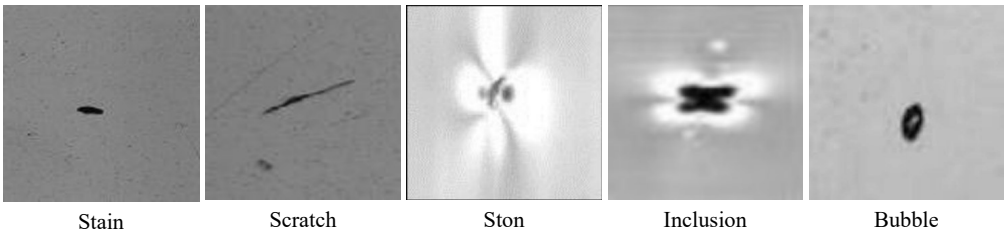


FIGURE 8. Picture of glass defects.

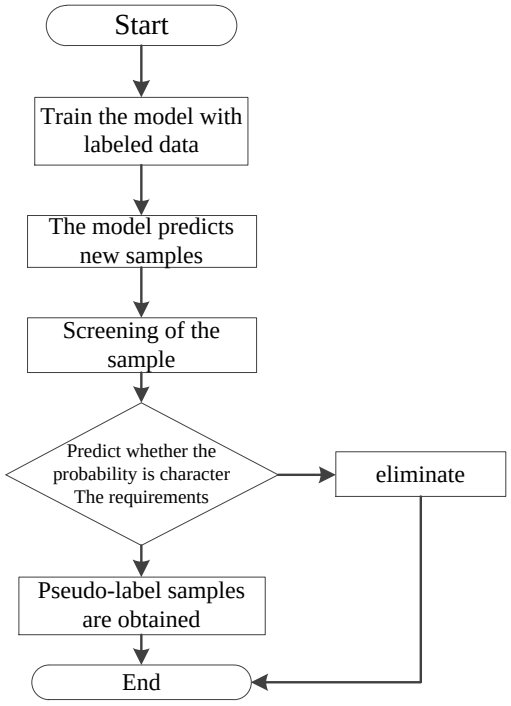


FIGURE 9. Pseudo-label sample acquisition.

TABLE 5. Number of images for each type of defect in the dataset

The type of defect	Training set/sheet	Validation set/sheet	Test set/sheet
Bubble	1667	186	206
Stain	1446	161	178
Inclusion	1330	148	164
Ston	1526	136	150
Scratch	1221	140	155
Total	7190	771	853

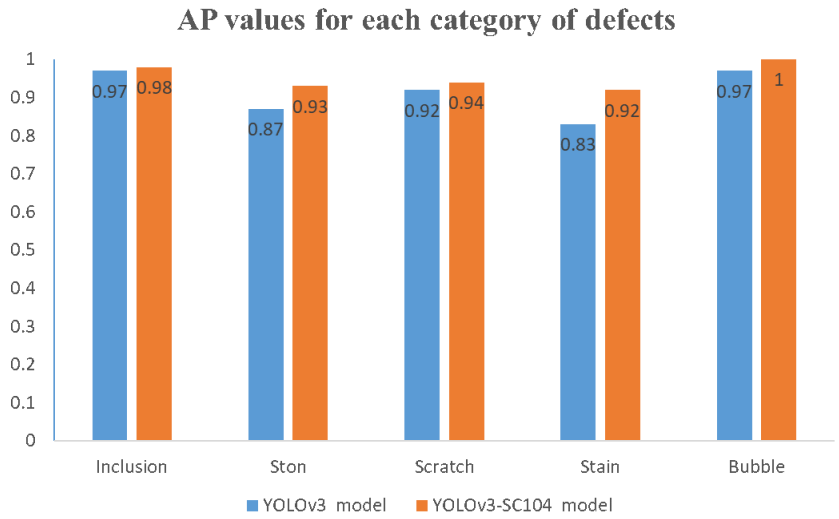


FIGURE 10. AP comparison chart.

trained and tested the self-built glass defect dataset. At the same time, in order to further verify the effectiveness of the optimization model in the detection of automotive glass defects, the mAP values of different models under different iterations are obtained, and the YOLOv3-SC104 model is compared with the original model, at the same time, add the F1 value to evaluate the quality of the model, and the formula for F1 is:

$$\mathrm{F1} = 2 * P * R/(P + R) \tag{3}$$

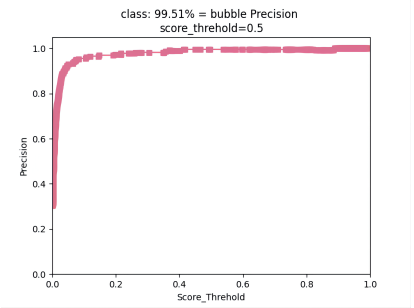
The two models are tested on the self-built data set, and the AP values of each type of defect are shown in Figure 10.

It can be clearly seen from Figure 10 that compared with the original model, the improved model is more suitable for the self-built glass defect data set, the AP value of each type of defect has been improved, and the detection accuracy of the small target defect of the stain is significantly improved. Other data pairs for both models are shown in Table 6.

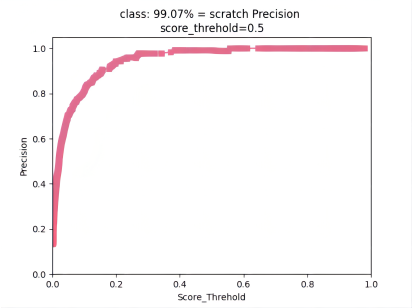
TABLE 6. Model comparison

model	Size	Backbone	mAP/%	Recall%	F1
YOLOv3	416×416	Darknet-53	91.34	74.01	0.83
YOLOv3-SC104	416×416	Darknet-53	95.19	77.21	0.86

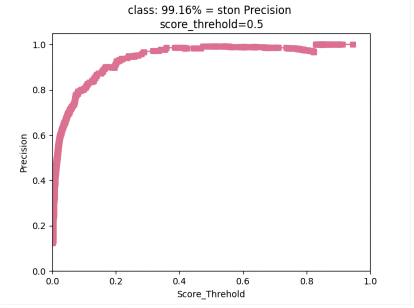
As can be seen from Table 6, the mAP of the improved model is 95.19%, an increase of 3.82% compared with YOLOv3, the Recall and F1 indicators are significantly higher than YOLOv3, and the accuracy of the test set is shown in Figure 11.



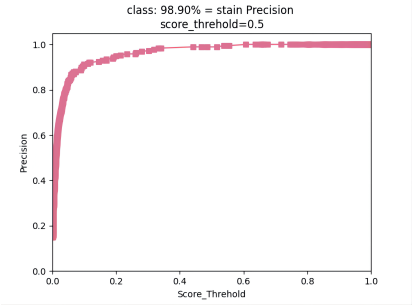
(a) Bubble



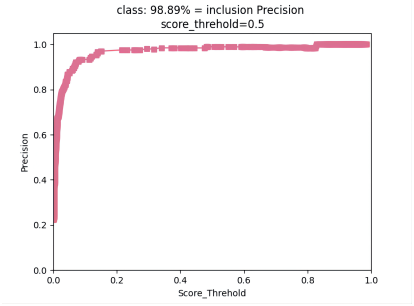
(b) Scratch



(c) Ston



(d) Stain



(e) Inclusion

FIGURE 11. Test set accuracy for the five defects.

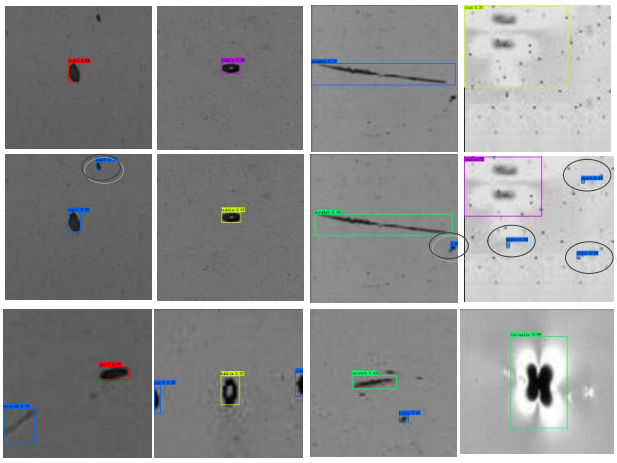


FIGURE 12. Visualisation of model test results.

It can be calculated from Figure 11 that the detection accuracy of the model on the test set is as high as 99.11%. Synthesizing the various indicators of the model on the data set, it can be seen that the use of the improved YOLOv3-SC104 model is more suitable for glass defect detection. The experiment visualizes the test results of the optimization model, as shown in Figure 12, the first behavior YOLOv3 recognition result, the second behavior optimization model recognition result, and the third behavior multi-objective recognition effect.

As can be seen from Figure 12, the detection ability of the improved model to detect small targets is significantly improved, and the target circled in the black circle in the figure is the small target detected by the improved model. The detection ability of large targets has also been significantly improved, and for multi-targets in the test set, the improved model also has a better recognition effect.

4. **Conclusion.** In order to improve the detection ability of YOLOv3 on targets, especially the problem of low accuracy in detecting small targets, the channel feature refinement mechanism, SPP module and CBAM attention mechanism are introduced on the YOLOv3 network structure, and the k-means algorithm is used to redesign 12 prior boxes to identify target defects. Experimental results show that compared with YOLOv3, YOLOv4 and other models, the mAP value of the improved model on the self-built glass dataset reaches 95.19%, which is superior in accuracy, and can better identify small targets, and the detection confidence of large targets is also significantly improved, and the detection effect of multi-target pictures in the test set is better, and the recognition accuracy is very high. This method provides reference value in the detection of automotive glass defects. However, using the improved model for detection, there will still be missed detection and false detection problems, and future work will be carried out in the direction of small object detection using YOLOv3 to overcome the complex background.

REFERENCES

[1] F.-Q. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, Z.-Y. Li, "Human Motion Recognition Based on SVM in VR Art Media Interaction Environment," *Human-centric Computing and Information Sciences*, vol. 9, no. 40, 2019.

[2] F.-Q. Zhang, T.-Y. Wu, G.-Y. Zhang, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP Journal on Image and Video Processing*, no. 58, 2019.

[3] E.-K. Wang, X. Zhang, F. Wang, T.-Y. Wu, C.-M. Chen "Multilayer dense attention model for image caption," *IEEE Access*, pp. 66358–66368, 2019.

[4] R. Girshick, J. Donahue, T. Darrell, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 580-587.

[5] Y. Lecun, Y. Bengio, G. Hinton, "Deep learning," *Computer Science*, vol. 5, no. 3, pp. 28–33, 2015.

[6] L.-D. Guo, S.-F. Ding "Research Progress on Deep Learning," *Computer Science*, vol. 5, no. 42, pp. 28–33, 2015.

[7] Y. Zhang, M.-L. Zhang, X.-L. Lv, C. Guo, Z.-H. Jiang, "Review of Research on Small Target Detection Based on Deep Learning," *Computer Engineering and Applications*, vol. 38, 2022.

[8] Y. Liu, Y.-W. Zhan, "Survey of Small Object Detection Algorithms Based on Deep Learning," *Computer Engineering and Applications*, vol. 2, no. 57, pp. 37–48, 2021.

[9] T.-Y. Huang, X.-J. Yang, G.-H. Xiang, and L. Chen, "Study on Small Target Pedestrian Detection and Ranging Based on Monocular Vision," *Computer Science*, vol. 11, no. 47, pp. 205–211, 2020.

[10] W. Ge, Z.-W. Shi, "Application of Improved YOLOV3 Algorithm in Pedestrian Identification," *Computer Engineering and Applications*, vol. 20, no. 55, pp. 128–133, 2019.

[11] C.-H. Hu, C. Chen, C. He, H. Pei, J.-X. Zhang, "SAR detection for small target ship based on deep convolutional neural network," *Journal of Chinese Inertial Technology*, vol. 3, no. 27, pp. 397–405+414, 2019.

[12] L.-S. Sun, J.-X. Wei, D.-M. Zhu, M. Shi, "Surface Defect Detection Algorithm of Aluminum Profile Based on AM-YOLOv3Model," in *Laser & Optoelectronics Progress*. vol. 24, no. 58, pp. 360–370, 2021.

[13] Y. Ma, Y.-G. Peng, T.-Y. Wu "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121–2133, 2022.

[14] F.-Q. Zhang, T.-Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, L. Liu "Application of quantum genetic optimization of LVQ neural network in smart city traffic network prediction," *IEEE Access*, vol. 8, pp. 104555–104564, 2020.

[15] M.-E. Wu, J.-H. Syu, C.-M. Chen "Kelly-based options trading strategies on settlement date via supervised learning algorithms," *Computational Economics*, vol. 4, no. 59, pp. 1627–1644, 2022.

[16] C.-M. Chen, L.-L. Chen, W.-S. Gan, L. Qiu, W.-P. Ding "Discovering high utility-occupancy patterns from uncertain data," *Information Sciences*, no. 546, pp. 1208–1229, 2021.

[17] L. Sommer, N. Schmidt, A. Schumann, "Search area reduction fast-rcnn for fast vehicle detection in large aerial imagery," *25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 3054–3058.

[18] S. Ren, K. He, R. Girshick, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.

[19] J. Redmon, S. Divvala, R. Girshick, "You only look once:Unified,real-time object detection," *Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[20] W. Liu, D. Ancuelov, D. Erhan, "SSD: single shot MuluBox detector," *European Conference on Computer Vision*, pp. 21–37, 2016.

[21] W.-Y. Liu, J.-M. Peng, B.-L. Wu, T.-F. You, "Improved YOLOv3 life article detection method for sorting," *Transducer and Microsystem Technologies*, vol. 6, no. 41, pp. 134–137, 2022.

[22] K.-Q. Huang, X.-R. Liu, M.-Y. Huang "Research on small target detection method based on improved YOLOv3," *Transducer and Microsystem Technologies*, vol. 4, no. 41, pp. 52–55, 2022.

[23] S.-Y. Xu, K.-B. Chu, J. Zhang, C.-T. Feng "An Improved YOLOv3 Algorithm for Small Target Detection," *Electronics Optics & Control*, pp. 1–8, 2022.

[24] X.-P. Wang, R.-F. Zhang, Y.-H. Liu, J.-H. Huang, Z.-X. Chen, "Improved YOLOv3 Garbage Classification and Detection Model for Edge Computing Devices," *Laser & Optoelectronics Progress*, vol. 4, no. 59, pp. 291–300, 2022.

[25] Q.-Q. Zhang, L.-Z. Liu, J.-M. Ning, G.-D. Wu, Z.-H. Jiang, M.-J. Li, D.-L. Li, "Tea buds recognition under complex scenes based on optimal YOLOV3 model," *Acta Agriculturae Zhejiangensis*, vol. 9, no. 33, pp. 1740–1747, 2021.

[26] F.-Y. Ren, X.-B. Pei, Z. Qiao, Y. Bai, "YOLOv4 Lightweight Detection Method Based on CBAM," *Journal of Chinese Computer Systems*, pp. 1–8, 2022.

[27] J.-W. Bo, C.-T. Zhang, C.-L. Fan, H.-J. Li, "Ore Conveyor Belt Sundries Detection Based on Improved YOLOv3," *Computer Engineering and Applications*, vol. 21, no. 57, pp. 248–255, 2021.

[28] P.-Y. Jiang, Q.-C. Tao, M.-Q. Ai, "STEEL SURFACE DEFECT IMAGE CLASSIFICATION BASED ON ATTENTION MECHANISM AND DEEP LEARNING," *Computer Applications and Software*, vol. 9, no. 38, pp. 214–219, 2021.

[29] X.-P. Liu, Y.-Q. Li, L. Li, Z. Wang, Y. Liu, "Improved YOLOV3 Target Recognition Algorithm with Embedded SENet Structure," *Computer Engineering*, vol. 11, no. 45, pp. 243–248, 2019.

[30] R.-X. Li, X. Lv, Y.-S. Zhang, Y. Li, P. Liu, "Research on Underground Personnel Detection Based on Improved YOLOv4 Algorithm," *Mining Research and Development*, vol. 11, no. 41, pp. 179–185, 2021.