Improvement and Application of YOLOv3 for Smartphone Glass Cover Defect Detection

Yuan Cheng, Jigang Wu, Jun Shaov, and Deqiang Yang

Abstract Smartphone glass covers defects detected by human, which is inefficiency, high costs, low detection accuracy and labour intensive, while the automatic detection methods based on traditional machine vision is poor flexibility, low yield and poor generalisation capability. Therefore, this paper introduces YOLO (You Only Look Once) v3 to smartphone glass cover defects for the first time. The YOLOv3 algorithm was improved for the actual characteristics and specific requirements of defect detection. First of all, the channel attention mechanism SENet (Squeeze and Excitation Networks) was added to the feature extraction network to detect inconspicuous defect features. Moreover, a 104 × 104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects. Finally, the scaling factor coefficient of the BN (Batch Normalization) layer in the convolutional network is used as the important factor for model pruning to improve the defect detection speed. The improved YOLOv3 algorithm is applied to smartphone glass cover defect detection, and a high accuracy and high detection speed method for smartphone glass cover defects is proposed. 15,914 production site images covering four types of defects, including chipped edges, pits point, soiling and scratches, were obtained from smartphone glass cover manufacturers, 14,321 were annotated as the training set and 1593 were used as the test set to compare and analyse the proposed method and the original YOLOv3 algorithm in this paper. These experiments showed that the mAP (mean average precision) of the detection was 81.0% and the detection speed was 43.1 sheets/s. Compared to the original YOLOv3 algorithm, the mAP of the detection increased by 3% and the detection speed increased by 6.7 frames/s, which meets the need for high precision and efficient detection of defects in the industrial production of smartphone glass covers.

Keywords Mobile phone glass covers · YOLOv3 · Channel attention mechanism · Model pruning

Y. Cheng (✉) · J. Wu · J. Shaov · D. Yang
Hunan Provincial Key Laboratory of Health Maintenance for Mechanical Equipment, Hunan University of Science and Technology, Hunan Province, Xiangtan 411201, China
e-mail: chenghnust@163.com

H. Zhang et al. (eds.), *Proceedings of IncoME-VI and TEPEN 2021*,
Mechanisms and Machine Science 117,
https://doi.org/10.1007/978-3-030-99075-6_10

1 Introduction

Smartphone screen is the core key component of human–computer interaction, generally consisting of three parts: display module, touch module and glass covers, the glass cover is located in the outermost layer of the screen, which is the solid shell and touch medium of the screen. With the rapid development of artificial intelligence and the arrival of the 5G era, smartphones have become a necessary tool [1], and people's quality requirements for smartphones are getting higher and higher. Smartphone glass covers inevitably produce all kinds of defects during the production process, such as chipped edges, scratches, soiling, pits point and so on. In order to meet the high quality requirements of users for smartphones, manufacturers must carry out 100% quality checks on smartphone glass covers to a high standard. At this stage of mass production, the recognition system depends on a manual inspection with the aid of such tools as bright lights and magnifying glasses. Restricted by human subjective awareness and experience, this inspection method is characterized by inefficiency, high costs, high false detection rates and labour intensive. Therefore, it is important to study the method of detecting defects in mobile phone glass covers to replace manual labour.

Smartphone glass cover defect detection puts forward the following requirements for automatic detection methods: (1) good flexibility of the detection method, which can adapted to various types of defects; (2) strong generalization ability of the detection method, which can adapted to the characteristics of defect features such as inconspicuous and multi-scale; (3) good real-time detection method, which can adapt to the beat of mass production; (4) high detection accuracy, which can completely replace manual detection. Currently, the defect detection of the smartphone cover glass is investigated with differential image method [2, 3], background elimination method [4] and threshold segmentation method [5, 6] in machine vision. The traditional methods generally only detect defects of a certain type or defects with periodic textures, which cannot meet the requirements of flexible inspection. At the same time, these methods are heavily influenced by noise, resulting in poor detection accuracy. In recent years, deep learning-based target detection algorithms [7–10] have made significant improvements in detection accuracy and efficiency relative to traditional methods by building a variety of different network structures, paired with the use of powerful training algorithms to adaptively learn the representation of high-level semantic information in images [11]. In the study of surface defect detection, the YOLOv3 [12] algorithm has shown better detection accuracy and detection speed. Zhang Guangshi et al. [13] used the YOLOv3 algorithm to detect smear marks and missing defects in gears. Weigang [14] and others used the YOLOv3 algorithm to detect defects such as pressed-in iron oxide, patches and cracks on the surface of strip steel. Hongcai et al. [15] used the YOLOv3 algorithm on pharmaceutical glass bottle defect detection and was able to effectively detect defects such as tube end residue, gas lines, bubbles, scratches, stains and stones on glass bottles. Although the YOLO v3 algorithm can provide flexible detection of surface defects, it requires distinctive defect features and a small scale span, and further improvements are needed in

terms of real-time detection. There is no research on mobile phone glass cover defect detection using the YOLOv3 algorithm.

This article introduces YOLOv3 to the defect of the smartphone glass cover for the first time. In view of the actual characteristics and specific requirements of defect detection, the YOLOv3 algorithm is improved. The channel attention mechanism SENet [16] is added to the feature extraction network to solve the problem of unobvious defect features, a 104 × 104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects, and the scaling factor coefficient of the BN (Batch Normalization) layer of the convolutional network is used as the importance factor for model pruning to improve the defect detection speed. On this basis, the improved YOLO v3 algorithm is applied to the defects of the smart phone glass cover, and a high-precision and high-speed detection method for the defects of the smart phone glass cover is proposed. From the smartphone glass cover manufacturer, 15,914 pictures of the production site covering 4 types of defects such as chipping, pits, dirt and scratches were obtained. 14,321 pictures were marked as training sets and 1593 pictures were used as test sets. The proposed method and the original YOLOv3 algorithm are compared and analyzed.

2 Introduction of YOLOv3

2.1 YOLOv3 Detection Principle

YOLOv3 was proposed by Redmon in 2018, the algorithm works by dividing the image containing the detected target into a $S \times S$ grid, with the width and height of the grid noted as c_x,c_y. When the centre of the target object falls into a grid cell, the coordinates of the relative centroid related to the upper left corner of the grid $(\sigma(t_x), \sigma(t_y))$, as well as the relative width t_w and relative height t_h, would be output by the grid. As a result, the final target prediction frame can be obtained with the actual position, including width and height of the grid [17], which as shown in Fig. 1.

In the figure, the red box is the actual prediction box and the dashed box is the priori bounding box. The center point coordinates b_x, b_y), width b_w and height b_h of the prediction box can be obtained by the arithmetic of YOLOv3, where:

$$b_x = c_x + \sigma(t_x) \tag{1}$$

$$b_y = c_y + \sigma(t_y) \tag{2}$$

$$b_w = p_w e^{t_w} \tag{3}$$

$$b_h = p_h e^{t_h} \tag{4}$$

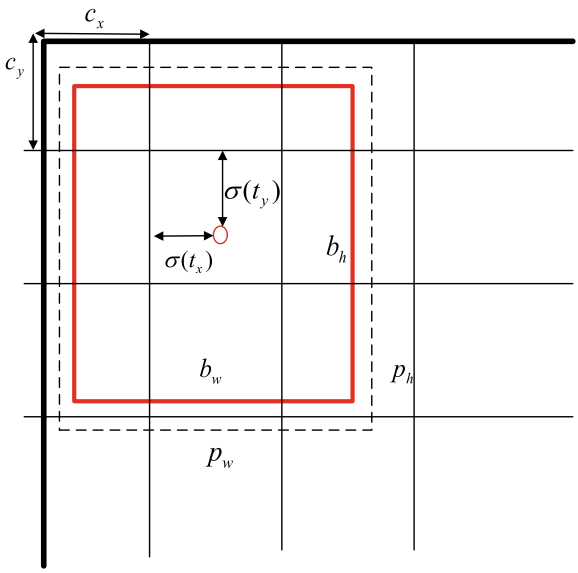


Fig. 1 Schematic diagram of the YOLOv3 prediction box

2.2 YOLOv3 Network Framework

The darknet-53 network is used to the feature extraction network of YOLOv3, and its network structure is shown in Fig. 2.

During the detection process, three feature maps with different scales would be generated by YOLOv3, whose sizes are 13 × 13, 26 × 26 and 52 × 52, with each feature map corresponding to a different field of perception and priori bounding box. The feature map with the size of 13 × 13 has a larger field of perception and the priori bounding box is relatively large, making it suitable for larger objects. The objects of medium size can be detected by the feature map with the size of 26 × 26. The feature map with the size of 52 × 52 has a smaller field of perception and the corresponding a priori frame is relatively small, which is suitable for the detection of small objects.

3 Improvements in YOLOv3

3.1 Channel Attention Mechanism SENet

The importance level of each channel is not considered in the feature extraction process of YOLOv3, which would induce the poor extraction of useful information. Therefore, the channel attention mechanism SENet is incorporated with a view to improving the feature extraction network of YOLOv3 algorithm.

	Type	Filters	Size	Intput	Output
	Convolutional	32	3×3/1	416×416×3	416×416×32
	Convolutional	64	3×3/2	416×416×32	208×208×64
1×	Convolutional	32	1×1/1	208×208×64	208×208×32
	Convolutional	64	3×3/1	208×208×32	208×208×64
	Residual				208×208×64
	Convolutional	128	3×3/2	208×208×64	104×104×128
2×	Convolutional	64	1×1/1	104×104×128	104×104×64
	Convolutional	128	3×3/1	104×104×64	104×104×128
	Residual				104×104×128
	Convolutional	256	3×3/2	104×104×128	52×52×256
8×	Convolutional	128	1×1/1	52×52×256	52×52×128
	Convolutional	256	3×3/1	52×52×128	52×52×256
	Residual				52×52×256
	Convolutional	512	3×3/2	52×52×256	26×26×512
8×	Convolutional	128	1×1/1	26×26×512	26×26×128
	Convolutional	512	3×3/1	26×26×128	26×26×512
	Residual				26×26×512
	Convolutional	1024	3×3/2	26×26×512	13×13×1024
4×	Convolutional	512	1×1/1	13×13×1024	13×13×512
	Convolutional	1024	3×3/1	13×13×512	13×13×1024
	Residual				26×26×1024
	Avgpool		Global		
	Connected		1000		
	Softmax				

Fig. 2 Structure model of the darknet-53 network

The convolutional feature channel interrelationship is adopted by SENet for modelling, and channel responses in specific layers of the convolutional neural network are reassigned to enhance the extraction of useful information. Three main components, namely Squeeze, Excitation and Weight Assignment, are contained in the module, with the basic structure shown in Fig. 3.

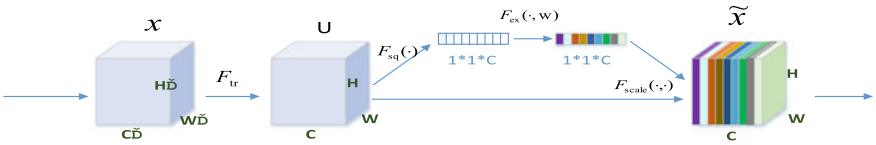


Fig. 3 Basic structure of the SENet

The channel attention mechanism initially performs a conversion operation, which is represented as a convolution operation in practice. In the conversion operation, the input features are x, the output features are U, and the convolution kernel is V.

Squeeze operation encodes the feature map, which is obtained from the conversion operation, compressing the two-dimensional feature map on each channel into a real number with a global perceptual field. The number, which is obtained from the global average pooling formula, represents the original weight of the channel, which can be calculated as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{5}$$

The obtained channel raw weights can be normalized by the excitation operation with a multilayer perceptron containing multiple layers, which is composed of a fully connected layer, ReLU activation function, a fully connected layer and Sigmoid activation function. The final weights of each channel can be expressed as s_c, which can be calculated as follows:

$$s_c = F_{ex}(z, \mathrm{w}) = \sigma(g(z, \mathrm{w})) = \sigma(\mathrm{w}_2 \delta(\mathrm{w}_1 z)) \tag{6}$$

where: δ is the ReLU activation function and σ is the Sigmoid activation function.

The weight assignment operation assigns weights s_c to each channel in the output feature map U, after the conversion operation obtains the final output $\tilde{x}$, which can be calculated as follows:

$$\tilde{x} = F_{scale}(u_c, s_c) = u_c \otimes s_c \tag{7}$$

$\otimes$ denotes element-by-element multiplication, and SENet enables the assignment of weights to channels in the above manner.

3.2 Improvement of Feature Detection Network

Smaller defects are generated in the manufacture process of the mobile phone glass covers. The smallest perceptual field of YOLOv3 corresponds to the feature map with the size of 52×52, which is obtained by downsampling the input image with a factor of 8. Therefore, YOLOv3 has a poor performance in detecting targets with pixels within the range of 8×8.

In this paper, the detection network of the YOLOv3 algorithm has been upgraded, so that its ability to detect small targets can be enhanced. The improved detection layer is based on the original feature maps and continues to perform 4-fold downsample to obtain 104×104 scale of feature maps. Combined with the channel attention module,

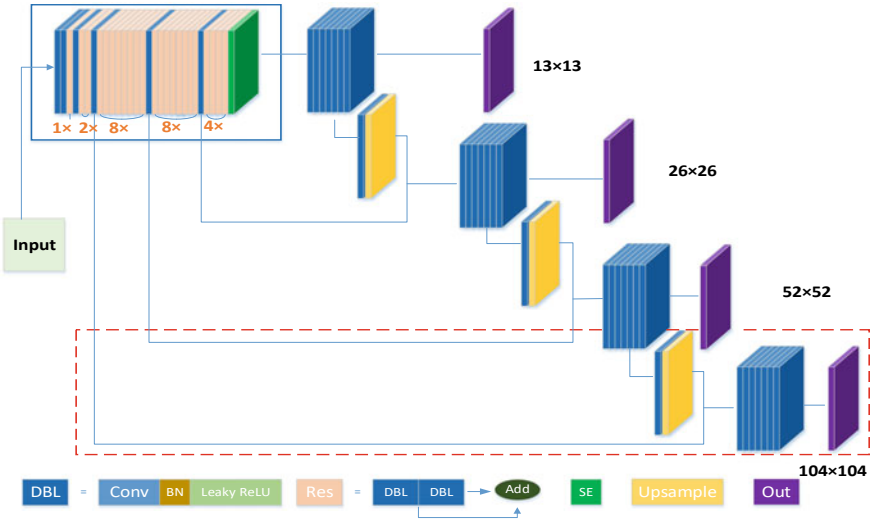


Fig. 4 YOLOv3-improve1 detection network

the improved YOLOv3 network can be obtained and denoted as YOLOv3-improve1, as shown in Fig. 4.

In the figure, Conv as a 5-layers convolutional layer, is composed of 3×3 and 1×1 convolutional with different convolutional kernel sizes. The solid blue boxes refer to the feature extraction network of YOLOv3 and the red dashed boxes refer to the added detection layers. There are 4 dimensional feature maps for the improved YOLOv3, namely 13×13, 26×26, 52×52 and 104×104, each of which is assigned with 12 anchor boxes in descending order. The feature map 104×104 is obtained after the network is improved, which combines the deep information contained in the 109th layer of the network and the shallow information in the 11th layer, thereby providing a further improvement in small target detection.

3.3 Model Pruning

In the industrial manufacture, a certain detection speed shall be achieved during the detection process with the aim of ensuring the balance in the assembly line. The parameters of YOLOv3 are large and computationally intensive, and the computing power has been restricted by the computer terminals on industrial sites. Therefore, the model operations shall be reduced and the detection speed shall be increased with guaranteed detecting accuracy. In this paper, the layer pruning and channel pruning methods proposed by Liu [18] et al. have been adopted.

In network channel pruning, the scaling factor coefficient γ of the BN layer in the convolutional network is regarded as the important factor. When γ is smaller,

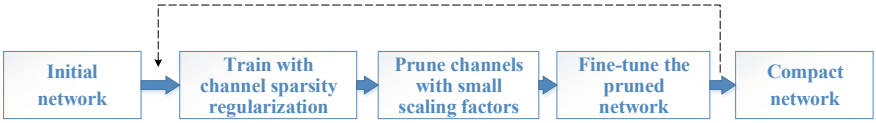


Fig. 5 Diagram of the pruning process of the YOLOv3 model

the corresponding channel is less important and can be pruned. The objective optimisation function of the algorithm as a whole is as follows:

$$L = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \tag{8}$$

where: the first term refers to the model prediction loss and the second term refers to the canonical term about γ. γ is a hyperparameter used for weighing the two terms, generally assigned as 1e−4 or 1e−5, $g(*)$ with the expression $g(s) = |s|$, and the L1 paradigm. The overall pruning process is shown in Fig. 5.

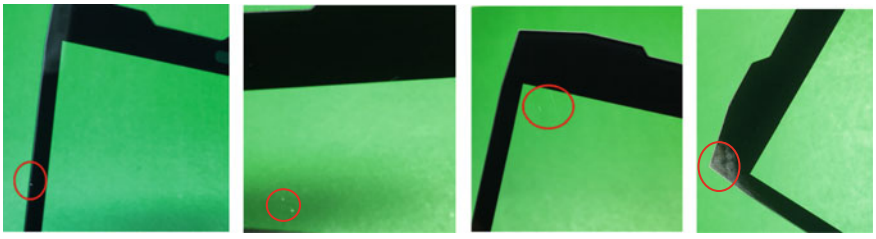
4 Experiments and Results Analysis

4.1 Experimental Platform

The experimental platform is Supermicro infreesys server, with operating system: Ubuntu 18.04LTS, CPU: Intel W2123, memory: 32G, graphics card: NVDIA Geforce RTX2080Ti × 2, and video memory: 16 GB × 2. Deep learning framework: Pytorch.

The dataset used in this experiment was shot at the production site, the types of defects were divided into four categories: chipped, pit point, scratching and soiling, with the specific defect images shown in Fig. 6.

A total of 15,914 images were collected in this dataset, and they were marked with labelImg software. The dataset for this study was generated according to the



(a) Chipped edge (b) Pit point (c) Scratching (d) Soiling

Fig. 6 Defect map of the mobile phone glass covers

Table 1 Priori bounding boxes sizes for each algorithm

Algorithms	The size of priori bounding boxes
YOLOv3	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20
YOLOv3-improve1	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20 16,50 23,39 28,58
YOLOv3-improve2	8,11 11,13 9,17 13,12 11,16 15,17 14,21 35,13 39,20 16,50 23,39 28,58

VOC dataset format required by YOLOv3. The dataset was divided into the training dataset and the test dataset in the ratio of 9:1, with a total of 14,321 photos in the training dataset and 1593 photos in the test dataset.

The algorithm obtained by model pruning based on YOLOv3-improve1 is named YOLOv3-improve2. The dataset was trained and tested with YOLOv3, YOLOv3-improve1 and YOLOv3-improve2 algorithms. The number of samples per batch was set to 16 with subdivision = 8. The input image size was set to $416 \times 416 \times 3$, where 3 was the number of image channels. Besides, the momentum was set to 0.9. The YOLOv3 algorithm contained 3 detection layers, with each layer assigned with 3 priori bounding boxes, and 9 priori bounding boxes were required. In contrast, the YOLOv3-improve1 and YOLOv3-improve2 algorithms contained 4 detection layers, with 3 priori bounding boxes per layer, 12 priori bounding boxes were required, besides, their priori bounding boxes were equal in size. According to the k-means clustering algorithm, the relevant of priori bounding box information is clustered as shown in Table 1.

4.2 Experimental Results

The experiments were conducted on three algorithms with training epochs of 500, and the Adam method was adopted as the parameter optimization method. The bounding box loss values, coordinate loss values, classification loss values, and confidence loss values were included in the trained loss values of the three algorithms, with the total loss value comparison curves of the three algorithms shown in Fig. 7.

In the above figure, the loss value of the YOLOv3-improve2 algorithm is obtained during fine-tuning training, so its initial loss values are relatively low compared to the other two algorithms; while, the trend is similar to the other two algorithms during the training process. During the training process, it can be seen that the YOLOv3-improve1 algorithm and the YOLOv3-improve2 algorithm have approximately similar loss values, and the decline process is gentle, while the YOLOv3 algorithm has an oscillating situation, with large oscillation fluctuations. From the figure, it can be seen that the YOLOv3-improve1 and YOLOv3-improve2 algorithms have overall lower loss values than the YOLOv3 algorithm during the training process. According to the above analysis, it can be concluded that the improved algorithm is more effective than YOLOv3.

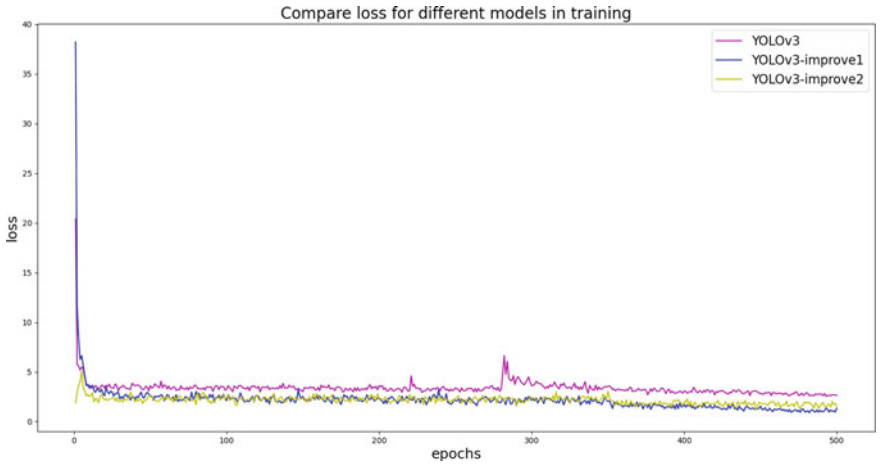


Fig. 7 Comparison of training loss values

Compare the mean average accuracy value mAP of YOLOv3, YOLOv3-improve1, YOLOv3-improve2, and their comparison graphs shown in Fig. 8. Information about the three algorithms, including the trained mAP values and the detection speed (how many sheets per second are detected) is shown in Table 2.

From Fig. 8, it can be seen that the YOLOv3-improve1 and YOLOv3-improve2 algorithms have an overall higher train mAP than the YOLOv3 algorithm. As per contents in Table 2, YOLOv3-improve1 algorithm has 3.3% increase in its mAP value than YOLOv3 algorithm. Due to its deeper network model, its detection speed is 6 sheets/s slower than the original algorithm. The YOLOv3-improve2 algorithm has

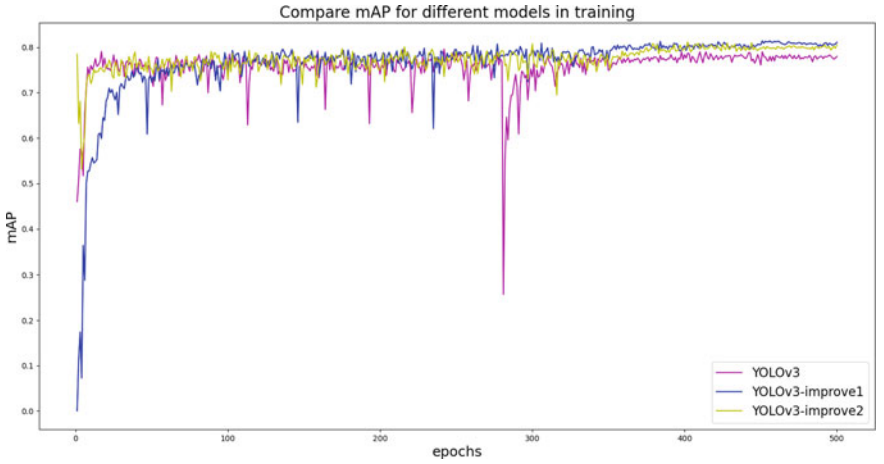
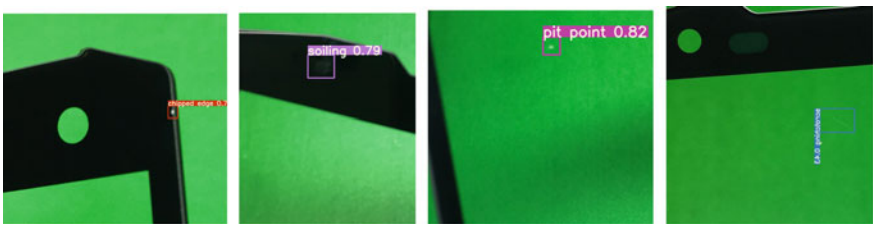


Fig. 8 Comparison of training mAP values

Table 2 Comparison of experimental results

Model	Feature detection layer	Attention mechanism	Network pruning	mAP (%)	Detection speed (sheets/s)
YOLOv3	3 layers	No	No	78.0	36.4
YOLOv3-improve1	4 layers	Yes	No	81.3	30.4
YOLOv3-improve2	4 layers	Yes	Channel crop 60%, layer crop 12 layers	81.0	43.1



(a) Soiling detection (b) Pit point detection (c) Scratching detection (d) Chipped edge detection

Fig. 9 Mobile phone glass covers defect detection effect

a 0.3% decrease in its mAP value than YOLOv3-improve1, but its detection speed increases by 12.7 sheets/s compared to YOLOv3-improve1, which is a significant improvement with less loss of accuracy. In addition, YOLOv3-improve2 algorithm improves detection accuracy by 3% compared to the YOLOv3 algorithm, while the detection speed also increases by 6.7 sheets/s. This demonstrates that the improved algorithm in this paper has a better performance in detection.

YOLOv3-improve2 training weights were adopted to detect mobile phone glass covers defects, with the results shown in Fig. 9.

According to the above figure, it can be seen that the final YOLOv3-improve2 algorithm provides a more accurate detection method of mobile phone glass covers.

5 Conclusion

The detection of smartphone glass cover defects using manual methods, which is inefficient, costly, low detection accuracy and labour intensive, while the detection methods using traditional machine vision is poor detection flexibility, low yield and poor generalisation capability. Therefore, this paper adopts the YOLOv3 algorithm for defect detection on smartphone glass covers, and improves the YOLOv3 algorithm for the specific requirements and practical characteristics of defect detection.

The channel attention mechanism SENet is added to the Darkent-53 network to solve the problem of inconspicuous defect features, a 104 × 104 scale detection layer was added to the YOLOv3 detection network to solve the problem of multi-scale defects, and the scaling factor coefficient of the BN layer of the convolutional network is used as the importance factor for model pruning to improve the detection speed. A large number of photographs covering chipped edges, scratches, pits and dirty defects were taken from a smartphone glass cover manufacturing company to make a training dataset and a validation dataset. The proposed method and the original YOLOv3 algorithm are compared and analysed, and the results showed that the algorithm outperforms the original YOLOv3 algorithm in all aspects, not only in terms of real-time performance, but also in terms of detection accuracy, meeting the need for high precision and efficient detection of defects in the industrial production site of smartphone glass covers.

References

1. Chengxuan, W., Qi, L., Ying, T.: Five trends in the development of glass for mobile phones. Glass **47**(04), 1–6 (2020)
2. Jian, C., Gao, J., Ao, Y.: Automatic surface defect detection for mobile phone screen glass based on machine vision. Appl. Soft. Comput. (2016)
3. Chuanxia, J., Jian, G.: Research on visual inspection method for surface defects of mobile phone glass covers. Packag. Eng. **39**(5), 16–21 (2018)
4. Chuanxia, J.: Research on machine vision detection and classification method for mobile phone glass covers surface defects. Guangdong University of Technology (2017)
5. Liang, L.-Q., Li, D., Fu, X., Zhang, W.-J.: Touch screen defect inspection based on sparse representation in low resolution images. Multimedia Tools Appl. **75**(5), 2655–2666 (2016)
6. Jianguo, Z., Li, Y., Qi, J.K., Ji, T., Liu, J.: Research on scratch detection on mobile phone screen surface based on machine vision. Appl. Opt. **41**, 1–5 (2020)
7. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. Neural Inf. Proc. Syst. 91–99
8. He, K., Gkioxari, G., Dollar, P., et al.: Mask R-CNN. In: International Conference on Computer Vision, 2980–2988 (2017)
9. Redmon, J., Divvala, S.K., Girshick, R., et al.: You only look once: unified, real-time object detection. Comput. Vis. Pattern Recognit. 779–788 (2016)
10. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.:SSD: single shot MultiBox detector. Lect Notes Comput Sci 21–37 (2016)
11. Li, C., Zhang, X., Huang, Y., Tang, C., Fatikow, S.: A novel algorithm for defect extraction and classification of mobile phone screen based on machine vision. Comput. Ind. Eng. 106530
12. Jiang, J., Cao, P., Lu, Z., Lou, W., Yang, Y.: Surface defect detection for mobile phone back glass based on symmetric convolutional neural network deep learning. Appl. Sci. **10**, 3621 (2020)
13. Hongjia, S.: Research on Mobile Phone Screen Defect Detection and Classification Algorithm Based on Depth Model. Zhejiang University (2018)
14. Wei, S.: Deep Convolutional Neural Network-Based Defect Detection for Mobile Phone Screens. University of Electronic Science and Technology (2019)
15. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 7132–7141 (2018)

17. Hui, Z., Kunfeng, W., Feiyue, W.: Progress and prospects of deep learning in target vision detection. Acta Automatica Sinica **43**(8), 1289–1305 (2017)
18. Liu, Z., Li, J., Shen, Z., et al.: Learning efficient convolutional networks through network slimming. International Conference on Computer Vision, 2755–2763