Journal of Network Intelligence ©2023 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **8**, Number **4**, November **2023**

Defect Detection of Aluminum Wine Bottle Cap Based on Improved YOLOv4

Jin Han*

College of Computer Science and Technology
Shandong University of science and technology
266590 Qingdao,Shandong Province,China
shnk123@163.com

Yujie Li

College of Computer Science and Technology
Shandong University of science and technology
266590 Qingdao,Shandong Province,China
lyjsku2021@163.com

Wenhao Wu

College of Computer Science and Technology
Shandong University of science and technology
266590 Qingdao,Shandong Province,China
2115065840@qq.com

*Corresponding author: Jin Han

Received March 10, 2023, revised May 26, 2023, accepted July 18, 2023.

Abstract. *The defect detection of wine bottle cap is an important part of industrial quality inspection in distillery. High detection accuracy, low missed detection rate and fast detection speed are important reference performance of wine bottle cap defect detection. At present, wine bottle cap detection based on deep learning hasn't been extensively used. Therefore, this paper introduces an improved YOLOv4 model, which aims to improve the detection accuracy and reduce the missed detection rate of the model without sacrificing the detection speed. Firstly, the Loss Function of YOLOv4 model is improved to accelerate its convergence speed; Secondly, BAM is drew into the backbone to enhance its capacity of extracting useful information and reduce the computational complexity of the model; In addition, the residual feature augmentation module is improved and introduced into the feature extraction layer to reduce the loss of feature information caused by convolution operation; Finally, the original image data of wine bottle cap defects comes from Alibaba Cloud Tianchi competition platform. The dataset is expanded through operations such as data flipping. The results of this experiment show that the improved YOLOv4 is superior than the original model, and the mAP reaches 94.17%, which can meet the requirements of industrial quality inspection.*

Keywords: wine bottle cap defect detection, YOLOv4, the loss function, BAM, dataset

1. **Introduction.** The defect of wine bottle cap will impact the quality of the wine at first hand and additionally affect its taste and sales [1]. As an indispensable part of the product, the bottle cap has defects such as deformation, fracture and scratch, which means that the product has major quality problems, which may lead to abnormal sales of the product in the market, and even food safety issues. Therefore, manufacturers' and users' requirements are getting higher and higher for the appearance quality of wine bottles. At

present, the inspection of wine bottle cap mainly depends on artificial inspection. This inspection method not only can not meet the needs of industrial production in terms of work efficiencies, but also is subject to the visual influence of inspectors. Some small defects are easy to be missed during fatigue inspection.The quick growth of computer vision has promoted the development of defect detection technology in industrial quality inspection, such as steel, cotton textile, aluminum products and so on.

Compared with manual detection, machine vision-based detection methods improve its detection efficiency and accuracy. However, the traditional machine vision detection algorithms have poor flexibility for the feature extraction, it is necessary to establish an effective feature extraction algorithm on the basis of the specific type of bottle cap defects. Due to the different shapes and sizes of bottle cap defects in industrial products, a large number of algorithmic design resources are needed for feature extraction, which indicates that it has poor generality for target objects. The bottle cap defect detection using deep learning algorithm not only has higher applicability and stability than the traditional visual detection algorithm, but also has higher detection accuracy for changing scenes and targets. However, defect detection algorithm based on deep learning has not been widely used in wine bottle cap defect detection. How to improve the accuracy of wine bottle cap detection and reduce the missed detection rate has become the research goal. In this paper, a defect detection method of wine bottle cap based on deep learning is proposed. YOLOv4 [2] is selected as the detection algorithm, a single-stage detector with excellent speed and accuracy as the baseline. By analyzing the defect characteristics of the wine bottle cap surface, we found that most of the defects that are difficult to identify are small defects, which inspired us to design a series of schemes to reduce the loss of features, improve its loss function, and introduce an attention mechanism and Residual Feature Augmentation module [3]. On this basis, we propose an improved version of YOLOv4, which is more suitable for the defect detection of industrial wine bottle caps by reducing its missed detection rate and improving the detection accuracy.

The improved YOLOv4 not only has the characteristics of high real-time, but also has better adaptability to the defects of wine bottle caps. The network is forced to pay closer attention to where salient features are located by the enhanced loss function and the introduced attention mechanism. The addition of residual feature augmentation reduces the loss of feature information and helps the model deal with difficult samples. The main work of this paper is as follows:

(1) The position loss function CIOU of YOLOv4 is improved. The improved position loss function makes the detection frame tend to the real frame, and improves the convergence speed and the detection performance of the model.

(2) We combine the attention mechanism including spatial attention and channel attention with the bottleneck of CSPDarknet-53, and make the network pay closer attention to the required parts through the weighted feature map, so as to further improve the detection accuracy of wine bottle cap defects.

(3) We add the improved residual feature augmentation module to PANet to reduce the loss of feature information, thereby reducing the missed detection rate of the model.

(4) Using the defect detection data pictures of bottled Baijiu from Alibaba Cloud Tianchi competition, this paper selects the defect data. The defects are mainly divided into five parts: bottle cap scratch, bottle cap deformation, bottle cap broken edge, bottle cap fracture and abnormal code spraying. The data set is expanded to 2661 pictures through mirror image turning, rotation and other operations. The target experiment verifies the effectiveness of this method, and the map of the five defects reaches 94.17%.

The remainder of this essay is organized as follows. The section 2(Related Work) introduces the research status of surface defect detection based on deep learning and the principle of YOLOv4. The network structure is shown in detail in Section 3(Methods). Section 4 (Experiment) shows the problems related to the design of ablation experiment and analysis module. The section 5(conclusion) summarizes this paper's work.

2. Related Work.

2.1. Research status of surface defect detection based on deep learning.

Last few years, due to the computer in rapid development, camera and image processing technology, computer vision technology has been specially applied to industrial production and other corresponding fields. Since then, people have studied a series of defect detection methods in industry.

Currently, deep learning is widely used in the fields of object detection, image segmentation [4], transfer learning [5], and other fields. Its great advantage lies in the use of convolutional neural networks.In the field of object detection, the detection performance of object detection algorithm using convolutional neural networks are better than that of manual detection. Object detection using convolution neural network has become a hot research topic in the field of object detection of defective products.At present, there are two kinds of object detection algorithms based on convolutional neural network: two-stage target detection algorithm and one-stage target detection algorithm. In 2013, Girshick et al. [6] proposed the CNN based detector R-CNN, and the two-stage target detection algorithm came out. Since then, representative algorithms such as Fast R-CNN [7] and Faster R-CNN [8] have emerged. Because the two-stage target detection algorithm has the characteristics of high precision, it is widely used in defect detection scenarios. Yang et al. [9] used the Fast R-CNN detection method of multi task FPN structure to detect steel surface defects. Geng et al. [10] proposed a PCB surface defect algorithm based on Faster R-CNN algorithm, which constructs feature pyramid for multi-scale fusion and uses focal loss as loss function. However, the real-time performance of the two-stage detection algorithm is poor. In 2016, Redmon et al. [11] proposed the You Only Look Once (YOLO) network, followed by a series of one-stage detection algorithms, such as SSD [12], YOLOv2 [13], YOLOv3 [14] and Retinanet [15]. While the detection accuracy is not reduced, it meets the requirements of high real-time, and then gradually replaces the two-stage detection algorithm in defect detection. For example, Chen et al. [16] introduced the bottleneck transformer (BOT) to construct a nonburr line cylinder block liner surface defect detection based on YOLOv4. Li et al. [17] introduced an insulator fault detection method based on YOLOv4.

2.2. YOLOv4.

The overall framework of YOLOv4 is shown in Figure 1. Where CMB represents a convolution combination consisting of convolution module, Mish activation function, and BN.The backbone combines the advantages of Darknet and CSPNet [18]. It consists of five CSP modules, and each CSP module consists of a different number of residual blocks. The gradient vanishing problem and over fitting problem of the model can be solved by using CSP module instead of ordinary convolution layer, improving the detection performance of the model.The feature extraction layer of YOLOv4 combines spatial pyramid pooling (SPP) [19] and PAN [20]. The SPP structure consists of three different max-pooling operations to further extract and fuse feature information. The feature layer from the backbone and SPP is fused through PAN and transmitted to the prediction network for detection. The difference between the neck of YOLOv4 and YOLOv3 is that a bottom-up feature transmission path is constructed. It can fuse the shallow position information with the deep semantic information, and enhance the detection ability of different scale feature layers.

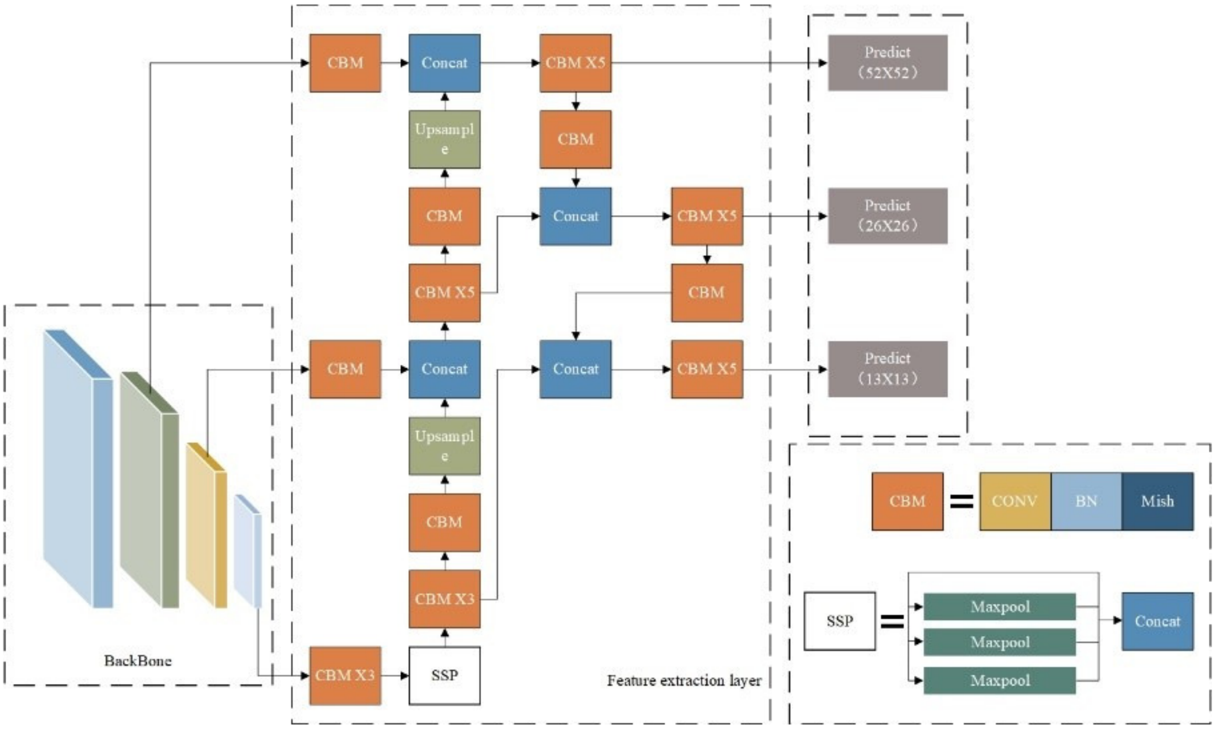


Figure 1. YOLOv4 model structure

3. **Method.** This section describes in detail the structure of the wine bottle cap defect detection network. Firstly, we introduce the improved loss function, and then add the BAM module to the backbone network. Finally, we propose the improved residual feature augmentation module.

3.1. **Improvement of loss function.** Suppose that the real frame and the prediction frame of target N are $G\{..., [x, y, w, h]\}$ and $P\{..., [x1, y1, w1, h1]\}$ respectively, where x, y represent the center point coordinates of the real frame, w, h represent the width and height of the real frame, $x1, y1$ represent the center point coordinates of the prediction frame, and $w1$ and $h1$ represent the width and height of the prediction frame respectively. Through the calculation of the loss function, the prediction frame P gradually moves to coincide with the real frame G, as shown in the Figure 2.

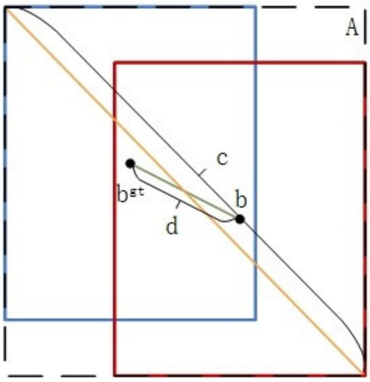


Figure 2. Schematic diagram of CIOU

The loss function of YOLOv4 is shown in Formula 1, which is composed of position loss, category loss and confidence loss, and is represented by L_{loc}, L_{cls} and L_{conf} respectively.

$$Loss = L_{loc} + L_{cls} + L_{conf} \tag{1}$$

The position loss function of YOLOv4 uses CIOU Loss (Complete Intersection over Union Loss) [21], and its expression is as follows:

$$L_{loc} = 1 - CIOU \tag{2}$$

$$CIOU = IOU - \left(\frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu\right) \tag{3}$$

$$\begin{aligned} c =& [(max(x + \frac{w}{2}, x_1 + \frac{w_1}{2}) - min(x - \frac{w}{2}, x_1 - \frac{w_1}{2}))^2 + \\ & (max(y + \frac{h}{2}, y_1 + \frac{h_1}{2}) - min(y - \frac{h}{2}, y_1 - \frac{h_1}{2}))^2] \end{aligned} \tag{4}$$

$$\nu = \frac{4}{\pi^2}\left(\arctan\frac{w}{h} - \arctan\frac{w_1}{h_1}\right) \tag{5}$$

$$\alpha = \frac{\nu}{(1 - IOU) + \nu} \tag{6}$$

Where, b and b^{gt} denote the center points of the prediction frame and the real frame, ρ denotes the euclidean distance between the center points of the two boxes, and c represents the diagonal region that can contain the minimum closure region of the prediction frame and the real frame at the same time, that is, the diagonal distance of box A in the Figure 2.

CIOU loss involves not only the coincidence degree of the prediction frame and the real frame, but also the distance and scale between the two frames. Even if the two frames do not overlap, the CIOU loss can still provide the next moving direction for the prediction frame.

We note that when the centers of the prediction frame and the real frame are on the same horizontal line, that is, in the four cases in the Figure 3, the next step of the prediction frame should move parallel to the real frame. Therefore, controlling the next movement of the prediction frame through the above CIOU has the problem of slow model convergence. Therefore, the position loss function is improved in this paper. When the center points of the two frames are on the same x horizontal line, c in CIOU is the y-direction length of the minimum closure of the two frames, this is

$$c = \left(max\left(y + \frac{h}{2}, y_1 + \frac{h_1}{2}\right) - min\left(y - \frac{h}{2}, y_1 - \frac{h_1}{2}\right)\right)^2 \tag{7}$$

When the center points of the two frames are on the same y horizontal line, c in CIOU is the x-direction length of the minimum closure of the two frames, that is

$$c = \left(max\left(x + \frac{w}{2}, x_1 + \frac{w_1}{2}\right) - min\left(x - \frac{w}{2}, x_1 - \frac{w_1}{2}\right)\right)^2 \tag{8}$$

Therefore, the pseudo code of the position loss function of the model is shown in algorithm 1. In Section 4.4.1, a detailed experiment is carried out on this part.

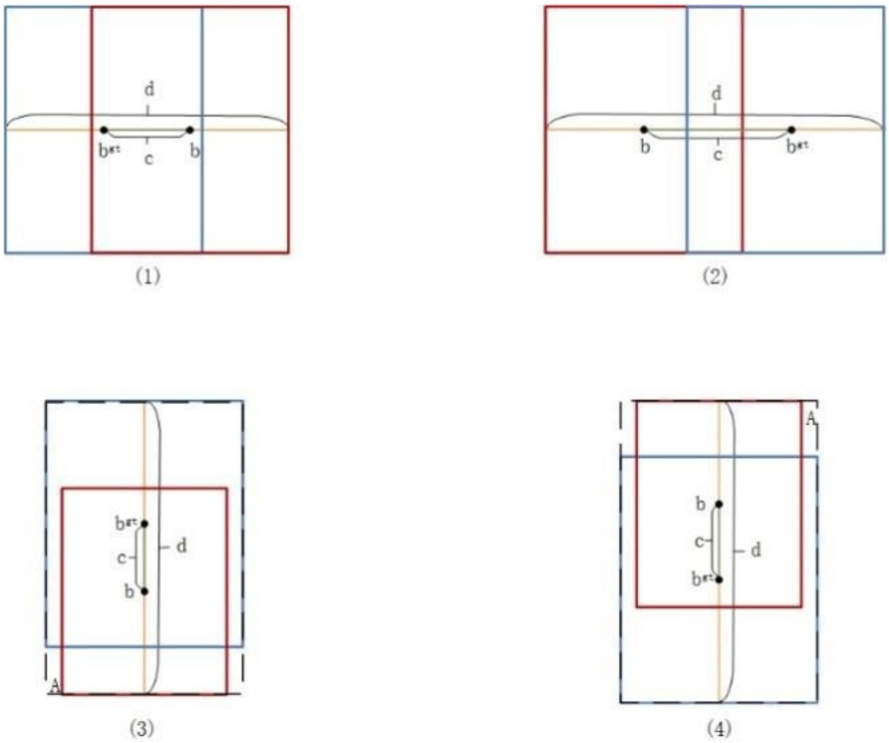


Figure 3. Four special cases of coincidence

```
Algorithm 1 CIOU Loss
1: Input: G, P
2: Output: CIOU Loss
3: if x == $x_1$
compute c using Eq.(7)
  else if y == $y_1$
      compute c using Eq.(8)
  else
      compute c using Eq.(4)
  compute CIOU using Eq.(3)
  compute CIOU using Eq.(2)
  return CIOU Loss
```

3.2. **The improvement of the Backbone network.** In order to improve the ability of backbone to extract feature information, this paper improves the backbone of YOLOv4 and Bottlenet Attention Moudle (BAM) is introduced, which is a mixed domain attention mechanism proposed by Park et al. [22] on BMVC2018. In this work, the author focuses on the influence of attention on the general depth neural network, and proposes a simple and effective attention model, which can be combined with any forward propagation convolution neural network to combine attention with the bottleneck idea.

Thus, the first four CSP modules in backbone are retained in this paper, and the three BAM modules are inserted between the residual modules of the last group of CSPNets, as shown in the Figure 4.

BAM structure is a mixture of channel attention [23] and spatial attention. This paper uses BAM structure to construct a new backbone network residual block, whose structure is shown in Figure 5. The purpose of inserting multiple BAM into the Bottlenet is to

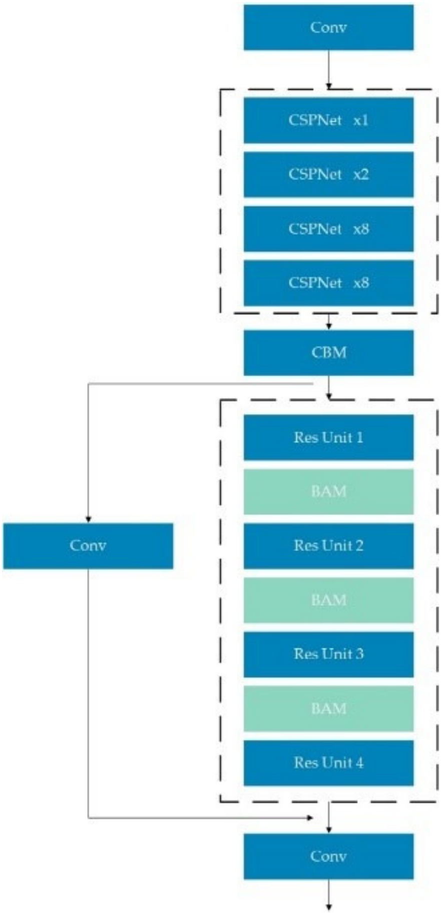


FIGURE 4. Backbone network after adding BAM module

build a hierarchical attention, focus the network on the required parts, and improve the expression performance of the feature graph.

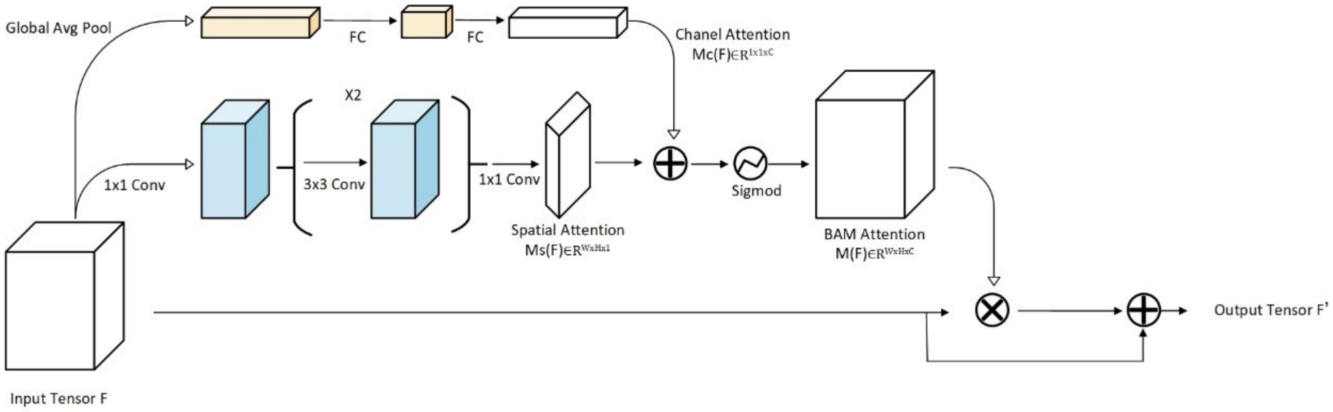


FIGURE 5. The structure diagram of BAM

3.3. **Improvement of feature extraction layer.** YOLOv4 adopts the path aggregation network (PAN), which is different from FPN [24] in that it adds a top-down transmission path to enhance the fusion ability of different feature scales. However, pan uses a large number of 1×1 and 3×3, while 1×1 the function of convolution is to change the number of channels. In this process, it may cause the loss of feature information. Guo Chaoxu et al. think that the loss of semantic information caused by the reduction of the number of channels can be compensated through spatial information, and constructed the Residual Feature Augmentation (RFA) module [3]. As shown in Figure 6.a, C5 is failed 1×1 convolution to reduce the characteristic layer of the number of channels, so C5 can be operated and then fused to M5 to reduce the loss of characteristic information. The operation process of residual feature augmentation is as follows: as shown in Figure 6.c, firstly, the C5 feature layer is downsampled into α_1×S, α_2×S, ..., α_n×S. Context features of different scales of S (S is the feature size of C5), each context feature is then taken as 1×1 convolution fixes the number of channels to the same number of channels and upsampling so that the size of the context feature is the same as that of C5. Finally, a module fusion called Adaptive Spatial Fusion (ASF) is used to adaptively combine these contextual characteristics, rather than just summation. Its structure is shown in Figure 6.d, a spatial weight graph is generated for each feature through the ASF structure to aggregate the context features into M6.

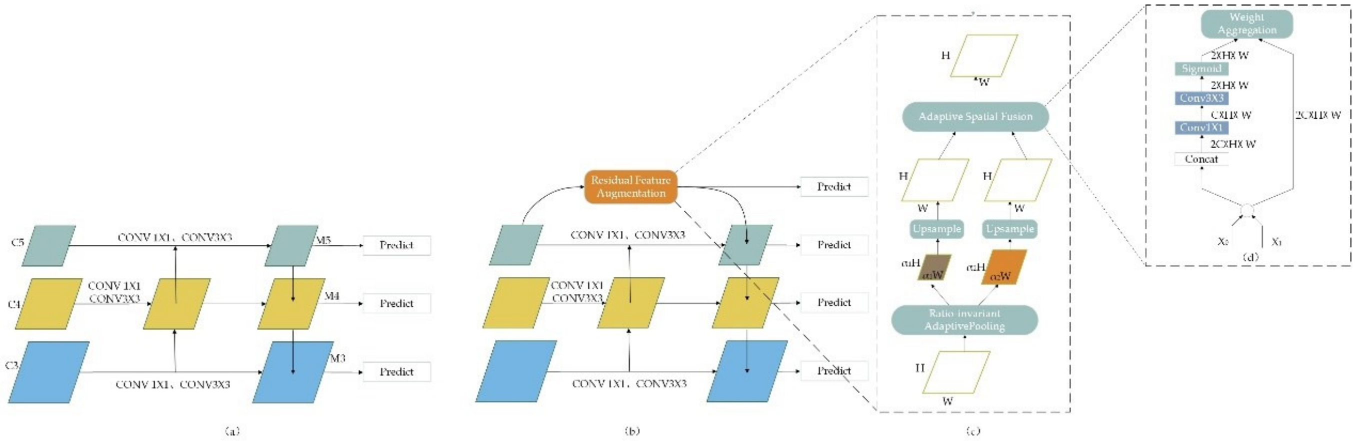


Figure 6. Structure diagram of original and improved feature extraction layer. (a) is the original feature extraction layer. (b) is the feature extraction layer after RFA is added. (c) is the structure diagram of RFA. (d) is the structure diagram of ASF.

In this experiment, we applied the residual feature augmentation to PAN, because the feature size of C5 is 52×52×256, which must be a multiple of 2 when used up and down, so n and channel take 2 and 128 respectively.

Considering that M6 feature layer has multi-scale context information, M6 and M5 are fused and M6 is used as a separate detection head, as shown in Figure 6.b.

4. Experiment and Analysis.

4.1. Experimental Platform and Parameters.

The system environment of this study is as follows: NVIDIA GeForce GTX2080Ti GPU and Windows 10 operating system. The epoch of all models in this paper is 250, the batch size is 4 and, an input image size is 416×416.

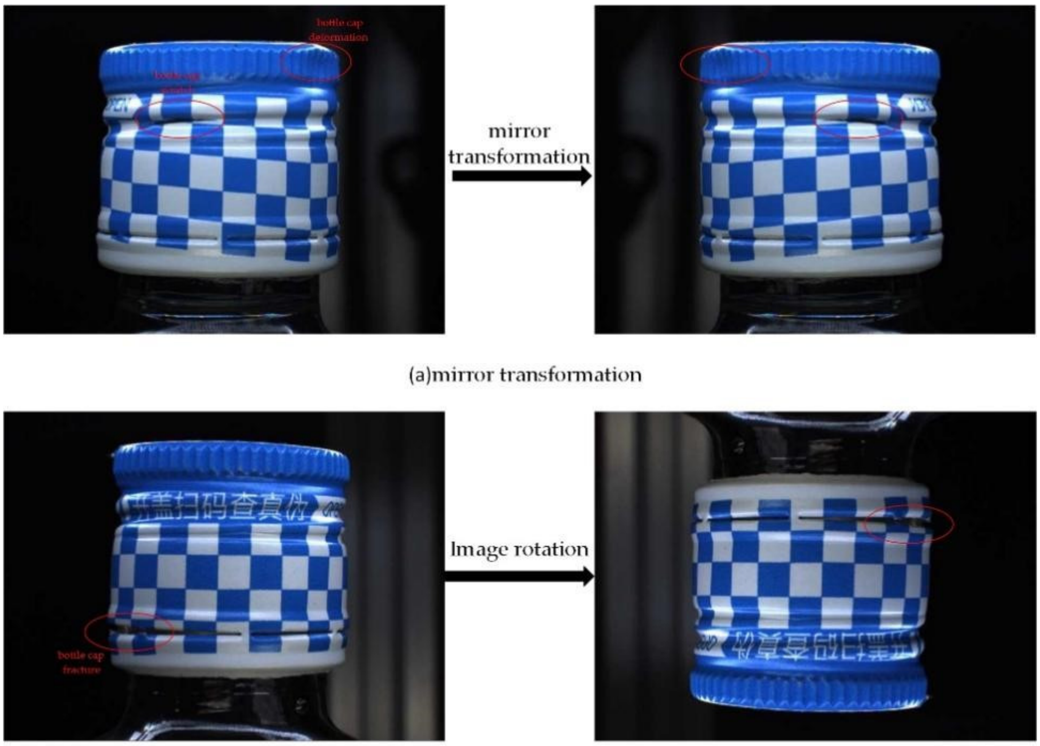
4.2. Production of dataset.

This paper uses the bottle cap defect as the detection target, and the data image is from the bottle Baijiu defect detection data set of Alibaba Cloud Tianchi competition. This paper selects 1000 pieces of defect data. Aimed at further expanding the data set, some image conversion technologies such as image flipping and image rotation are used.

Image flipping is usually divided into two types: Mirror conversion at horizontal position and mirror conversion at vertical position. Considering that most of the defects in the wine bottle cap image extend in the horizontal direction, in order to better cluster and get the detection frame, this paper adopts the mirror transformation of the horizontal position.The obtained image data after image mirroring is as shown in the Figure 7.a. Image rotation: All pixels on the image rotate the same angle with the image center as the origin.The mathematical formula for image rotation is as follows:

$$\begin{bmatrix} t_x \\ t_y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9}$$

In order to make the size of the image the same as that before the rotation, this paper uses 180° as the rotation angle. After the image is rotated, the defect image of the wine bottle cap is as shown in the Figure 7.b.



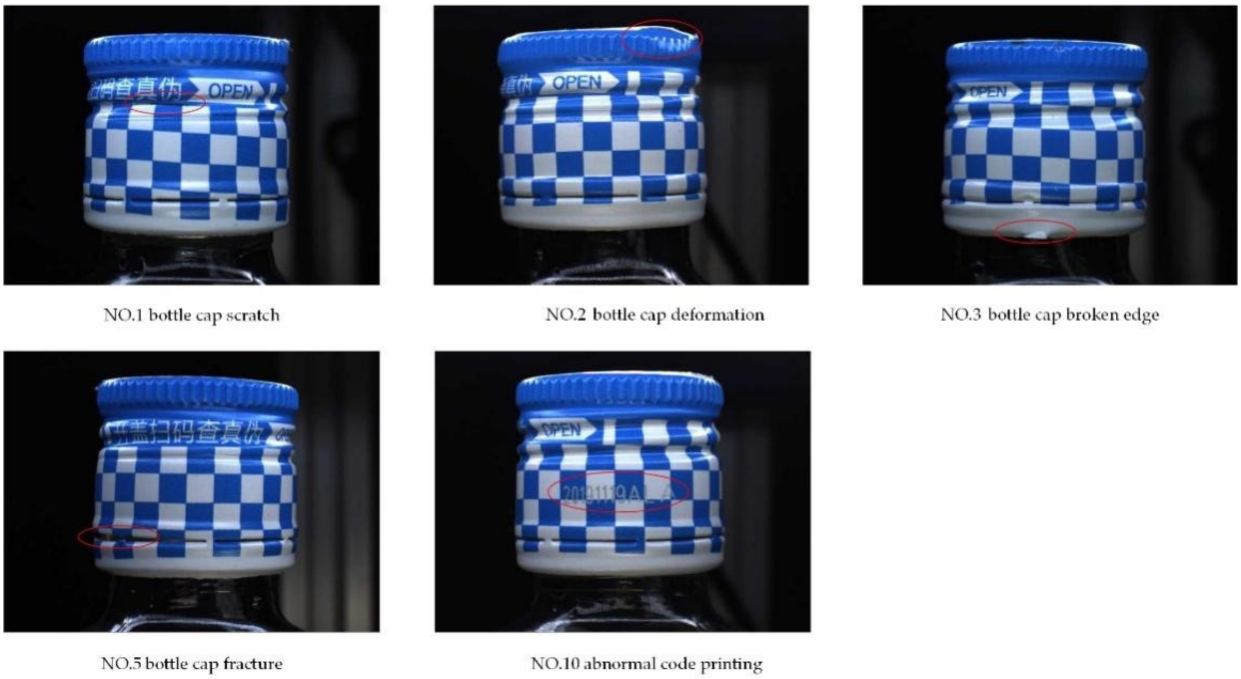
(a)mirror transformation

(b)Image rotation

FIGURE 7. Data preprocessing

After image enhancement, 2661 bottle cap defect images were obtained, and the bottle cap defect data set was constructed. This article sets the input size of the image to 416×416. Defects are mainly divided into five types: bottle cap scratch, bottle cap deformation, bottle cap broken edge, bottle cap fracture and abnormal code printing. As shown in the Figure 8, the five types of defects are represented by production labels 1, 2, 3, 5 and 10 respectively in the dataset. The ratio of training set, verification set and test set is 8:1:1. In this experiment, the tool LabelImg is used to label image defects. The annotation file in Pascal VOC format is saved in the form of XML.

4.3. **Experimental Results and Analysis.**



NO.1 bottle cap scratch

NO.2 bottle cap deformation

NO.3 bottle cap broken edge

NO.5 bottle cap fracture

NO.10 abnormal code printing

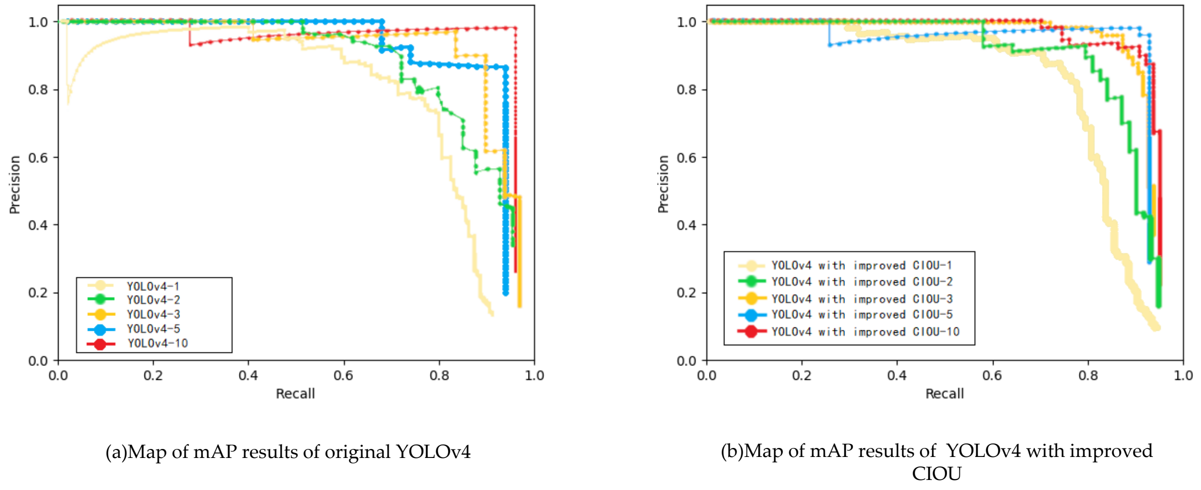
Figure 8. Display diagram of various defects

Table 1. Comparison of YOLOv4 and improved YOLOv4 with BAM

Model	AP_1	AP_2	AP_3	AP_5	AP_{10}	mAP	GFLOPs
YOLOv4	76.58	85.79	90.75	90.49	92.46	88.95	119.964
YOLOv4(+BAM)	80.24	93.49	90.53	92.12	94.23	92.67	105.792

4.3.1. *Analysis of the influence of the position loss function.* We trained the original YOLOv4 and the improved position loss function YOLOv4. The experimental results are as shown in the Figure 9. The P-R curve area of each type of defect detection of the improved CIOU loss model is larger than that of the original YOLOv4 model mAP from 88.95% to 90.01%. In addition, Figure 10 shows the loss change process during the training period. From the Figure 10, it can be seen that the loss value of YOLOv4 with improved loss function decreases faster after 15 rounds of iterations and tends to be stable after 80 iterations. Therefore, YOLOv4 with improved CIOU loss has better convergence effect and faster convergence speed, greatly improves the matching effect between the detection frame and the real frame, thus improving the detection accuracy of the model.

4.3.2. *Analysis of impact of improved backbone with BAM.* The Table1 shows that, after adding BAM module, the mAP from 88.95% to 92.67%, various types of AP have also been improved, the NO.1 AP increasing from 76.58% to 80.24%, the NO.2 AP increasing from 85.79% to 93.49%, the NO.5 AP increasing from 90.49% to 92.12%, and the NO.10 AP increasing from 92.46% to 94.23%. However, the NO.3 AP has not been improved, reducing by 0.22%, which may be the loss caused by the model paying more attention to small targets.The calculation cost of the model has been reduced from 119.964GFLOPs to 105.792GFLOPs, which improves the detection speed of the model. Therefore, through experimental verification, BAM module can improve the ability of backbone network to extract feature map, make the network pay more attention to the required parts, and improve the detection ability of the model.



(a)Map of mAP results of original YOLOv4

(b)Map of mAP results of YOLOv4 with improved CIOU

Figure 9. The P-R curve of YOLOv4 with improved CIOU and original YOLOv4

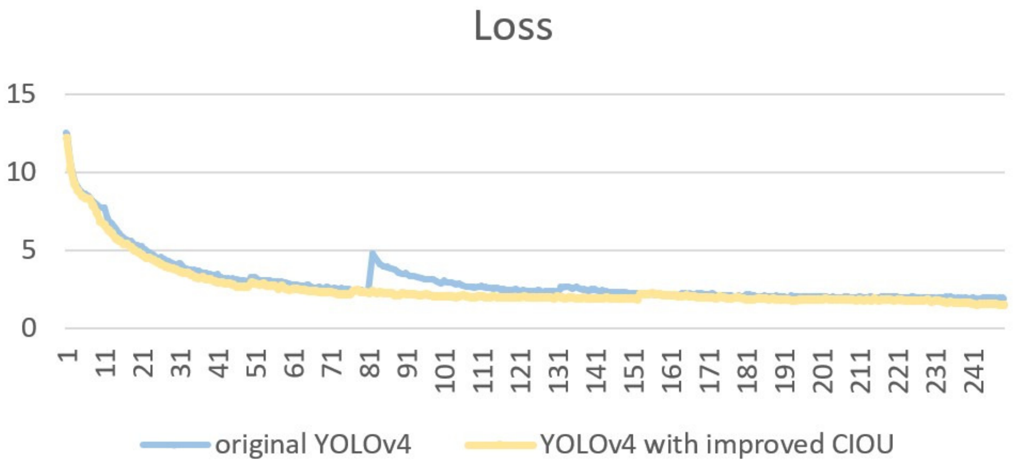


Figure 10. The curve of loss changes

Table 2. Comparison of missed detection rate between YOLOv4 and improved YOLOv4 with residual feature augmentation.

Model	AP_1	AP_2	AP_3	AP_5	AP_{10}
YOLOv4	36.75	33.94	8.0	11.2	10.0
YOLOv4(+RFA)	33.24	31.85	7.88	9.67	9.21

4.3.3. *Analysis of impact of improved feature extraction layer.* The Table2 illustrates the missed detection rate of the original YOLOv4 model and the YOLOv4 model with residual feature augmentation. It can be seen from the table that after adding the residual feature augmentation module, the missed detection rate of the model for small targets such as class numbers 1, 2 and 5 is reduced.The missed detection rate of No. 1, No. 2 and No. 5 defects decreased by 3.51%, 2.09% and 1.53% respectively. The impact on the missed inspection rates for NO.3 and NO.10 is not significant, with the missed inspection rates respectively reduced by 0.12% and 0.79%.Therefore, through experimental verification,it can be concluded that residual feature augmentation can reduce the loss of feature information, thus reducing the missed detection rate of the model and improving the detection performance of the model.

4.3.4. *Overall Performance analysis.* As shown in the Figure 11, we analyzed the performance of the overall improved YOLOv4 in this experiment.The AP@0.5 is 94.17%, AP@0.5:0.95 is 59.4%, which is 5.22% and 5.8% higher than the original YOLOv4, respectively. As shown in the yellow circle of Figure 11, for the same defect detection, the improved YOLOv4 model has better detection accuracy and some small defects that cannot be detected by YOLOv4 can be detected by the improved YOLOv4. In addition, in order to keep the detection speed of the model unchanged, we changed the residual module proportion of CSPNet from the original [1:2:8:8:4] to [1:2:4:4:8], and the model parameters only increased by less than 6m, but the detection accuracy of our improved model is higher. Therefore, our improved method is effective for detecting the defects of wine bottle caps.

4.3.5. *Analysis of performance comparison with other models.* This improved model proposed by us is aimed at improving the detection accuracy of small targets such as scatch defects and bottle cap fractures. We compare the improved YOLOv4 with some typical object detection models and the original YOLOv4 to verify its advantages.The comparison results are shown in the table 3. It can be seen from the table that the mAP of the improved YOLOv4 model is 10.64%, 7.23%, 13.89% and 5.22% higher than that of Fast R-CNN, YOLOv3, SSD and original YOLOv4, respectively.The detection speed is second only to YOLOv4, 0.001s faster than YOLOv3, 0.017s faster than Faster R-CNN, and 0.006s faster than SSD. In general, although the detection speed of the improved YOLOv4 model proposed by us is slightly lower than that of the original YOLOv4, it can meet the real-time requirements of industrial quality inspection and has better detection accuracy.

5. **Conclusions.** This paper presents a defect detection algorithm of wine bottle cap based on the improved YOLOv4 model. To reduce the missed detection rate, we improved the position loss function CIOU in the YOLOv4 model. To improve the accuracy of bottle cap defect detection, we introduce BAM into the CSP module of the backbone network to improve the network's attention to useful information. To reduce the loss of features in the process of feature information transmission, we designed a new detection head based on the residual feature augmentation module in the feature extraction layer.

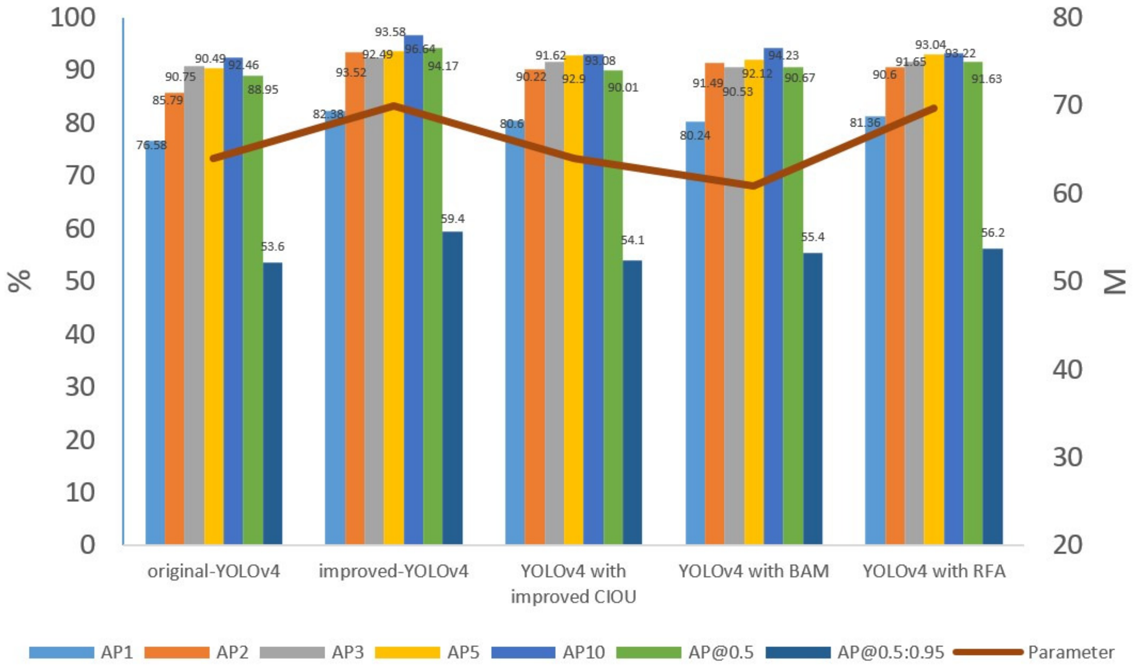


Figure 11. Overall performance comparison

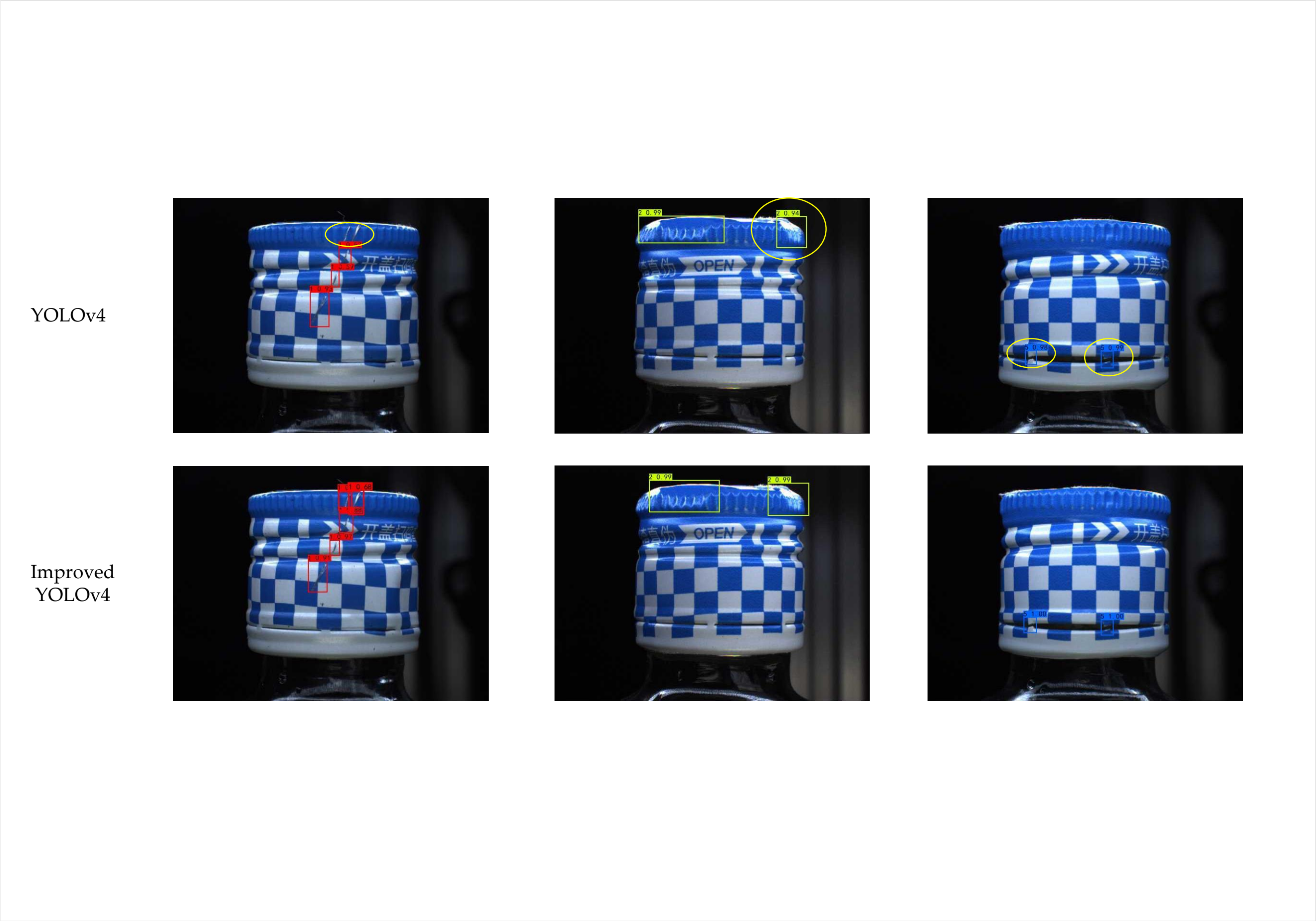


Figure 12. Comparison of test results

Table 3. Comparison table of various models

Model	mAP	Average time(s)	Input size
Faster R-CNN	83.52	0.101	416×416
YOLOv3	86.94	0.085	416×416
SSD	80.28	0.090	416×416
YOLOv4	88.95	0.079	416×416
Our model	94.17	0.084	416×416

In the bottle Baijiu defect detection data set of Alibaba Cloud Tianchi competition, 2661 images of bottle cap defects were collected, and the data set was expanded through image flipping, image flipping and other operations. The experiments on this data set show that the proposed model based on the improved YOLOv4 can be effectively applied to the industrial wine bottle cap quality inspection scene. The map of the improved YOLOv4 model reaches 94.17%, which is 5.22%, 7.23%, 13.89% and 10.64% higher than that of the original YOLOv4, YOLOv3, SSD and Faster R-CNN respectively. The detection speed of the improved YOLOv4 is 0.084s. Although it is slightly 0.005s lower than the original YOLOv4, it is 0.001s, 0.006s and 0.017s higher than that of YOLOv3, SSD and Faster R-CNN, and the detection accuracy is higher than that of the other four models. Therefore, the defect detection of wine bottle cap based on the improved YOLOv4 model proposed in this paper provides a technical reference for the convenience of industrial quality inspection.

Acknowledgment. This work is supported by The Key Project of Shandong Provincial Natural Science Foundation (ZR2020KE023), Natural Science Foundation of Shandong Province (ZR2021MD057) and Shandong Graduate Education Quality Improvement Plan Construction Project(SDYAL17034). At the same time, this paper thanks CSC for its support.

REFERENCES

[1] Q. Lin, "Added value for innovation of packaging design," *Packaging Engineering*, vol. 31, no. 106-109, 2010.

[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[3] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 595–12 604.

[4] E. K. Wang, C.-M. Chen, M. M. Hassan, and A. Almogren, "A deep learning based medical image segmentation technique in internet-of-medical-things domain," *Future Generation Computer Systems*, vol. 108, pp. 135–144, 2020.

[5] Y. Ma, Y. Peng, and 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*, no. Preprint, pp. 1–13, 2022.

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

[7] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 1137–1149, 2015.

[9] L. Yang, Y. Zhang, T. Wang, and T. Liu, "New method for steel surface defect detection based on improved faster r-cnn," *Journal of Jilin University(Information Science Edition)*, no. 409-415, 2021.

[10] Z. Geng and T. Gong, "Pcb surface defect detection based on improved faster r-cnn," *Modern Computer*, no. 19, pp. 89–93, 2021.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[13] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.

[14] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.

[16] Y. Chen, Q. Fu, and G. Wang, "Surface defect detection of nonburr cylinder liner based on improved yolov4," *Mobile Information Systems*, vol. 2021, pp. 1–13, 2021.

[17] L. Zhao, Z. Jiang, Z. Teng, and Y. Jia, "Fault detection method for insulators using improved yolov4," *Journal of Network Intelligence*, vol. 7, no. 4, pp. 2414–8105, 2022.

[18] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "Cspnet: A new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 390–391.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[20] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8759–8768.

[21] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 658–666.

[22] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: Bottleneck attention module," *arXiv preprint arXiv:1807.06514*, 2018.

[23] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[24] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.