



Research article

Detection of cigarette appearance defects based on improved YOLOv4

Guowu Yuan, Jiancheng Liu, Hongyu Liu, Yihai Ma, Hao Wu and Hao Zhou*

School of Information, Yunnan University, Kunming 650504, China

* **Correspondence:** Email: zhouhao@ynu.edu.cn; Tel: 8687165033748.

Abstract: Appearance defects are visible factors that affect the quality of cigarettes. Most of the consumer complaints received by tobacco companies are caused by appearance defects of cigarettes. Therefore, it is of great significance to reduce cigarettes with appearance defects. At present, tobacco factories mainly detect the appearance quality of cigarettes through manual sampling inspection. The manual method has low detection efficiency, it is difficult to unify the judgment standard, and it is easy to cause secondary pollution to cigarettes. According to the features of cigarette appearance defects, the YOLOv4 (You Only Look Once Version 4) model was improved for cigarette appearance defect detection. We have improved the following: 1) the channel attention mechanism was introduced into YOLOv4 to improve the detection precision; 2) the K-means++ algorithm was used to optimize the selection of clustering centers; 3) the spatial pyramid pooling (SPP) was replaced with atrous spatial pyramid pooling (ASPP) to improve the defect detection ability with different sizes; 4) the α -CIoU loss function was used to improve the detection precision. The mAP of our improved method reached 91.77%, the precision reached 93.32%, and the recall reached 88.81%. Compared with other models, our method has better comprehensive performance and better detection ability.

Keywords: appearance defect; cigarette; object detection; YOLOv4; deep learning

1. Introduction

The tobacco industry is an important industry in China and an important source of national and local fiscal revenue. In recent years, consumers have raised the requirements for cigarette quality. The appearance quality of cigarettes is most easily noticed by consumers. Therefore, tobacco companies need to reduce appearance defects and avoid cigarettes with appearance defects from entering the market.

At present, a high-speed cigarette production line can produce 150–200 cigarettes per second. With manual inspection of appearance defects, it has been difficult to meet the requirements. Tobacco companies are eager to automatically detect the appearance defects of cigarettes through computer vision. According to the detection results, the cigarettes with appearance defects can be automatically removed in the production line. Next, according to the statistical data of defect detection, the production line can be adjusted to reduce the probability of defective cigarettes. These operations can improve cigarette quality and reduce production costs.

With the development of deep learning, AlexNet [1], visual geometry group network (VGG16) [2], residual network (ResNet) [3] and other networks have been applied in many detection and classification applications. Automatic detection for product quality has been applied to bamboo strips, textiles, steel strips, circuit boards, etc. Gao et al. [4] used the improved CenterNet network to classify 10 appearance defects of bamboo strips, and the average detection accuracy (mean Average Precision, mAP) reached 76.9%. Liu [5] proposed a detection method based on improved Faster Regions with CNN features (R-CNN), which classified nearly 20 kinds of defects on cloth, and the mAP reached 63.4%. Ding et al. [6] added a hole convolution layer to the AlexNet network to increase the receptive field, and the average accuracy and average recall rate of cloth defect classification reached 85%. Kou et al. [7] proposed a Faster-RCNN-based steel strip defect detection model, FRDNet, which achieved a mAP of 67.7% on the GC10-DET steel strip defect data set, which was 4.9% higher than the original model. Xu et al. [8] applied the improved YOLOv3 model to the surface defect detection of steel plates, and the accuracy of the test set was improved by 23.3% compared with the original YOLOv3. Lawal [9] applied the spatial pyramid pool and mish activation function to YOLOv3, and the improved model improved the recognition accuracy of tomatoes to 96.4%. Roy et al. [10–12] added a dense module to the YOLOv4 backbone network and modified the PANet and activation functions. The improved model achieved fast speed and high accuracy in the detection of plant diseases and insect pests, in the detection of mango growth period and in the detection of apple diseases and insect pests.

Some scholars have also studied the detection and classification of cigarette appearance defects. Xiao [13] analyzed the area ratio of the incomplete part to judge the defect, but this method has a high rate of false detections. Li et al. [14] used the maximum contour area determination method to detect obvious appearance defects of cigarettes and then used template matching to detect cigarettes with small defects. Yuan et al. [15] proposed a classification method for cigarette appearance defects based on the ResNeSt model, and the classification accuracy reached 92.04%. However, only the classification was performed, and the location of the defects was not given. Liu et al. [16] proposed a detection method based on improved YOLOv5s for cigarette appearance defects, and the detection accuracy reached 90.9%. Liu et al. [17] proposed an improved CenterNet-based cigarette appearance defect detection method. The average detection accuracy mAP was 95.01%, but the detection speed is only 45 fps and needs to be further improved.

The cigarette samples are long and narrow images, and the defects belong to small targets. To obtain statistical information such as the category and location of cigarette defects, we regard it as a target detection problem. The classification models, such as VGG16, ResNet, Xception and EfficientNet, cannot locate the defects. The object detection model, such as YOLO, can not only classify the defects but also locate them. The location is helpful to reduce the probability of defective cigarettes by adjusting the production lines. YOLO is one of the most important target detection models, and it has advantages in precision and speed. In this paper, we improved YOLOv4 for

detecting the appearance defects of cigarettes. Our method improved the generation method of a priori boxes, then introduced the attention mechanism, replaced the spatial pyramid pooling (SPP) structure with the atrous spatial pyramid pooling (ASPP) structure and improved the loss function. The experimental results showed that our improved model achieved 91.77% mAP, 93.32% precision rate and 88.81% recall rate.

2. Overview of the appearance defects of cigarettes

During industrial production, some defective products will inevitably occur for various reasons. In the process of cigarette production, cigarette defects are related to many factors. For example, problems such as high-speed operation of the assembly line and poor quality of cigarette raw materials will cause cigarette defects.

The cigarette appearance images used in our experiments are from Yunnan branch of China Tobacco Industry Company, Limited. The images are captured by the high-speed industrial cameras on the automated production line. The front and back images of cigarettes can be captured at different positions of the production line. A standard cigarette is 84 mm in length and 7.8 mm in diameter. Therefore, the aspect ratio of the sample image is about 10:1.

Cigarettes are generally composed of two parts: the longer part with shredded tobacco is called the cigarette stick, and the shorter part with filter material is called the filter tip. According to the location and the appearance defect cause, appearance defects can be divided into four categories: “Dotted”, “Folds”, “Untooth” and “Unfilter.” “Untooth” refers to the misalignment of the wrapping paper of the cigarette during the production process, which is mainly caused by the production machine. “Folds” refers to a wrinkle-like shape on the cigarette, which is mainly caused by the production machine rolling the filter tip with the filter paper or rolling the cigarette with the cigarette stick paper. “Dotted” refers to spots of different sizes on the cigarette, which are mainly formed by unqualified printing of cigarette paper and filter tips, or dyeing in the later stage. “Unfilter” defects are mainly caused by running out of filter paper or a failure of the production machine to pack the filter paper. Images of the appearance defects of cigarettes are shown in Figure 1. We define normal cigarettes as those without appearance defects; see Figure 2.

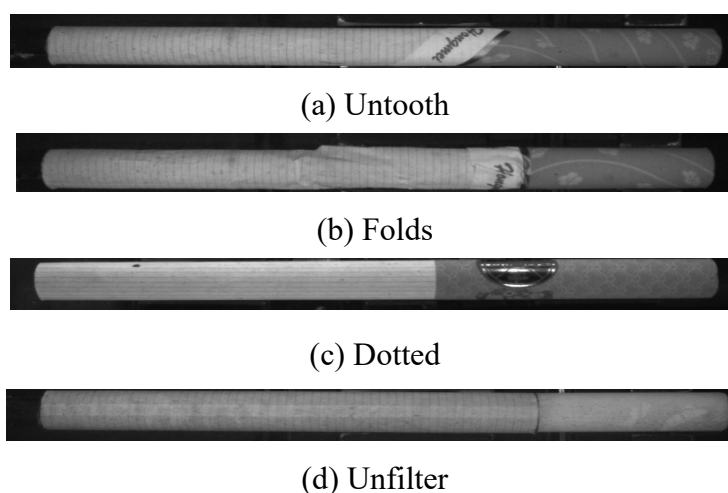


Figure 1. Cigarettes with a defective appearance.

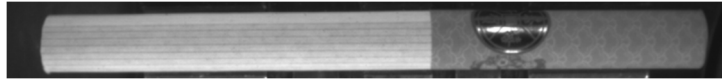


Figure 2. Cigarette with a normal appearance.

3. Materials and methods

3.1. Introduction to the YOLOv4 model structure

The YOLO network was proposed by Redmon et al. [18] in 2016. YOLOv2, YOLOv3 and YOLOv4 are improved versions. Through comparative experiments on cigarette defect datasets, we have found that YOLOv4 is more effective than others. Therefore, we chose YOLOv4 for defect detection.

The YOLOv4 network can be divided into four parts: the backbone feature extraction network, SPP structure [19], path aggregation network (PANet) [20] and detection head. The network structure of YOLOv4 is shown in Figure 3.

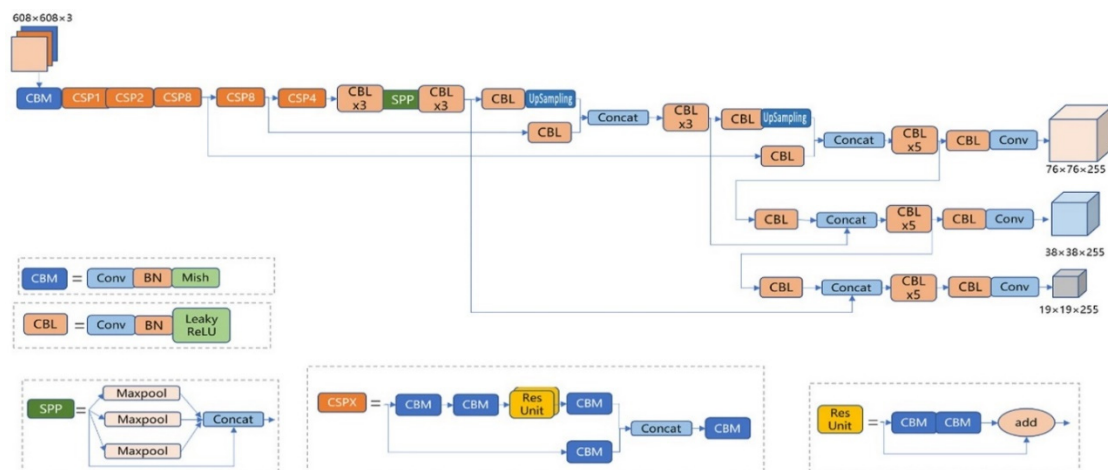


Figure 3. The network structure of YOLOv4.

3.2. Adding the channel attention mechanism module

The principle of channel attention mechanism is similar to how people can focus on an important feature when they look at pictures [21]. This method can improve the effectiveness of feature extraction. In computer vision, it can better learn the relevant features to improve the detection accuracy of the target.

The YOLOv4 model cannot automatically learn the importance of different channel features, and it cannot make full use of the extracted features. These disadvantages affect the classification and regression effect. We integrate a channel attention mechanism (SENet) into the backbones of YOLOv4, which can better focus on the relationship between different channels of the feature map, thus improving the effectiveness of feature extraction. This module can improve the detection accuracy of cigarette appearance defects. The SENet structure is shown in Figure 4:

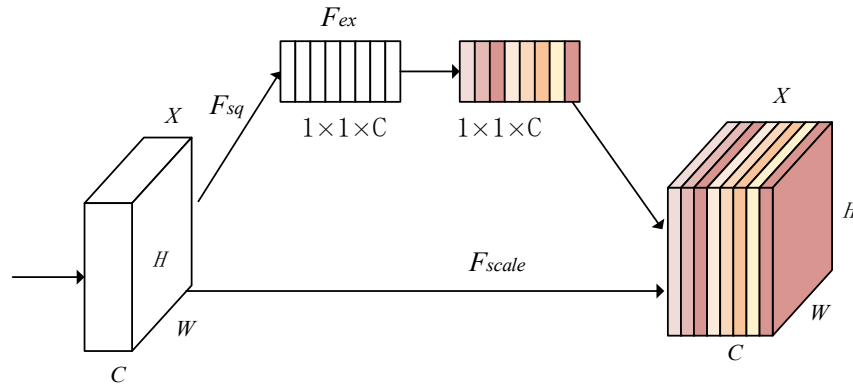


Figure 4. SENet structure.

As shown in Figure 4, the channel attention module mainly included three processes: squeeze, excitation and scale. The SENet module gives more weight to important information and less weight to unimportant information. This can save resources, quickly obtain the most effective information and make better use of image features.

3.3. Prior box

The prior box is a rectangle designed according to the common sizes and proportions of the detected objects. It has a significant impact on the accurate prediction of the targets. The YOLOv4 model has a large difference in the prior box size, and the size is suitable for microsoft common objects in context (COCO), visual object classes (VOC) and other datasets but not for small targets, such as cigarette appearance defects. To make the prior box more suitable for cigarette appearance defects, we introduced K-means++ clustering to further adjust the size of the prior box. When the K-means++ algorithm selects the initial cluster center, the distance between the cluster centers should be as far as possible. The algorithm can make the model obtain the optimal prior box and improve detection accuracy.

The steps of the K-means++ algorithm are as follows:

Algorithm 1: K-means++ clustering algorithm

Input: Data set $R = \{x_1, x_2, \dots, x_n\}$, n is the number of data

Output: Cluster center points $\{c_1, c_2, \dots, c_k\}$, k is the number of center points

Algorithm steps:

- 1) Select one point randomly from R as the initial cluster center point c_1 ;
 - 2) The minimum distance $d(x_i)$ of each sample from the nearest cluster center is calculated;
 - 3) Calculate the probability $P(x_i) = \frac{d(x_i)^2}{\sum_{x_j \in R} d(x_j)^2}$ that each datum x_i is selected as the next cluster center;
 - 4) The datum with the largest probability $P(x_i)$ is selected as the next cluster center;
 - 5) It ends when k cluster centers are selected; otherwise, jump to 2);
 - 6) Clustering is done according to the classical K-means algorithm, until convergence.
-

3.4. ASPP

SPPNet extracts the features through several multi-scale pooling operations, and it then combines the features and inputs them into the later fully connected layer. SPP was used in the YOLOv4 algorithm. This method can obtain the features of different receptive fields, but the features cannot reflect the grammatical relationship between the local and the overall. Therefore, we have adopted the ASPP [22] module, which can gather the features of multi-scale context and improve the detection ability for different sizes targets.

Atrous spatial pyramid pooling network (ASPPNet) was proposed in 2017, and it uses atrous convolution [23]. The principle of atrous convolution is shown in Figure 5. It effectively increases the receptive field. Its implementation uses the dilation rate. Figure 5(a) is a schematic diagram of an ordinary convolution (dilation rate = 1), and its receptive field after convolution is 3. Figure 5(b) is a schematic comparison of a dilated convolution (dilation rate = 2), and its receptive field after convolution was 5.

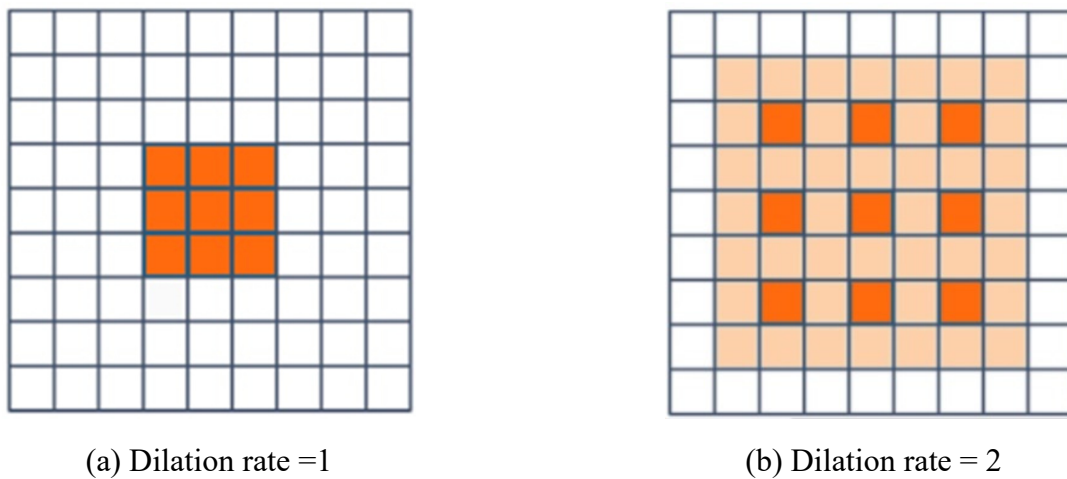


Figure 5. Schematic diagram of the dilated convolution.

The size of the dilated convolution kernel k' can be calculated by Eq (1):

$$k' = k + (k - 1)(r - 1) \quad (1)$$

where k is the size of the initial convolution kernel, and r is the dilation rate.

The size of the corresponding receptive field f_m can be calculated by Eq (2):

$$f_m = f_{m+1} + (k' - 1) \prod_{i=1}^{m-1} S_i \quad (2)$$

where m refers to m layer, and S_i refers to the step length of i layer.

ASPPNet inputs the image into several dilated convolutional layers with different dilation rates, as shown in Figure 6. Then, the feature obtained by these convolutions is fused with the result of the

input image after global average pooling. This method can effectively extend the feature channel.

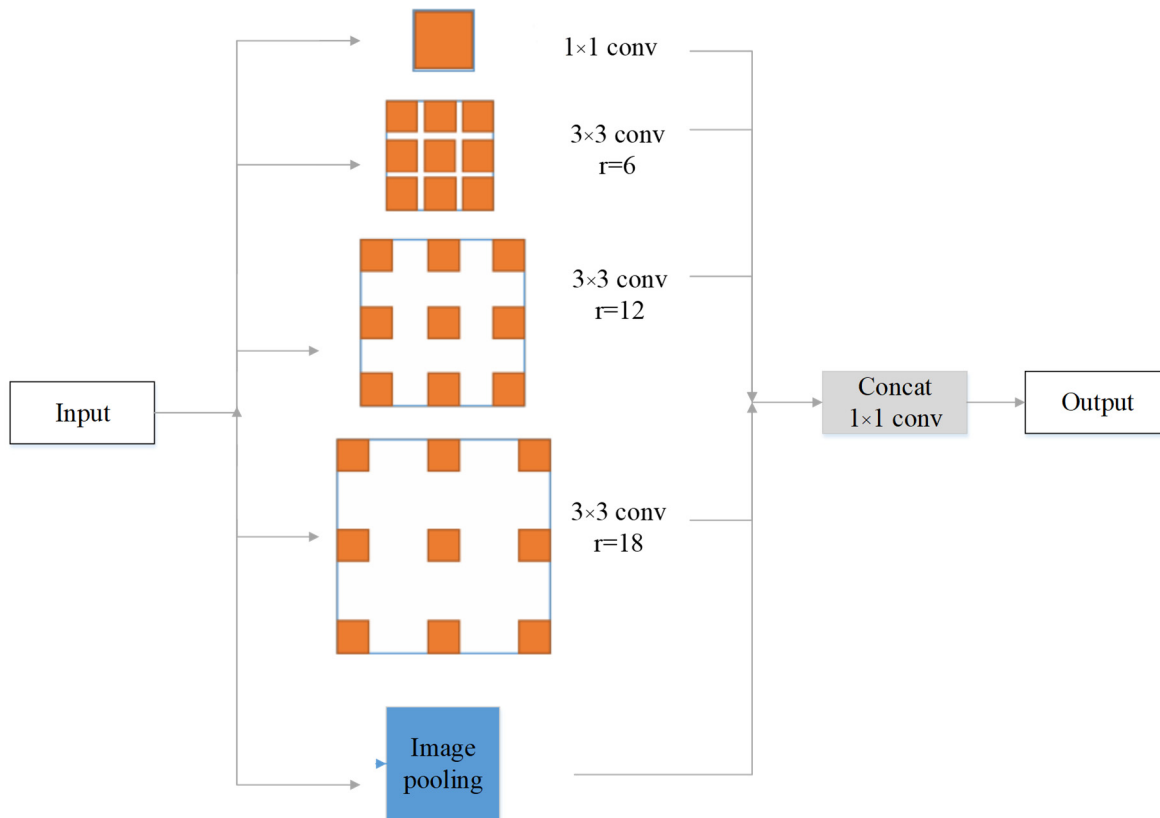


Figure 6. Schematic diagram of the ASPP structure applied in our model.

We replaced the SPP in the YOLOv4 with the ASPP. The ASPP can reduce the missed detection rate of cigarette appearance defects and learn the characteristics of defect targets with different sizes by increasing the receptive field.

3.5. Improvement of the loss function

The choice of the loss function has a certain impact on the performance of the network, such as affecting the convergence of the loss function and the detection accuracy of the model. In the YOLOv4 network, the complete intersection over union (CIoU) is used to define the loss function. This loss function was proposed by Zheng et al. [24], along with distance intersection over union (DIoU). In the field of target detection, the most basic loss functions are the intersection over union (IoU) and the generalized intersection over union (GIoU) [25].

The IoU calculates the intersection and parallel ratio of the prediction box and the real box. Using the IoU, the loss function L_{IoU} is calculated as follows:

$$L_{IoU} = 1 - IoU = 1 - \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

where A represents the area of the real box, and B represents the area of the prediction box.

The IoU calculates the intersection ratio, which can reflect the detection effect. However, the IoU has many shortcomings. The GIoU, DIoU and CIoU were proposed on the basis of the IoU. The GIoU added the closure area as a penalty item. Furthermore, the DIoU considered the Euclidean distance between the center points of the prediction box and the regression box on the basis of the GIoU. The final CIoU considered the aspect ratio on the basis of DIoU.

Using the CIoU, the loss function L_{CIoU} is calculated as follows:

$$L_{CIoU} = 1 - CIoU = 1 - \frac{|A \cap B|}{|A \cup B|} + \frac{\rho^2(b, b^{gt})}{c^2} + \beta v \quad (4)$$

where b and b^{gt} represent, respectively, the central point of the prediction box and the real box, ρ represents the Euclidean distance of the two central points, c represents the diagonal distance of the minimum closure region containing both the prediction box and the real box, β is the weight function, and v is used to measure the similarity of the aspect ratio. β and v are calculated as follows:

$$\beta = \frac{v}{(1 - IoU) + v} \quad (5)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (6)$$

where w^{gt} and h^{gt} represent the width and height of the real box, and w and h represent the width and height of the prediction box.

He et al. [26] proposed the α -CIoU in 2021, which is an improvement on the CIoU. The loss function $L_{\alpha-CIoU}$ is calculated as follows:

$$L_{\alpha-CIoU} = 1 - CIoU^\alpha \quad (7)$$

where α generally takes an integer greater than 1.

This power α can increase the gradient of IoU to improve the regression accuracy. In this paper, we adopt $L_{\alpha-CIoU}$ to improve the detection accuracy. In the experiment, by adjusting with different values of α , the detection accuracy will change. In Section 4.7, the experimental results are presented.

3.6. Improved YOLOv4 model structure

After the above four improvements, the model structure of our network is shown in Figure 7. Compared with Figure 3, we replaced the SPP structure in the YOLOv4 network with the ASPP structure, and we then added the SENet module to the PANet structure. These enable our model to better extract the features of different-size images and pay more attention to important features. The prior box selection and loss function in the network are also improved to make them more suitable for the cigarette appearance defect dataset.

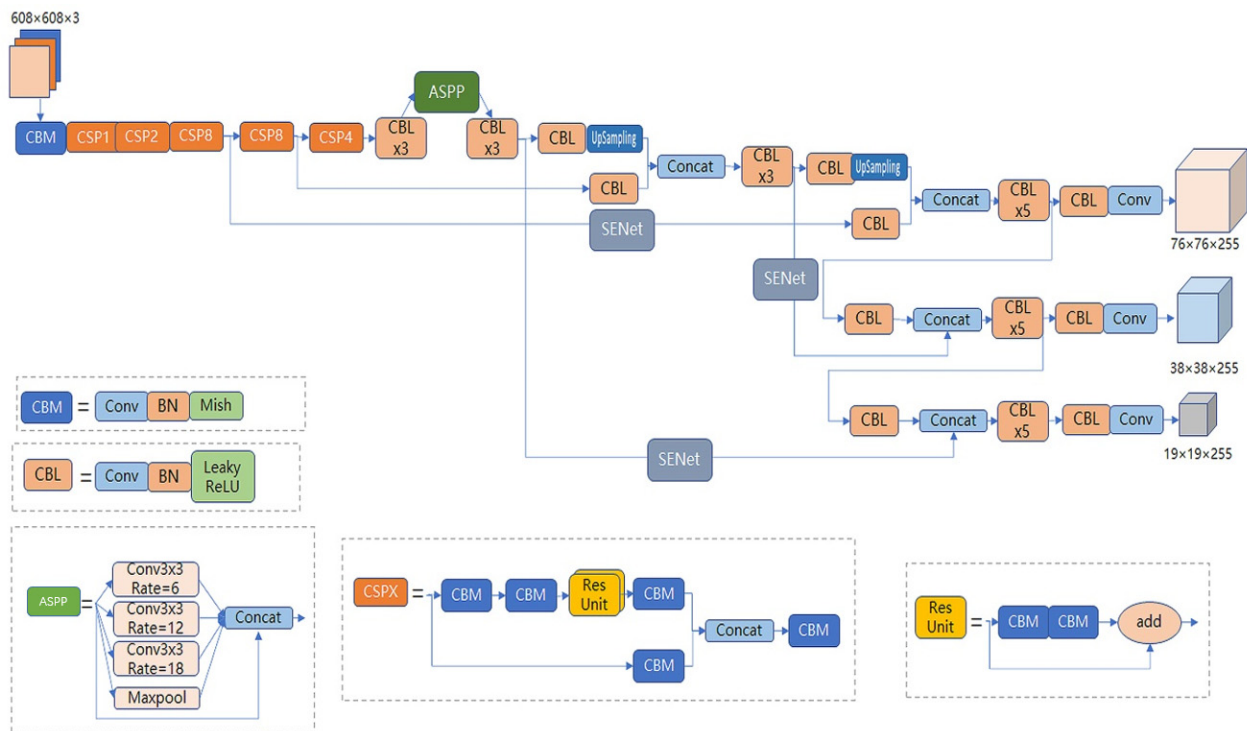


Figure 7. Schematic diagram of our improved model.

4. Results and discussion

4.1. Experimental dataset

The image dataset of cigarette appearance defects used in the experiment was from the Yunnan branch of China Tobacco Industry Company, Limited. The images were captured by high-speed industrial cameras on the automated production line, and they were grayscale images.

After data enhancement, the dataset contained 16,200 images, including “Normal”, “Dotted”, “Folds”, “Untooth” and “Unfilter.” First, all images were labeled, and then they were randomly divided according to a ratio of about 6:2:2. The numbers of each category are shown in Table 1.

Table 1. Statistics of the cigarette appearance image dataset.

Appearance type	Training set	Validation set	Test set
Normal	1800	700	540
Dotted	2100	640	560
Unfilter	2000	650	700
Folds	1900	650	710
Untooth	1920	600	730

As can be seen from Table 1, after data enhancement, the samples in all categories were roughly

balanced, which is more conducive to network training.

4.2. Training process analysis

Figure 8 shows the loss function curves for our YOLOv4 model and the original YOLOv4 model under equal conditions. In this paper, an epoch was set to 300. From Figure 13, the original model only converged when the epoch was about 235, while our improved model converged when the epoch was about 225. Therefore, it can be seen that our training time was shorter, the loss value was lower, and the effect of our model was better.

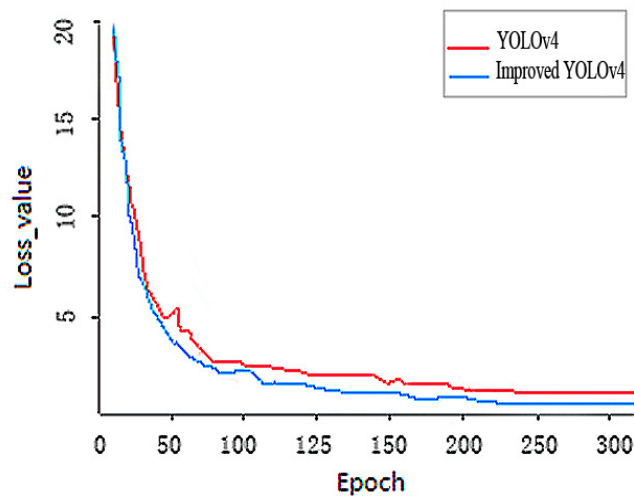


Figure 8. Comparison of loss curve.

4.3. Experiment configuration and evaluation index

In the experimental software, the operating system was Windows 10, the programming platform was PyCharm, and the architecture was based on PyTorch. For the hardware, the CPU was an Intel Core i7-10700k, the memory was 32 GB, and the GPU was an RTX 2080Ti. During training, the batch size was 16, the iterations epoch was 300, and the learning rate was 10^{-4} .

The evaluation indexes used in the experiments were accuracy, precision, recall, average precision (AP), mAP and processing frames per second (FPS). The accuracy (A), precision (P) and recall (R) are as follows:

$$A = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (8)$$

$$P = \frac{T_p}{T_p + F_p} \quad (9)$$

$$R = \frac{T_p}{T_p + F_N} \quad (10)$$

where T_p is the number of samples that were positive and also correctly classified as positive, and F_p is the number of samples that were negative but incorrectly classified as positive. T_N is the number of samples that were negative and also correctly classified as negative, and F_N is the number of samples that were positive but classified as negative.

After obtaining the P and R of each category, a precision-recall (P-R) curve can be shown. AP is represented by the area surrounded by the P-R curve and coordinates, and mAP is the average of the AP values of all categories. The AP and mAP are calculated as follows:

$$AP = \int_0^1 PRdR \quad (11)$$

$$mAP = \frac{1}{N} \sum_{k=1}^N AP(k) \quad (12)$$

where N represents the total number of categories, and $AP(k)$ represents the AP of the category k .

4.4. Experimental comparison of defect detection effects

Figures 9–13 show the detection effects of five cigarette appearance types. Figure 9 shows the detection results of normal cigarettes. Since normal cigarettes have no special features, to identify more features, the labeling frame was arbitrarily marked, so the detection frame was also distributed in any possible position. This allowed us to compare and maximize the elimination of irrelevant parameters, so that the network model could learn better results, and the appearance of the cigarettes could be better detected. As shown in Figure 9, the detection confidence of the original method was 0.77, and it was 0.86 after the improvement. The detection confidence of the original algorithm in Figure 10 was 0.90, and it was 0.98 after the improvement. In Figure 11, the original algorithm missed detection, and it was 1.0 after the improvement. The detection confidence of the original algorithm in Figure 12 was 0.84, and it was 1.0 after the improvement. The detection confidence of the original algorithm in Figure 13 was 0.95, and it was 0.96 after the improvement.

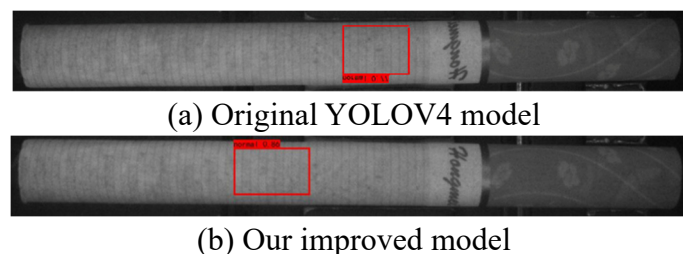
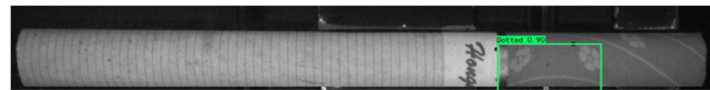
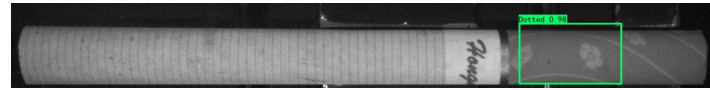


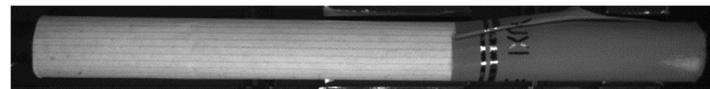
Figure 9. Comparison of the original model and our improved model in the normal type.



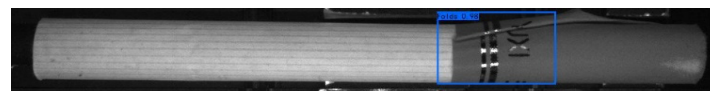
(a) Original YOLOV4 model



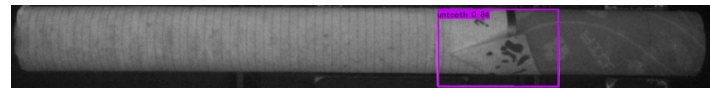
(b) Our improved model

Figure 10. Comparison of the original model and our improved model in the dotted type.

(a) Original YOLOV4 model



(b) Our improved model

Figure 11. Comparison of the original model and our improved model in the folds type.

(a) Original YOLOV4 model



(b) Our improved model

Figure 12. Comparison of the original model and our improved model in the untooth type.

In general, the original algorithm had a low detection accuracy for cigarette appearance defects, and there was leakage detection, while the improved algorithm had a higher detection accuracy of cigarette appearance defects, and the positioning was more accurate, while leakage detection and error inspection were rare.

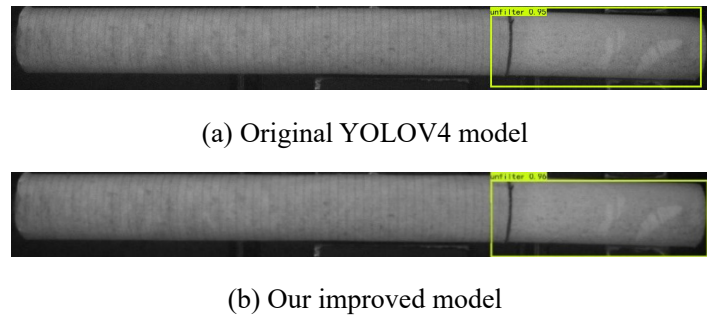


Figure 13. Comparison of the original model and our improved model in the unfilter type.

4.5. Comparison of detection accuracy in various cigarette samples

The various types of defect detection by our improved model are shown in Figure 14. The AP of “normal” and “untooth” was low, and other defects were nearly 100%. The main reasons are that several defects with high AP are more obvious, while normal cigarettes have no obvious characteristics and thus have low AP.

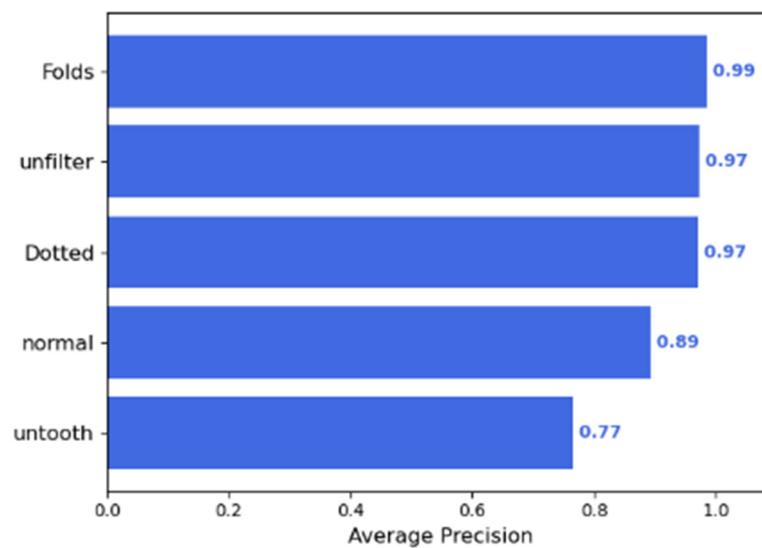
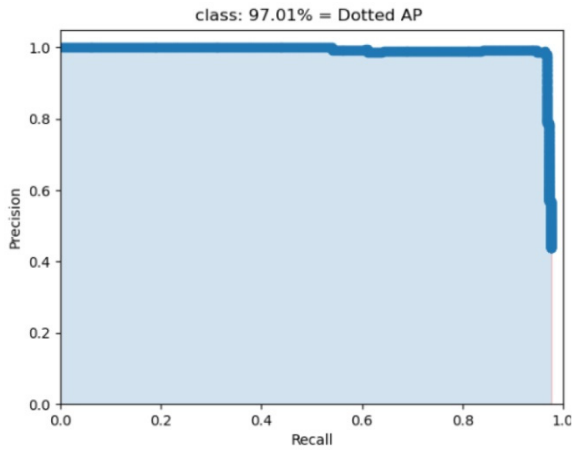


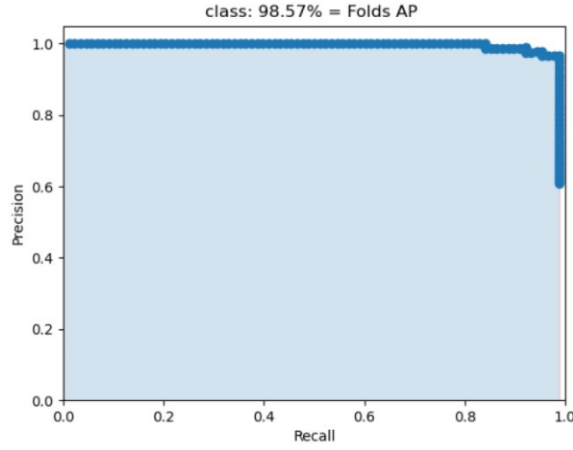
Figure 14. The AP of our improved model in various cigarette samples.

The P-R curves of defect detection by our improved model are shown in Figure 15. It is obvious that our improved model has achieved good detection results in the classes of “dotted,” “folds” and “unfilter” but unsatisfactory detection performance for “normal” and “untooth.”

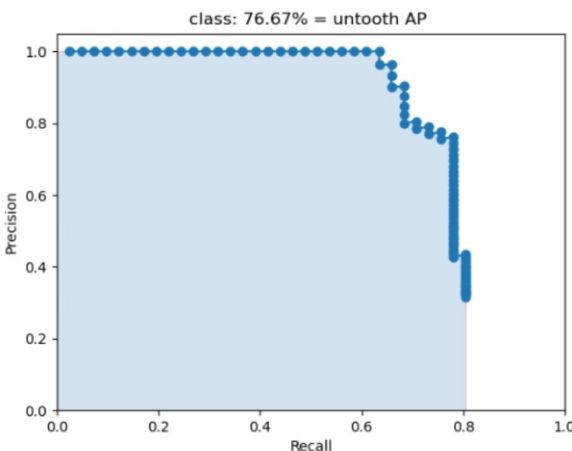
Figures 16 and 17 show the precision and recall curves of our improved model. The “Dotted,” “Folds,” and “Unfilter” types had higher precision and recall rates, while the “Untooth” and “normal” types had lower precision and recall rates.



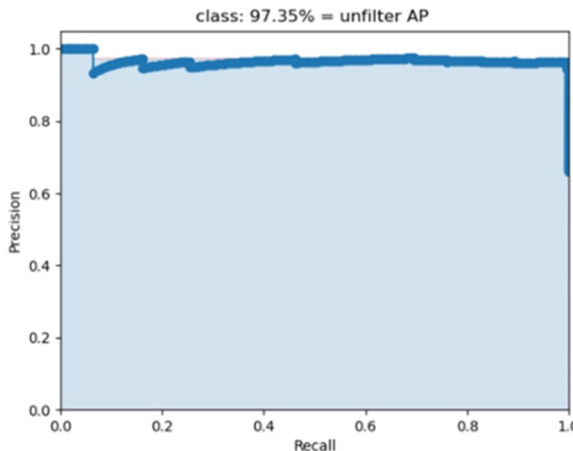
(a) Dotted



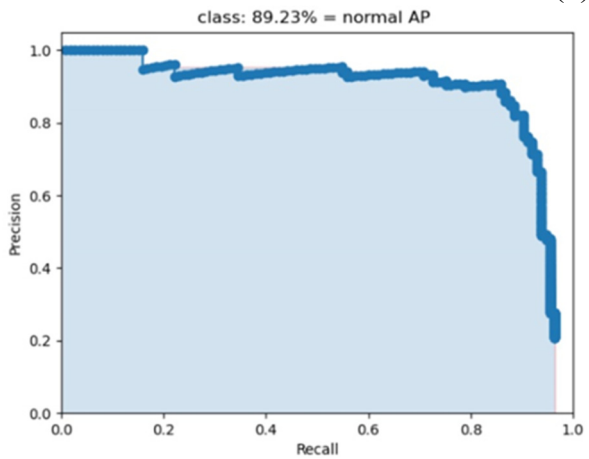
(b) Folds



(c) Untooth

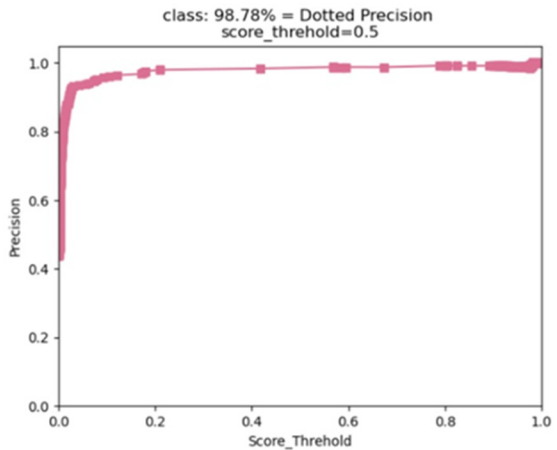


(d) Unfilter

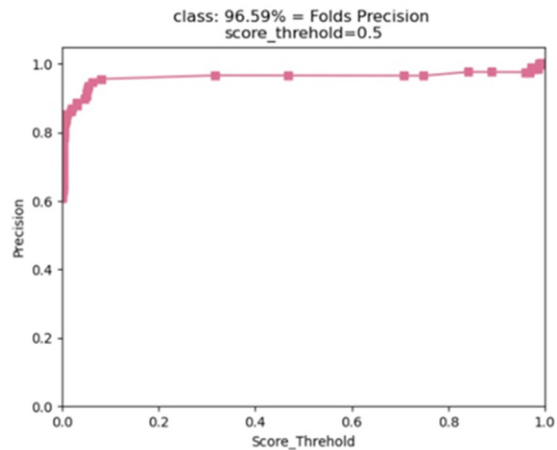


(e) Normal

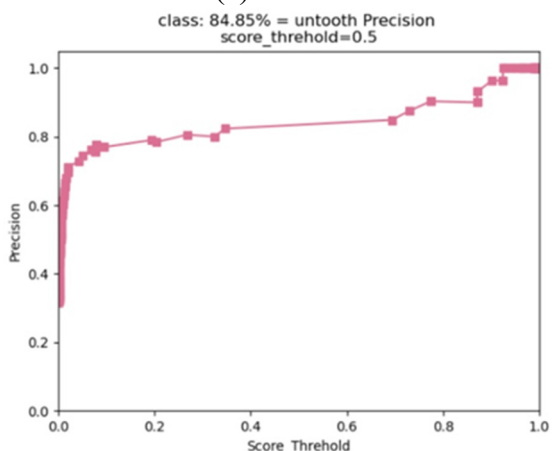
Figure 15. The P-R curves of our improved model.



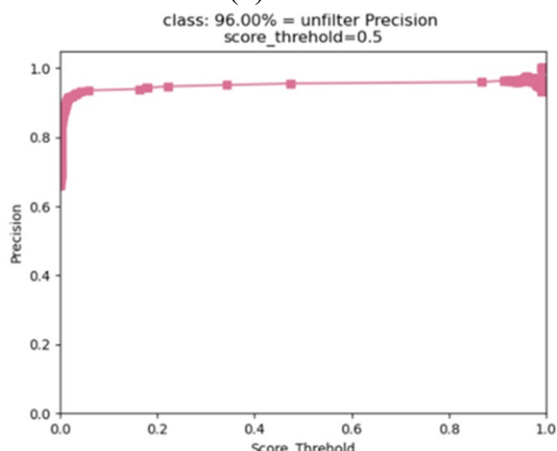
(a) Dotted



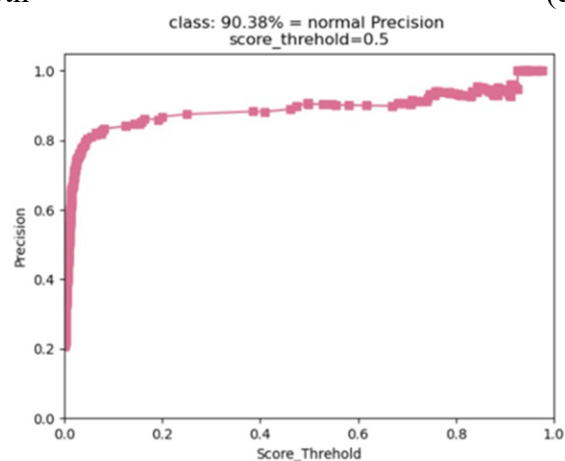
(b) Folds



(c) Untooth



(d) Unfilter



(e) Normal

Figure 16. Precision curves of our improved model.

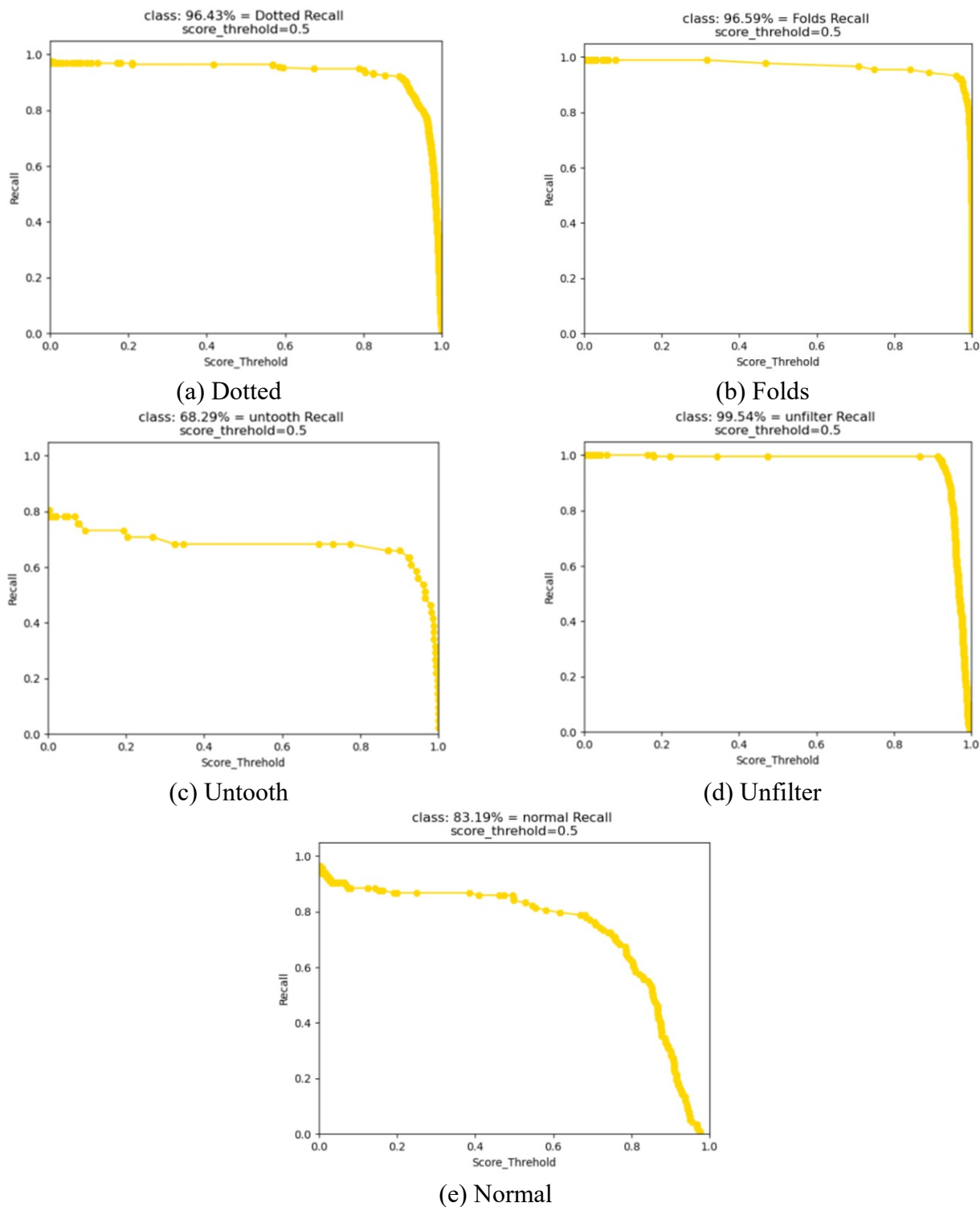


Figure 17. The recall curves of our improved model.

4.6. Experimental comparison of data augmentation

Mosaic is a built-in data augmentation of the YOLOv4. This method randomly cuts four images and splices them into a new image. New images are used as training data. Because the aspect ratio of the cigarette image is about 10:1, we found that the mosaic is not suitable for the cigarette dataset.

Therefore, we used additional data augmentation, such as image inversion, Gaussian blur, horizontal mirror inversion, affine transformation and brightness transformation.

In the four experiments shown in Table 2, we compared the mAP of our improved model with YOLOv4's using four data enhancement methods. Among them, the data augmentation of experiment 1 is mosaic, experiment 2 had no data augmentation, experiment 3 had our data augmentation, and experiment 4 combined the mosaic method and our data augmentation.

From Table 2, we can see that the two models' *mAP* values both decreased when the mosaic data enhancement was added. When only using our data augmentation method, the effect is the best. So, the original Mosaic was removed. Therefore, we replaced the mosaic with our data augmentation in the experiment.

Table 2. Comparison of data augmentation in the YOLOv4 model.

Experiment	Data augmentation	mAP /% (original YOLOv4)	mAP /% (Our improved YOLOv4)
1	Mosaic	85.05	88.92
2	No data augmentation	86.32	90.58
3	Our data augmentation	87.71	91.77
4	Mosaic + Our data augmentation	86.89	90.93

4.7. Comparison of different powers α in the loss function

In Formula 7 of Section 4.6, the loss function $L_{\alpha-CIoU}$ is calculated using the power α . Table 3 shows comparative experiments on different powers α in the loss function.

Table 3. Comparison of detection performance using different powers α in the loss function.

α	Precision/%	Recall/%	mAP/%
1	92.38	88.29	91.23
2	92.43	88.10	91.52
3	93.87	88.81	91.77
4	92.41	88.13	91.46

We found that the detection performance is the best when the power α is 3. Therefore, we finally chose L_{3-CIoU} as the loss function of this study.

4.8. Ablation experiment

Table 4 shows the ablation experiment results. These experiments were based on the YOLOV4 which replaced the mosaic with our data augmentation.

Table 4. Comparison of the results of adding different modules.

Experiment	K-means++	SENet	ASPP	3-CIoU	mAP/%
1	–	–	–	–	87.71
2	√	–	–	–	88.02
3	√	√	–	–	89.99
4	√	√	√	–	91.23
5	√	√	√	√	91.77

First, experiment 1 was the YOLOV4 with our data augmentation, and its mAP is 87.71%. Second, when the K-means++ algorithm was introduced to select the initial cluster center, the model could obtain the optimal prior box, which led to a 0.31% rise in mAP. Third, when the SENet module was introduced to learn the relevant features, it led to a 1.97% rise in mAP. Fourth, the ASPP module was introduced, and it led to a 1.24% rise in mAP. Fifth, the 3-CIoU was introduced, and it led to a 0.54% rise in mAP. It can be seen from Table 5 that the mAP was improved when the four different modules were gradually added.

4.9. Comparison experiment with other models

To verify the progressiveness of our improved model, we compared the main performance index with other models on the cigarette appearance image dataset, such as Faster R-CNN, YOLOv4, YOLOv5, YOLOP, YOLOX, SSD and CenterNet. The experimental results are shown in Table 5.

Table 5. Comparison experiment with other models.

Algorithm	Precision/%	Recall/%	mAP/%	FPS
Faster R-CNN	83.12	82.65	82.83	39
YOLOv4	87.87	84.76	86.89	60
YOLOv5	89.89	82.98	90.73	76
YOLOP	89.23	81.25	84.31	84
YOLOX	89.91	83.02	90.75	77
SSD	81.47	79.91	81.03	47
CenterNet	92.65	78.30	88.87	54
Ours	93.32	88.81	91.77	53

It can be concluded from the results in Table 5 that our model is the best in precision, recall and mAP, but the average detection speed is not optimal, and it is slower than YOLOv4, YOLOv5, YOLOP, YOLOX and CenterNet.

In the detection of cigarette appearance defects, detection accuracy is the most important. Because a high-speed cigarette production line can produce 150-200 cigarettes per second, all models above cannot achieve real-time detection in our experimental software and hardware platform. In our experimental platform, the CPU is an Intel Core i7-10700k, the memory is 32 GB, and the GPU is NVIDIA GeForce RTX 2080Ti. Due to experimental conditions, we cannot test our

model in a better hardware environment. On better hardware, such as NVIDIA GeForce RTX 3090Ti and NVIDIA H100, we believe that detection speed can be improved.

5. Conclusions

This paper proposed a defect detection method for the cigarette appearance dataset. The main work aimed to discuss how to optimize the network of the original YOLOv4 algorithm and improve the detection accuracy of the model.

In this paper, ASPP is used instead of SPP, and an SE attention mechanism is added to the network to help extract features. Then, we replaced K-means with K-means++ and replaced the Mish activation function with α -CIoU activation function to improve convergence speed and detection accuracy. Finally, according to the characteristics of the cigarette data set, the Mosaic data enhancement method of the original model was replaced. The ablation experiment shows that the improvement in this paper has a positive contribution to the accuracy improvement of YOLOv4 on the cigarette data set. Comparative experiments show that the improved model achieved 91.77% mAP, 93.32% precision and 88.81% recall on the cigarette data set.

The method proposed in this paper is helpful to control the outflow of defective cigarettes and to improve factory efficiency. It can further replace traditional manual detection methods, improve large-scale industrial production efficiency and further realize automatic detection.

Our improved model has a significant improvement in various accuracy index, but the detection speed is not optimal. In the future, we will further improve the model under the premise of ensuring accuracy. We will mainly focus on reducing the amount of calculation and model size and improving the detection speed. For example, the convolution is replaced by a depthwise separable convolution, and the backbone network is CSP Darknet53, which can be replaced with a lighter network. If we are lucky to have new scientific research funding, we will also update the experimental equipment and improve the detection speed of our method.

Acknowledgments

This research was funded by the Natural Science Foundation of China (Grant No. 62061049, 12263008), the Yunnan Provincial Department of Science and Technology-Yunnan University Joint Special Project for Double-Class Construction (Grant No. 202201BF070001-005), the Key R&D Projects in the Yunnan Province (Grant No. 202202AD080004) and the Application and Foundation Project of the Yunnan Province (Grant No. 202001BB050032), with the participation and funding of the Practice Innovation Fund Project for Professional Degree Graduates of Yunnan University (Grant No. ZC-22221382).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM*, **60** (2017), 84–90. <https://doi.org/10.1145/3065386>
2. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, preprint, arXiv: 1409.1556.
3. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
4. Q. Q. Gao, B. C. Huang, W. Z. Liu, T. Tong, Detection method of bamboo strip surface defects based on improved CenterNet, *J. Comput. Appl.*, **31** (2020), 1–8. <https://doi.org/10.11772/j.issn.1001-9081.2020081167>
5. Y. Y. Liu, *Research on cloth defect detection method based on deep learning*, Master thesis, Harbin Institute of Technology in Harbin, 2020.
6. G. X. Ding, H. Huang, Y. Ma, Automatic detection of cloth defects based on laws texture filtering, in *Proceedings of 2019 2nd International Conference on Intelligent Systems Research and Mechatronics Engineering (ISRME 2019)*, (2019), 148–152.
7. X. P. Kou, S. J. Liu, Z. R. Ma, Steel strip defect detection method based on Faster-RCNN, *China Metall.*, **31** (2021), 77–83. <https://doi.org/10.13228/j.boyuan.issn1006-9356.20200506>
8. Q. Xu, H. J. Zhu, H. H. Fan, H. Y. Zhou, G. H. Yu, Study on detection of steel plate surface defects by improved YOLOv3 network, *Comput. Eng. Appl.*, **56** (2020), 265–272. <https://doi.org/10.3778/j.issn.1002-8331.2003-0232>
9. M. O. Lawal, Tomato detection based on modified YOLOv3 framework. *Sci. Rep.*, **11** (2021), 1447. <https://doi.org/10.1038/s41598-021-81216-5>
10. A. M. Roy, R. Bose, J. Bhaduri, A fast accurate fine-grain object detection model based on YOLOv4 deep neural network, *Neural Comput Appl.*, **34** (2022), 3895–3921. <https://doi.org/10.1007/s00521-021-06651-x>
11. A. M. Roy, J. Bhaduri, Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4, *Comput. Electron. Agr.*, **193** (2022), 106694. <https://doi.org/10.1016/j.compag.2022.106694>
12. A. M. Roy, J. Bhaduri. A deep learning enabled multi-class plant disease detection model based on computer vision, *AI*, **2** (2021), 413–428. <https://doi.org/10.3390/ai2030026>
13. Z. Y. Xiao, *Research and Implementation of Cigarette Defect Detection Algorithm*, Master Thesis, Yunnan University in Kunming, 2018.
14. J. Li, H. H. Lu, X. Wang, J. H. Hong, S. Wang, L. X. Shen, et al., Online inspection system for cigarette tipping quality based on machine vision, *Tob. Sci. Technol.*, **52** (2019), 109–114. <https://doi.org/10.16135/j.issn1002-0861.2018.0562>
15. G. W. Yuan, J. C. Liu, H. Y. Liu, R. Qu, H. Zhou, Classification of cigarette appearance defects based on ResNeSt, *J. Yunnan Univ.: Nat. Sci. Ed.*, **44** (2022), 464–470. <https://doi.org/10.7540/j.ynu.20210257>
16. H. Y. Liu, G. W. Yuan, Cigarette appearance defect detection method based on improved YOLOv5s, *Comput. Technol. Dev.*, **32** (2022), 161–167. <https://doi.org/10.3969/j.issn.1673-629X.2022.08.026>

17. H. Y. Liu, G. W. Yuan, L. Yang, K. X. Liu, H. Zhou, An appearance defect detection method for cigarettes based on C-CenterNet, *Electronics*, **11** (2022), 2182. <https://doi.org/10.3390/electronics11142182>
18. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 779–788, <https://doi.org/10.1109/CVPR.2016.91>
19. K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, *IEEE Trans Pattern Anal. Mach. Intell.*, **37** (2015), 1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
20. S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 8759–8768. <https://doi.org/10.1109/CVPR.2018.00913>
21. J. Hu, L. Shen, G. Sun. Squeeze-and-excitation networks, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>
22. M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang, DenseASPP for semantic segmentation in street scenes, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2018), 3684–3692. <https://doi.org/10.1109/CVPR.2018.00388>
23. F. Yu, V. Koltun., Multi-scale context aggregation by dilated convolutions, preprint, arXiv:1511.07122.
24. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2020), 12993–13000. <https://doi.org/10.1609/aaai.v34i07.6999>
25. H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2019), 658–666. <https://doi.org/10.1109/CVPR.2019.00075>
26. J. He, S. Erfani, X. Ma, J. Bailey, Y. Chi, X. Hua, α -IoU: A family of power intersection over union losses for bounding box regression, preprint, arXiv:2110.13675



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)