Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 1, No. 1, 2022

Research on Detection Algorithm of Tablet Surface Defect Based on Yolov3

Yao Sun*, Shiquan Shao

Electrical Engineering, Southwest Minzu University, Chengdu, Sichuan, 610041, China

Abstract: Tablet surface defect detection is an important part of tablet quality inspection, and manual detection and traditional pattern recognition methods are difficult to achieve the expected results. In this regard, this paper proposes a method for detecting tablet defects based on YOLOV3, which firstly uses industrial cameras to complete the collection of tablet defect images and creates a data set; then uses Daeknet-53 as the backbone network to initially extract features; secondly, constructs FPN feature pyramid enhancement Extraction of features; finally use yolo-head to obtain prediction results; detection mAP reaches 92.97% on the test set of the self-built data set. The experimental results show that the method has certain applicability and feasibility.

Keywords: YOLOV3, Target detection, Tablet surface defect detection.

1. Introduction

Tablets go through a series of complex assembly line processes from production to packaging, usually including ingredients, sieving, mixing, granulation, drying, whole granulation, total mixing, tableting, coating, packaging and other steps [1]. In the entire production line, it is inevitable that there will be tablet defects, impurities and foreign matter mixed into it, etc., which will affect the quality of drugs and lead to increased production costs.

At present, the detection of surface defects of tablets in my country is mainly divided into manual detection and traditional machine vision detection. Manual detection often leads to eye fatigue due to long-term work in this method, the detection accuracy is reduced, and manual removal is labor-intensive, low-efficiency, and prone to false detection and missed detection [2]; traditional machine vision The extraction effect of the technology on the feature is difficult to achieve the expected effect, and this method relies on the manual setting of the discriminant threshold, or on the manual selection of pattern features. Either way, operators are required to have rich professional knowledge, which is more professional for production practitioners.

With the development of deep learning [3] in recent years, object detection technology has been widely used in various fields. Deep learning detection algorithms are mainly divided into two categories in terms of ideas: one is the two-stage method, also known as the region-based target detection algorithm. This type of method divides target detection into two parts: generating candidate frames and identifying target categories, such as Mask R-CNN[4], Faster R-CNN[5] and other algorithms, the advantage is that the detection accuracy is high. The other is the one-stage method, which directly predicts the category of the object from the picture, such as YOLO[6], SSD[7] and other algorithms, the advantage is that the detection speed is faster. In this paper, the YOLOV3 algorithm with balanced accuracy and real-time performance is used to detect tablet defects, and the trained model is tested using the test set, which has high accuracy.

2. YOLOV3 Algorithm

YOLOV3[8] is the third generation of the YOLO series of algorithms. Compared with YOLOV1 and YOLOV3, the detection accuracy and speed have been greatly improved. Its structure is shown in Figure 1. The overall idea of the YOLOV3 algorithm is to divide an image into s×s grid cells for detection. Each grid is responsible for detecting the targets falling into it, and predicting the bounding boxes and positions of the targets contained in all grids at one time. information, and probability vectors for all classes.

2.1. Backbone Feature Extraction Network Darknet-53

The Darknet-53 network is shown in the left part of Figure 1. It is based on the Darknet-19 proposed by yolov2 and has two main improvements:

One is that Darknet-53 is a fully convolutional structure with a total of 53 convolutional layers, so it is named as such, each of which is composed of convolution + BatchNormalization normalization + LeakyReLU activation function. The formula of the LeakyReLU activation function is shown in Equation 1.

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \frac{x_i}{a_i} & \text{if } x_i < 0, \end{cases} \tag{1}$$

The second is to introduce the residual module in Resnet (Residual Network) [9]. As shown in Figure 2, the residual module consists of two convolution operations on the input and an add operation on a shortcut link, where the input x It can be expressed as the result of convolution with a convolution kernel size of 3 × 3 and a stride of 2, which replaces pooling to achieve downsampling. The function of the residual network is to set the shortcut link [19] between the convolutional layers, which can effectively reduce the difficulty of training the deep network and enable the network to converge better.

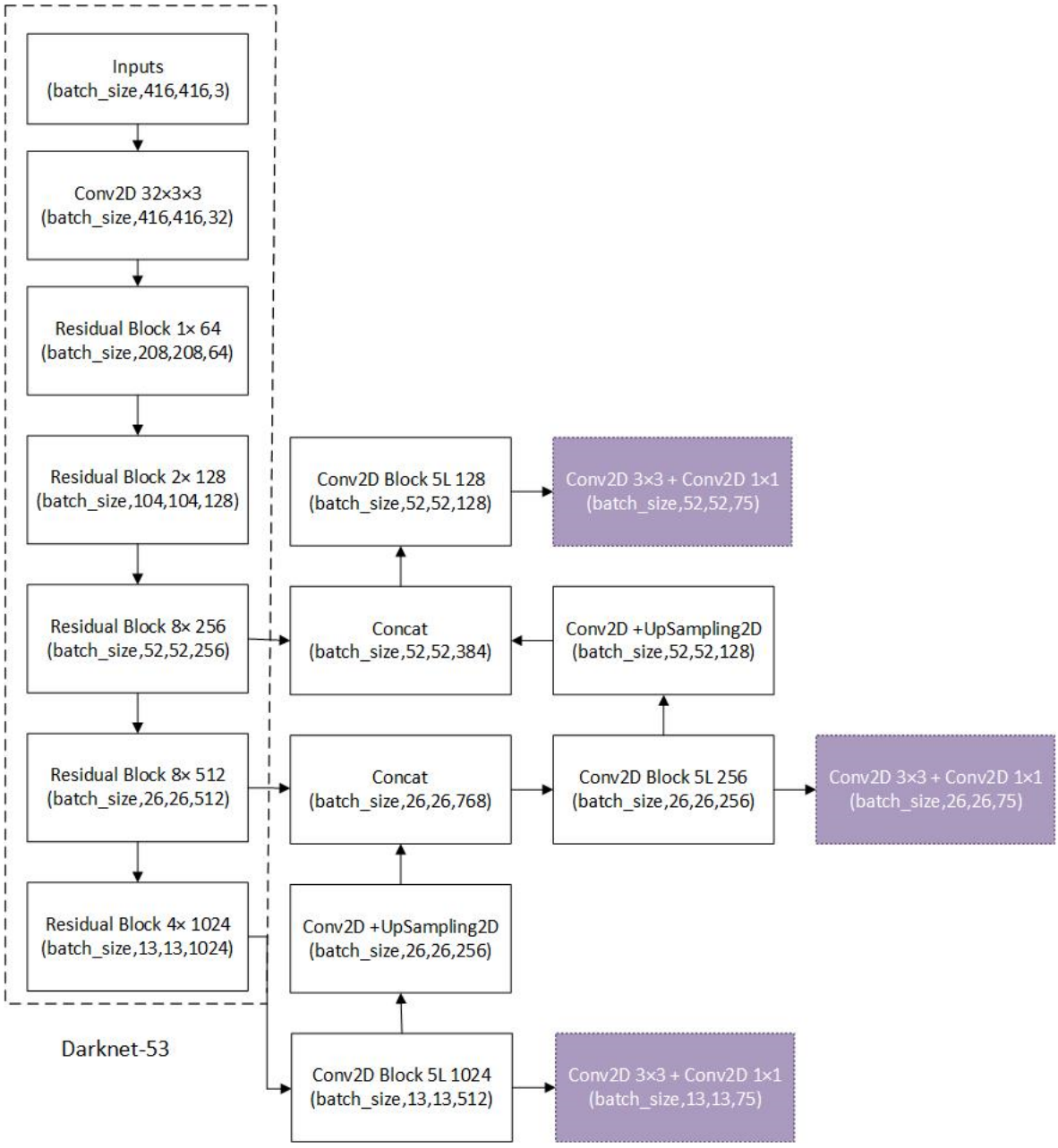

Figure 1. YOLOV3 structure

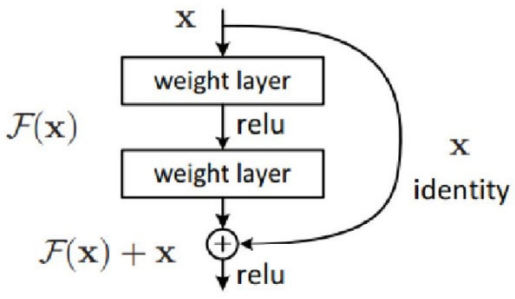

Figure 1. Residual structure

It is worth noting that from the Residual Block module in the Darknet-53 section in Figure 1, 1, 2, 8, 8, and 4 represent several repeated residual modules, and retain the different sizes of the last three modules. output, so that the output can be used in subsequent Concat operations for feature fusion.

2.2. Multi-scale Feature Fusion

YOLOV3 borrows the idea of FPN (Feature Pyramid) [10], and uses three feature maps of different scales for object detection. As can be seen in the Darknet-53 part of Figure 1, a total of three feature layers are extracted, and the shapes of the three feature layers are (52, 52, 256), (26, 26, 512), (13, 13, 1024). After upsampling, Concat and other operations on the three features, the fusion of features is realized, which further improves the ability to detect small targets.

2.3. Bounding Box

2.3.1. Anchor Box

Before predicting the bounding box, you need to understand anchors. Anchors are actually anchor boxes with suitable size and shape clustered according to the target parameters in the positive samples in the training set. The function of anchors is mainly used to predict the boundary. box (boundingbox).

YOLOV3 uses the K-means clustering algorithm to cluster the label boxes in the data set while following the techniques of YOLOV2 a priori box, and obtains three types of a priori boxes of large, medium and small. The a priori box of the COCO dataset is shown in Table 1. There are 9 anchors in total, each 3 anchors corresponds to an output, and an output contains 3 predicted bounding boxes.

Table 1. Prior box assignments

	Assignment result		
Featuremap	13*13	26*26	52*52
Feeling Wild	Big	Middle	Small
priori box	(116,90);(156,198);(373 ,326)	(30,61);(62,45);(59,119)	(10,13);(16,30);(33,23)

2.3.2. Prediction of Bounding Boxes

When the input feature map and a priori box are obtained, the bounding box can be calculated. The regression process of the bounding box is shown in Figure 3. The grid cell in the second row and the second column is responsible for predicting the target of the area. The coordinates of its upper left corner are (1, 1), then for a grid cell, C_x =1,C_y=1. The blue box in the figure is the predicted bounding box. t_x and ty are the offsets of (0, 1) after being processed by the sigmoid function respectively. Then according to the formulas, the offsets of t_x and t_y are added to C_x respectively. C_y can get the center coordinates (b_x, b_y) of the bounding box. P_w, P_h are the width and height of the anchors mapped to the feature map, t_w, t_h are the scaling values of the scale, because the width and height are non-negative numbers, so first do the exponential operation on t_w, t_h. According to the formula in the figure, the values of the four coordinates of b_x, b_y, b_w, and b_h of the bounding box can be predicted.

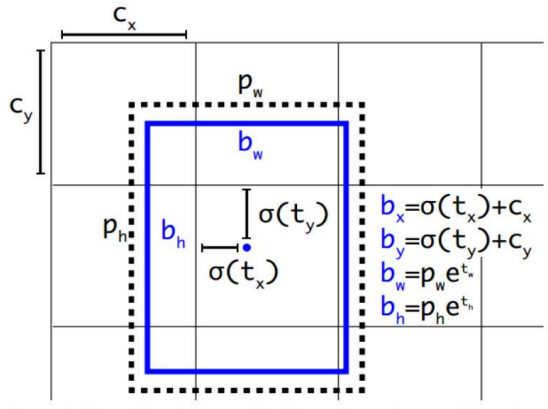

Figure 2. Bounding box calculation

2.4. Loss Fuction

The loss function of YOLOV3 mainly consists of three parts: coordinate error, confidence error, and classification error. Calculated as follows:

$$Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(x_i - \hat{x}_i^j)^2 + (y_i - \hat{y}_i^j)^2] +$$
$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(\sqrt{w_i^j} - \sqrt{\hat{w}_i^j})^2 + (\sqrt{h_i^j} - \sqrt{\hat{h}_i^j})^2] -$$
$$\sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] -$$
$$\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] -$$
$$\sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in classes} ([\hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - P_i^j)]$$

In the formula: S is the division coefficient of the image, B is the number of bounding boxes predicted in each grid, C is the confidence, p is the class probability, c=0,1,...,classes is the class number, i=0,1,...,S^2 is the grid number, j=0,1,...B is the frame number, x_i is the abscissa of the center point of the bounding box in the ith grid, and y_i is the ith grid The ordinate of the center point of the bounding box in the grid, w_i is the width of the bounding box in the ith grid, h_i is the height of the bounding box in the ith grid, λ_{coord} is the weight coefficient, and λ_{noobj} is the penalty weight coefficient.

3. Experimental Process and Result Analysis

3.1. Data Set Production

The tablet defect data set is produced according to the manual manufacturing method in the paper of Yang [11] according to the guidance of the workers of the pharmaceutical enterprise, and then the defect images are collected by industrial cameras. The types of defects are mainly divided into three categories: stains, scratches, and missing, such as shown in Figure 4.

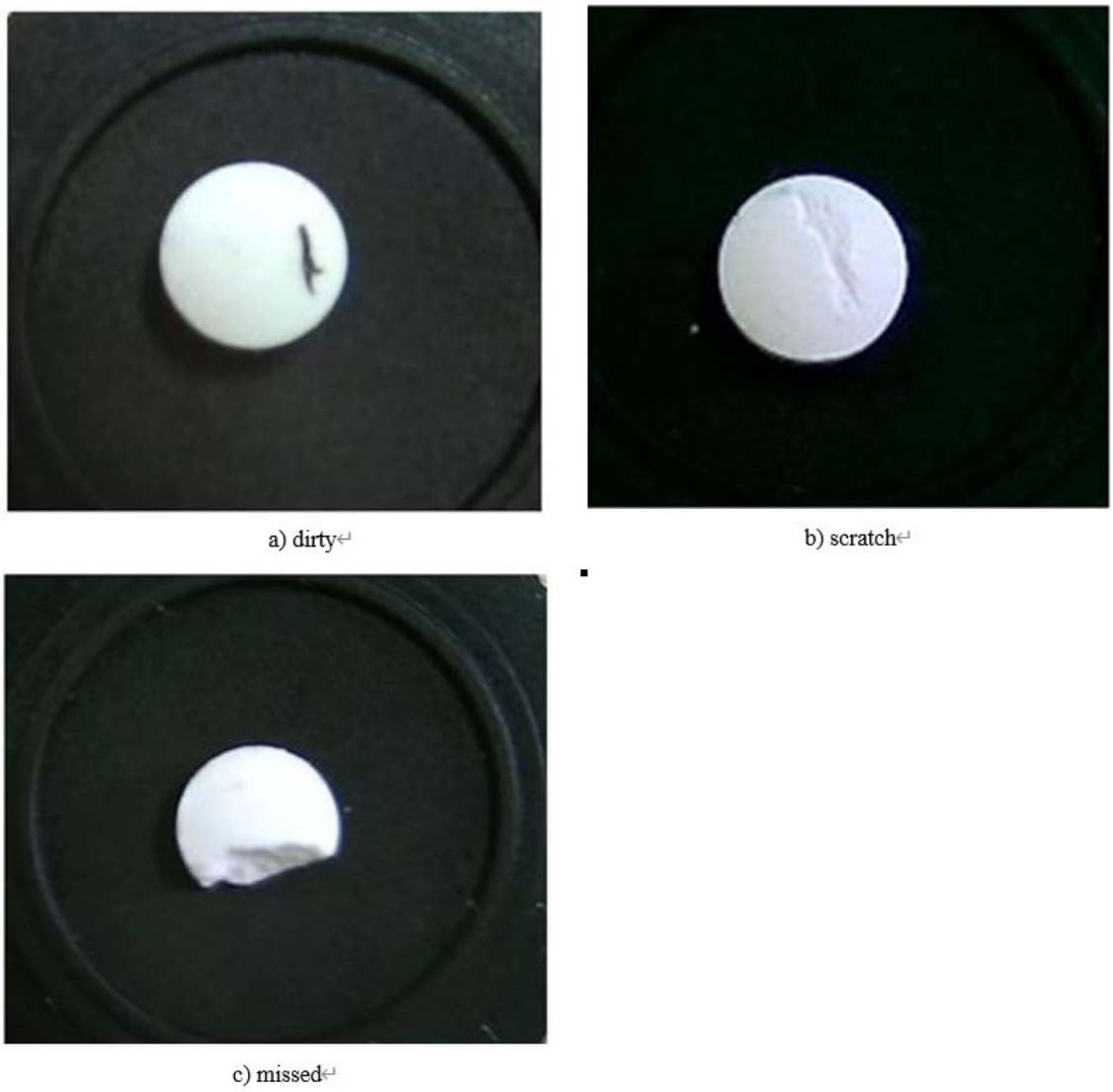

Figure 3. Data image

The classification of the dataset is shown in Table 2. The label of the dataset is implemented by labelimg software. The stain class is marked as dirty, the missing class is marked as missed, and the scratch class is marked as scratch. Finally, the data set is divided into training set, validation set and test set according to 8:1:1.

Table 2. Dataset Classification

type	quantity
dirty	453
missed	300
scratch	303
total	1056

3.2. Experimental Environment

The experimental environment adopts the Linux operating system.

Processor: Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz,
GPU: RTX 3060 12.6 GB,
RAM: 30.1 GB
Solid State Drive: 483.2 GB

3.3. Training Process

The initialization parameters Momentum of the training model are set to 0.9, Decay to 0.0005, Batch size to 16, using mini-batch stochastic gradient descent for optimization, the initial learning rate (Learning rate) is 0.001, epoch Set to 100. The change trend of training loss is shown in Figure 5. The loss fluctuated in the 50th round because the trunk was frozen in the first 50 rounds, and the thaw training began in the 51st round, which improved both time and resource utilization. After the training is completed, the weight with the smallest loss value is used as the final weight to detect the tablet defect images.

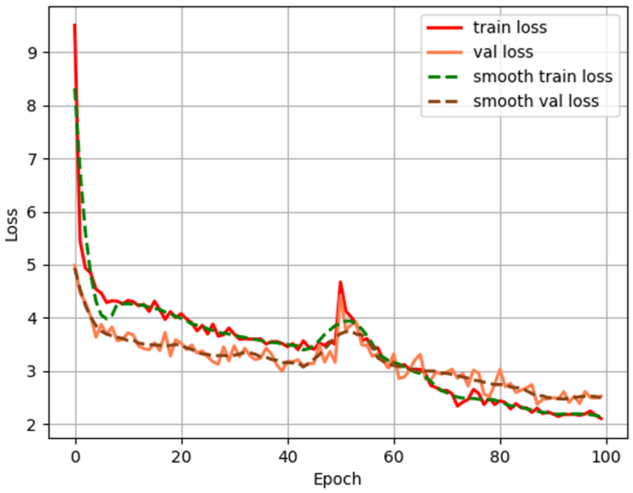

Figure 4 training loss

3.4. Analysis of Results

After the network model is trained, save the weight with the lowest loss value, and then input the pre-detected image into it, and finally get the detection result. The test result is shown in Figure 6.

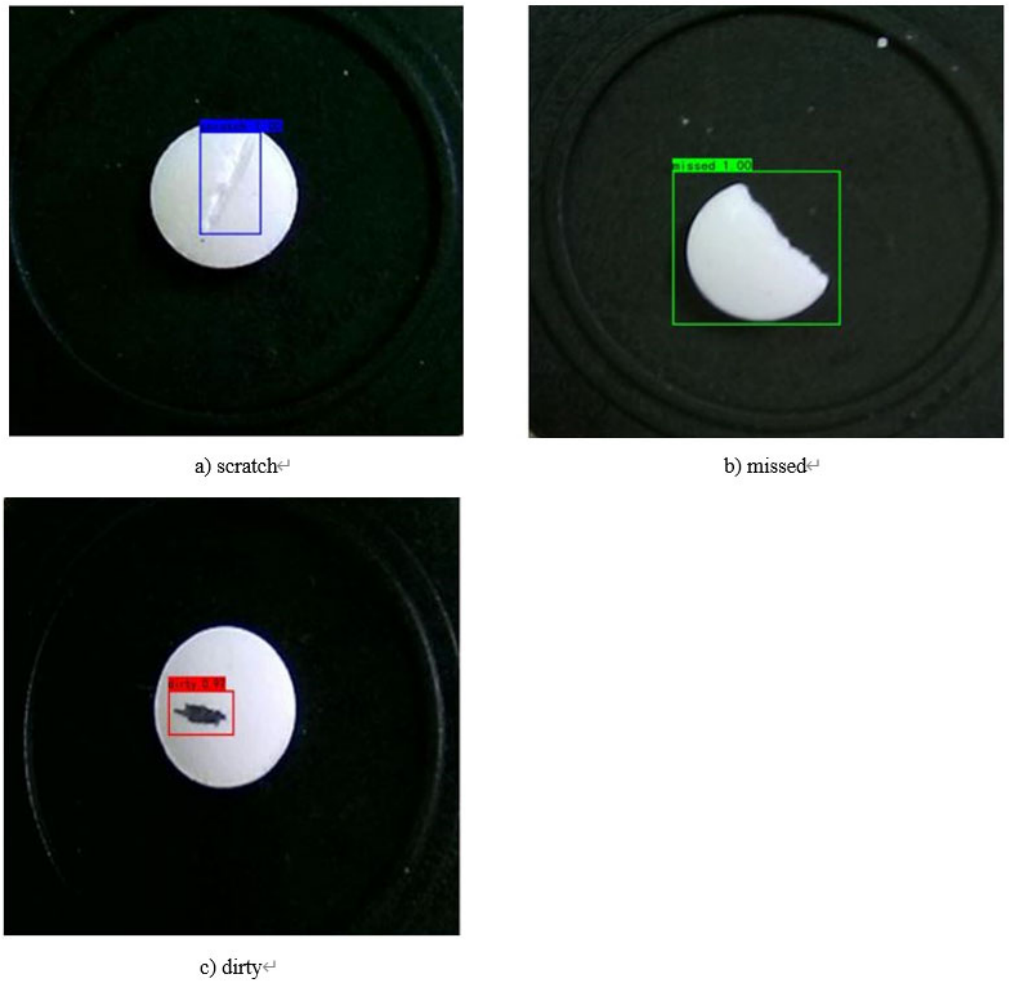

Figure 5. Test results

The mean Average Precision (mAP) calculated by the precision rate P (Precision) and the recall rate R (Recall) is used as the performance evaluation standard of the network model. mAP is the average of the average detection accuracy across all categories and is used to evaluate the overall performance of the detection model. P, R and mAP are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

In the formula, TP represents the number of positive samples that are correctly identified as positive samples, FP represents the number of negative samples that are incorrectly identified as positive samples, and FN represents the number of positive samples that are incorrectly identified as negative samples.

AP is expressed as the area under the precision and recall curves, and mAP is expressed as the mean of various APs. The formula is as follows:

$$AP = \sum_{i=1}^{n-1}(r_{i+1} - r_i)P_{inter}(r_i + 1) \quad (5)$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \quad (6)$$

In formula (5), r1, r2...rn are the Recall values corresponding to the first interpolation in the Precision interpolation segment arranged in ascending order, and in formula (6), k represents the number of categories.

Table 3. Comparison of different detection frameworks

Detection framework	mAP/ %	frame rate /(frame · s^{-1})
YOLOV3	92.97	43.9
SSD	85.52	48.8
Faster R-CNN	89.70	23.8

As can be seen from Table 3, the accuracy of yolov3 is higher than that of SSD and Faster R-CNN in the detection of surface defects of tablets; in terms of detection speed, YOLOV3 and SSD are two-stage detection algorithms and two-stage detection algorithm Faster R-CNN. Compared with CNN, it shows the advantage of speed. Although YOLOV3 is slightly lower than SSD in frame rate, its detection accuracy can reach 92.97%, so YOLOV3 is more applicable.

4. Conclusions

In order to solve the problem of difficulty and low efficiency in the detection of surface defects of tablets by manual and traditional machine vision technology, YOLOV3 target detection algorithm is used to detect surface defects of tablets. It can reach 43.9/(frame s-1), which is more applicable than SSD and Faster R-CNN.

Acknowledgment

This work was financially supported by Postgraduate innovative project of Southwest Minzu University fund(NO.CX2021225).

References

[1] Hu Anxiang. Research on the key technology of defective tablet detection based on machine vision [D]. Shandong University.

[2] Yao Jiangtao. Research on tablet detection and recognition algorithm based on machine vision [D]. Harbin Institute of Technology.

[3] Lecun Y , Bengio Y , Hinton G . Deep learning[J]. Nature, 2015, 521(7553):436.

[4] He K, Gkioxari G, Dollár P, et al. Mask R-CNN[C]. in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980-2988.

[5] Girshick R. Faster R-CNN[C]. in: Proceedings of the IEEE international conference on computer vision, 2015.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[7] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. in: Proceedings of European conference on computer vision, 2016.

[8] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun.

[10] Lin T Y , Dollar P , Girshick R , et al. Feature Pyramid Networks for Object Detection[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[11] Yang Xudong. Detection of surface defects of tablets based on deep learning [D]. South China Agricultural University.