Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

doi:10.3233/FAIA220021

Spoon Surface Defect Detection Based on Improved YOLO V3

Can Wu[a] and Zhiqiang Zeng[a,1]
[a] *College of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China*

Abstract. At present, there may be some problems in the production process of spoon, such as the lack of material on the surface of spoon. In order to effectively detect the surface defects of spoon, a defect detection method based on improved YOLO V3 model is proposed in this paper. Firstly, the output layers of the second and third residual blocks in the backbone network Darknet-53 are selected to build the feature pyramid network, which shortens the transmission path of feature information. In this case, we can better retain the feature information of small target defects. Secondly, the anchor boxes is adjusted to strengthen the ability of the model for small target defects detection. We test the proposed method on one spoon defect dataset, which is collected from the real-world industry manufactory scenario. The results show that the average precision of our algorithm reaches at 95.14%, which is better than the conventional YOLO V3 algorithm by 9.35%. Meanwhile, our algorithm is 9.12% faster than YOLO V3 with a 32.3 fps detection speed, demonstrating its efficiency and effectiveness for spoon defect detection.

Keywords. Defect detection, YOLO V3, Small target defects

1. Introduction

Spoon, one of the most frequently used tableware, plays an important role in our daily life. Due to the complexity of industry production, the spoon is prone to defects in the production process, such as the lack of material on the surface, which will affect people's normal use in the process of diet. Traditional automatic defect detection algorithms need to manually design defect features and use sliding windows for region selection. However, the defect forms are complex and diverse, and there are often few defect targets in the image, so the detection accuracy and efficiency are difficult to be guaranteed. In contrast, the deep learning algorithm extracts the abstract features of the image through a large number of convolution operations and carries out autonomous learning. It does not need to design a specific approach for specific defects. With high precision and fast speed, it can solve the shortcomings of the traditional machine vision type detection algorithms. At present, there are two kinds of widely used target detection algorithms: one is the target detection algorithm based on candidate regions, such as Faster R-CNN[1] and Mask R-CNN[2], etc, In this kind of algorithm, the first step is to use Region Proposal Network (RPN) to generate candidate regions, and the second step is to use detection network to classify and regression candidate regions. The other is the target detection algorithm based on end-to-end learning represented by

[1] Corresponding Author; E-mail: zqzeng@xmut.edu.cn.

YOLO[3] and YOLO V3[4]. This kind of algorithm eliminates the process of generating candidate regions. It directly predicts the target category and target boundary box in the input image, greatly accelerating the detection speed.

In recent years, target detection algorithm based on convolutional neural network[5] has been widely introduced into the field of surface defect detection. For example, Reference[6] proposes a leather defect detection system based on Mask R-CNN. Reference[7] proposes a weighted ROI pooling, which improves the accuracy of Faster R-CNN model in steel surface defect detection. Reference[8] proposes a method of fusing multi-layer features of convolutional neural network, and achieves good detection results in the task of strip surface defect detection.

In this paper, we aims at detecting the defects on the surface of spoons. Through observation, it is found that the size of defect targets is generally small, containing less feature information. YOLO V3 model is one widely-used efficient tool for object detection and has make a lot of successes in many applications. Despite its success, it cannot be directely applied to the spoon defect detection due to the small-size defect problem. In order to improve the detection performance of YOLO V3 model for spoon defect targets, the following two improvements are proposed in this paper. Firstly, the output layers of the second and third residual blocks of the backbone network Darknet-53 are selected to build the feature pyramid network(FPN)[9], which can improve the detection speed while retaining more target feature information. Secondly, more dense anchor boxes are imposed to detect more small targets.

2. YOLO V3 Model

2.1. YOLO V3 network

The architecture of YOLO V3 is shown in Figure 1. It uses darknet-53 as the backbone

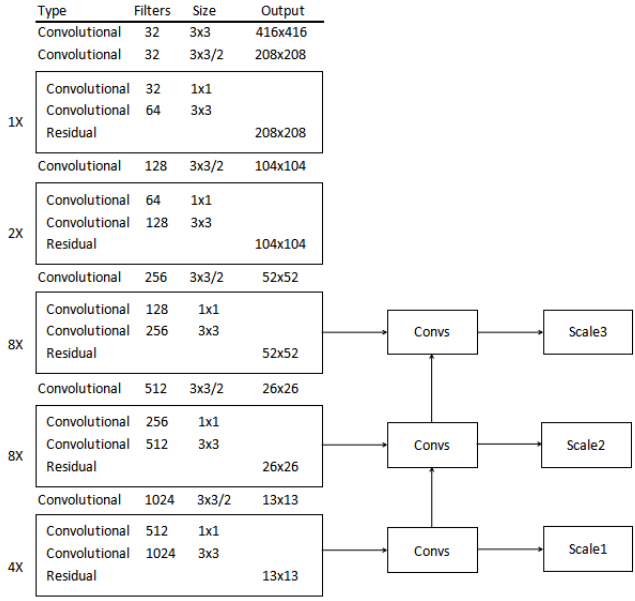


Figure 1. YOLO V3 network

network and FPN to extract target features from multiple different scales. Darknet-53 draws on the approach of ResNet[10] and adds a short-cut link between layers, which alleviates the gradient vanishing problem when training the model with the increased number of network layers. Darknet-53 contains five residual blocks in total. YOLO V3 selects the feature maps output by the last three residual blocks to build a feature pyramid network, and finally transmits the output feature maps to the later detection network for target prediction.

The detection network first divides the input picture into SxS grids of the same size, and sets multiple anchor boxes on each grid to predict the targets, Let (c_x, c_y) be the central coordinate of the anchor box relative to the feature map and (p_w, p_h) be the width and height of the anchor box. The output content of the grid is the confidence information and position information (t_x, t_y, t_w, t_h) of the prediction bounding box. Here, (t_x, t_y) represents the offset of the central coordinate of the prediction bounding box. (t_w, t_h) refers to the scaling ratio of the width and height of the prediction bounding box. The center coordinates (b_x, b_y), width (b_w) and height (b_h) of the prediction bounding box can be calculated by formula (1):

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_x = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases} \quad (1)$$

The function σ(x)is a sigmoid function. It can ensure that the center offset of the prediction boundary box is between 0 and 1, which is conducive to the convergence of the network model. Finally, the prediction bounding box with the highest confidence is selected by NMS algorithm and used as the final detection result.

2.2. Improved YOLO V3 network

Through observation, it is found that the size of spoon defect target and is generally small. Therefore, if the model possesses a deep network structure, it may be not conducive to the transmission of feature information of small target. This paper solves this problem mentioned above by improvimg the feature pyramid network of YOLO V3 model. Concretely, we selects the output of the second and third residual blocks to construct the feature pyramid network, so as to shorten the transmission path of feature information and better retain the feature information of small target defects. The architecture of Improved YOLO V3 network is shown in Figure 2. The size of the input image is 544x544. First, we perform five-convolution-operations, i.e., sequent operations with 1x1, 3x3, 1x1, 3x3, 1x1 convolutional kernels, on the output feature map of the third residual block to obtain a new feature map, namely (FM1). Then, after performing 1x1 and Unsampling convolution operation on FM1, we connect it with the feature map output from the second residual block, and perform five-convolution-operations on the results of the previous operation to obtain another feature map, namely (FM2). Finally, two 3x3 and 1x1 convolution operations are performed on FM1 and FM2 to generate two output feature maps_with sizes of 68x68 (scale1) and 136x136 (scale2), respectively, which are sent to the detection network for target prediction.

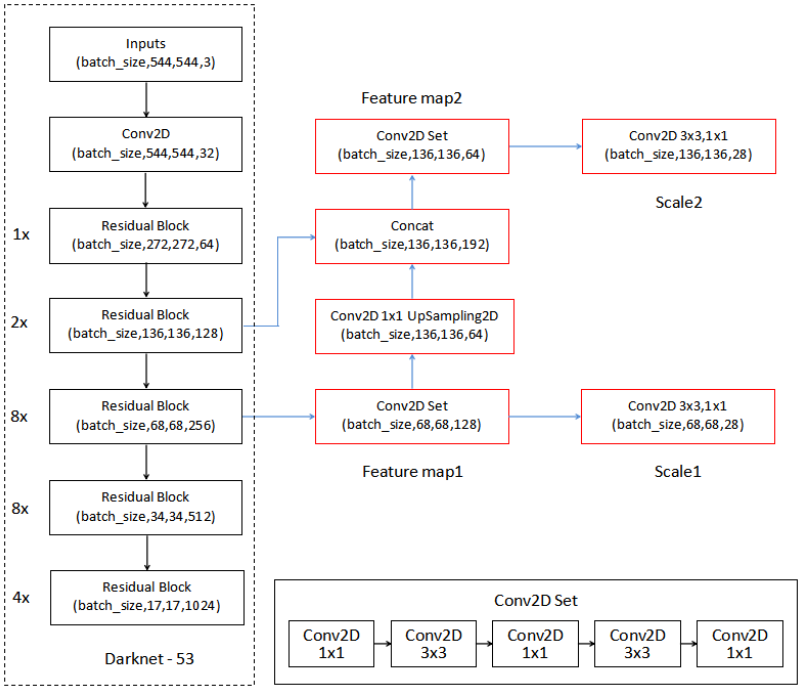


Figure 2. Improved YOLO V3 network

YOLO V3 is an anchor-based target detection algorithm. The initial 9 anchor boxes are obtained by clustering, and 3 anchor boxes are set on each feature layer. The distribution of anchor boxes in the above contents may not be able to deal with the situation that the defect targets are generally small. The IOU values of some defect targets and anchor boxes are lower than the preset threshold, and then they are ignored as negative samples.

In the above content, we have added a larger shallow feature map for YOLO V3 model, which can better obtain the feature information of small target defects. In order to detect more small-size defects, firstly, the step size of anchor box on shallow feature map is set to 4; Secondly, the K-means clustering algorithm is used to cluster the size of the target bounding box in this paper, and the obtained clustering centers are used as the anchor boxes. The sizes of the eight anchor boxes are: (7,6), (18,18), (22,20), (28,23), (34,24), (43,28), (57,35), (72,40). The first four boxes are placed on the feature map scale2 to detect small targets, The last four boxes are placed on the feature map scale1 to detect larger targets.

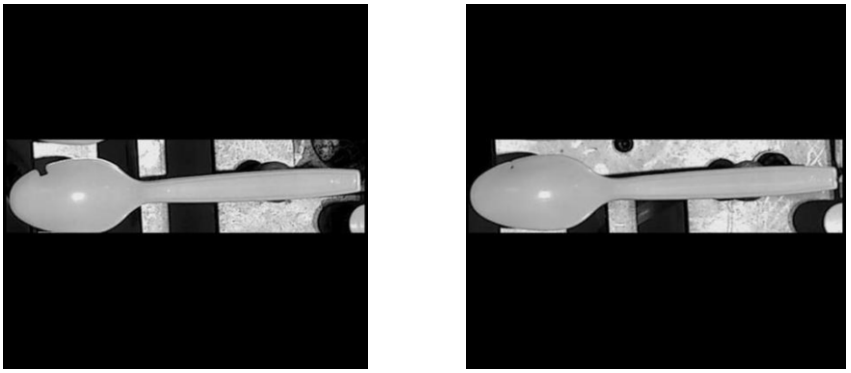
The improved network structure can better adapt to the situation that the defect target size is generally small in the spoon defect detection task, and then the detection accuracy of the model is improved. In addition, because the overall structure of the network is reduced, the detection speed is also increased.

3. EXPERIMENT

3.1. Experimental setting

The experiment of this paper is carried out under CentOS 7.5.1804 operating system; The CPU is Intel Xeon Gold 6242 and the main frequency is 3.1 GHz; The memory size is 502 GB; GPU is Geforce RTX 2080Ti with 12 GB memory size; Our method is implemented in the pytorch framework with python 3.8.10.

In this paper, the spoon image is taken with an industrial camera. Through observing the image, it is found that the lack of material on the spoon surface is a common defect. According to the defect size, it can be divided into two kinds of defects, namely "large defect" and "small defect". Its typical samples are shown in Figure 3.



(a)large defect (b)small defect

Figure 3. Typical spoon defect image

The size of the collected image is 1920x1080. In order to facilitate the training of the model, this paper uses the mask based method to extract the spoon image from the initial image. Due to the scarcity of defect samples, we expand the extracted defect samples by horizontal flip, vertical flip and diagonal flip, and finally obtains 240 defect samples and 180 defect free samples. Labelimg software is used to label the above images in Pascal VOC format. Finally, the positive samples and negative samples are divided into Training set, Validation set and Testing set according to the ratio of 7:1.5:1.5, respectively.

3.2. Model training

In order to prove the effectiveness of the content proposed in this paper, we select conventional YOLO V3 as the baseline model. All the models are trained and tested on the same data set in this paper. Firstly, the pretrained darknet-53 network is introduced as the backbone network, and then 80 epochs are trained in the data set of this paper. The specific parameters are as follows: the model optimization method is Stochastic Gradient Descent, and the momentum is 0.9;the weight decay is 0.0001; the initial learning rate is 0.001; the learning rate begins to decay at the 50th epoch, and the attenuation multiple is 0.1; In the initial stage of training, we also used the learning rate

preheating to train model, the learning rate slowly increased from 0.001 to 0.01, and the warm-up iteration was 20 epochs.

The weight file size and training time obtained by training the model through the above training methods are shown in Table 1. Through the improvement of YOLO V3, the reduction of network weight files and training time makes the model more portable and can be transplanted to devices with poor computation resources.

Table 1. Comparison table of model weight file and training time information

Detection algorithm	Weight file/MB	Training time/s
YOLO V3	469.85	1253
Our Method	176.93	1011

3.3. Evaluation index

In this paper, we use mAP (Average Precision) and FPS (Frame Per Second) as the evaluation indexes of the model performance. AP is calculated by the P (Precision) and R (Recall). The PR coordinate system is established with the P (Precision) as the ordinate and the R (Recall) as the abscissa. The area of the area below the PR curve is the AP of the category. mAP is obtained by summing and averaging the AP values of each category. The calculation formulas of P and R are formulas (2) and (3).

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{2}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{3}$$

Here, N_{TP} represents the number of correctly identified defect targets, N_{FP} and N_{FN} represent the number of incorrectly identified and unrecognized defect targets, respectively.

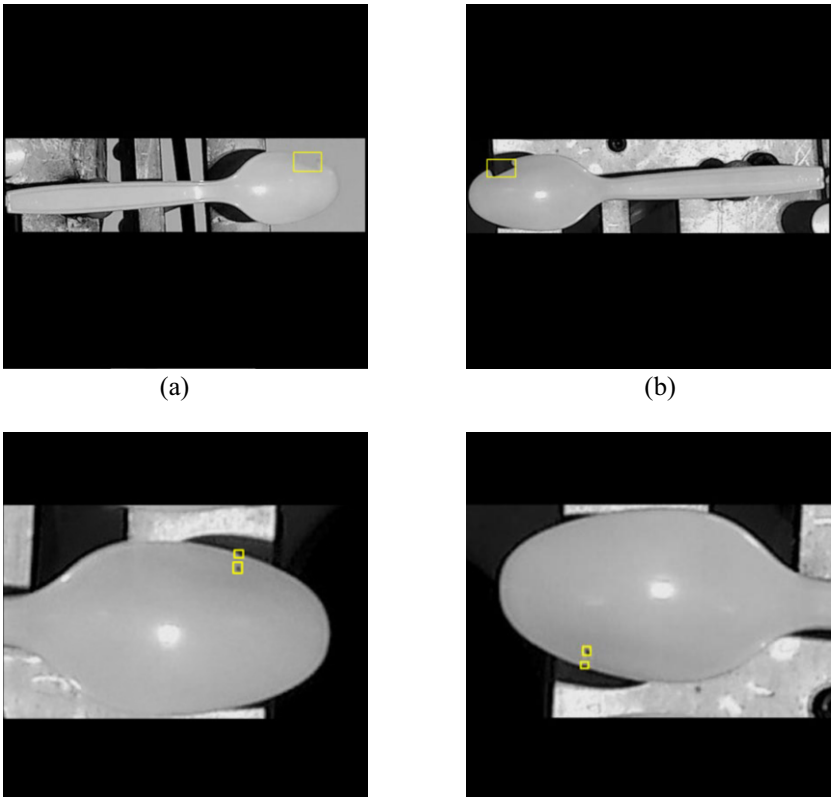
3.4. Results analysis

For all the compared methods, the model with the best performance in the validation set is selected for testing. The experimental results are shown in Table 2. The mAP of the our model is 95.14% with an increase of 9.35% over YOLO V3. Besides, the detection speed reaches 32.3fps, which is faster than YOLO V3 9.12%. Therefore, compared with the YOLO V3 model, our model has our model has better performance in the spoon surface defect detection task.

Table 2. Performance comparison of different models. Here, the "blm" and "slm" refer to the large defect and small defect, respectively.

Detection Algorithm	AP/%		mAP/%	fps
	blm	slm		
YOLO V3	90.5	81.1	85.79	29.6
Our Method	90.9	99.4	95.14	32.3

The detection results of our method are shown in Figure 4. Among them, figure (a) and figure (b) are the detection results of the first type of defects, i.e., the large defect, while figure (c) and figure (d) are the detection results of the second type of defects, i.e., the small defect.



(a) (b)

(c) (d)

Figure 4. Detection results. (a-b) blm (c-d) slm

By observing the above detection results, we can see that our model has a good detection effect on the defects in the spoon image. We can clearly see that there are many black spots in the image, of which the morphology are very similar to the small-size defects, making a great challenge for the defect detection. Facing this challenge, our model can still distinguish them well and detect defect targets with high confidence. This shows that our model has good robustness and can deal with some interference caused by complex detection environment.

4. CONCLUSION

In this paper, for the task of spoon surface defect detection, we propose an efficient algorithm to detect spoon defects. By adjusting the distribution of feature pyramid network and candidate boxes of YOLO V3 model, the model can obtain more small target defect features and detect more small target defects. Through comparing the experimental results, it can be seen that the proposed method improves the detection

precision and speed of the model to a certain extent. However, there is still a certain gap from the needs of practical engineering. The next step is to continue to optimize the model and further improve the detection speed of the model without affecting the detection precision.

ACKNOWLEDGEMENT

This paper was supported by National Natural Science Foundation of China (Grant Nos.61871464), National Natural Science Foundation of Fujian Province (Grant Nos.2020J01266, 2021J011186), the "Climbing" Program of XMUT (Grant No.XPDKT20031), Scientific Research Fund of Fujian Provincial Education Department (Grant No. JAT200486), Program of XMUT for high-Level talents introduction plan (Grant No.YKJ19003R).

References

[1] Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[2] He K , Gkioxari G , P Dollár, et al. Mask R-CNN[C]// IEEE. IEEE, 2017.

[3] Redmon J , Divvala S , Girshick R , et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.

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

[5] Ketkar N . Convolutional Neural Networks[J]. Springer International Publishing, 2017.

[6] Liong S T , Gan Y S , Huang Y C , et al. Automatic Defect Segmentation on Leather with Deep Learning[J]. 2019.

[7] Wei R , Song Y , Zhang Y . Enhanced Faster Region Convolutional Neural Networks for Steel Surface Defect Detection[J]. ISIJ international, 2020, 60(3):539-545.

[8] He Y , Song K , Meng Q , et al. An End-to-end Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, PP(99):1-1.

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

[10] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.