International Journal of Innovative
Computing, Information and Control
Volume **18**, Number **4**, August **2022**

ICIC International ©2022 ISSN 1349-4198
pp. **1329–1338**

REAL-TIME DEFECT DETECTION OF METAL SURFACE BASED ON IMPROVED YOLOV4

Yanju Liu[1], Qiuji Wang[2], Huiyu Zhang[2], Yanzhong Liu[2,*] and Kaifeng Zhao[2]

[1]School of Mathematics and Information Science
Nanjing Normal University of Special Education
No. 1, Shennong Road, Nanjing 210038, P. R. China
yanjuliu@mail.njts.edu.cn

[2]School of Computer and Control Engineering
Qiqihar University
No. 42, Wenhua Avenue, Jianhua District, Qiqihar 161000, P. R. China
{ qiujiwang; huiyuzhang; kaifengzhao }@qqhru.edu.cn
*Corresponding author: yanzhongliu@qqhru.edu.cn

Received October 2021; revised January 2022

Abstract. *It is important for the surface quality of metal to process final product. Therefore, it is necessary to strictly control the defects on the surface of metal. Aiming at the current YOLOv4 algorithm with low detection accuracy and poor performance on small-scale information and slow detection speed, an improved YOLOv4 automatic detection method is proposed. First, in order to enhance detection target feature extraction and reduce gradient vanishing, the feature extraction network CSPDarkNet53 in YOLOv4 is replaced with lightweight deep neural network MobileNetv3 in this paper. Secondly, in order to improve the learning efficiency and accelerate the convergence speed, K-means clustering is adopted to generate a prior box to suit for this experiment. Finally, the confidence loss is redefined and a loss function is proposed that can adapt to the multi-scale to solve the problem of poor detection effect due to the positive and negative sample imbalance. The experimental results show that the mAP value and the speed are improved about 7.94% and 4.52 f/s, compared with the original YOLOv4 model for the surface defect detection of the metal. The accuracy of this model is improved effectively based on ensuring the detection speed.*

Keywords: Defect detection, YOLOv4, MobileNetv3, K-means

1. **Introduction.** Metal production technology is an important criterion for measuring the industrial development level of a country. In the early stage of applying such technology, manual online detection method was adopted in most enterprises. However, the detection by traditional classifier is time-consuming without that good effect and cannot meet the current industrial demand.

In recent years, machine vision has become a popular topic, especially CNN (Convolutional Neural Network) [1]. There are two types of CNN frameworks for deep learning based target detection algorithms: two-stage target detector and single-stage target detector. In terms of the two-stage detection framework, Girshick et al. [2] put forward the first CNN-based target detection framework RCNN (Region Convolutional Neural Network) [3]. Soon afterwards, a series of excellent two-stage target detectors come out one after another, such as Fast RCNN [4], Faster RCNN [5], and Mask RCNN [6]. Although with high detection precision, two-stage target detector cannot be used in practical industrial

DOI: 10.24507/ijicic.18.04.1329

production due to its slow detection speed, this shortcoming is overcome by single-stage target detector. Regarding the single-stage target detector, Liu et al. [7] analyzed the anchor box setting strategy of Faster RCNN as per the regression approach and proposed an SSD (Single Shot multi-box Detector) [8]. This way increased the detection speed while guaranteeing the precision, but cannot realize real-time detection. In contrast, Redmon et al. [9] proposed the YOLO (You Only Look Once) [10] using the entire image as the network input, regressed the position and category of bounding box in the output layer directly, and set target detection process as a regression task. This way increases the detection speed greatly and can realize real-time detection basically, but it has low precision. Later, Redmon and Farhadi proposed the YOLOv2 [11] version, which further improved the detection speed, but it was still not suitable for detecting small targets. Therefore, Redmon and Farhadi proposed the YOLOv3 [12] version based on the idea of SSD, introduced a BN (Batch Normalization) layer [13], and designed a multi-layer convolutional layer to improve the feature extraction ability. Recently, YOLOv4 [14] replaced the backbone network of YOLOv3 and added many practical techniques on the basis of YOLOv3, so that the detection precision and speed were further improved.

In this paper, YOLOv4 is selected as the basic network to achieve a good balance between detection precision and speed. Meanwhile, an improved real-time target detection algorithm for metal surface defect is proposed to solve the current problems in YOLOv4. For the low detection precision and poor performance on small-scale information, this paper replaces the backbone network, and uses K-means clustering algorithm [15] to adjust the a priori frame and redefine the confidence loss. Compared with the original YOLOv4, the proposed algorithm in this paper is featured by its high detection speed and mAP value, and its accuracy in identifying the position, shape and size of a defect, which meets the current demand of industrial production.

The remainder of the paper is organized as follows. We firstly present improved YOLOv4 algorithm in Section 2. Experimental results are demonstrated by the dataset of steel surface defects disclosed in Section 3. Finally, we conclude the paper with some future work in Section 4.

2. Improved YOLOv4 Algorithm.

YOLOv4 is an end-to-end target detection algorithm, and it is mainly composed of three major parts: backbone feature extraction network, add-on modules and feature processing layer. The network structure diagram of YOLOv4 is as shown in Figure 1.

Based on the YOLOv4 algorithm, an improved YOLOv4 is proposed in this paper in order to improve the performance of the model while solving the missing detection, false

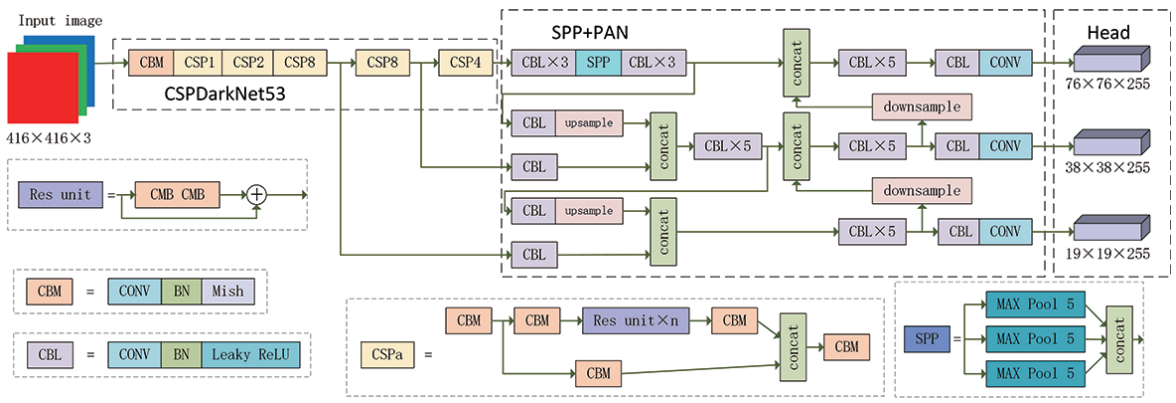


FIGURE 1. YOLOv4 network structure diagram

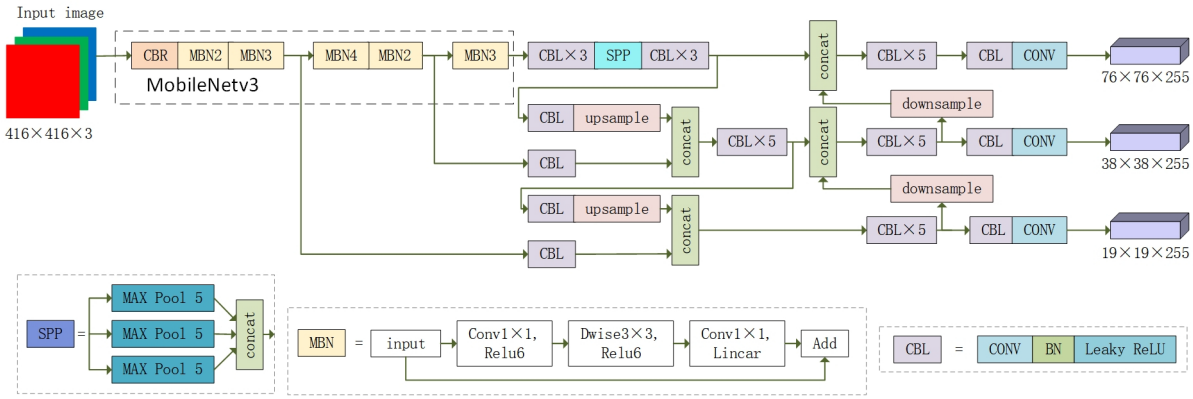


FIGURE 2. Improved YOLOv4 network structure diagram

detection, and other problems of YOLOv4 caused by the insufficient extraction of target features and the vanishing of gradient. The network structure diagram of the improved YOLOv4 is as shown in Figure 2.

2.1. **Replacing the feature extraction network.** In this paper, the backbone network CSPDarkNet53 of YOLOv4 is replaced with MobileNetv3. Compared with CSPDarkNet53, MobileNetv3 is a lighter deep neural network with small volume, but still maintains strong feature extraction ability.

MobileNetv3 inherits the depth-wise separable convolution of MobileNetv1 and the same residual structure with linear bottleneck as MobileNetv2. In order to better adapt to different scenarios, MobileNetv3 is also launched with two versions, large and small. The overall structure of MobileNetv3 is illustrated in Figure 3.

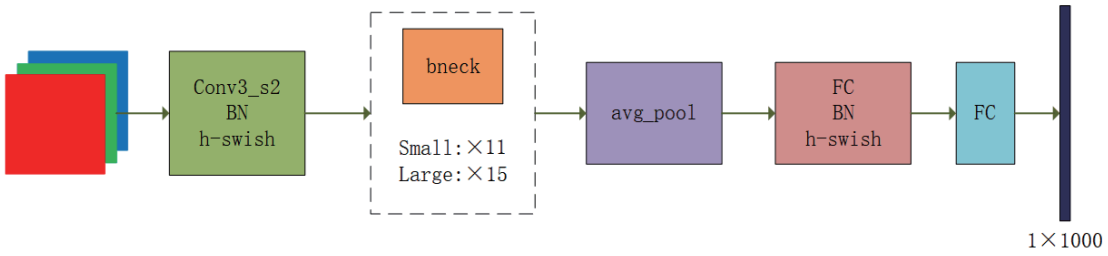


FIGURE 3. MobileNetv3 network structure diagram

2.2. **Using *K*-means clustering algorithm to adjust the anchor.** YOLOv4 adopts an anchor to predict the position of a target, so the anchor setting is very important. The original anchor of YOLOv4 is designed for 80 categories of objects in the COCO dataset. However, the defects of targets in the dataset of this research have large differences in aspect ratio. The default anchor is not suitable for the metal surface defects studied in this paper, and otherwise may result in a poor detection effect.

For this reason, K-means clustering algorithm is used in this research to adjust the size of the anchor so as to improve the precision in detecting defects on metal surface. This algorithm is commonly used in current machine learning, mainly to get each point distributed to a class cluster represented by the nearest class cluster center point provided that the K value and K center points of initial class clusters are given, and recalculate the center point of a class cluster as per all points in the class cluster, then iteratively distribute the points and renew the center point of class cluster, till the center point has little change or is iterated to the specified frequency.

The calculation equation of the K-means clustering algorithm is as shown in Equation (1), where $\mu_i = \frac{1}{|C_i|}\sum_{a\in C_i}$. a is the mean vector of cluster C_i, and the smaller the E value is, the higher the similarity of the samples in the cluster is.

$$E = \sum_{i=1}^{k}\sum_{a\in C_i} \|a - \mu_i\|_2^2 \tag{1}$$

Then, a clustering analysis is made on the dataset of this experiment by K-means clustering algorithm. Deep feature images were detected by large anchor, while shallow feature images were detected by small anchor.

2.3. **Redefining the confidence loss function.** In YOLOv4, three different scales of anchors are finally generated to predict. However, the ratios of positive and negative samples measured by different scales of anchors are different and the prediction results obtained by different scales of anchors with the same loss coefficients and are not exactly the same, so that the model can hardly obtain the optimal parameters. In view of this, a multi-scale loss function is put forward in this paper to adapt to different scales of anchors in YOLOv4. Taking the anchor scale of 38×38 as the reference point, the loss coefficients of anchors with scales of 19×19 and 76×76 are added with different scale balance coefficient β. Assuming that A_1, A_2 and A_3 are the ratios of positive and negative samples measured by three different scales of anchors, M represents the number of grids, and the value of β is as shown in Equation (2).

$$\beta = \begin{cases} \dfrac{A_2}{3A_1}, & M = 76 \\ 1, & M = 38 \\ \dfrac{3A_2}{A_3}, & M = 19 \end{cases} \tag{2}$$

As revealed in the equation, β balances different scales by adaptive selection of the ratio of positive to negative samples. The problem of extreme imbalance between positive and negative samples is solved in this paper by adding sample balance coefficient α before confidence loss of positive sample and adding coefficient $(1-\alpha)$ before confidence loss of negative sample. Here, the higher the α value is, the higher the weight of positive sample is and the lower the weight of negative sample is, and vice versa. Beyond that, considering the low identification rate of complex samples caused by over many simple samples in the training set, positive sample is added with modulation coefficient $(1-C_x^y)^\gamma$ and negative sample is added with modulation coefficient $(C_x^y)^\gamma$. The improved confidence loss is calculated as per Equation (3). Here, M represents the number of grids, each scale of feature image is divided into M^2 grids, N is the number of anchors generated in each grid, I_{xy}^{obj} indicates whether it is a positive sample or not, I_{xy}^{noobj} indicates whether it is a negative sample or not, $\widehat{C}_x^y$ represents the true confidence, and C_x^y represents the predicted confidence.

Experiments have proved that when α is 0.6 and γ is 2, the experimental result reaches the optimal.

$$\begin{aligned} Loss_c = &-\sum_{x=0}^{M^2}\sum_{y=0}^{N} I_{xy}^{obj}\left[\widehat{C}_x^y\beta\alpha(1-C_x^y)^\gamma\log(C_x^y) + \left(1-\widehat{C}_x^y\right)\log(1-C_x^y)\right], \\ &-\sum_{x=0}^{M^2}\sum_{y=0}^{N} I_{xy}^{noobj}\left[\widehat{C}_x^y\log(C_x^y) + \left(1-\widehat{C}_x^y\right)(1-\beta\alpha)(C_x^y)^\gamma\log(1-C_x^y)\right] \end{aligned} \tag{3}$$

3. **Experimental Results and Analysis.** This paper adopts the dataset of steel surface defects disclosed by Professor Song et al. in Northeastern University [16]. This dataset mainly includes two parts: NEU-CLS and NEU-DET. This dataset contains five types of defects, i.e., rolled-in_scale, patches, pitted_surface, inclusion, and scratches. Each type of defect involves 300 pictures available for making target detection. The operating environment of the experiment is as shown in Table 1.

Table 1. Experiment operating environment

Category	Environmental conditions	Category	Environmental conditions
CPU	Intel(R)Core i7-9750H@2.60GHz	CUDA version	CUDA 9
VGA card	Nvidia GeForce GTX 1660 Ti	Deep learning framework	PyTorch
Memory	16 GB	Operating environment	PyCharm
OS	Windows 10 64 bits	Scripting language	Python3.6

3.1. **Evaluation indicators.** Mean Average Precision (mAP) is a commonly used evaluation indicator in target detection. In this experiment, mAP and FPS (Frames Per Second) are used to measure precision and speed. P-R curve chart is plotted, with Recall (R) as the horizontal axis and Precision (P) as the vertical axis. The value of mAP is calculated by averaging the area AP in each category of P-R curve chart, as shown in the following steps. Equation (4) and Equation (5) show the calculation methods of P and R, respectively. Equation (6) is for calculating AP, and mAP is calculated as Equation (7).

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \int_0^1 Psmooth(s)ds \tag{6}$$

$$mAP = \frac{sum(AP)}{n} \tag{7}$$

In the equations, TP is true positive sample, FN is false negative sample, and FP is false positive sample. $\int_0^1 Psmooth(s)ds$ is the area in P-R curve chart. $sum(AP)$ is the sum of all categories of APs. n is the number of all categories. P and R are generally contradictory, namely, when P is high, R is often low, and vice versa. However, when judging the performance of the model, it is not accurate to merely measure the precision by IOU (Intersection over Union). Therefore, it is needed to calculate the mAP based on the area APs in the P-R curve charts.

3.2. **Result analysis.** The experimental dataset is divided into training set, test set, and verification set at a ratio of $8 : 1 : 1$. In the training settings, the value of epochs is 100, batch size and mini-batch size are 4 and 2, respectively. The initial learning rate is 0.001, the final learning rate is 0.00001, the training time is about 300 min, and the threshold value of confidence in the test is 0.5. The curve chart of the final training loss function is as displayed in Figure 4, where the horizontal axis represents the number of iterations, and the longitudinal axis represents the loss value of the training.

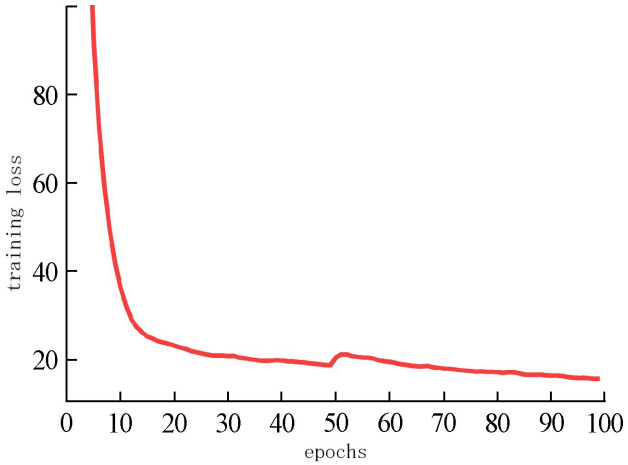
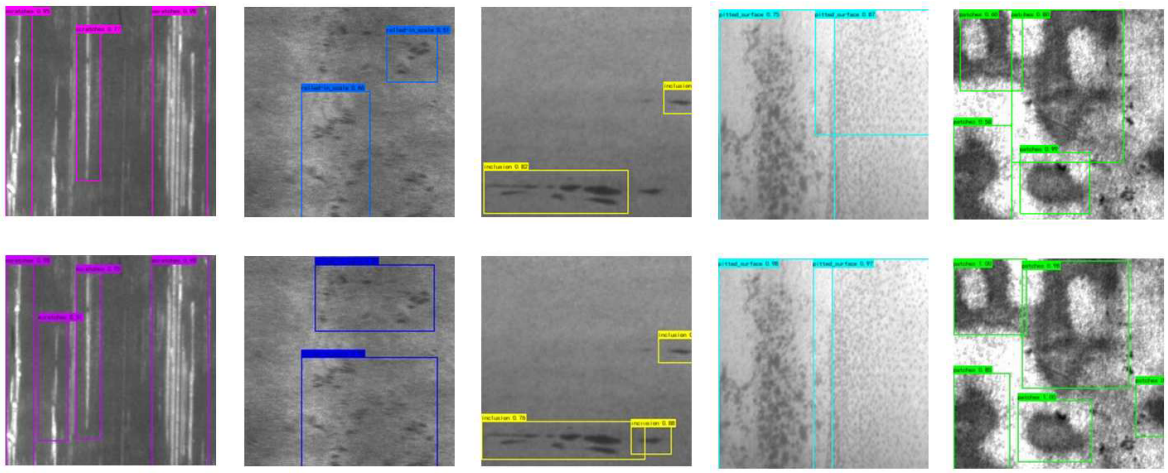


FIGURE 4. Variation curve for training



(a) (b) (c) (d) (e)

FIGURE 5. Comparison of test results: (a) show the test results comparison of the scratches class; (b) show the test results comparison of the rolled-in_scale class; (c) show the test results comparison of the inclusion class; (d) show the test results comparison of the pitted_surface class; (e) show the test results comparison of the patches class

As can be seen from Figure 4, totally 100 iterations were conducted in the training, where the loss of the algorithm model decreases constantly before 20 iterations. After exceeding 20 iterations, the loss basically stabilizes and the detection model reaches convergence.

In this paper, each type of defects is selected with partial test sets as the detection effect pictures, as shown in Figure 5. In Figure 5, there are five groups of data divided as per the number of defect types and representing different defect types, respectively. The detection effect pictures of the original YOLOv4 are shown on the top of each group of data, while those of the improved YOLOv4 are shown on the bottom.

As discovered in the detection effect pictures shown in Figure 5, the original YOLOv4 shows missing detection and low defect identification degree as observed in groups (a), (b), and (e), and also fails to get the defect part fully detected as presented in groups (b) and (d), while the improved YOLOv4 shows obvious improvement in the detection effect.

The improved version is not only friendly to detecting small targets, but also has high rate in identifying middle and large targets significantly. It can position the target more accurately and shows few false detection and missing detection.

Figure 6 shows each category of P-R curve chart before and after the algorithm is improved, where, the data in group (a) show each category of P-R curve charts measured by the original YOLOv4, while the data in group (b) show each category of P-R curve charts measured by the improved YOLOv4.

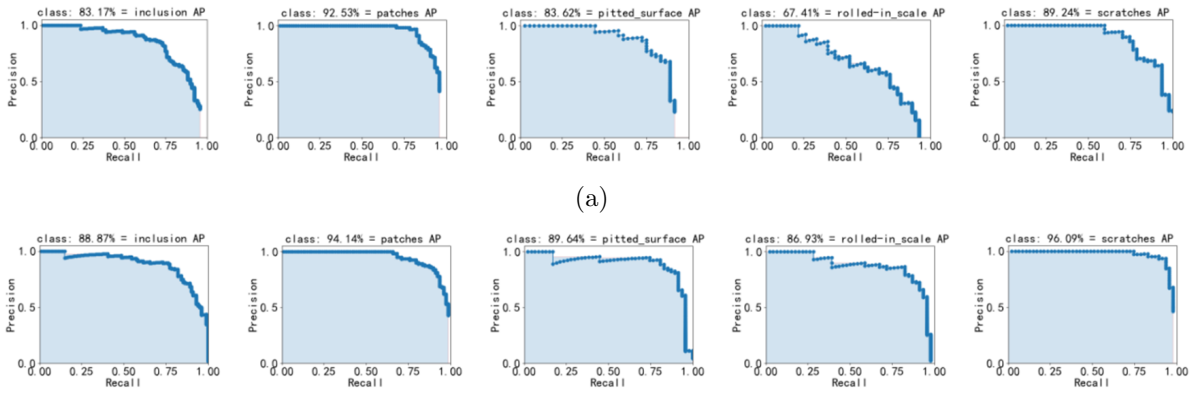


FIGURE 6. P-R curve chart comparison: (a) show the P-R curves of various classes based on YOLOv4; (b) show the P-R curves of various classes based on improved YOLOv4

As demonstrated in Figure 6, compared with the original YOLOv4, the AP of each defect type measured by the improved version is improved to different extent. Especially for the defect type of "rolled-in_scale", the AP is improved by 19.52%. The abilities of identifying the other three types of defect, i.e., "inclusion", "pitted_surface" and "scratches", are also improved to a large extent, with AP values improved by 5.7%, 6.02%, and 6.85%, respectively. Finally, the ability of identifying patches type of defect is also improved somewhat, with AP value improved from 92.53% to 94.14%, a higher identification precision.

To further verify whether the approach proposed in this paper is effective to the improvement of YOLOv4 algorithm, multiple groups of experiments were made by taking different improvement strategies. The final comparison results of the experiments are as illustrated in Table 2.

In Table 2, Model 1 is the original YOLOv4 algorithm, Model 2 is the algorithm obtained after getting the backbone network of YOLOv4 replaced with MobileNetv3,

TABLE 2. Comparison experiment of each model

Model	Model 1	Model 2	Model 3	Model 4
inclusion/%	83.17	86.39	86.81	88.87
patches/%	92.53	92.87	93.16	94.14
pitted_surface/%	83.62	82.19	87.31	89.64
rolled-in_scale/%	67.41	69.08	77.26	86.93
scratches/%	89.24	93.65	94.18	96.09
FPS/(f/s)	21.87	26.15	26.37	26.39
mAP/%	83.19	86.28	89.66	91.13

Model 3 is the algorithm added with K-means clustering algorithm on the basis of model 2, and Model 4 is the algorithm obtained after redefining the confidence loss function of model 3. Compared with the original YOLOv4, mAP of the improved YOLOv4 is significantly increased from 83.19% to 91.13%, and FPS increased from 21.87 f/s to 26.39 f/s basically meeting the current industrial demand.

4. **Conclusion.** In this paper, an improved algorithm based on traditional YOLOv4 is proposed for detecting defects on metal surface. The extraction of features of the detection target is enhanced by replacing the original basic network CSPDarkNet53 of YOLOv4 with MobileNetv3. In addition, K-means clustering algorithm is adopted to adjust the size of anchor and the confidence loss function is redefined, which further improves the detection precision of the model. Therefore, the improved algorithm can be used for rapid and accurate identification of defects on metal surface and improving the automation level.

However, some defect types are susceptible to environmental light or may undergo dramatic changes in background, and the defect features are complex and varied, resulting in a decrease in the identification precision. Therefore, in subsequent research, the data set can be further expanded to enhance the robustness of the model. And the network model can be further optimized to improve the identification precision and speed of the model.

Acknowledgment. This work is partially supported by National Natural Science Fund Youth Fund Project of China (grant no. 61403222), Heilongjiang Provincial Department of Education (grant no. 135309466). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] S. Inoue, I. Hidaka and T. T. Zin, An intelligent vision-based approach for work group identification through helmet detection, *ICIC Express Letters, Part B: Applications*, vol.13, no.5, pp.511-517, 2022.

[2] R. Girshick, J. Donahue, T. Darrell and J. Malik, Region-based convolutional networks for accurate object detection and segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.38, no.1, pp.142-158, 2016.

[3] C. Zhou, J. Zhou, C. Yu, W. Zhao and R. Pan, Multichannel sliced deep RCNN with residual network for text classification, *Chinese Journal of Electronics*, vol.29, pp.880-886, 2020.

[4] G. Vaca-Castano, N. da Vitoria Lobo and M. Shah, Holistic object detection and image understanding, *Computer Vision and Image Understanding*, vol.181, no.4, pp.1-13, 2019.

[5] Z. Liu, Y. He, C. Wang and R. Song, Analysis of the influence of foggy weather environment on the detection effect of machine vision obstacles, *Sensors*, DOI: 10.3390/s20020349, 2020.

[6] A. Date and S. Shah, Implementation of modified mask RCNN, *International Journal of Innovative Technology and Exploring Engineering*, vol.9, pp.4167-4172, 2019.

[7] Z. B. Liu, M. Subburaman and Y. Shoshana, Does the GH/IGF-1 axis contribute to skeletal sexual dimorphism? Evidence from mouse studies, *Growth Hormone & IGF Research*, vol.27, pp.7-17, 2016.

[8] H.-T. Choi, H.-J. Lee, H. Kang, S. Yu and H.-H. Park, SSD-EMB: An improved SSD using enhanced feature map block for object detection, *Sensors*, vol.21, DOI: 10.3390/s21082842, 2021.

[9] J. Redmon, S. Divvala and R. Girshick, You only look once: Unified, real-time object detection, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, pp.779-788, 2016.

[10] K. Woosuk, C. Hyunwoong, K. Jongseok, K. Byungkwan and L. Seongwook, YOLO-based simultaneous target detection and classification in automotive FMCW radar systems, *Sensors*, vol.20, no.10, DOI: 10.3390/s20102897, 2020.

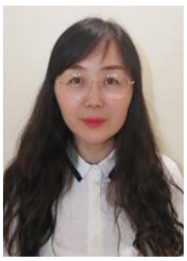
[11] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7263-7271, 2017.

[12] J. Redmon and A. Farhadi, Yolov3: An incremental improvement, *arXiv e-prints*, DOI: 10.48550/arXiv.1804.02767, 2018.

[13] R. Lan, H. Zo, C. Pang, Y. Zhong, Z. Liu and X. Luo, Image denoising via deep residual convolutional neural networks, *Signal, Image and Video Processing*, vol.15, pp.1-8, 2021.

[14] H. Xin, Z. Chen and B. Wang, PCB electronic component defect detection method based on improved YOLOv4 algorithm, *Journal of Physics: Conference Series*, DOI: 10.1088/17426596/1827/1/012067, 2021.
[15] M. Horvat, J. Alan and K. Burnik, Assessing the robustness of cluster solutions in emotionally-annotated pictures using Monte-Carlo simulation stabilized K-means algorithm, *Machine Learning and Knowledge Extraction*, vol.3, pp.435-452, 2021.
[16] Y. He, K. C. Song, Q. G. Meng and Y. H. Yan, An end-to-end steel surface defect detection approach via fusing multiple hierarchical features, *IEEE Trans. Instrumentation and Measurement*, vol.69, no.4, pp.1493-1504, 2020.

Author Biography



Yanju Liu received her B.Sc. degree in computer application from Harbin Institute of Technology, China, 1998; her M.Sc. degree in computer science and technology from Qiqihar University, China, 2007; and her Ph.D. degree in machinery manufacturing and automation from Harbin University of Science and Technology, China, 2013.

Prof. Liu is currently a full-time professor at the School of Mathematics and Information Science, Nanjing Normal University of Special Education, China. Her main research interest is machine learning, defect detecting and three-dimensional reconstruction. She has published over 60 papers in journals and conferences. She is hosting some research projects funded from National Natural Science Foundation of China, China Ministry of Education, etc.

Qiuji Wang received her B.Sc. degree in software engineering from Jiangsu Normal University Kewen College, China, 2018; and her M.Sc. degree in computer technology from Qiqihar University, China, 2022. She has published 4 papers in journals and conferences. Her main research interest is computer vision, machine learning and object detecting.

Huiyu Zhang received her B.Sc. degree in computer science and technology from Qiqihar University, China, 2022. She has gained admission to The University of Auckland. She has published 6 papers in journals and conferences. Her main research interest is machine learning.

Yanzhong Liu received his B.Sc. degree from Qiqihar University, China, 1994, and his M.Sc. degree from Qiqihar University, China, 2009. Now he is an associate professor in Qiqihar University. His main research interest is machine learning and error analysis and defect detecting.

Kaifeng Zhao received his B.Sc. degree in in software engineering from Jining Medical University, China, 2018; and his M.Sc. degree in computer technology from Qiqihar University, China, 2022. He has published 3 papers in journals and conferences. His main research interest is image recognition and image classification.