Surface Defect Detection Algorithm of Aluminum Sheet Based on Improved Yolov3

Liu Yang[1], Guoxiong Hu[1(✉)], and Li Huang[2]

[1] School of Software, Jiangxi Normal University, Nanchang 330027, China
huguoxiong@126.com
[2] School of Education, Jiangxi Normal University, Nanchang 330027, China

Abstract. The surface defect detection of aluminum sheet is of great significance to ensure the appearance and quality of aluminum sheet. The surface defects of aluminum sheets have the characteristics of different shapes, obvious size differences, and difficult to obtain defect samples, which make defect detection challenging. In order to solve this problem, we make the following improvements to YOLOv3: Adding attention mechanism modules after the three feature layers output by the model backbone and after neck upsampling; Freezing the model backbone and using pretrained for transfer learning. The proposed YOLOv3 + ECA model is compared with the target detection models such as YOLOv3 and Faster-RCNN. It is found that the mAP of our model reaches 96.22%, which is higher than the current conventional algorithm. The AP values for different types of defects have good detection results.

Keywords: Defect detection · Attention Mechanism · Few Shot · Small Target Detection

1 Introduction

Aluminum alloys are widely used in transportation, electronics, machinery and other fields [1]. In the process of production and processing, it is easy to have pinhole, scratch, dirty, fold and other surface defects. These surface defects will not only affect the appearance and fatigue strength of products, but also increase the production cost of enterprises. More seriously, it will threaten people's lives. Therefore, it is of great significance to detect defects on the surface of aluminum alloys in time.

In the defect detection of metal industrial products, the commonly used detection methods include manual inspection, magnetic flux leakage testing, radiography testing, traditional machine vision detection and deep learning detection [3–6]. Among them, manual inspection generally has the disadvantages of low precision, strong subjectivity and high labor intensity [2]; Magnetic flux leakage testing is not suitable for detecting cracks with complex shapes and narrow cracks, especially closed cracks; Radiography testing can easily determine the nature of defects, but it is expensive and harmful to human body; Compared with the previous detection methods, machine vision detection

Y. Cao and X. Shao (Eds.): MONAMI 2022, LNICST 474, pp. 127–137, 2023.
https://doi.org/10.1007/978-3-031-32443-7_9

has the outstanding advantages of high detection accuracy, fast detection speed and low cost. However, the traditional machine vision relies on artificial design features, and the robustness of the algorithm is poor. With the advent of the big data era and the improvement of computer computing power, deep learning has developed rapidly. Compared with traditional machine vision technology, deep learning achieves higher accuracy in image classification, image segmentation, target detection and other fields [5].

Owing to the superiority of deep learning algorithm in defect detection, the detection method based on deep learning has been widely concerned and studied. However, there are not many applications in actual industrial scenarios. The main reason is determined by the characteristics of defect data. The details are as follows: (1) The defect shapes are different. (2) Small target defect samples are few. (3) The size of different types of defects is obviously different. (4) It is difficult to obtain defect samples. This makes the practical application of surface defect method based on deep learning very challenging.

The attention mechanism has the properties of plug-and-play nature, which improves the detection accuracy of the model without significantly increasing the amount of computation. And it can also enable the network model to pay attention to more valuable information for the task, thereby improving the efficiency of task processing and the ability of feature extraction.

Therefore, this paper focuses on the attention mechanism and improves the YOLOv3 algorithm. The contribution of this paper are as follows:

1. Building defect dataset. The industrial defect images on the surface of aluminum sheet were collected by industrial camera and classified into four types of defects. Namely Pinhole, Scratch, Dirty and Fold.
2. Improving YOLOv3 model structure. We add attention mechanism modules to the neck of the network model to improve the feature extraction ability. It effectively solves the problem that small target defects are difficult to detect.
3. Model training with transfer learning. In view of the problem that it is difficult to obtain defect dataset, this paper uses pretrained weights to perform transfer learning by freezing the backbone of the model. This method not only speeds up the training speed of the model, but also enables the model to achieve better detection results on small datasets.

This paper is structured as follows: In the first section, we expound the background and significance of defect detection on the surface of aluminum sheets, enumerate the current main defect detection methods, and analyze their deficiencies. Section 2 introduces the related research on object detection algorithms and attention mechanisms. Section 3 proposes solutions to the problem that the surface defects of aluminum sheets are difficult to detect. Section 4 conducts a comparative experiment on the improved YOLOv3 algorithm and analyzes the performance of the algorithm. Section 5 briefly summarizes the content of the previous sections and draws conclusions.

2 Related Work

Object detection can obtain accurate positioning and category information of targets. At present, it is mainly divided into two-stages algorithms and one-stage algorithms. The former first generate candidate boxes that may contain targets, and then adjust the boxes and classify the targets. The latter does not generate candidate boxes, and the final detection result can be obtained after a single detection.

The main algorithm of the two-stage detection method is Faster-RCNN [7]. For example, Ding *et al*. [8] proposed a network dedicated to tiny defect detection (TDD-net). This method strengthens the fusion of information from the underlying structure and improves the detection accuracy of small defects. He *et al.* [9] proposed a method based on Faster R-CNN, using FPN, Soft Non-Maximum Suppression (NMS) and Region of Interest (ROI) alignment to improve the accuracy of the model. But the detection speed is slow. The main representatives of one-stage detection methods are YOLO and SSD. For example, Guo *et al.* [10] proposed MFST-YOLO model. Based on YOLOv5, this method combines Trans module designed by transformer and multi-scale feature fusion structure. It can quickly detect defects on steel surface, but the average detection accuracy still has great room for improvement.

Attention mechanism is widely used in the field of object detection due to its plug-and-play characteristics and the ability to capture valuable information. For example, Yao *et al.* [11] proposed an enhancing region and boundary Awareness Network (ERBANet). The author combines ERBANet with attentional feature enhancement (AFE) module. The method achieves faster detection speed and higher detection accuracy. Li *et al.* [12] proposed a hybrid attention mechanism, which is mainly composed of channel attention, spatial attention and aligned attention modules. Then the hybrid attention mechanism is combined with the single-stage target detection algorithm HAR to achieve target detection. The detection accuracy of the algorithm has been significantly improved.

Similarly, for the object detection task, the acquired image quality is also related to the accuracy of defect detection. Currently, the popular image quality assessment (IQA) methods are the no-reference(NR) techniques. For example, Hu *et al.* [14] proposed parametric models that describe general characteristics of chromatic data in natural images, which are unified in a common NR IQA metric. The proposed metric provides solutions to various color image processing problems. In addition to color images, night-time images are also used as objects for object detection. A night-time blind image quality assessment (BIQA) method based on support vector regression (SVR) [15] can effectively predict night-time image quality.

All of the above studies provide more possibilities for the detection of surface defects of aluminum sheets. At the same time, it can be seen from the above research that defect detection has produced a relatively large contradiction in accuracy and speed. This contradiction can be effectively alleviated by combining attention mechanism.

3 Methods

3.1 Improved YOLOv3 Model

YOLOv3 is a classic one-stage target detection algorithm. Compared with the two-stage detection algorithm, it can directly generate the object category probability and position

coordinate value without going through the candidate region generation stage. So it has faster detection speed, but there is still room for improvement in detection accuracy. Therefore, while maintaining the detection speed, this paper introduces attention mechanism ECA module to improve the detection accuracy. The YOLOv3 network model combined with ECA module as show in Fig. 1.

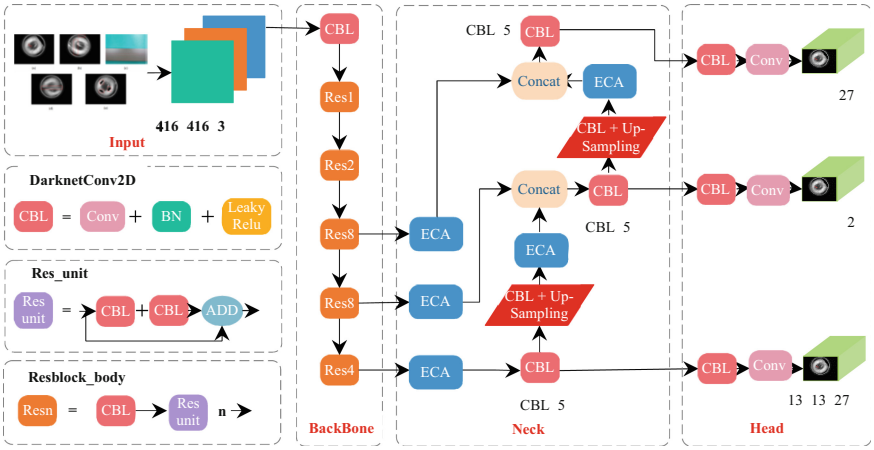

Fig.1. Improved YOLOv3 model

BackBone. The model can be divided into four parts: input, backbone, neck and head. Among them, the backbone feature extraction network is Darknet-53. This part has a total of 6 separate convolution layers and 23 residual blocks. As can be seen from Fig. 1, the input features of the residual block first go through a convolutional layer CBL with a kernel size of 1×1 and a stride of 1. Its number of channels is reduced to $\frac{1}{2}$ of the original, which is $\frac{Inchannels}{2}$. Then enter a 3×3 convolutional layer for feature extraction, and the number of channels is restored to *Inchannels*.

The output of the 3×3 convolution is directly added to the skip connection to obtain the output. After residual operation, the shape of the input feature map remains unchanged. The skip connections in it alleviate the gradient vanishing problem caused by increasing depth in deep neural networks. Secondly, the convolution part of Darknet-53 uses the unique darknetconv2d structure. The structure first performs a convolution and L2 regularization. After completing the convolution, Batch Normalization and LeakyReLU activation function are performed. The definition of LeakyReLU is shown in formula (1).

$$\text{LeakyReLU (x)} = \begin{cases} x & x > 0 \\ \alpha x & x \leqslant 0 \end{cases} \tag{1}$$

Neck. After Darknet-53 feature extraction, three effective feature layers are obtained. Their sizes are (52, 52, 256), (26, 26, 512), (13,13, 1024). Attention mechanism is a plug-and-play module. In order to successfully use the pretrained weight for transfer

learning in the future, we do not add the attention module to the backbone network, but add the attention module after the three feature layers extracted from the backbone network and after two upsampling of the neck of the model. The introduction of the attention mechanism can make the model pay attention to the more valuable information for the task among the numerous information, so as to improve the efficiency and accuracy of task processing. Attention mechanism can be roughly divided into channel attention mechanism, spatial attention mechanism, and channel and spatial mixed attention mechanism.

ECA module is an implementation of channel attention mechanism. Adding this module to the model not only does not increase the complexity of the model, but also effectively improves the detection effect of the model. The structure of the ECA module is shown in Fig. 2. We first perform global average pooling(GAP) operation on the input feature $\chi \in R^{H*W*C}$ and compress it into $Z \in R^{1*1*C}$. The calculation formula is shown in formula (2). Then, we input Z into 1D convolution with a kernel size of k for feature learning to get the channel weights. The calculation formula of k is shown in formula (3). Finally, the results can be obtained by multiplying the weight to the feature directly.

$$z_c = \frac{1}{H * W} \sum_{i=1}^{H} \sum_{j=1}^{W} \chi_c(i, j) \tag{2}$$

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{3}$$

where C indicates the number of channels; $|t|_{odd}$ represents the nearest odd number of t; $\psi(\cdot)$ indicates nonlinear mapping.

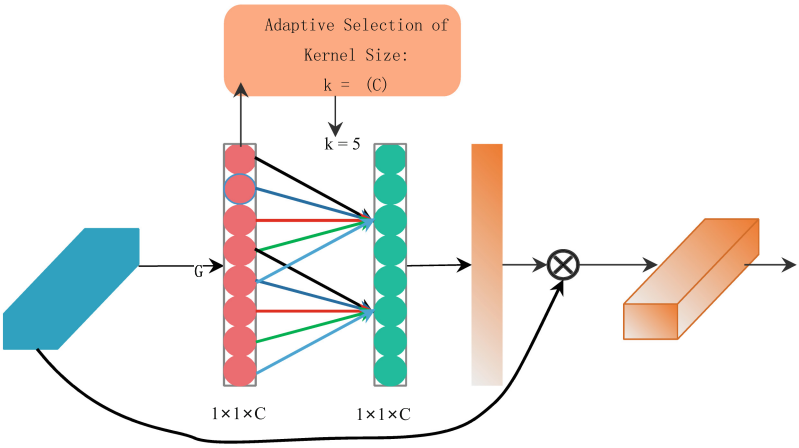

Fig. 2. ECA module.

Head. After the input features are passed through the neck network, three enhanced features can be obtained. Then these three enhanced features are passed into Head through

the convolutional layer to obtain the prediction result. The shapes of the prediction results are (13, 13, 27), (26, 26, 27), (52, 52, 27). However, these prediction results do not correspond to the position of the final prediction box on the picture, so they still need to be decoded. The decoding process is to calculate the coordinates of the last displayed bounding box: b_x, b_y, and the width and height:b_w, b_h. In this way, we actually get the position of the predicted box. Each prediction box includes the position of prediction box, category information and confidence information. The calculation process is as follows:

$$b_x = \sigma(t_x) + c_x \tag{4}$$

$$b_y = \sigma(t_y) + c_y \tag{5}$$

$$b_w = p_w e^{t_w} \tag{6}$$

$$b_h = p_h e^{t_h} \tag{7}$$

where c_x and c_y are the coordinate of the upper left corner of the cell; p_w and p_h denote the edge length of priori box; $\sigma(t_x)$ and $\sigma(t_y)$ are the offset based on the grid point coordinates of the upper left corner of the center point of the rectangular box; t_w and t_h denote the width and height of the prediction box.

3.2 Loss Function

To calculate the model loss, we should first calculate the Intersection over Union (IOU) of all ground-truth and prediction box in each image (see Fig. 3). When the IOU is greater than ignore thresh, we take out the priori box with the largest IOU in each network point. The formula for calculating IOU is as follows:

$$IOU = \frac{A \cap B}{A \cup B} \tag{8}$$

Then we can calculate the model loss. The loss of this model can be divided into location loss, confidence loss and category loss. The calculation formulas correspond to formulas (9), (10) and (11) respectively.

$$lbox = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} (2 - w_i \times h_i)[(x_i - \hat{x}_i^j)^2 + (y_i - \hat{y}_i^j)^2 + (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] \tag{9}$$

$$lobj = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} (c_i - \hat{c}_i)^2 \tag{10}$$

$$lcls = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} \sum_{c \in classes} p_i(c) \log(\hat{p}_i(c)) \tag{11}$$

where S is the grid size; B denotes the number of candidate boxes; I_{ij}^{obj} denotes whether there is a target object in the jth prior box of the ith grid. If there is, $I_{ij}^{obj} = 1$. Otherwise, $I_{ij}^{obj} = 0$.

The loss of the whole model is the sum of the above three. The calculation formula is as follows:

$$Loss = lbox + lobj + lcls \tag{12}$$

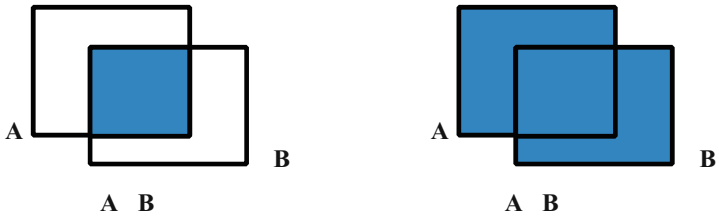

Fig. 3. IOU principle diagram.

4 Experimental Results and Analysis

4.1 Transfer Learning of Pretrained Mode

Transfer learning [13] is a technical means to apply the machine learning model trained in one field to another. It increases the utilization of the model to some extent, shows the superior performance when the amount of training data is small, and saves the training time and storage cost. For the neural network, it only needs to cut a trained neural network from the middle, and then splice it to other networks to realize the transfer learning. In the training phase, the topological structure and all super parameters of the migration module can remain unchanged, and the weight can decide whether to retrain according to needs. Transfer learning is divided into pretrained mode and fixed value mode according to whether to update the weight of the old module. Faced with the scarcity of training data for aluminum sheet surface defects, we have more advantages in using the pretrained transfer method.

We divide the training into two stages: freezing stage and unfreezing stage. In the freezing phase, the backbone of the model is frozen, and the feature extraction network is not changed, only the network is fine tuned. We set the initial learning rate to 0.01. During the training, we can adjust the batch size appropriately according to the use of the video memory. After unfreezing, all parameters of the network will be changed.

4.2 Experiment Description

All experiments were carried out on windows10 operating system with pytorch based on Python 3.8. The pretrained model is used in the training process. The batch size of

the freeze phase is 16. The batch size in the unfreeze phase is 8. The initial learning rate is 0.001, and its decline mode is cos. Momentum is set to 0.937, IOU threshold is set to 0.5, and Adam optimizer is used.

4.3 Datasets

In the actual industrial production, the product yield is relatively high. So it is difficult for us to obtain a large number of defect image samples. The dataset in this paper is collected by industrial camera, including 4 types of aluminum sheet surface defects (see Fig. 4). We have a total of 401 images with more than 1000 defects. Quantitatively, this is far from sufficient for object detection model training. Therefore, in order to reduce the possibility of over fitting, this paper performs data enhancement operations, which include randomly removing some pixels, sharpening, affine transformation, changing brightness, tone following and horizontal flipping to increases the number of images to 2459. Then, we randomly select 10% of the data as the test set, and divide the rest of data into training and validation sets in a 9:1 ratio.

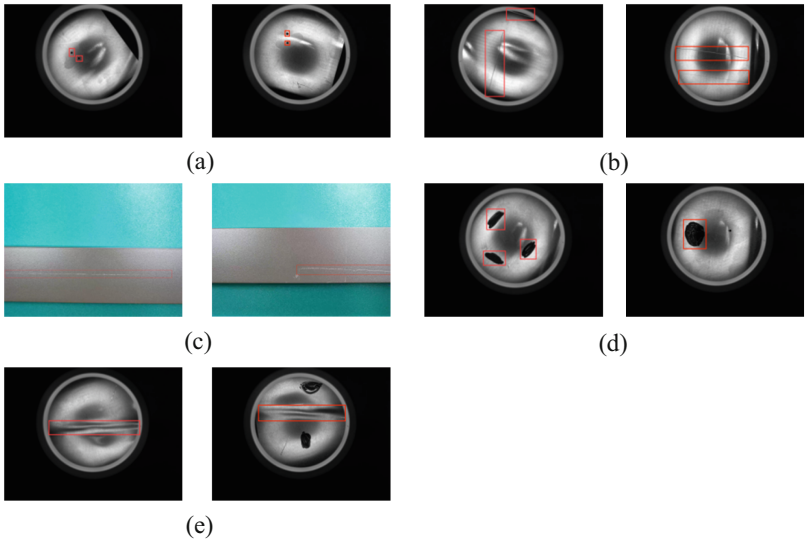

Fig. 4. Four types of aluminum sheet surface defects and corresponding label frames. (a) Pinhole: a circular small hole with an aperture less than or equal to 2 mm; (b) and (c) Scratch: linear scar with different length and depth; (d) Dirty: irregular black brown block; (e) Fold: part of the surface metal of the aluminum sheet is folded into the aluminum sheet, causing the metal to form overlapping layer defects.

4.4 Experimental Results

In order to verify the effectiveness of the attention adding mechanism proposed in this paper. We compare YOLOv3 + ECA with target detection models such as YOLOv3 and

Faster-RCNN. The evaluation indicators include Precision, Recall, Average Precisio (AP), mean Average Precision (mAP) and FPS. The definitions of Precision and Recall are shown in formulas (13) and (14) respectively:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

where *TP* is the number of samples that are predicted to be positive and are actually positive; *TN* denotes the number of samples that are predicted to be negative and actually negative; *FP* denotes the number of samples with positive prediction and negative actual prediction; *FN* denotes the number of samples with negative prediction and positive actual prediction.

AP is the area enclosed by the precision-recall curve (P-R) and the coordinate axis. mAP denotes the mean value of AP, which reflects the comprehensive performance of the object detection algorithm. The definitions of AP and map are shown in formulas (15) and (16) respectively:

$$\text{AP} = \int_0^1 \text{p}(r)dr \tag{15}$$

$$\text{mAP} = \frac{1}{\text{n}} \sum \text{AP} \tag{16}$$

where $p(r)$ denotes the P-R curve,; n denotes the number of types of defects.

As can be seen from Table 1 and Fig. 5, the mAP of SSD is 92.56%, but the AP for the unclear defect such as scratch is only 78%. The mAP of Faster-RCNN reaches 94%, but the AP for small target defects such as pinhole is only 79%. The mAP of YOLOv3 is 66.42%, and the AP for different types of defects are all lower than the other three models. The mAP of YOLOv3 + ECA reached 96.22%. Detection speed is 20 FPS. Compared with the original YOLOv3, its detection rate has been greatly improved. At the same time, compared with SSD, YOLOv3 + ECA is slightly inferior in detection speed, but it can still meet the needs of industrial real-time detection. Although the AP and mAP of Faster-RCNN can also reach high values, their detection speed is only 6 FPS, which cannot meet the needs of industrial real-time detection. It can be seen that the effectiveness and superiority of the YOLOv3 + ECA.

Table 1. Performance comparison of different defect detection models.

Model	AP (%)				mAP (%)	FPS
	Pinhole	Scratch	Dirty	Fold		
SSD	94	78	100	98	92.56	**33**
Faster-RCNN	79	**98**	100	**100**	94.32	6
YOLOv3	89	10	100	95	73.54	8
YOLOv3 + Transfer Learning	60.70	7.60	99.96	97.41	66.42	21
YOLOv3 + Transfer Learning + ECA	**95**	91	**100**	99	**96.22**	20

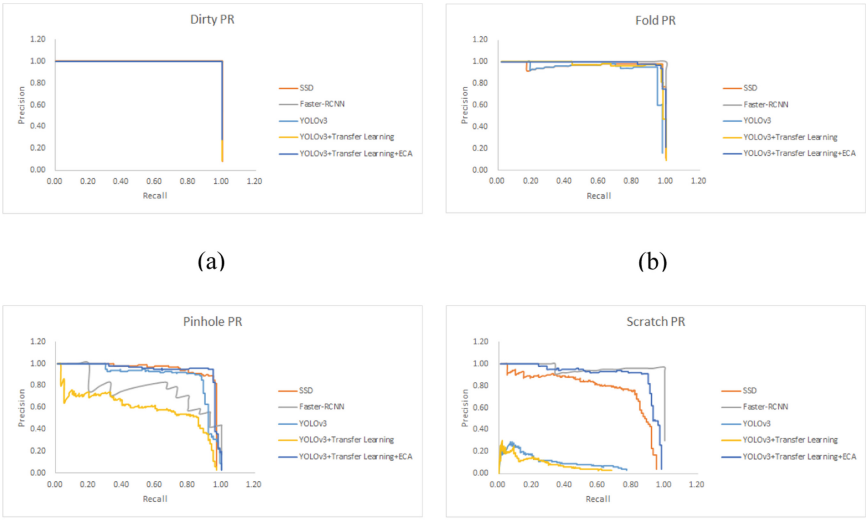

Fig. 5. PR curves of different target detection algorithms and different defects.

5 Conclusions

In this paper, we have improved YOLOv3 to detect surface defects of aluminum sheets. On the one hand, in order to solve the problem of small dataset size, this paper trains through transfer learning. On the other hand, in order to improve the detection accuracy of the algorithm for small target defects and unclear defects, this paper adds attention modules based on the original YOLOv3. The experimental results show that the mAP of the improved YOLOv3 algorithm is improved. Especially for scratch detection, the AP of the original YOLOv3 is only 10%, while the improved model improves to 91%. Moreover, the detection time of this model is 20 FPS, which can meet the needs of industrial real-time detection. Since there are not only four types of defects in aluminum

sheets in reality, we will continue to study the automatic detection of different types of defects in the future.

Funding Statement. This work was supported by the Science and Technology Project of Jiangxi Provincial Department of Education under Grant no. GJJ200305 and GJJ191689, the Natural Science Foundation of Jiangxi Province under Grants no. 20202BABL202016.

References

1. Zhang, J., Song, B., Wei, Q., et al.: A review of selective laser melting of aluminum alloys: processing, microstructure, property and developing trends. J. Mater. Sci. Technol. **35**(2), 270–284 (2019)
2. Tao, X., Zhang, D., Ma, W., et al.: Automatic metallic surface defect detection and recognition with convolutional neural networks. Appl. Sci. **1575**, 1–15 (2018)
3. Dehui, W., Lingxin, S., Wang, X., et al.: A novel non-destructive testing method by measuring the change rate of magnetic flux leakage. J. Nondestr. Eval. **36**(2), 1–11 (2017)
4. Malarvel, M., Singh, H.: An autonomous technique for weld defects detection and classification using multi-class support vector machine in X-radiography image. Optik **231**, 166342 (2021)
5. O'Mahony, N., et al.: Deep learning vs. traditional computer vision. In: Arai, K., Kapoor, S. (eds.) CVC 2019. AISC, vol. 943, pp. 128–144. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-17795-9_10
6. Yang, J., Li, S., Wang, Z., Dong, H., Wang, J., Tang, S.: Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges. Materials **13**(24), 5755 (2020)
7. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
8. Ding, R., Dai, L., Li, G., et al.: TDD-net: a tiny defect detection network for printed circuit boards. CAAI Trans. Intell. Technol. **4**(2), 110–116 (2019)
9. He, D., Wen, J., Lai, Z., et al.: Textile fabric defect detection based on improved faster R-CNN. AATCC J. Res/ **8**(1_suppl), 82–90 (2021)
10. Guo, Z., Wang, C., Yang, G., et al.: MSFT-YOLO: improved YOLOv5 based on transformer for detecting defects of steel surface. Sensors **22**(9), 3467 (2022)
11. Yao, Z., Wang, L.: ERBANet: enhancing region and boundary awareness for salient object detection. Neurocomputing **448**, 152–167 (2021)
12. Li, Y.-L., Wang, S.: HAR-Net: Joint learning of hybrid attention for single-stage object detection. IEEE Transactions on Image Processing **29**, 3092–3103 (2020)
13. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
14. Hu, R., Liu, Y., Gu, K., Min, X., Zhai, G.: Toward a no-reference quality metric for camera-captured images. IEEE Trans. Cybern. (2021)
15. Hu, R., Liu, Y., Wang, Z., et al.: Blind quality assessment of night-time image. Displays **69**, 102045 (2021)