Machine vision problem for fast recognition of surface defects of thermoelectric cooler components based on deep learning method

Z Q Yu¹, M Zhao¹, J L Huang², T X Wen³ and T D Liao⁴

¹ School of Computer Science and Engineering, Central South University, Changsha 41000, China
² Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, Fujian, 362000, China
³ College of Engineering, Huaqiao University, Quanzhou, Fujian, 362001, China
⁴ Research Center for Photonics Technology, Quanzhou Normal University, Quanzhou, Fujian, 362000, China

Email: yuz997@csu.edu.cn

Abstract. During thermoelectric coolers (TEC) production, a complex industrial manufacturing process must be experienced, which may cause defects on the surface of the TEC component. To improve the efficiency of TEC component defect inspection, we propose a machine vision technology based on deep learning for surface defect detection. In order to make the deep learning method based on the you only look once (YOLO) model more efficient, first of all, we use a more lightweight network ResNet34 to improve the original network structure. Then, the loss function is improved to complete intersection over union (CIoU) loss. Experiments performed using the proposed model, show an obvious reduction in the number of parameters, the detection speed is as high as 6.5pcs/s, and the detection accuracy is 97.61%. This method lays a good foundation for the further application of deep learning methods in the field of industrial detection. The experimental results verify the feasibility and effectiveness of the model.

1. Introduction
Thermoelectric coolers (TEC) are widely used in industrial, telecommunications, aerospace equipment, and semiconductor manufacturing processes for precise temperature control. It can be said that the development of high-tech industries is inseparable from TEC. In the manufacturing process of TEC, it is necessary to go through the process of part slicing, welding, testing and packaging, which will inevitably produce some defects, such as surface cracks, dirt, and damage [1]. The existence of these surface defects may cause the temperature control device to malfunction. Therefore, it is necessary to detect defects in the production and processing of TEC parts. At present, the defect inspection in the industry generally adopts manual visual inspection. However, this method has high labor cost but low efficiency. Therefore, it is inevitable to develop new technologies for intelligently detecting surface defects of TEC parts.

Although many researchers have proposed some machine vision algorithms to achieve the task of defect inspection. Zhang et al. proposed a vision inspection system for the surface defects of strongly reflected metal based on multi-class support vector machine (SVM) [2]. The research developed a
computer vision-based system for inspection of strongly reflected metal surface defects based on wavelet transform, spectral measure, and SVM. The classification results show that this method can effectively identify metal surface defects, but it cannot guarantee the real-time inspection requirements in the industry.

In recent years, deep learning technology has achieved great development in the field of image vision, and its practical application effects in many fields far exceed traditional technologies. The object detection method based on deep neural network technology can actively learn target features, and achieve the dual effects of positioning and classification while maintaining accuracy and speed. The current object detection algorithms can be roughly divided into two categories: "two-stage detection" and "one-stage detection". "Two-stage detection" means that the detection algorithm needs to be completed in two steps. First, it needs to obtain candidate regions and then classify. The representative of this type of algorithm is the region-conventional neural network (R-CNN) series, such as R-CNN, Fast R-CNN and Faster R-CNN, etc [3-5]. The opposite is "one-stage detection", which can be understood as one step, no need to find candidate areas separately, but an algorithm that uses regression to directly detect the location and category of the target. Typical examples are SSD and YOLOv3 [6-7].

Surface defect inspection is a branch of object detection which applies the algorithms to the actual detection scene. For instance, Yin et al. used the YOLOv3 network as the target detector in the sewage pipeline defect detection system to detect six types of defects (fractures, holes, sediments, cracks, fractures, and roots) [8]. The average f1 scores of the verification set and the test set were respectively for 0.876 and 0.882, mean average precision (mAP) reached 85.37%, which verified the performance of the model.

Taking into account the requirements of TEC parts defect inspection, we choose a deep learning model based on YOLOv3. Compared with the two-stage detection algorithm, it maintains a good balance between accuracy and speed. But in actual detection scenarios, this method still has some problems. The complexity of network training is relatively high and the number of model parameters after training is large when using the original YOLOv3 model. The high cost of use in industrial production lines makes it impossible to put into practical application. For the problems give above, we proposed an effective improvement to the original algorithm, which can further apply to the unique defect inspection scenario of TEC components.

2. Methodology

The main methodology of this research is to quickly and accurately detect defects in TEC components based on the YOLOv3 model. However, it is only used to detect defects of TEC parts, and the parameters of the feature extraction layer in the original YOLOv3 network are too redundant, which may lead to overfitting. That is, the training accuracy is high, but the test accuracy is low. Therefore, we redesigned a lighter model and replaced the feature extraction network in the original model with ResNet34 composed of residual modules [9]. While reducing the computational complexity, the original accuracy is maintained to the greatest extent. In addition, the bounding box regression loss function is optimized, and complete intersection over union (CIoU) loss is used to improve the back propagation process of the network [10]. The improved model is more suitable for the specific scene of TEC part defect inspection. Finally, the model structure of this article is shown in Figure 1.
Figure 1. Our network structure diagram. (1) Input TEC component image 1920*1200. (2) Reset image size 416*416. (3) Feature extraction. (4) Multi-scale training. (5) YOLO detection. (6) Prediction output.

2.1. Backbone
Object detection consists of two parts, one part is migrated by the image classification network and used as a feature extractor. This part, we call it the backbone of the model. Another part of the subsequent network is responsible for detecting the location and category of the object from these features. Backbone can effectively extract the category features of the target in the image, which is crucial for the detection task.

As the backbone of the YOLOv3 algorithm, the Darknet-53 network extracts deeper feature information with high accuracy. However, in a limited computing unit, it is likely that the amount of calculation will increase as the network depth increases, thereby reducing the speed of detection. In our study, in order to reduce the number of parameters in the actual detection model, we chose the ResNet34 as a new backbone, which is excellent in speed and accuracy.

ResNet is the most widely used CNN feature extraction network, which can be understood as a classification network composed of many residual units. ResNet constructs a residual learning unit by superimposing an identity mapping layer on a shallow network, which reduces the redundancy of information in the data while maintaining high precision.

As shown in Figure 2, the residual module is divided into two parts: direct mapping and residual mapping: the residual mapping part generally consists of two or three convolution operations. The direct mapping part is to directly transmit the output of the upper layer of the residual network to the next layer of the network without adding additional parameters. At the same time, the gradient is directly passed to the upper layer of the network during the back propagation process. Its expression is as follows:

\[ x_{i+1} = f(h(x_i) + F(x_i, W_i)) \]  

Among them: \( x_i \) is the input, \( W_i \) is the convolution operation, \( f \) is the activation function, \( h(x_i) \) is the direct mapping part, \( F(x_i, W_i) \) is the residual mapping part. There are 32 layers of residual units in ResNet34 used in this article. The number of residual blocks corresponding to block1, block2, block3, and block4 are 3, 4, 6, 3 respectively, plus a 7x7 convolutional layer and a fully connected layer, a total of 34 floors. The pooling layer is not counted as one layer.

2.2. Loss
The loss function of the model also consists of two parts: classification loss and bounding box regression loss. In the training process, the loss function of the original model can be divided into three main parts according to the function, namely classification loss, confidence loss and coordinate loss. These functions facilitate the back propagation process of the network during training:

\[ loss = loss_{coord} + loss_{conf} + loss_{class} \]
In this article, the bounding box regression loss function of the original model is improved, and CIoU loss is selected to replace the IoU loss of the original model. IoU actually measures the relative size of the overlap of the two bounding boxes. The larger the overlap between the prediction box and the real box, the better the position prediction effect of the model.

\[ \mathcal{L}_{\text{IoU}} = 1 - \frac{|B \cap B^\text{gt}|}{|B \cup B^\text{gt}|} \quad (3) \]

Where \( B^\text{gt} \) is the ground truth, and \( B \) is the prediction box. Traditionally, \( \ell_n \) - norm loss (for example, \( n = 1 \) or \( 2 \)) is used to measure the distance between bounding boxes on the coordinates of \( B \) and \( B^\text{gt} \). Redmon used IoU loss in YOLOv3 to improve the union index. But IoU Loss cannot optimize the situation where two boxes do not intersect.

On the basis of IoU, CIoU adds the consideration of the euclidean distance and scale loss of the smallest bounding rectangular, so that the predicted box will be more in line with the real box. Accordingly, we choose CIoU loss to make the model converge faster which can be put into production faster after training.

\[ \mathcal{L}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b,b^\text{gt})}{c^2} + \alpha v \quad (4) \]
\[ \alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (5) \]
\[ v = \frac{4}{\pi^2} \left( \arctan \frac{w^\text{gt}}{h^\text{gt}} - \arctan \frac{w}{h} \right)^2 \quad (6) \]

where \( b \) and \( b^\text{gt} \) denote the central points of \( B \) and \( B^\text{gt} \), \( \rho(\cdot) \) is the euclidean distance, and \( c \) is the diagonal length of the smallest enclosing box covering the two boxes. where \( \alpha \) is a positive trade-off parameter, and \( v \) is used to measure the consistency of the aspect ratio.

3. Experiment and analysis
The YOLOv3-ResNet-CIoU model proposed in this study is trained and verified based on the Darknet framework. In the training process, the network parameters are gradually optimized and adjusted. In the final training process, the network is trained for a total of 2000 times, the batch size is 32, the learning rate is set to 0.001, and the steps mode is selected to update the learning rate. Finally, the original YOLOv3, improved YOLOv3-ResNet and YOLOv3-ResNet-CIoU models were trained for nearly 400 epochs respectively. The experimental environment is as follows: Intel (R) Core (TM) i7-9700 CPU @ 3.00GHz, 32 GB RAM, GeForce RTX 2080 super GPU, and training based on the Ubuntu 18.04.4 operating system.

3.1. Dataset
In order to develop a database containing defect images of multiple TEC components, a total of 180 images (with a resolution of 1920 × 1200 pixels) were collected using the Haikang MV-CA020-20GM camera. Collect pictures of the same surface of the same TEC component sample under different motion speeds, different exposure times, and different subdivisions of the feeder stepping motor to drive the turntable. Three speeds of 6 grains/sec, 8 grains/sec, and 10 grains/sec were selected respectively, the pulse subdivision numbers were 6400, 12800, and 25600, and the exposure time was 60us, 100us, 200us, and 300us. Except for the first type, which is qualified samples, the rest are defective samples, including four types of defects, such as special-shaped, cracks, smudge, and hydrolysis.

3.2. Analysis of results
The models compared in this study were trained using the same TEC data set. The experimental curve in Figure 3 shows that the overall trend of the three models is roughly the same, that is, the loss value decreases rapidly in the early training stage, and the jitter range is larger. In the middle and late stages of training, the change of the loss value tends to stabilize while the rate of decline becomes slow.
However, the loss curves of these three models also have some differences. The most obvious difference that can be seen from the figure is that the violent jitter time of the loss value of the YOLOv3-ResNet-CIoU model is shorter than that of the other two models, and the jitter amplitude is smaller. The entire loss curve is relatively smoother which shows that the improved model can converge faster, and the loss value after convergence is also lower than the other two models.

Figure 3. The loss curves of the three YOLOv3 models during training

Figure 4. Comparison of the precision recall curves for three models

In the object detection task, the average precision (AP) is calculated using precision and recall, and finally the average precision (mAP) is used as the performance evaluation standard of the detection model. The accuracy of a certain category is calculated by the following formula:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (7)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (8)
\]

TP is a true number, indicating that the detection frame is correct and the IoU is greater than the threshold. FP is the number of false positives, which represents an error in the detection frame, and the IoU is less than the threshold, while FN is the number of false positives, which represents a real frame, but it is not detected by the model.

It can be seen from the P-R graph in Figure 4 that as recall increases, precision begins to fluctuate, and the overall trend is declining. When the recall rate is about 0.63, the accuracy of the YOLOv3 model drops to about 0.91, and the YOLOv3-ResNet model drops to 0.92 when the recall rate is about 0.76, but the accuracy of YOLOv3-ResNet-CIoU remains at about 0.98. In other words, the improved model has obvious advantages under the same recall rate, which shows that using CIoU as the loss function can train the model more fully and improve the detection performance of the model.

Table 1. Performance comparison between YOLOv3 and the two improved models

|        | P (%) | R (%) | mAP (%) | Speed (pcs/s) | Parameter Size (M) |
|--------|-------|-------|---------|---------------|-------------------|
| Origin YOLOv3  | 88.36 | 72.82 | 84.81   | 5.1           | 234.98           |
| YOLOv3-ResNet  | 98.44 | 54.09 | 88.66   | 6.8           | 159.30           |
| YOLOv3-ResNet-CIoU | 83.93 | 67.96 | 97.61   | 6.5           | 159.30           |

The final improved YOLOv3-ResNet-CIoU model used in this study reached an accuracy of about 97% and detection speed of 6.5pcs/s shown in Table 1. Compared with the original YOLOv3 model, the accuracy is improved by about 12.8%, and the speed is also increased by about 1.4pcs/s. First, the increase in accuracy is related to the replacement of the loss function. By directly minimizing the normalized distance between two center points, the CIoU loss can make the model converge faster and better. Secondly, the increase in detection speed benefits from the improvement of the network structure. We replaced the original Darknet53 convolutional network with a more lightweight backbone—ResNet34, and the model parameter size was reduced from 235M to 159M. In general, compared with
the original model, the YOLOv3-ResNet-ClIoU model has a better detection effect on the TEC components data set.

Figure 5 demonstrates a sampling of the results. It can be seen that this model can simultaneously identify and locate different types of defects on TEC component images.

![Detection result of TEC defects based on YOLOv3 model](image)

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

**Figure 5.** Detection result of TEC defects based on YOLOv3 model (a) crack (b) shaped (c) smudge (d) hydrolysis.

4. Conclusion

Our study applied deep learning method to the surface defect inspection of TEC components and improves the basic convolutional network framework and loss function of the YOLOv3 model. The improved model has reached the expected results from the experimental results. However, we also found some problems in the experiment. For instance, the test accuracy is too high which means overfitting may exited. In future work, we will conduct a more in-depth study of this method. Cross-validation and data expansion may be considered to increase the diversity of training samples which can better serve the inspection of surface defects of TEC components.

References

[1] Zhen H and Kefei S 2011 Design of thermostat system based on Proteus simulation software. Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology pp 1901-1904.

[2] Xue Z, Yan D, Yan L, Ai S and Rui L 2011 A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM. Expert Systems with Applications 38 5930-5939.

[3] Girshick R, Donahue J, Darrell T and Malik J 2014 Rich feature hierarchies for accurate object detection and semantic segmentation Proceedings of the IEEE conference on computer vision and pattern recognition pp 580-587.

[4] Girshick R 2015 Fast r-cnn Proceedings of the IEEE international conference on computer vision pp 1440-1448.

[5] Ren S, He K, Girshick R and Sun J 2015 Faster r-cnn: Towards real-time object detection with region proposal networks arXiv preprint.

[6] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C 2016 Ssd: Single shot multibox detector European conference on computer vision pp 21-37

[7] Redmon J and Farhadi A 2018 Yolov3: An incremental improvement. arXiv preprint

[8] Yin X, Chen Y, Bouferguene A, Zaman H, Al-Hussein M and Kurach L 2020 A deep learning based framework for an automated defect detection system for sewer pipes Automation in Construction 109 102967.
[9] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition Proceedings of the IEEE conference on computer vision and pattern recognition pp 770-778.

[10] Zheng Z, Wang P, Liu W, Li J, Ye R and Ren D 2020 Distance-IoU loss: Faster and better learning for bounding box regression Proceedings of the AAAI Conference on Artificial Intelligence 34 (7) pp 12993-13000.