Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230835

Morphology and Transformer-Based YOLOv5 for Workpiece Surface Crack Detection

Xinghua REN[a], Shaolin HU[a,b1], Yandong HOU[a], Ye KE[b]
[a] *School of Artificial Intelligence, Henan University, Zhengzhou, China*
[b] *School of Automation, Guangdong University of Petrochemical Technology, Maoming, China*

Abstract: The surface crack of the workpiece is similar to the background and the background is complex. In the detection process, problems such as missed detection, false detection, and difficulty in detection are prone to occur. To solve the above problems, this research proposes a workpiece surface crack detection technique based on morphology and improved YOLOv5. To improve the ability of the model to extract global information, the erosion method in morphology is used to improve the crack feature of the data set. Next, the global and local information in the YOLOv5 backbone network is fused by MobileViTv3. And finally, SIoU is used as the loss function of the bounding box regression to improve the accuracy of the bounding box localization. After comparative experiment verification, the designed model achieves 73.1% and 74.6%accuracy on the original and corrosion datasets, respectively. Accuracy of the model improves by 13.6% and 15.1%, respectively. The results of the example experiment show that the method proposed in this paper has a good detection effect, realizes the accurate detection of cracks on the device surface. And it provides a novel idea for using deep learning to detect cracks in real scenes.

Keywords: Crack detection; Morphology; YOLOv5; MobileViTv3; SIoU

1 Introduction

Cracks are the early manifestation that an object has structural damage. When the crack is on the surface of the workpiece, and the depth of the crack is much smaller than the thickness of the workpiece, the crack is called the surface crack of the workpiece. The surface cracks of workpieces have brought huge hidden dangers to industrial production and economic activities. Taking train rails as an example, my country's "Rail Damage Classification" clearly states that the fractures of the rails are due to the accumulated damage on the surface of the rails. When the rails are broken, there are certain dangers in the train running. Therefore, the timely detection of cracks on the surface of workpieces can reduce the safety hazards of industrial production, which is a realistic requirement for safe production.

The machine vision method is a common method for crack detection because of its fast detection speed and high precision. In paper [1], the red channel of the ceramic tile

[1] Corresponding Author: Shaolin HU, Henan University; Guangdong University of Petrochemical Technology; E-mail: hfkth@126.com

image is extracted to be used as the pre-processed image, reconstructed the image, and detected the cracked area using a binarization algorithm based on the crack information. A. Landstorm et al. implemented automatic online crack detection using morphological processing and logistic regression statistical classification on 3D profile data of steel plate surfaces [2]. Since the above method first processes the image to highlight the target features and then manually extracts the features for detection, the detection effect is not satisfactory in the actual environment due to the irregularity of the crack structure. Chen et al. designed a naive Bayes fully convolutional network (NB-FCN) for crack detection on the metal surface of underwater components [3]. The network finds cracks by fusing multiple video frames that meet the temporal and spatial correlation, and achieved good detection results on the self-built metal parts dataset. In paper [4], Tiejun He et al. proposed to introduce a Space-to-depth layer based on YOLOv5 in order to adapt to the detection task with small ground resolution and pavement disease target. The above method automatically extracts crack features through the convolutional neural network, which improves the detection accuracy, but the convolutional neural network only focuses on local information and cannot establish long-distance connections of the global image.

To solve these problems, this paper proposes a detection method based on morphology and improved YOLOv5. Firstly, this paper performs morphological processing on the data set to enhance the crack features, then uses the MobileViTv3 module to replace the CBS module in the YOLOv5 network and uses SIoU as the bounding box loss function of the model. And finally, the modified model is trained using the morphologically processed dataset.

2 Methodology Improvement

2.1 Morphological Processing

Morphology is one of the most widely used techniques in image processing, mainly used to extract shape features from images that downstream tasks can learn. The most basic method of morphology is corrosion operation, and the operation process of this method is shown in Fig.1. It can be seen from the corrosion process that corrosion can reduce the brightness of the image and increase the area of the darker area in the image [5]. Because the crack is darker than the context in the image, the area of the crack will be enlarged after the image is corroded, the characteristics of the crack will be enhanced, and the learning ability of the model will be improved. The comparison before and after crack corrosion is shown in Fig.2.

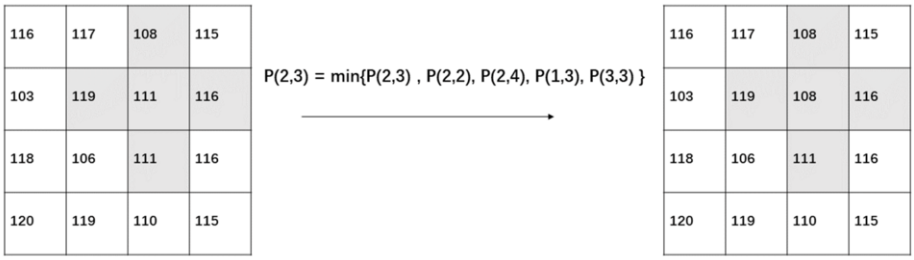

Fig.1 Schematic diagram of corrosion process

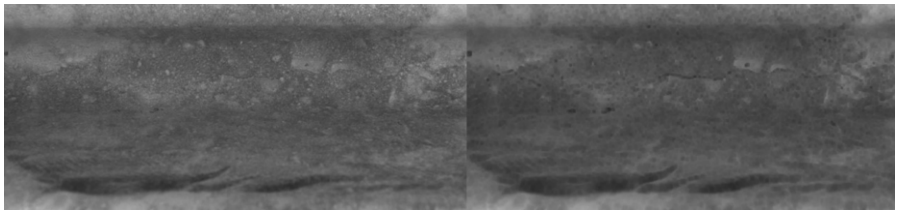

Fig.2 Comparison diagram before and after corrosion

2.2 Model Improvement

In 2020, the Glenn Jocher team proposed a detection model YOLOv5 with high detection accuracy and fast speed. As shown in Fig.3, the model consists of three parts: the backbone network, the neck network and the output, which are mainly composed of CBS, C3, SPPF and conv2d modules. The backbone network extracts rich information features from the input image; the neck network uses FPN The feature fusion method recombines the image features; the detection head predicts the bounding box and category according to the transmitted image features, eliminates redundant prediction boxes through NMS, and finally outputs the predicted category and frame coordinates with the highest confidence [6]. There are four versions of the model: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. YOLOv5m, YOLOv5l, and YOLOv5x are all derived from the network deepening by controlling the scaling factor on the basis of YOLOv5s. Although the larger the model is, the better the model detection effect is, the more resources it consumes, and the more difficult it is to deploy in real industrial scenarios, so YOLOv5s is selected as the base model for workpiece surface cracks in this paper.

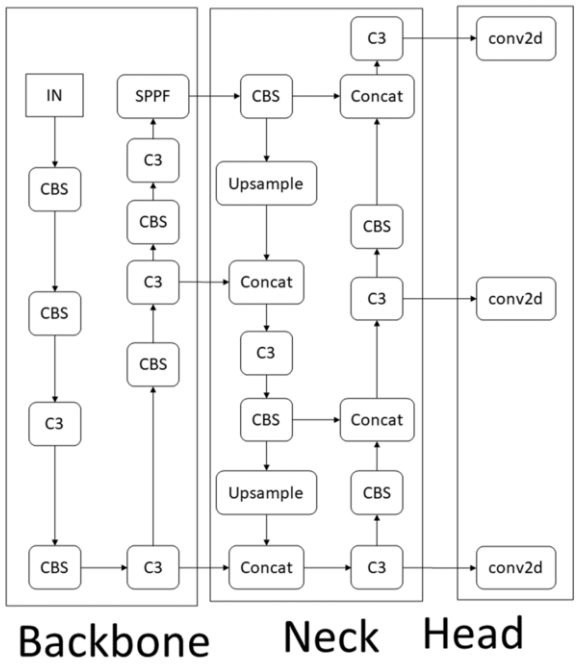

Fig.3 YOLOv5 structure diagram

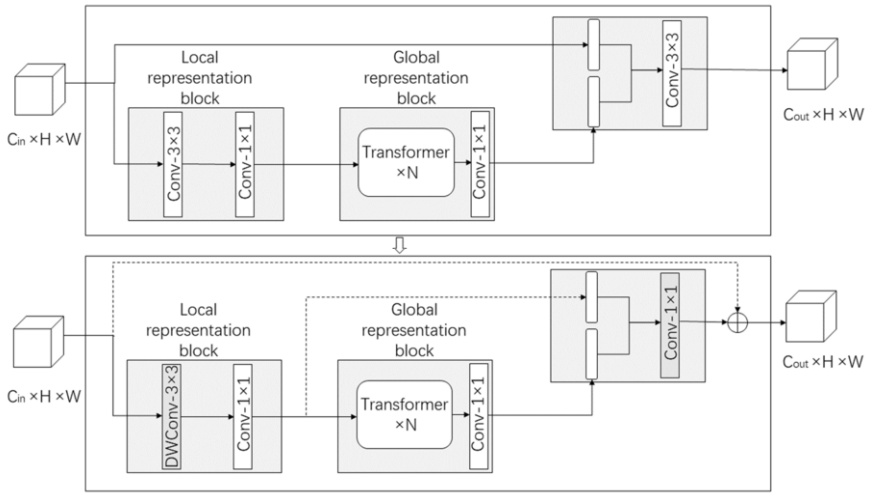

Fig.4 MobileViTv1&v3 structure diagram

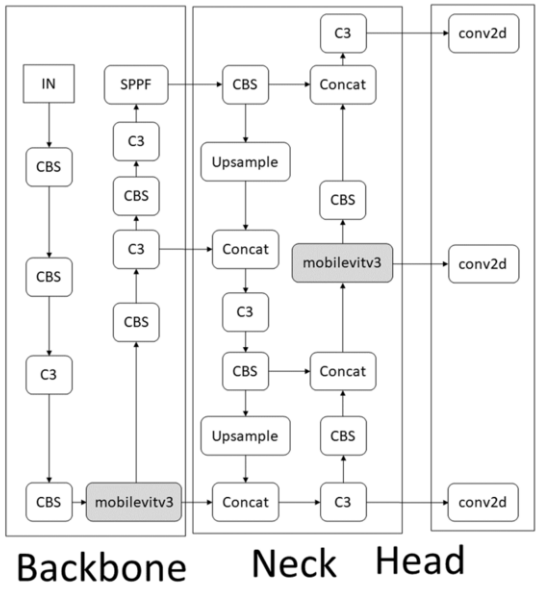

Fig.5 Modified YOLOv5 structure diagram

Although the convolutional neural network is suitable for computer vision tasks, the convolution operation can only capture local information, and cannot capture the long-range context information inside and between targets; ViT can realize input adaptation and long-range dependence through self-attention, is aimed at the extraction of a global understanding of visual scenes, of high-order spatial interactions, and of competing modelling capabilities. The role of ViT in target detection is to expand the receptive field of the image through global relationship modeling, obtain more contextual information, calculate the possible location of the target, and extract the feature information of the detected target [7]. Although ViT is very powerful, it has the shortcomings of too many model parameters and high computing power requirements. In addition, due to the lack

of spatial bias items in ViT, ViT training and task migration are more difficult. In order to solve the above problems, Apple published a hybrid architecture model of CNN and ViT in 2021, which combines the lightweight and high efficiency of CNN and the global information capture of ViT, and the network training process is faster and more stable [8]. As shown in Fig.4, MobileViTv3 is obtained by simply modifying the fusion block on the basis of the MobileViTv1 module, which solves the scaling problem of the MobileViTv1 module and simplifies the learning task. Therefore, this paper chooses the MobileViTv3 module to replace some modules of YOLOv5 to improve the detection effect of the model. In this paper, considering computing resources and model complexity, two residual feature modules C3 are selected to replace them with MobileViTv3. The positions of these two modules are the second C3 module of the backbone network and the C3 module connected to the medium-sized detection head. The network structure diagram after the replacement is illustrated in Fig.5.

The detection performance of target detection depends on the design of the loss function. The detection model locates the target through bounding box regression. Bounding box regression refers to using a rectangular bounding box to predict the position of the target object in the image, and then continuously refines the prediction through the bounding box loss function. The position of the bounding box, and finally accurately localize the target position. Therefore, a good definition of the bounding box loss function will bring performance improvements to the object detection model. The existing bounding box loss function of YOLOv5 is CIoU, which contains the IoU loss of the area of the overlapping area of the predicted and real boxes, the normalized distance loss between the centroids of the predicted and real boxes, and the loss of the aspect ratio between the predicted and real boxes, as far as possible to ensure that the width and height aspect ratio of the predicted box and the real box are closer [9]. However, CIoU has the disadvantage that w and h cannot be raised or lowered at the same time during the regression process of the predicted box, and it cannot reflect the relative direction of the predicted box and the real box. In addition, when the width and height of the predicted box are the same as the ratio of the real box, the relative penalty item of the ratio will not work and cannot reflect the real width and height of the real box. To solve this problem, SIoU introduces the vector angle between the real box and the predicted box as the angle loss of the model, and uses the width and height of the predicted box and the real box as the shape loss of the model [10]. Therefore, this paper uses SIoU as the bounding box loss function of the model.

3 Experimental Results and Analysis

In this paper, we first select workpiece surface cracks for labeling to create a dataset, train the model and conduct various comparative ablation experiments. Then the corrosion operation is performed on the dataset, and the corroded dataset is trained on the modified model and compared with the model trained using the original dataset. In this experiment, mAP@0.5 was selected as the detection index of detection accuracy. mAP@0.5 is the average precision of all pictures in each category when the IoU is 0.5. The higher the value of mAP@0.5, the better the detection effect of the model on the given data set. In this experiment, mAP@0.5 was selected as the detection index of detection accuracy. mAP@0.5 is the average precision of all pictures in each category when the IoU is 0.5. The higher the value of mAP@0.5, the better the detection effect of the model on the given data set. In order to show the complexity of the model, this

experiment selects parameters and GFLOPs as detection indicators, where parameters are the amount of model parameters, and GFLOPs is an indicator to measure the complexity of the model. In order to prove the rationality of choosing YOLOv5s in this article, this article did a comparative experiment between YOLOv5s and Yolov7-tiny in a given data set. In order to show the improvement effect of this article more intuitively, this article conducted an ablation experiment on the basis of the comparative experiment. In this paper, Yolov7-tiny and YOLOv5s are denoted as v7-tiny and 5s respectively, the model using SIoU as the bounding box loss function is recorded as 5s-SIoU, and the model after MobileViTv3 replaces C3 is recorded as 5s-MobileViTv3, and the modified model trained with the original dataset is denoted as ours-original, and the modified model trained with the corrupted dataset is denoted as ours-new. The experimental results are shown in Table 1.

Table 1: Experimental Results

model	mAP@0.5	parameters	GFLOPs
v7-tiny	53.4%	6007596	13.0
5s	59.5%	7012822	15.8
5s-SIoU	60.2%	7012822	15.8
5s-MobileViTv3	72.8%	9856854	30.7
Our-original	73.1%	9856854	30.7
Ours-new	74.6%	9856854	30.7

It can be seen from the experimental results in Table 1 that the detection accuracy of YOLOv5s is higher than that of Yolov7-tiny, so this paper selects YOLOv5s as the basic model for improvement. According to the experimental results of 5s, 5s-SIoU and 5s-MobileViTv3, using SIoU as the bounding box loss function of the model improves the detection accuracy of the model without increasing the complexity of the model, but the improvement effect is limited. The MobileViTv3 module can significantly enhance the detection effect of the model. After replacing the C3 module, the detection accuracy of the model has increased by 13.3% compared with YOLOv5s. Although the number of model parameters has increased by 2844032, it still meets the requirements of real-world applications. The last two sets of experiments show that the erosion operation can improve the detection accuracy of the model. By comparing the final method with the experimental results of YOLOv5s, the detection accuracy of the method proposed in this paper is 15.1% higher than that of YOLOv5s, which proves the effectiveness of the method proposed in this paper.

In this paper, two sets of comparative experiments are done to prove the feasibility of the improved model in actual scene detection. we first use the original image to conduct an example experiment on the above model to check the actual detection effect of the model. The results are shown in Fig.6. It can be seen from Fig.6 that v7-tiny detects the cracks in the picture, and the model after the replacement of the MobileViTv3 module is much better than the model before the replacement of the MobileViTv3 module. The experimental results in the last line show that the detection performance of the final model is the best, which also proves the effectiveness of adopting SIoU as the bounding box function. In order to verify the influence of corrosion on the test results, we first conduct example experiments on the improved model before and after to verify the usability of the improved model in real scenes; then we train the same model with the dataset before and after corrosion to get 2 kinds of training results, finally we corrode the same image and detect the image before and after corrosion with 2 kinds of training results to get 4 kinds of detection results, and the results are shown in Fig.7. The (A) shows the results of the model trained on the original dataset to detect the original image,

the (B) shows the results of the model trained on the original dataset to detect the corrupted image, the (C) shows the results of the model trained on the corrupted dataset to detect the corrupted image, and the (D) shows the results of the model trained on the corrupted dataset to detect the original image. The comparison between (A) and (C) proves that the confidence level of the detection box becomes higher after corrosion, but there may be a case of missed detection. The detection results in (C) and (D) prove that the detection of the original image using the model trained with the corroded dataset is not satisfactory, and thus the detection target must first be corroded before being applied to a real-world scenario. The above experiments demonstrate the feasibility of the proposed method applied in practical scenarios.

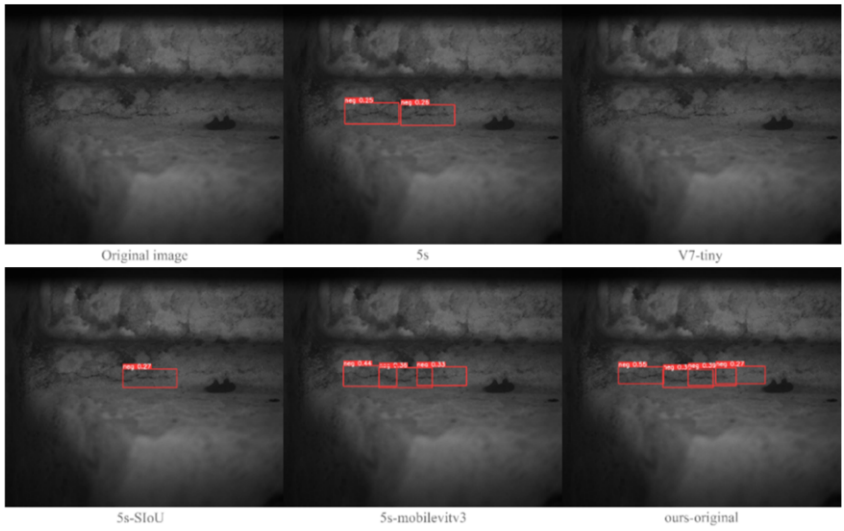

Fig.6 Crack detection results

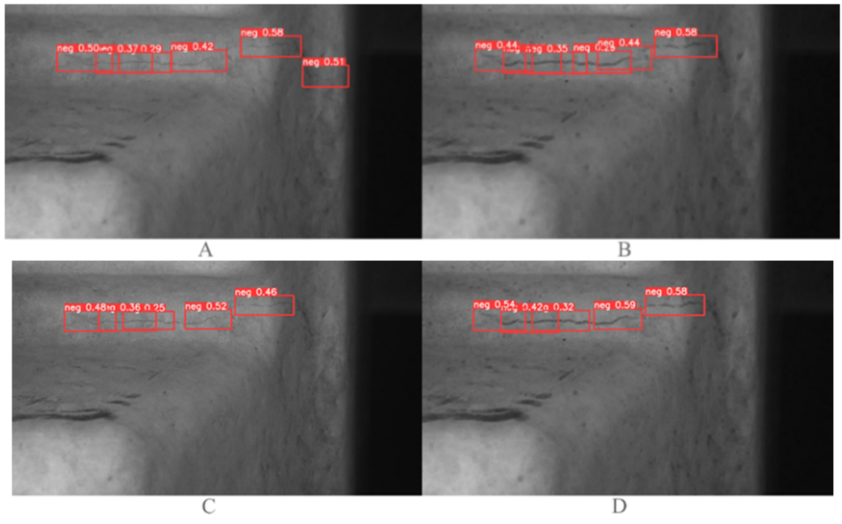

Fig.7 Comparison of test results

4 Conclusion and Outlook

This paper proposes a new detection algorithm based on morphology and YOLOv5 for the surface crack detection of workpieces. The crack features are enhanced through corrosion, such that the global information can be better captured in combination with YOLOv5 and MobileViTv3. In this model, the SIoU is utilized as bounding box regression function to improve the positioning accuracy of the prediction box. Therefore, the detection effect of the model can be enhanced, providing a new idea for industrial surface crack detection. Experiments have proved that the method proposed in this paper has improved by 13.6% on the original data set and 15.1% on the corroded data set compared with the original model.

Although the method proposed in this paper is significantly improved compared with the original method, the model becomes more complex, and the model can be light weighted later by using lightweight modules or knowledge distillation. In addition, due to the quantity and quality of the dataset images, the detection accuracy of the model has more room for improvement, and the detection accuracy of the model can be further improved by enhancing the sample and quality of the dataset.

Through experiments, this paper found that although the model trained with the corrosion dataset can improve the detection accuracy, but still needs to detect the target also corrosion operation to achieve better detection results, the application in the real scene is limited, and the future can try to improve the generalization ability of the model through the method of mixed dataset. Corrosion, the sole morphological technique employed in this study, is used to enhance the fracture characteristics and process the image. In the future, we can try to utilize other morphological processing techniques to improve the characteristics of the cracks in order to improve the model's detection ability.

Acknowledgement

This work is supported by National Natural Science Foundation of China (61973094) and National Natural Science Foundation of Guangdong Province.

References

[1] Qiang Li, Shuguang Zeng, Yanshan Xiao, Shaowei Zhang, Xiaolei Li. (2020), “Machine vision-based crack detection method for ceramic tile surface”, *Advances in Laser and Optoelectronics*, Vol.57, No.08, pp.43-49, doi: 10.3788/LOP57.081004

[2] A. Landstrom and M. J. Thurley. (2012), "Morphology-Based Crack Detection for Steel Slabs," in *IEEE Journal of Selected Topics in Signal Processing*, Vol.6, No.7, pp.866-875, doi: 10.1109/JSTSP.2012.2212416.

[3] F. -C. Chen and M. R. Jahanshahi. (2020), "NB-FCN: Real-Time Accurate Crack Detection in Inspection Videos Using Deep Fully Convolutional Network and Parametric Data Fusion," in *IEEE Transactions on Instrumentation and Measurement*, Vol.69, No.8, pp.5325-5334, doi: 10.1109/TIM.2019.2959292.

[4] Tiejun He,Huaen Li. (2023), “A pavement disease detection model based on improved YOLOv5”, *Journal of Civil Engineering*, pp.1-12, doi: 10.15951/j.tmgcxb.22101073.

[5] Qitong Xiao. (2022). “Research on Edge Detection Method of Remote Sensing Image Based on Adaptive Mathematical Morphology”, *Liaoning Technical University*, doi: 10.27210/d.cnki.glnju.2022.000629.

[6] Xiaowen Wang,Bo Liang, Fangfang Liu. (2023), “YOLOv5 Pedestrian Fall Detection Algorithm Based on Attention Mechanism and Weighted Box Function”, *Journal of Shanxi University (Natural Science Edition)*, pp.1-9, doi: 10.13451/j.sxu.ns.2022067.

[7] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). “End-to-End Object Detection with Transformers”, *European Conference on Computer Vision (ECCV)*, Vol.12346. Springer, doi: 10.1007/978-3-030-58452-8_13

[8] Sachin Mehta and Mohammad Rastegari. (2022), " MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer", Preprint at https://doi.org/10.48550/arXiv.2110.02178

[9] Zhaohui Zheng and Ping Wang and Wei Liu and Jinze Li and Rongguang Ye and Dongwei Ren. (2019), “Distance-IoU loss: Faster and better learning for bounding box regression”, Preprint at https://doi.org/10.48550/arXiv.2110.02178

[10] Zhora Gevorgyan. (2019), “SIoU Loss: More Powerful Learning for Bounding Box Regression”, Preprint at https://doi.org/10.48550/arXiv.2205.12740