Y. LIANG, J. WU

ISSN 0543-5846
METABK 64(1-2) 94-96 (2025)
UDC:669.13:620.197.6:62.01:338.82=111

SURFACE DEFECT DETECTION ALGORITHM OF HIGH TEMPERATURE CASTING SLAB BASED ON IMPROVED YOLOv5s

Received – Primljeno: 2024-06-06
Accepted – Prihvaćeno: 2024-08-10
Preliminary Note – Prethodno priopćenje

In order to improve the accuracy of surface defect detection of high temperature casting slab, an improved YOLOv5s surface defect detection algorithm is proposed. Firstly, Swin Transformer network structure is added to enhance the ability of feature extraction.Secondly, a coordinate attention mechanism is introduced to increase the sensitivity of position and direction information. Finally, a target detection layer is added to better realize feature fusion and enhance the generalization ability of the network. The improved algorithm has performed ablation experiments on the data set, which shows the effectiveness of the algorithm.

Keywords: casting slab, surface defect, detection, deep learning, YOLOv5s

INTRODUCTION

The surface defect detection of high temperature casting slab is of great significance for improving the production efficiency, product quality, and economic benefit of the iron and steel industry [1]. Due to the interference of dust and scrap iron in the casting slab production process and the complex background of the casting slab, high temperature radiation and scale have great interference on surface imaging, resulting in low defect identification, and the traditional machine vision method cannot meet the actual requirements in terms of accuracy and real-time performance [2]. With the development of artificial intelligence technology, surface inspection technology based on machine vision is widely used. By virtue of its advantages, the deep learning technology develops rapidly, and some achievements have been achieved in the surface defect detection of casting slab [3].

Based on the target detection algorithm YOLOv5s, an improved surface defect detection algorithm for high temperature casting slab is presented.

NETWORK STRUCTURE OF YOLOv5s

YOLOv5 is one of the classic versions of the deep learning model YOLO series of real-time object detection algorithms. It has the characteristics of high detection speed, high accuracy, easy to use, and strong expansibility. YOLOv5 has four different model variants: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These models have a significant impact on performance and resource dissipation; different trade-offs are provided to adapt to different application scenarios. In this paper, YOLOv5s, a model of the YOLOv5 series with obvious speed advantages, strong flexibility, and smallest volume, is selected.

The network structure of YOLOv5s consists of four modules: Input, Backbone, Neck, and Head. Backbone is composed of CBS, CSP1_X, and SPPF, and its function is to extract image features. The Neck consists of CSP2 module, upsampling module, and Concat to fuse the features extracted by Backbone. The Head contains three feature detectors with different scales to predict the target of the output feature map.

IMPROVED YOLOv5s ALGORITHM

For YOLOv5s algorithm, firstly, Swin Transformer network structure is added to enhance the ability of feature extraction. Secondly, coordinate attention (CA) is introduced to increase the attention of position and orientation information and reduce the interference of invalid information. Finally, a detection layer is added to better implement feature fusion and enhance the generalization ability of the network. The SIOU loss function proposed by Zhora Gevorgyan was used to evaluate the test results. The SIOU loss function improves the convergence speed and reasoning accuracy of the model in the training process by redefining the penalty index and considering the angle of regression vector.

Add a Swin Transformer fabric

Swin Transformer solves the huge difference between text words and high resolution image pixels by constructing hierarchical feature maps and sliding win-

Y. Liang, School of Applied Technology,University of Science and Technology Liaoning, Anshan, Liaoning China. E–mail: liangyan_00128@163.com
J. Wu,School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, Liaoning China.
Corresponding author: J. Wu. E-mail: wujieaa@163.com

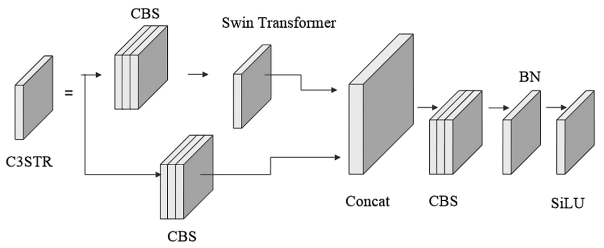

Figure 1 C3STR structure

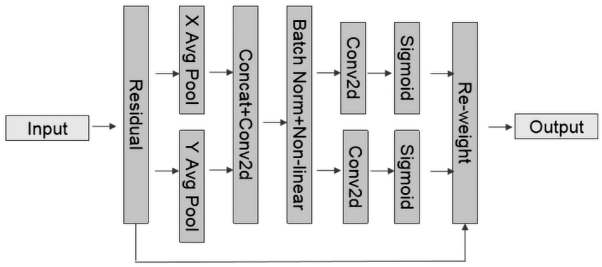

Figure 2 CA principle

dows, and achieves excellent results in both object classification and object detection tasks. In order to enhance the feature extraction capability, in the backbone network Backbone, integrate the window self-attention in Swin Transformer into the C3 (CSP+CBS) module to form the C3STR structure, as shown in Figure 1.

Swin Transformer Block consists of Window Multi-head Self-attention (W-MSA) and Shifted-window Multi-head Self-attention (SW-MSA) is composed of MultilayerPerceptron (MLP), and the connection between modules adopts a residual structure, W-MSA and SW-MSA are paired, and the multi-head self-attention mechanism is calculated as follows:

$$A(Q,K,V) = SM\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{1}$$

where Q is the query matrix. K is an index matrix. V is the value matrix. d is a dimension of Q or K. B is the relative position offset.

Compared with the traditional Transformer, the C3STR structure adopts the form of sliding window, which contains different pixels to realize the information transfer between adjacent windows, which reduces the calculation amount and improves the running efficiency of the network.

Introduce Coordinate Attention mechanism (CA)

The coordinate attention mechanism has the ability to capture cross-channel information and enhance the perception of position and direction information while simply merging into the core module of a lightweight network. In a large number of casting slab surface defect images, due to the complex image background, the difficulty of inspection is increased to a great extent, which can be suppressed by introducing CA. The principle is shown in Figure 2.

The CA module is mainly divided into two parts: the coordinate information embedding part and the coordinate attention generating part. Firstly, the output of channel c with height h can be expressed as:

$$z_c^h(h) = \frac{\sum_{0\le i<W} x_c(h,i)}{W} \tag{2}$$

Where $x_c(h,i)$ is the input of channel c in the horizontal direction. Output of channel c with width w can be represented as:

$$z_c^w(w) = \frac{\sum_{0\le j<H} x_c(j,w)}{H} \tag{3}$$

Where $x_c(j,w)$ is the output of channel c in the horizontal direction. In addition, the two kinds of changes aggregate features along two spatial directions respectively to obtain a pair of directional sensing feature maps. Secondly, using the 1×1 convolution transformation function F, make the transformation:

$$f = \delta\left(F\left[z^h, z^w\right]\right) \tag{4}$$

Where δ is the Sigmoid activation function. Finally, get the output of the CA:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{5}$$

Where $g_c^h(i)$ is the attention weight in the horizontal direction. $g_c^w(j)$ is the attention weight in the vertical direction. The feature map is enhanced with two weights to bolster its ability to represent features effectively.

Add detection layer

A group of CBS and C3 modules are added after the last C3 layer of the backbone network Backbone, and a group of up sampling and down sampling are added to the corresponding position of the feature fusion network Neck to better realize the feature fusion function and significantly enhance the generalization ability of the network.

EXPERIMENTAL RESULTS AND ANALYSIS

There are 4670 images in the data set of casting slab surface defects, including 6 categories of scar, split crack, slag skin, seam, scratch and cutting slab. The data set is divided randomly according to the ratio of 6:2:2 to obtain 2802 images in the training set, 934 images in the verification set and 934 images in the test set. The data set distribution is shown in Table 1.

In order to test the effectiveness of the improved YOLOv5s algorithm, the ablation test was conducted on the sample data set, and the performance of various defects and the overall network model was evaluated by using the mean average precision mAP, and the preci-

Table 1 **Sample data set**

Category	Training set	Verification set	Test set
Scar	381	126	126
split crack	639	218	218
Slag skin	394	131	131
Seam	392	128	128
Scratch	701	235	235
Cutting slab	295	96	96

Table 2 **Ablation experiment results**

C3STR	CA	Detection	mAP/%	P/%	R/%
–	–	–	72.8	69.8	71.2
+	–	–	73.6	73.2	61.3
–	+	–	73.4	73.5	71.6
–	–	+	73.3	71.7	71.6
+	+	+	76.8	73.1	73.5

sion P, recall rate R and other parameters were considered. The experimental results are shown in Table 2. Where, "+" indicates that the module is filled in the YOLOv5s network, and "–" indicates that the module is not filled.

Table 2 shows that for the improved YOLOv5s algorithm, when C3STR structure is introduced into the network, the mAP is increased by 0.8%.Adding CA, mAP increases by 0.6%; The detection layer is added, and the mAP is improved by 0.5%; When all improvements are added, the overall mAP improves by 4.0%.

CONCLUSION

An improved YOLOv5s algorithm is proposed to detect surface defects of high temperature casting slab. Swin Transformer network structure is added to enhance the capability of feature extraction. The coordinate attention mechanism CA is introduced to increase the sensitivity of position and direction information.The target detection layer is added to achieve better feature fusion, enhance the generalization ability of the network, and improve target detection accuracy. The improved algorithm performs ablation experiments on the data set. The mean average accuracy mAP is improved by 4.0%, which indicates the effectiveness of the algorithm.

REFERENCES

[1] Y. S.Weng, J. Q. Xiao, Y. Xia, Strip surface defect detection based on improved mask R-CNN algorithm. Computer engineering and applications.57(2021)19, 235-242.

[2] Y. Wu, Y. Z. Yang, X. L. Su, et al, Surface defect detection method of steel plate based on faster R-CNN. Journal of dongHua university (natural science) 47(2021)3, 84-89.

[3] J. F. Dai, H. Z. Qi, Y. W. Xiong, et al, Deformable convolutional networks.16th IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, 2017, 764-773.

Note: The responsible translators for English language is Y. Liang – University of Science and Technology Liaoning, China