Improved Yolov5 Algorithm for Surface Defect Detection of Strip Steel

Sun Zeqiang[1] and Chen Bingcai[1,2](✉)

[1] College of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054, China
cbc9@qq.com

[2] School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, Liaoning, China

Abstract. Aiming at the problems of low detection accuracy and slow detection speed in the traditional method of detecting strip steel surface defects, this paper proposes an improved yolov5 algorithm for detecting strip steel surface defects. Firstly, the data set of strip surface defects was constructed, and the K-means algorithm was used to cluster the defect samples, and the prior box parameters of different sizes were obtained. Secondly, the attention-yolov5 algorithm is proposed, which draws on the item-based Attention mechanism, adds channel Attention and spatial Attention mechanism to the feature extraction network, and uses the filtered weighted feature vector to replace the original feature vector for residual fusion. Finally, In order to improve the ability of defect feature extraction, the convolution layer is added after the main feature is extracted from different feature layers of the network output and after the pooling structure of spatial pyramid. The experimental results show that the mAP value of the improved yolov5 algorithm on the test set is as high as 87.3%, which is 5% higher than the original yolov5 algorithm. The average detection time of a single image is 0.0219s, which is basically the same as the original algorithm, and the detection performance is also better than the Faster RCNN and yolov3.

Keywords: Algorithm · yolov5 · Steel strip · Surface defect detection · Deep learning

CLC Number: TP391.4 Document code A

1 Introduction

Strip steel products are widely used in China's national economy and iron and steel industry important products, material performance, geometric size and surface quality is the main judging basis to measure the material efficiency. At present, the material and size of cold-rolled strip products can basically meet the requirements, but the surface quality problem is often the main influencing factor [1]. In recent years, the construction

Q. Liang et al. (Eds.): Artificial Intelligence in China, LNEE 854, pp. 448–456, 2022.
https://doi.org/10.1007/978-981-16-9423-3_56

of infrastructure is higher and higher demand for the accuracy of strip steel, strip steel is affected by the structure and production process led to a variety of surface defects, such as inclusion, patches, pitting, pressure into the scale and scratches, and so on, these defects light to reduce the service life of steel strip and uses, or may affect the construction of the building safety construction accidents. Therefore, this paper hopes to effectively detect the defects on the strip steel surface through the method of computer vision object detection, and then can separate the defective products.

Deep learning as an important technology of image recognition field, has a broad prospect of application and research on image recognition technology to promote the development of computer vision and artificial intelligence has important theory value and practical significance [2], the convolutional neural network and the introduction of deep learning, realized the high precision and high efficiency in the defect detection of advantage, It has gradually become the main research direction of defect detection. Currently, object detection algorithms are mainly divided into two categories. The first category is two-stage object detection algorithms represented by RCNN series, such as RCNN algorithm [3], Fast RCNN algorithm [4], Faster RCNN algorithm[5], Mask RCNN algorithm [6], The other is the one-stage object detection algorithm represented by yolo algorithm, such as OverFeat algorithm [7], yolo algorithm [8], yolov2 algorithm [9], yolov3 algorithm [10], and SSD algorithm [11]. The two-stage object detection algorithm requires the algorithm to generate the target candidate box, that is, the target position, first. Then do the classification and regression to the candidate box. The one-stage object detection algorithm only uses a convolutional neural network CNN to directly predict the categories and locations of different targets. The first kind of method is more accurate, but the speed is slow, but the second kind of algorithm is faster, but the accuracy is lower.

In recent years, many scholars have begun to devote themselves to defect detection. Li Bin et al. [12] used the improved yolo algorithm to detect the surface defects of aerospace engine parts, and the mAP value of the detection result was as high as 82.67%, and the average detection time of a single picture was 0.1240s.Chang Jiang et al. [13] solved the problem of insufficient samples in deep learning by using improved generation adversation network to generate more realistic strip defect images. In order to realize defect detection of bullet appearance, Ma Xiaoyun et al. [14] used k-means++ algorithm to generate sliding window anchor in Faster RCNN. Cheng Song et al. [15] realized image detection and recognition of welds through an improved yoloV4 model, and its mAP value was 88.52%, and the detection speed was also improved to 24.47fps.Weng Yushang et al. [16] improved the data of MaskRCNN in the detection of strip steel defects, which greatly improved the detection accuracy, but the detection speed was only 5.9 fps.

In this study, on the basis of deep learning, attention modules are added to the optimized yolov5 network to add weight to find more important feature channels, and experiments and tests are carried out to improve the accuracy of intelligent identification of strip surface defects under different signal-to-noise ratios.

2 Deep Convolutional Network Model and Attention Mechanism

2.1 Yolo Algorithm

Yolo algorithm regards the object detection framework as a spatial regression problem, and a single neural network can get the prediction of the boundary box and category probability from the complete image after a single operation. Its detection process is shown in Fig. 1. Firstly, the image size is adjusted, and then the image is sent into the convolutional network. Finally, the target detected by the network prediction is processed. The specific operation method is to divide the whole image into S × S grids implicitly, and the prediction can be made according to the grid in which the object center appears.

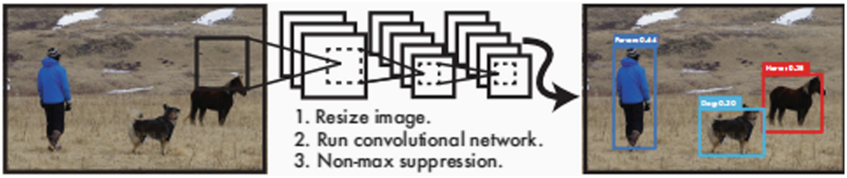

Fig. 1. Detection process of yolo algorithm

2.2 SE Attentional Mechanism

There are two main types of attention mechanisms. One is the channel attention, such as the SE module. The other is the space attention, such as the space converter network. And features can be tuned on a per-channel basis, so that the network can learn to selectively reinforce features that contain useful information and suppress unwanted features through global information. Spatial attention aims to understand more detailed features. Spatial converter network STN explicitly allows spatial manipulation of data and processing of data to enhance geometric invariance of the model. The attention mechanism can be intuitively interpreted using human visual mechanism. For example, our visual system tends to focus on parts of an image that help with judgment and ignore irrelevant information.

The basic structure of the SE Block used in this paper is shown in Fig. 2. The first step is the squeeze operation, which takes the global spatial features of each channel as the representation of the channel to form a channel descriptor. The second step is to learn the dependence on each channel and adjust the feature graph according to the dependence. The adjusted feature graph is the output of the SE Block.

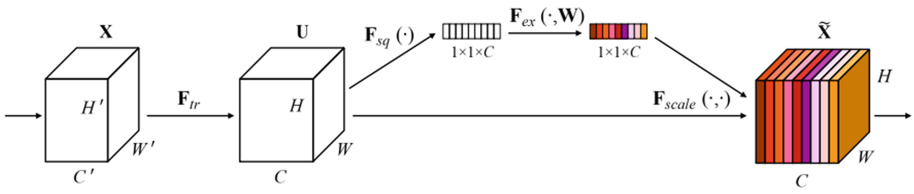

Fig. 2. A squeeze-and-excitation block.

2.3 Improved Network Architecture

The yolov5 algorithm borrowed yolov4's CSPDarkNet53 as the backbone feature extraction network and added some new features, such as FOCUS, CONV, LIBRARY CSP, SPP, etc., in which FOCUS: First, four input copies are copied and divided into four slices by slicing operation. Then, they are spliced by concat layer. Here, splicing refers to the combination of the number of channels to increase the number of features of the image, while the information under each feature remains unchanged. Then through the CBL layer, that is, through the convolution layer (CONV) first, extract the different features of the input, which is helpful to find specific local image features. Secondly, through the Batch Norm layer, the gradient distribution of each batch is controlled near the origin to achieve the normalization of the results, so that the deviation of each batch will not be too large. Finally, the leaky_relu activation function is used to enter the result to the next level of convolution. The Focus module is designed to increase speed by reducing the amount of computation and reducing the number of layers, not by increasing the mAP. SPP: space pyramid pooling consists of three parts: CONV, Maxpooling and CONCAT. First, CONV is used to extract the feature output, and then the maximum pooling layer of three different kernel_sizes is used to conduct the subsampling, and the respective output results are splicing and fused and added to their initial features. Finally, the output is restored to the same as the initial input through conv.

However, in order to solve the problem that the feature distinction between defect types and defect and component structure is not obvious, this paper uses the design idea of convolutional layer in the original yolov5 algorithm to add a layer of SE attention module between CSP2_1 and CBL, and the SE attention mechanism models the dependency relationship between channels. The characteristic response values of each channel can be adjusted adaptively, and the improved network structure is shown in the dotted box in Fig. 3. After the improved network structure processing, on the one hand, SE Block dynamically adjusts the characteristics of each channel according to the input, so as to enhance the expression ability of the network; on the other hand, it only increases a small amount of model complexity and calculation expense, so as to extract the features of defect targets more quickly.

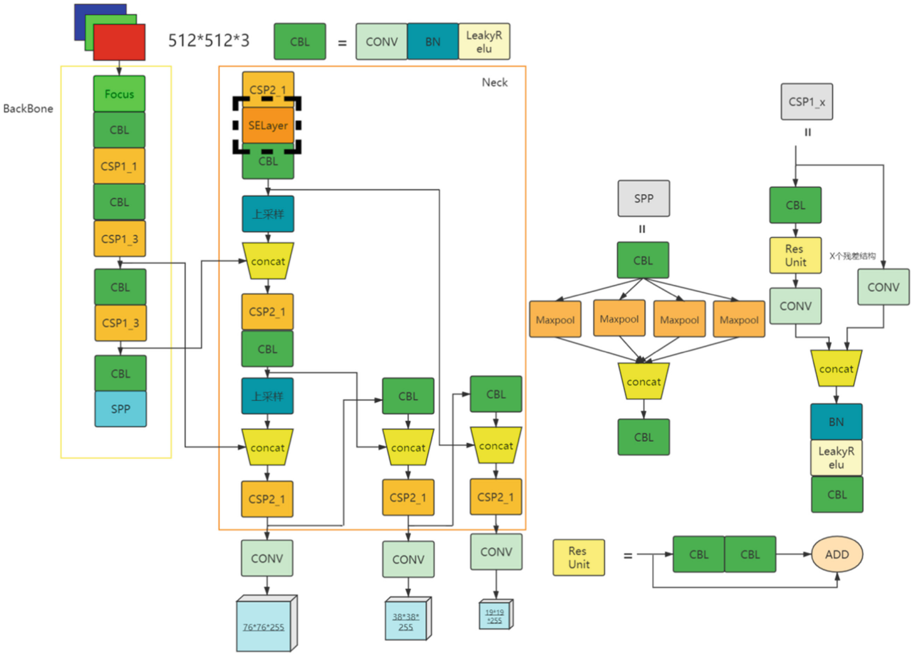

Fig. 3. Improved yolov5 network structure

3 Experimental Results and Analysis

3.1 Experimental Platform

The experiment carried out in this paper was completed under the Windows10 operating system. The hardware configuration used in the experiment is as follows: CPU: Intel(R) Core(TM) I3-10100F CPU@3.60GHz;Graphics processor (GPU): NVIDIA GeForce RTX 2080 TI.

3.2 Experimental Data

In this paper, NEU-DET dataset released by Northeastern University was used for experiments. The data set collected 300 pictures of surface defects of 6 types of strip steel. However, the crazing type defect pictures were not obvious, and the defect ambiguity was not good for the experiment, so the crazing type defect was deleted. In the end, the defect pictures used in the experiment in this paper include Inclusion, Patches, Pitted surfaces, Rolled in Scale and Scratches. This data set has a total of 1500 images. The distribution table of strip surface defect data set is shown in Table 1.The defect picture sample of data set is shown in Fig. 4.

Table 1. Distribution table of strip surface defect data set

Types of defects	In	Pa	Ps	Rs	Sc
Number of images	300	300	300	300	300

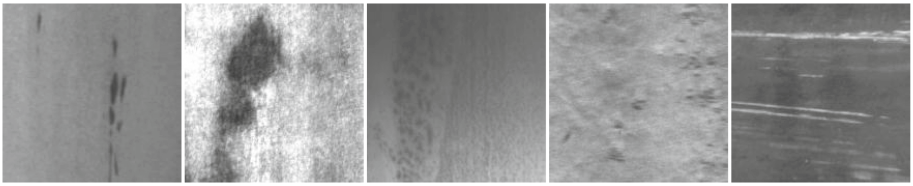

Fig. 4. A graphic example of a data set

3.3 Evaluation Criteria

MAP value was used as the evaluation index in the experiment. mAP is the mean of the average accuracy of all categories, and its calculation formula is

$$mAP = \frac{\sum AP}{N(\text{class})} \tag{1}$$

Where, N is the number of category detection, and AP can comprehensively consider the influence of accuracy and recall rate.PR curve can be obtained by taking accuracy as the vertical axis and recall rate as the horizontal axis. For continuous PR curve, AP in Eq. (1) is

$$AP = \int_0^1 P(R)\,\mathrm{d}R \tag{2}$$

Where P is the accuracy rate and R is the recall rate

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{\mathrm{TP}}{TP + \mathrm{FN}} \tag{4}$$

When training the model, NEU-DET data set is first transformed into COCO data set format. In the training, the transfer learning method is firstly adopted to load the 'yolov5x.pt' pre-weight file to train the NEU-DET data set, so as to generate the weight file of its own model. And set epochs to 100 and batch_size to 8.Set the learning rate as 0.01 and the attenuation coefficient as 0.2.

In this paper, the average detection Time (TIME) of a single image is used as the evaluation index of detection speed. The smaller Time is, the faster the detection speed is, and it is easier to achieve real-time detection. Equation (5) is the calculation formula of Time.

$$Time = \frac{TotalTime}{NumFigure} \tag{5}$$

Where, TotalTime is the TotalTime of detection, and NumFigure is the number of images detected.

3.4 Experiment Settings

In this experiment, SE attention mechanism was added to the original yolov5 model to compare it with the original yolov5 model. The experimental results are shown in Table 2.

Table 2. Comparative results of the improved yolov5 experiment

Model	Single defect recognition accuracy/%					mAP/%	Detection speed/fps
	In	Pa	Ps	Rs	Sc		
yolov5	84.3	90.9	87	48.1	99.5	82.0	**33.299**
yolov5_se	**87.4**	**93.9**	**94.5**	**60.9**	**99.5**	**87.3**	32.91

As can be seen from Table 2 above, after the addition of attention mechanism, the mAP value of yolov5_se generated was increased by 5%, and the detection speed was slightly decreased, but the detection accuracy of each defect was greatly improved.

For better comparison shows that after joining attention mechanism accuracy improvement, this experiment to further introduce other model for comparison are shown in Table 3 below, the one stage detection speed and phase detection speed significantly faster than the two stage detection speed of the algorithm, it also illustrates the stages of the algorithm is better real-time performance, however, It can also be found that the detection performance of yolo algorithm for the pressed oxide skin RS is poor. This is because the pressed oxide skin RS has some defects of small targets, so the accuracy is low. Even the detection accuracy of yolov3_se for RS is only 39.1%, which seriously lowers the average accuracy mAP value. As for the scratch SC defect, it is a big target and easy to detect. Moreover, the contrast between the scratch and the background is obvious, which further reduces the difficulty of detection. Therefore, the detection rate of yolov3_se, yolov5 and yolov5_se on this target reaches 99.5% (Fig. 5).

Table 3. Detection results of different algorithms

Model	Single defect recognition accuracy/%					mAP/%	Detection speed/fps
	In	Pa	Ps	Rs	Sc		
FasterRCNN	70.2	85.6	74.3	63.9	81.7	75.14	2.5
MaskRCNN	72.3	86.5	79.3	**77.8**	89.5	81.02	2.3
ResNetDNN	83.9	89.5	88.3	74.9	89.1	85.14	2.9
SSD	65.1	84.9	69.1	58.8	87.3	73.04	41
yolov3	60.5	83.1	73.2	60	86.1	75.28	50
yolov3_se	**93.1**	85.8	**99.5**	39.1	99.5	83.4	**53.47**
yolov5	84.3	90.9	87	48.1	99.5	82.0	33.299
yolov5_se	87.4	**93.9**	94.5	60.9	**99.5**	**87.3**	32.91

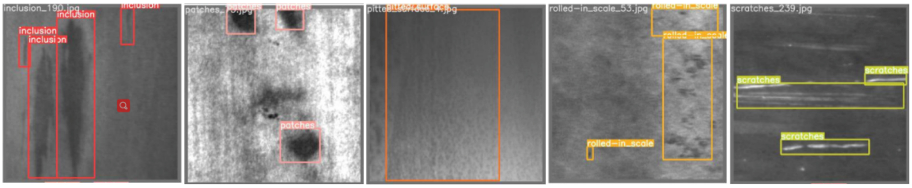

Fig. 5. Improved yolov5 detection results

4 Conclusion

In this paper, a strip surface defect detection method based on improved yolov5 algorithm is proposed. Firstly, k-means algorithm was used to cluster the surface defect data, and nine groups of different prior boxes were obtained to increase the size difference of prior boxes. Then, an SE attention layer is added after the CSP network to realize the channel attention mechanism, and the expression ability of the network is increased by adaptive adjustment of the characteristic response value of each channel. Experimental results show that compared with the original YOLOv5 algorithm, The improved YOLOv5 algorithm can effectively improve the detection accuracy of strip surface defects and realize the intelligent and efficient detection of strip surface defects under the condition that the detection speed is basically flat. Compared with FasterR-CNN and YOLOv3, the detection accuracy of the improved YOLOv5 algorithm is better than both of them, but the detection speed is slower than that of YOLOv3.The next step is to preprocess the original images and expand the data set through data enhancement to further optimize the accuracy of the network. At the same time, the network structure is optimized to improve the speed of defect detection and achieve a better real-time intelligent detection of strip surface defects.

References

1. Li, X.: Study on surface quality defects of cold rolling hood furnace annealing strip, **4**(05), 11–12 (2021)
2. Zheng, Y., Li, G., Li, Y.: A review on the application of deep learning in image recognition. Comput. Eng. Appl. **55**(12), 20–36 (2019)
3. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587. IEEE Computer Society Press, LosAlamitos (2014)
4. Girshick, R.: Fast RCNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448. IEEE Computer Society Press, Los Alamitos (2015)
5. Ren, S.Q., He, K.M., Girshick, R., et al.: Faster RCNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
6. He, K.M., Gkioxari, G., Dollár, P., et al.: Mask RCNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969. IEEE Computer Society Press, Los Alamitos (2017)
7. Sermanet, P., Eigen, D., Zhang, X., et al.: OverFeat: integrated recognition, localization and detection using convolutional neworks[OL]. https://arxiv.org/abs/1312.6229. 03 July 2019
8. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788. IEEE Computer Society Press, Los Alamitos (2016)
9. Redmon, J., Farhadi, A.: yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271. IEEE Computer Society Press, Los Alamitos (2017)
10. Redmon, J., Farhadi, A.: yolov3: an incremental improvement[OL]. https://arxiv.org/abs/1804.02767. 03 July 2019
11. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
12. Li, B., Wang, C., Wu, J., Liu, J., Tong, L., Guo, Z.: An Improved yoloV4 algorithm for surface defect detection of AEROENGINE parts [J/OL]. Laser Optoelectron. Prog. {3}, {4}, {5}, 1–17. http://kns.cnki.net/KCMS/detail/31.1690. 18 July 2021. TN. 20210105.1048.006. HTML
13. Chang, J., Guan, S., Shi, H., Hu, L., Ni, Y.: Defect classification of strip steel based on improved generative adversation network and mobilenetv3. Adv. Laser Optoelectron. 201, **58**(04), 221–226
14. Ma, X., Zhu, D., Jin, C., Tong, X.: Detection of bullet appearance defect based on improved RCNN. Laser Optoelectron. Prog. **56**(15), 117–124 (2019)
15. Cheng, S., Dai, J., Yang, H., Chen, Y.: Weld image detection and recognition based on improved yolov4 [J/OL]. Laser Optoelectron. {3}, {4} {5}, 1–12. http://kns.cnki.net/kcms/detail/31.1690. 18 July 2021 The tn. 20210712.1728.088. HTML
16. Weng, Y., Xiao, J., Xia, Y.: Improved mask RCNN algorithm for surface defect detection of strip steel [J/OL]. Comput. Eng. Use. http://kns.cnki.net/kcms/detail/11.2127.TP.20210420.0937.014.html