Research on Surface Defect Detection Algorithm of Strip Steel Based on Improved YOLOV3

Zhang Ning¹a, Zhiwei Mi¹b*
¹School of Mechanical and Electrical Engineering and Automation, Shanghai University, Shanghai, China
aemail: zhn123@shu.edu.cn,
*bcorresponding author: bemail: zwmi@mail.shu.edu.cn

Abstract. Aiming at the problems of slow detection speed and low detection accuracy in the surface defect detection of hot rolled strip steel, an improved YOLOV3 algorithm model was proposed in this paper. Firstly, the data sets is expanded by using the data enhancement algorithm to improve the robustness of the algorithm. Secondly, K-means++ with less randomness was used to perform clustering analysis on defect labels instead of K-means algorithm, and appropriate Anchor Boxes were selected as the initial candidate boxes for the improved YOLOV3 network. Finally, this study changed the detection network of YOLOV3, added a layer of prediction box and performed feature fusion to improve the detection ability of the network for small targets. The experimental results show that the improved YOLOV3 algorithm achieves 89.5% mean average precision on NEU-DET data set, which is 14.7% higher than the original YOLOV3 algorithm. Detection speed in 34 FPS at the same time, can meet the needs of industrial production.

1. Introduction

Strip steel is an indispensable basic material in automobile manufacturing, military industry and other fields. However, due to the influence of factors such as production technology and equipment of rolling, surface defects in the process of strip steel production will inevitably produce, such as pitting, inclusion, scratches [1]. The strip properties, such as corrosion resistance, fatigue resistance severely because of these surface defects will be reduced. Therefore, in order to improve the surface quality of strip steel, it is very important for researchers to detect its surface defects.

For the defect detection of strip steel, the traditional method is to manually extract the characteristic factors and then classify the defects, but this method is difficult to meet the requirements of detection accuracy, robustness and real time in the actual production. In recent years, due to the development of deep learning, convolutional neural network has gradually become the mainstream method in the surface detection tasks. Convolution neural network is now in general use steel strip surface defect detection feature extraction and classification. Among them, Literature [2] used SSD algorithm to establish the detection model of hot-rolled steel coil edge defects, which achieved good results. However, there were only 2 types of defects for testing, which only made some useful exploration. Literature [3] used Defect Detection Network(DDN) to achieve end-to-end surface Defect Detection of strip steel and achieved a 70%-80% accuracy. Although this method has a good Detection accuracy, it has a slight deficiency in Detection speed and is difficult to meet the requirements of real-time Detection in industrial production.
Deep learning target detection algorithms are mainly divided into two categories at present. One is two-stage detection algorithm, including Faster R-CNN\(^4\) and Mask R-CNN\(^5\). The other is the one-stage detection algorithm represented by YOLOV3\(^6\) and SSD\(^7\). As one of the most excellent detection algorithms at present, YOLOV3 algorithm achieves a good balance between detection accuracy and detection speed. However, due to the numerous defects in small areas on the surface of hot rolled strip steel, and the limited detection ability of YOLOV3 for small targets, this paper improved YOLOV3 algorithm on the basis of: First, data enhancement algorithm was used to expand the data set to improve the robustness of the algorithm; Secondly, K-means++ with less randomness was used to perform clustering analysis on defect labels instead of K-means algorithm, and appropriate Anchor Boxes were selected as the initial candidate boxes for the improved YOLOV3 network. At the same time, the network structure of YOLOV3 was adjusted to add a large-scale feature map and fuse it with the shallow features to improve the detection ability of the algorithm for small defect targets. Experimental results show that the mAP of the improved YOLOV3 algorithm on NEU-DET\(^8\) data sets reaches 89.5%, which is 14.7% higher than the original YOLOV3 algorithm. Detection speed in 34 FPS at the same time, can meet the needs of industrial production.

2. Yolov3 Fundamentals

2.1. Detection principle

Yolov3 is a new end-to-end target detection algorithm proposed by Redmon et al in 2018. The YOLOV3 network model consists of two parts, the feature extraction network and the prediction network, and its network structure is shown in Figure 1. The network structure of YOLOV3 is a full convolutional network without pooling layer, and its subsampling operation is completed by the convolutional layer with a stride of 2. It also uses the idea of jump connection in ResNet to increase the number of network layers by using the residual structure, forming a 53-layer feature extraction network, also known as Darknet-53. With reference to Feature Pyramid Networks, Yolov3 designed multi-scale prediction Networks \((13\times13, 26\times26, 53\times52)\) and fused these Feature maps. The accuracy of location and category prediction can be improved effectively by using three feature maps of different scales.

![Figure 1 Yolov3 network structure diagram](image)

2.2. Loss function

The task of target detection is to get the position information and category information of the target. The loss function of two-stage detection algorithm includes the classification of the target category and the regression of the target position. In this way, end-to-end detection cannot be carried out, and it will make the network difficult to train. The YOLOV3 algorithm solves the two problems of
classification and location through a loss function, which can achieve end-to-end detection and improve the detection speed of the algorithm. The loss function is shown in Equation (1):

\[
    \text{Loss} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
    + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \left[ (\omega_i - \hat{\omega}_i)^2 + (h_i - \hat{h}_i)^2 \right] \\
    + \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
    + \sum_{i=0}^{S^2} I_{i}^{\text{noobj}} \sum_{c=1}^{c_{\text{classes}}} \left[ p_i(c) - \hat{p}_i(c) \right]^2
\]

(1)

Where, S is the image partition coefficient, B is the number of prediction boxes in each grid, C is the total number of categories, p is the category probability, \(x_i, y_i, \omega_i\) and \(h_i\) are the horizontal, vertical coordinates, width and height of the center point of the prediction box in the ith grid, \(\lambda_{\text{coord}}\) is the weight coefficient, and \(\lambda_{\text{noobj}}\) is the penalty weight coefficient.

3. Improvement of YOLOV3 algorithm

3.1. Data sets enhancement

The datasets used in this paper is the Northeastern University Open Source Hot-Rolled Strip Surface Defects Datasets (NEU-DET). The data sets is the strip Surface defect inspection standard data sets, collected the Inclusion, Scratches, Rolled In Scale, Crazing, Pitted Surface and Patches 6 kinds of defects, 300 copies of each defect, image size is 200 x 200.

Sufficient training samples are needed for the deep model to learn the characteristics of defects. If the number of samples is insufficient, the phenomenon of overfitting will occur. In order to enhance the robustness of the system, this paper proposes a dynamic data expansion method to simulate the real situation, and expands the 1800 sample data in NEU-DET data sets by randomly adopting five ways: shifting, changing the brightness, adding noise, rotating angle, and clipping.

Figure 2  effects of different ways of data enhancement

Taking scratches as an example, figure 2 shows the image enhancement effects of the above five ways. Each image was enhanced twice, and the enhancement method was randomly selected. After data amplification, the total number of samples is 5400 images.
3.2. Network structure improvement

In the yolov3 network, the size of the input image is $416 \times 416$, and the resolution of the largest prediction box is $52 \times 52$. The step size of subsampling between these is 8, so theoretically the smallest target that the yolov3 network can detect is $8 \times 8$. However, there are many small target defects on the surface of hot rolled strip steel and some of them are smaller than $52 \times 52$, so it is easy to cause the missed detection of small targets. Therefore, this paper redesigned the target detection network, added a layer of large scale feature map ($104 \times 104$), and integrated it into the feature pyramid structure. In this way, the target detection network has 4 feature maps, which contain shallow high-resolution information and deep high-semantic information, and can significantly improve the detection effect of small target defects. The improved multi-scale network structure is shown in the figure 3.

![Figure 3 Improved yolov3 network structure diagram](image)

3.3. Dimensional clustering of anchor box

Yolov3 uses the K-means clustering algorithm to get the anchor boxes with width and height informations. K-means algorithm will randomly select $k$ points as the initial clustering centers, and this randomness will have a negative impact on the clustering effects. In this paper, k-means++ algorithm with less randomness is used to replace k-means algorithm. K-means++ optimizes the selection of $k$ initialization centers, which can effectively reduce the clustering result deviation caused by the randomness of k-means algorithm. Figure 4 shows the clustering effect of k-means algorithm and k-means++ algorithm on the strip data set. Since the improved network structure has 4 feature maps, 12 anchor values are finally determined and 3 prior boxes are maintained on each feature map.

![Figure 4 clustering results of anchor boxes](image)
4. Experimental results and analysis

The algorithm in this paper is running on pytorch, a deep learning framework. The experimental environment is configured as follows: Intel(R) Core(TM) i7-9700 CPU @3.00GHz processor, 16GB of RAM, NVIDIA GeForce GTX 1660 Super graphics card, CUDA version 10.0.130, CUDNN version 7.4.15, and Windows10.

Network parameters are configured as follows: adam optimization algorithm is adopted, momentum is 0.9, weight decay is 0.0005, iteration number is 300, initial value of learning rate is set as 0.01, which gradually drops to 0.0005 during training, and its curve is cosine function.

After data enhancement, 5,400 images were obtained. The ratio of training Validation set and test set was randomly selected to be 9:1, and that of training and Validation set was 9:1, namely 4,374 training sets, 486 Validation sets and 540 test sets. The image uses XML format to record the location and category information of all defects.

When evaluating the performance of the network model, Precision and Recall should be taken into account simultaneously. The calculation methods of Precision and Recall are respectively expressed in Equations (2) and (3):

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (2)

$$\text{Recall} = \frac{TP}{TP + FN}$$  \hspace{1cm} (3)

Where, TP represents the number of positive samples correctly identified as positive samples, FP represents the number of negative samples identified as positive samples, and FN represents the number of positive samples identified as negative samples.

The average value of precision under different recall is Average Precision (AP), which is an index to evaluate the detection accuracy of a certain class. The mean value of detection precision of all target classes is mean average precision (mAP). In this paper, mAP is used as the precision index of the network model, and FPS is used to evaluate the detection speed of the network model, that is, the number of images that the network model can detect per second.

Table 1 shows the test results after various improvements on the original algorithm. When the clustering algorithm is changed from K-means algorithm to K-means++ algorithm, the detection accuracy and detection speed have been improved to a certain extent. The improved YOLOV3 algorithm adjusts the network structure and adds a large scale detection layer, which greatly improves the detection accuracy. Its mAP improves by 7.8 percentage points compared with the original YOLOV3 algorithm. Finally, the improved YOLOV3 algorithm, which adopts K-means++ algorithm and adds a layer of feature map at the same time, achieves 89.5% Mean Average Precision on 5400 images, and the detection speed is 34fps, which can meet the requirements of fast and accurate detection of strip steel surface defects.

| Network Structure | Clustering algorithm | mAP(%) | FPS |
|-------------------|----------------------|--------|-----|
| Original YOLOV3   | K-means              | 74.8   | 35  |
| Original YOLOV3   | K-means++            | 78.4   | 37  |
| Improved YOLOV3   | K-means              | 82.6   | 32  |
| Improved YOLOV3   | K-means++            | 89.5   | 34  |

Faster R-CNN, SSD, YOLOV3 and the improved YOLOV3 network were trained separately, and the detection performance of each model was tested on the test set. The detection results of different algorithm models were shown in Table 2. According to the results, the detection speed of the improved YOLOV3 algorithm is basically the same as that of the original YOLOV3 algorithm, but the detection accuracy is improved by 14.7%, which indicates that the network has improved the detection ability of small targets. Faster R-CNN algorithm based on candidate box classification achieved a
mAP of 76.7%, slightly higher than the original YOLOV3 algorithm of 74.8%, but lower than the improved YOLOV3 algorithm. Moreover, the detection speed of Faster R-CNN is far lower than that of regression-based YOLO series algorithms, and its FPS value is only 16, unable to meet the real-time requirements in actual production. The detection accuracy and detection speed of SSD algorithm are worse than YOLOV3 algorithm. Therefore, the improved YOLOV3 algorithm in this paper achieves a good balance in the detection speed and detection accuracy, and the effect is better.

Table 2  Comparison of detection results of different algorithms

| Algorithm   | FPS | mAP(%) |
|-------------|-----|--------|
| Faster R-CNN| 16  | 76.7   |
| SSD         | 28  | 62.8   |
| YOLOv3      | 35  | 74.8   |
| Improved YOLOv3 | 34  | 89.5   |

5. Conclusions
In this paper, YOLOV3 algorithm is applied to the surface defect detection problem of hot rolled strip steel. In view of the problem that the strip data sets is less, data enhancement operation is used to expand the data sets and improve the generalization of the algorithm. Aiming at the problem that there are many small targets on the surface of strip steel with different defect sizes, K-means++ algorithm is used to optimize the parameters of anchor boxes and adjust the detection network structure. A large scale detection layer is added to improve the detection ability of small targets under the condition of ensuring real-time detection. The improved network model is 14.7% higher than the mAP of the original YOLOV3 algorithm, which not only speeds up the convergence rate, but also improves the detection accuracy. In the future research, we will consider embedding the network model into the actual production equipment of strip steel, and further improve the algorithm model according to specific scenes to improve the detection speed and accuracy.

References
[1] Ghorai S, Mukherjee A, Gangadaran M, et al. (2013) Automatic Defect Detection on Hot-Rolled FlatSteel Products. IEEE Transactions on Instrumentation and Measurement, 62(3): 612–621.
[2] Zhang Z Y, Chen X B. (2019) Exploration on Recognition of Steel Coil Edge Defects Based on Deep Learning Model. Metallurgical Automation, 43(3):13.
[3] Song K C, Yan Y H. (2013) A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Applied Surface Science, 285 (part_PB):858–864.
[4] Ren S Q, He K M, Girshick R B, et al. (2017) Faster R- CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1137–1149.
[5] He K M, Gkioxari G,Dollar P, et al. (2017) Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. Venice. 2961 - 2969.
[6] Redmon J, Farhadi A. (2018) YOLOv3: An incremental improvement. Computer Vision and Pattern Recognition. arXiv Preprint, arXiv:1804. 02767.
[7] Liu W, Anguelov D, Erhan D, et al. (2016) SSD: single shot multibox detector.In: Proceedings of the IEEE European Conference on Computer Vision. Amsterdam. pp. 21 - 37.
[8] He Y, Song K C, Meng Q G, et al. (2019) An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, pp( 99):1-1.