Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 4, No. 1, 2023

Improved YOLOv5l-based Detection of Surface Defects in Hot Rolled Steel Strips

Zhiren Zhu *, Liming Zhou, Fankai Chen, Chen Liu, Fanrun Meng

College of Electrical Engineering, Tianjin University of Technology and Education, Tianjin, China
* **Corresponding author**: Zhiren Zhu (Email: 710203401@qq.com)

Abstract: To address the problems of complex background, different sizes and easy to miss and mis-detect in the detection of surface defects in hot-rolled strip, an improved YOLOv5l-based method for detecting surface defects in hot-rolled strip is proposed. Firstly, by adding the SimAM attention mechanism module to the aggregation network, the important information is focused with high weights to improve the recall rate of the original algorithm; secondly, by replacing all C3 modules in the YOLOv5l structure with C2F, a richer gradient of information flow is obtained to improve the accuracy rate of the original algorithm. The experimental results show that the average detection accuracy using the improved YOLOv5l improves by 5.3% and the accuracy rate by 8.3% compared to the original network, resulting in higher detection accuracy and lower error and miss detection rates, meeting the requirements of hot-rolled strip steel inspection in industrial manufacturing.

Keywords: Hot Rolled Strip; SimAM Attention Mechanism; C2F Module; Defect Detection.

1. Introduction

Hot-rolled strip [1] is an economical 'green steel' commonly used in industry. Hot-rolling is rolling above the recrystallisation temperature and is widely used in mechanical applications because of its low energy consumption, low cost, good vibration resistance and high production efficiency. In actual manufacturing, the hot rolling process will produce a variety of different defects on the surface of the strip, the common ones are Crazing, Rolled-in Scale, Scratches, Inclusion, Patches, Pitted Surface, six kinds of surface defects These defects have a serious impact on the qualification rate of hot rolled products [2]. Traditional strip inspection methods include manual sampling, magnetic fluxleakage testing, eddy current testing and infrared detection [3]. Manual sampling method refers to the use of the naked eye to distinguish defects, the method not only wastes manpower but also has the problems of leakage and low accuracy; magnetic fluxleakage testing [4] uses magnetic sensors to detect surface defects, but cannot detect closed cracks and limit the types of defects; eddy current detection testing [5] uses the principle of electromagnetic induction to detect metal surface defects, the method requires professional analysis and judgement, customized solutions and high detection costs; infrared detection method [6] uses the surface temperature of defective materials to detect defects, but the detection sensitivity is related to thermal emissivity, affected by time, temperature, location and size, and cannot accurately distinguish the types of defects.

With the rapid development of computer vision and deep learning [7], target detection algorithms based on deep neural networks [8] are widely used in defect detection. At this stage, target detection algorithms [9] can be divided into two categories according to the existence of candidate regions: one is the two-stage target detection algorithm represented by RCNN [10], SPPNeT [11], Fast RCNN [12], Faster RCNN [13]; the other is the single-stage target detection algorithm [14] represented by SSD [15], YOLO series [16], RetinaNet [17].

At present, the development of deep learning-based surface defect detection technology for hot-rolled strip steel has advanced rapidly, and the method not only improves the detection accuracy and precision, but also saves a lot of labor costs, and is widely used in practical production in. An improved YOLOv3 algorithm model is proposed in the literature [18], using a weighted K-means clustering algorithm to improve the matching graph of the a priori frame and the feature layer, which improves the inspection accuracy of the algorithm. Wang Daolei et al [19] proposed an improved algorithm based on YOLOv4-tiny, which combined multi-scale detection and attention mechanism to improve lightweight target detection accuracy. For the problems of small size of strip steel surface defects, fuzzy features and easy to miss detection. Zhou Jinwei et al [20] proposed an improved algorithm based on YOLOv5 by designing a new feature extraction module and modifying the confidence loss function to improve the stability of the algorithm convergence. Liu Jinchuan et al [21] added a small target detection layer to address the problem of small target miss and error detection; and introduced the Transformer encoder block module and the Convolutional block attention model (CBAM) attention mechanism module to address the problems of image crossover and overlap, improving the detection capability of the algorithm in complex backgrounds. Pan Meng et al [22] introduced 1x1 convolutional side branches by reconstructing convolution to improve the feature extraction capability of the network; added an attention mechanism with channels to retain more spatial information; and switched to a weighted bidirectional feature pyramid network to improve small target detection. Wang Bo et al [23] enhanced the fusion of image information by combining the Transformer layer with the BiFPN network structure; replaced the convolutional layer in the backbone network with a lightweight network, RepVGG, to enhance the feature extraction capability of the backbone network; and added a prediction layer to improve the multi-scale target detection capability.

To address the problems of varying size and uneven distribution of strip surface defects, multiple types of defects and complex backgrounds, this paper proposes an improved YOLOv5l-based algorithm for detecting strip surface defects, which improves the detection accuracy of the model while satisfying high detection accuracy and basically unchanged number of parameters and computational complexity, adds the SimAM attention mechanism module at the Head end and replaces the C3 module with the C2F module to improve the detection performance of the algorithm and meet the needs of industrial deployment.

2. YOLOv5l Algorithm and Improvements

2.1. Introduction to the YOLOv5l Algorithm

The YOLO series of algorithms are deep learning-based regression methods that use only a single convolutional neural network (CNN) network to directly predict the class and location of different targets. YOLOv5 is an improvement on the network structure of YOLOv4, with four models of target detection networks, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, based on the network structure of YOLOv5l version 6.0 is divided into four modules: Input, Backbone, Neck and Head, which improves the detection accuracy and learning speed compared with YOLOv4 [24]. The specific structure of the network in version YOLOv5l 6.0 is shown in Figure 1.

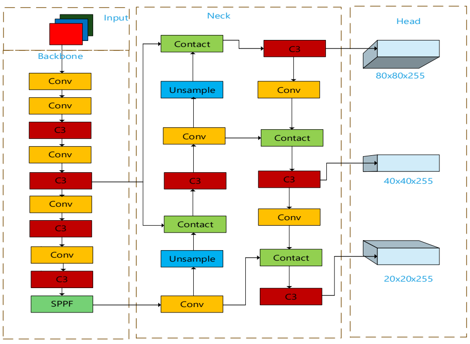

Figure 1. YOLOv5l structure diagram

2.1.1. Input

Compared with Yolov4, Yolov5l uses Mosaic data augmentation on the input side of its network structure to randomly scale, crop and arrange the dataset, thus improving the small target detection accuracy; secondly, it adds adaptive anchor frames to calculate the best anchor frame values for different training sets; finally, it uses adaptive image scaling to adaptively add the least black edges to the original image to reduce information redundancy, reduce computational effort and improve inference speed.

2.1.2. Backbone

YOLOv5l version 6.0 of Backbone is mainly divided into Conv module, CSPDarkNet53 and SPPF module. Among them, the Conv module replaces Focus in the old version to improve model efficiency while facilitating model export; CSPNet reduces computation, improves inference speed and obtains richer gradient combination information by segmenting the gradient flow, while ensuring no degradation in model detection and recognition accuracy; The SPFF module uses multiple small size pooling nucleus cascade instead of a single large size pooling nucleus in the SPP module, which improves the operation speed by merging the feature map of different receptive fields and enriching the expression ability of the feature map.

2.1.3. Neck

The Yolov4 and Yolov5l Neck modules both use an FPN+PAN structure to perform multi-scale feature fusion of strip surface defects by FPN and PAN (see Figure 2) to obtain feature maps at three different scales; FPN conveys top-down semantic feature enhancement and PAN conveys bottom-up localization features; the difference is that Yolov4 uses a normal convolution operation and Yolov5l uses a C3 module to enhance feature fusion.

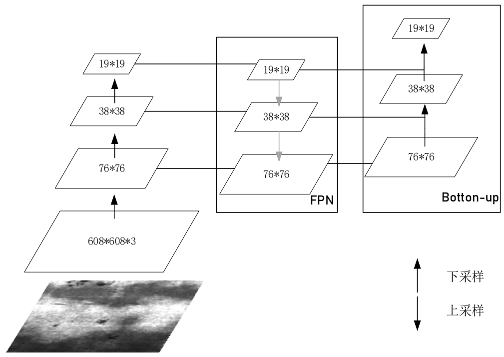

Figure 2. Structural diagram of FPN + PAN

2.1.4. Head

Yolov5l Head contains three detection layers, corresponding to the three different sizes of feature maps obtained in Neck, and three anchors with different aspect ratios are preset for the grid divided on each feature map to predict and regress targets; CIoU_Loss is used as the loss function of Bounding box, and for the screening of multi-target boxes, a weighted nms is used on the basis of DIoU_Loss to enhance the detection accuracy of occluded overlapping targets.

2.2. Improvements to the YOLOv5l Algorithm

2.2.1. Introduction of the SimAM Attention Mechanism Module

Adding an attention mechanism can effectively enhance the model's ability to extract features from images. The SimAM attention mechanism module [25] is a simple and very effective attention module for convolutional neural networks based on neuroscience theory. Unlike existing channel or null-field attention modules, this module derives 3D attention weights in the network layers without adding any parameters, and a schematic of 3D attention weight assignment is shown in Figure 3. The module aims to find important neurons by optimizing the energy function, using a linearly branchable metric between neurons. The energy function as defined by each neuron is:

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}(y_0 - \hat{x}_i)^2 \qquad (1)$$

among:

$$\begin{cases} \hat{t} = w_t t + b_t \\ \hat{x}_i = w_t x_i + b_t \end{cases} \qquad (2)$$

where t and xi are the target neuron and other neurons of the input feature tensor X, X ∈ RC × H × W,C, H and W are the number of channels, height and width of the feature tensor, respectively; i is the index of neurons on a channel; M is the number of all neurons on a channel, M = H × W; w_t and b_t are the weights and biases of the target neuron when transformed, respectively; all values in Eq.(1) are scalars, where y_t and y_0 are different values, and Eq.(1) is minimized when $\hat{t}$ =y_t and has $\hat{x}_i = y_0$ for all other neurons. The minimization formula is equivalent to finding the linear differentiability of the target neuron t and other neurons within the same channel. Using binary labels and adding regular terms, the final energy function is:

$$e_t(w_t,b_t,y,x_i) = [1-(w_t t+b_t)]^2 + \frac{1}{M-1}\sum_{i=1}^{M-1}[-1-(w_t x_i+b_t)]^2 + \lambda w_t^2 \tag{3}$$

Among:

$$\begin{cases} w_t = -\dfrac{2(t-\mu_t)}{(t-\mu_t)^2+2\sigma_t^2+2\lambda} \\ b_t = -\dfrac{1}{2}(t+\mu_t)w_t \end{cases} \tag{4}$$

$$\begin{cases} \mu_t = \dfrac{1}{M}\sum_{i=1}^{M} x_i \\ \sigma_t^2 = \dfrac{1}{M}\sum_{i=1}^{M}(x_i-\hat{\mu}_t)^2 \end{cases} \tag{5}$$

Where λ is the regularization factor; w_i is the weight of the ith neuron when transformed; from Eq.(4) it can be inferred that other neurons in the same channel satisfy the same distribution, so the mean and variance of all neurons can be calculated, replacing μ_t and σ_t^2 is the mean and variance of all neurons in the corresponding channel after removing neuron t, and all neurons on the same channel are multiplexed with this mean and variance, reducing the computational complexity of each location, the lower the energy, the greater the difference between neuron t and the surrounding neurons, and ultimately the minimum energy e_t^* at each location is calculated as follows:

$$e_t^* = \frac{4(\hat{\sigma}^2+\lambda)}{(t-\hat{\mu})^2+2\hat{\sigma}^2+2\lambda} \tag{6}$$

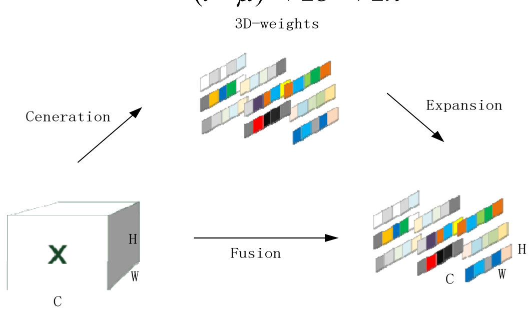

Figure 3. 3D Schematic representation of the attention weight allocation

2.2.2. Introduction of the C2F Module

A C3 module containing three standard convolutional layers (Conv+BN+SiLU) and n Bottleneck modules was designed in YOLOv5l with the help of the idea of CSPNet to extract the divergence and residual structure, which is the main module for learning on residual features, with two types of structure, one using multiple Bottleneck stacks and three standard convolutional layers; The other class uses only one basic convolution module, and the two classes are combined for concat operations. In this paper, the YOLOv5l model is improved by replacing the C3 module with the C2F module, so that the improved model can obtain richer gradient flow information and improve the detection accuracy of the model while ensuring its light weight. The C3 and C2F module structure pairs are shown in Figure 4.

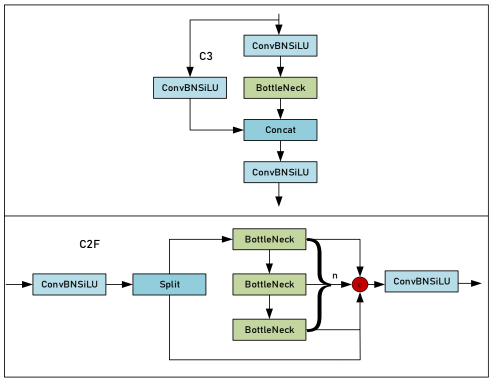

Figure 4. Structural comparison of C 3 and C2F models

3. Experimental Results and Analysis

3.1. Experimental Environment Setup

The hardware environment for the experiments is Windows 10, the CPU is Intel Core (TM) i7-9700K, the memory is 32GB, the GPU is NVIDIA GeForce RTX2080 Ti, and the software environment is Pytorch (1.10.0); CUDA (12.0); Numpy (1.24.3); Python 3.8; Pycharm 2022.1.

3.2. Dataset

This dataset was obtained from the public tape steel NEU-DET file of Northeastern University, which contains 6 different types of Crazing, Rolled-in Scale, Scratches, Inclusion, Patches and Pitted Surface defects, with an image size of 200x200 and a total of 1800 grey-scale images. After screening the data set in this paper, 1400 images were selected randomly according to the ratio of training set: validation set: test set 6:2:2. An example of the dataset is shown in Figure 5.

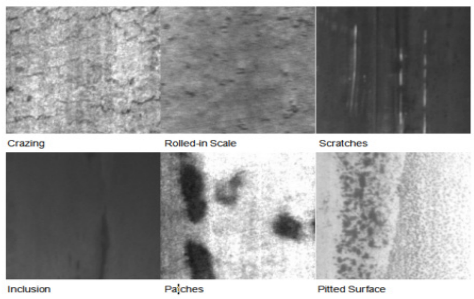

Figure 5. Dataset example

3.3. Parameter Setting and Evaluation Index

In this paper, we set epoch=200, batch size=4, conf_thres=0.5, initial learning rate is 0.01, learning rate momentum is 0.937, weight decay coefficient is 0.005, SGD algorithm is used for training, and precision rate P (Precision), recall rate R (Recall) and average precision value mAP (Mean Average Precision) as the model performance index. The formula is as follows.

$$P = \frac{TP}{TP+FP} \tag{7}$$

$$R = \frac{TP}{TP+FN} \tag{8}$$

$$\mathrm{m}AP = \frac{1}{C}\sum_{i=1}^{C} AP_i \tag{9}$$

Where TP (True Positive) indicates a positive sample with positive prediction; FP (False Positive) indicates a negative sample with positive prediction; FN (False Negative) indicates a positive sample with negative prediction; AP is the average precision of a single target category; mAP is the average precision value of AP for all categories; the P-R curve generated by the improved YOLOv5l is shown in Figure 6, where P is the vertical coordinate, R is the horizontal coordinate, and the area enclosed by the P-R curve and the coordinate axis is AP.

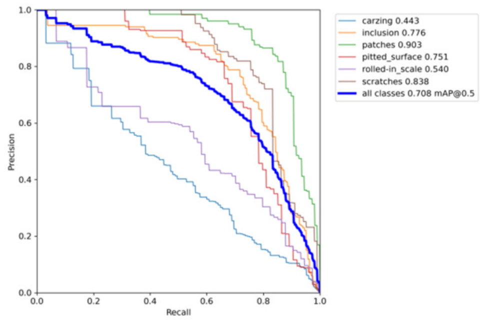

Figure 6. Analysis of the results of the improved P-R curves

3.4. Analysis of Experimental Results

In this paper, the improved YOLOv5l model is compared with the original YOLOv5l model, and the improved before and after models are trained on the same dataset for epoch times respectively, and the comprehensive evaluation results are shown in Table 1.

Table 1. Comprehensive assessment comparison

Network	P/%	R/%	mAP@0.5/%	mAP@0.95/%
YOLOv5l	62.6	64.2	65.7	30.5
YOLOv5l-SimAM	59.8	69.7	68.1	29.5
YOLOv5l-C2F	67.8	63.3	66.1	33.5
YOLOv5l-SimAM-C2F	70.9	64.3	70.9	36.4

The experimental results show that the detection accuracy is significantly improved after the introduction of C2F and SimAM attention mechanism in the original YOLOv5l network. This paper is based on the improved YOLOv5l model can detect strip surface defects more effectively compared to the original YOLOv5l, and the comparison graph of detection effect is shown in Figure 7.

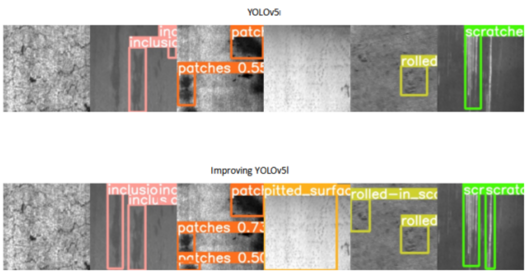

Figure 7. Comparison of the detection effects before and after the algorithm improvement

4. Summary

In the process of detecting defects on the surface of hot rolled strip steel, due to the problems of irregular target shape, different scales, complex background and easy false detection and omission, this paper designs a new detection algorithm based on the YOLOv5l algorithm. Firstly, the SimAM attention mechanism module is added to the Head side to improve the recall of the original algorithm without affecting the model parameters and computational complexity; secondly, all C3 modules in Backbone and Head are replaced with C2F modules in the original YOLOv5l model, on the basis of ensuring its lightweight, the improved model obtains more abundant gradient flow information and improves the identification accuracy of the model on the defects; Finally, the experimental results on the fused SimAM and C2F module on the NEU-DET dataset show that this model achieves 5.3% improvement in the accuracy of detecting strip surface defects compared to conventional neural networks.

References

[1] Li, Y., et al. "Progress in surface defect detection methods for strip steel." Journal of Iron and Steel Research . doi: 10. 13228/ j. boyuan. issn1001-0963.20220363.

[2] Wang Meng." A multi-scale feature map-based method for detecting defects in strip steel." Digital Technology and Applications 40.04(2022):36-39. doi:10.19695/j.cnki.cn12-1369. 2022.04.12.

[3] Zhang Yan,and Feng Feng." Exploration of strip steel surface defect detection technology." Information and Computer (Theoretical Edition) 33.11(2021):19-22.

[4] Pan Meng, Zhou Deqiang,and Chang Xiang." Characterization of surface defect detection by a novel pulsed leakage magnetic detection method." Sensors and Microsystems 36.12 (2017): 32-35. doi:10.13873/J.1000-9787(2017)12-0032-04.

[5] Wang B, et al. "A new eddy current detection method and its detection effect." Metallurgy of China 31.02(2021):50-54. doi:10.13228/j.boyuan.issn1006-9356.20200350.

[6] Ma, K., et al. "Study on SF_6 decomposition under pin-plate defects based on infrared detection method." High Voltage Electronics 48.12 (2012): 70-74. doi:10.13296/j.1001-1609. hva. 2012.12.015.

[7] Tian W. Research on Curriculum Design Method of Teaching Resource Library based on Deep Learning Technology [C]// Wuhan Zhicheng Times Cultural Proceedings of 6th International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2023). [publisher unknown], 2023:148-152. doi:10. 26914/ c. cnkihy. 2023. 010734.

[8] Zhang Zemiao, Huo Huan,and Zhao Fengyu." A review of target detection algorithms for deep convolutional neural networks." Small Microcomputer Systems 40.09(2019):1825-1831.

[9] Ning, J., et al. "A review of target detection algorithms for deep learning." Information Recorded Materials 23.10 (2022): 1-4. doi:10.16009/j.cnki.cn13-1295/tq.2022.10.057.

[10] Wen, Xinglin,and Bai, Tao." Design of audio emotion recognition and classification model based on improved multimodal RCNN." Modern Electronics 46.11(2023):114-118. doi:10.16652/j.issn.1004-373x.2023.11.021.

[11] Yang, H. Zhou,and Li, D. Li." YOLOv4 target detection based on improved SPPnet." Electronic Fabrication .22(2021):52-54. doi:10.16589/j.cnki.cn11-3571/tn.2021.22.018.

[12] Wang, Yanqing, et al. "Application of FastRCNN and CNN techniques in the morphological identification of three parasite eggs." Journal of Qiqihar Medical College 43.03(2022):234-237.

[13] Bai Chenshuai, et al. "An improved target detection algorithm based on Faster-RCNN (in English)." Journal of Measurement Science and Instrumentation .

[14] Wang, Dao-Lei, et al. "Improved SSD method for detection of hot spot defects in photovoltaic modules." Journal of Solar Energy 44. 04 (2023):420-425. doi: 10. 19912/ j.0254-0096. tynxb. 2021-1470.

[15] Zhou, Jinwei,and Wang, Jianping." A review of YOLO object detection algorithm research." Journal of Changzhou Institute of Technology 36.01(2023):18-23+88.

[16] Liu, J. Chuan, et al. "Improved RetinaNet for UAV small target detection." Science Technology and Engineering 23.01 (2023): 274-282.

[17] Zhang, Zhengchao." Improving YOLOv5 for lightweight strip steel surface defect detection." Computer Systems Applications . doi:10.15888/j.cnki.csa.009162.

[18] Li, W.G., et al. "Surface defect detection of strip steel based on improved YOLOv3 algorithm." Journal of Electronics 48.07 (2020):1284-1292.

[19] Zou, Wang,and Ji, Chang." An improved YOLOv4-tiny method for real-time detection of strip steel surface defects. " Mechanical Science and Technology . doi:10. 13433/ j. cnki. 1003-8728.20230034.

[20] Li, Shun,and Yang, Ying." Improved YOLOv5-based defect detection for hot-rolled strip steel." Computer Simulation .

[21] Wu Di, et al. "Improved YOLOv5-based surface defect detection in steel." Journal of Shaanxi University of Science and Technology 41.02(2023):162-169. doi: 10. 19481/j. cnki. Issn 2096-398x.2023.02.009.

[22] Zhang, Yang, Liu, Xiaofang,and Li, Wenwei." An improved YOLOv5n-based algorithm for strip steel surface defect detection." Journal of Sichuan University of Light and Chemical Technology (Natural Science Edition) 35.05 (2022): 60-67.

[23] Ma Yanting, et al. "Improved surface defect detection method for strip steel with YOLOv5 network." Journal of Electronic Measurement and Instrumentation 36.08(2022):150-157. doi:10.13382/j.jemi.B2205354.

[24] Dong, Yanhua,and Li, Jia'ao." Research on improving YOLOv5s remote sensing image recognition algorithm." Journal of Jilin Normal University (Natural Science Edition) 44.02 (2023):117-123. doi:10. 16862/ j. cnki. issn1674-3873. 2023. 02.017.

[25] Chuanjun Zhu, et al. "Hybrid defect detection model based on SimAM module and ResNet34 network." Modern Manufacturing Engineering .02(2023):1-9. doi:10.16731/ j. cnki. 1671-3133.2023.02.001.