Improving the Steel Surface Defect Detection Algorithm of YOLOv5 Network

Qian Zhang*

School of Software, Beijing University of Aeronautics and Astronautics, Beijing 100191

Abstract: ***Automated surface defect detection in industrial steel production is critical for ensuring product quality, but remains a challenging task due to the prevalence of small, low-contrast defects and the multi-scale nature of these anomalies. Current detection systems often struggle with insufficient accuracy, particularly for small targets, and a lack of robustness across varying defect sizes. To address these limitations, this paper proposes a series of targeted improvements to the YOLOv5s algorithm. First, at the input stage, the K-Means++ algorithm is employed to recluster the dataset and generate optimized initial anchor boxes, which provides a better prior for the model to learn from and improves localization, especially for small defects. Second, an attention mechanism is integrated into the backbone network to enhance feature representation. This module enables the model to focus computational resources on more informative spatial regions and channel features associated with defects, effectively suppressing irrelevant background noise and amplifying subtle defect signatures. Comprehensive experiments were conducted on a dedicated industrial steel defect dataset. The results demonstrate that the improved algorithm achieves a mean Average Precision (mAP@0.5) of 83.3%, representing a significant 6.2% increase over the baseline YOLOv5s model. Crucially, this performance gain is achieved without sacrificing inference speed; the enhanced model maintains a real-time detection rate of 96.5 frames per second (fps) on a standard GPU. These findings confirm that the proposed enhanced YOLOv5s algorithm successfully balances high precision with real-time processing capabilities, making it a viable and effective solution for automated visual inspection in demanding industrial environments such as steel manufacturing.***

Keywords: Surface Defect Detection, YOLOv5, Attention Mechanism, K-Means++, Industrial Steel, Small Target Detection, Real-time Inspection, Computer Vision.

1. INTRODUCTION

In the past, manual visual inspection was often used to detect surface defects in steel. With the development of machine vision, traditional image processing algorithms have gradually become mainstream, which extract defect features for detection. However, their speed is slow and their generalization and adaptability are weak.

With the application of deep learning, many models have emerged for steel surface defect detection based on deep learning to overcome the limitations of traditional image processing methods in feature extraction, providing more possibilities for enterprises. In the highly popular era of artificial intelligence, from Siri to Xiaodu, from Xiaobing to Xiaona, and then to Xiaoai Tongxue, these intelligent voice assistants are integrating into people's lives. The application fields of speech recognition technology are very wide, including smart homes, mobile devices, intelligent customer service, in car systems, intelligent healthcare, industrial control, intelligent toys, etc. Its core is to interact with machines through voice and enable them to complete related tasks. Zeng, Yuan, et al. (2025) investigated the relationship between education investment, social security, and household financial market participation, providing insights into how socioeconomic factors influence financial behaviors [1]. In the field of recommendation systems, Wang, Hao (2025) proposed a joint training approach for propensity and prediction models using targeted learning, addressing challenges posed by data missing not at random [2]. Ding, C., and Wu, C. (2024) conducted a systematic review on self-supervised learning for biomedical signal processing, focusing on ECG and PPG signals, highlighting its potential for improving healthcare analytics [3]. Similarly, Restrepo, D., et al. (2024) introduced a multimodal deep learning framework for low-resource healthcare settings, utilizing vector embedding alignment to enhance application scalability [4]. In data visualization and system development, Xie, Minhui, and Shujian Chen (2025) developed InVis, an interactive neural visualization system designed to facilitate human-centered data interpretation [5]. Zhu, Bingxin (2025) proposed RAID, an intelligent detection system aimed at improving reliability automation in large-scale advertising systems [6]. Zhang, Yuhan (2025) introduced InfraMLForge, a developer tooling framework for rapid large language model (LLM) development and scalable deployment, addressing efficiency challenges in AI development [7]. Hu, Xiao (2025) presented GenPlayAds, a generative model-based system for creating procedural playable 3D advertisements, offering innovative solutions for interactive marketing [8]. In healthcare research, Qin, Haoshen, et al. (2025) focused on optimizing deep learning models to combat amyotrophic lateral sclerosis (ALS) disease progression, contributing to advancements

in digital health interventions [9]. Wang, Yang, and Zhejun Zhao (2024) advanced abstract reasoning in artificial general intelligence by proposing a hybrid multi-component architecture, enhancing AI's cognitive capabilities [10]. Fu, Lei, et al. (2025) explored adversarial prompt optimization in LLMs, introducing HijackNet's approach to robustness and defense evasion, which has implications for improving AI security [11]. Lei, Fu, et al. (2025) developed a teacher-student framework for short-context classification, incorporating domain adaptation and data augmentation techniques to enhance model performance [12]. Finally, Zheng, Haoran, et al. (2025) introduced FinGPT-Agent, an advanced framework for multimodal research report generation, featuring task-adaptive optimization and hierarchical attention mechanisms, which improves the quality and efficiency of report generation [13].

2. YOLOV5 ALGORITHM STRUCTURE

YOLOv5 includes four different versions of models, which have the same structure but with adjustments in network depth and width. As shown in Figure 1, YOLOv5s consists of an input terminal, a backbone model, a neck network, and a prediction head. After image input, Mosaic data augmentation was used for data preprocessing to increase the sample size. Then, an adaptive algorithm is used to calculate the prior box center [6].

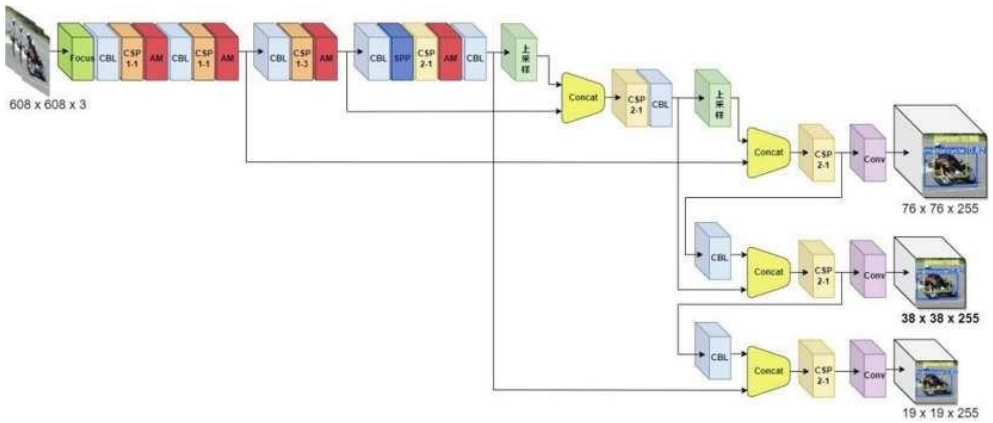

Figure 1: Improved YOLOv5s Network Structure

In the backbone structure, a feature pyramid structure is used to downsample and extract image features. This also includes convolution and pooling operations such as Focus structure and spatial pyramid pooling layer. The neck network performs feature fusion after Backbone feature extraction, and in addition to FPN, it also incorporates a bottom-up PANet structure [7] to transmit positional information. Perform full convolution on the prediction head and concatenate to output the detection result. Use GIoU to calculate the loss of the bounding box on the loss function [8].

3. IMPROVED DEFECT DETECTION ALGORITHM

3.1 Re clustering of prediction boxes

This article improves the prediction box calculation method at the input end of the YOLOv5s algorithm. The original YOLOv5 adaptive clustering method is affected by the initial clustering center and requires multiple regression calculations to converge, which increases the calculation time and reduces the detection accuracy. Therefore, this article uses K-Means++clustering prior box centers instead. The K-Means++ clustering algorithm uses weighted distance to better select initial centers and adapt to multi-scale targets. The clustering process is shown in Figure 2. Firstly, K is set to 9, which is consistent with the original algorithm. The predicted box centers obtained from regression are (25.1, 37.8), (35.6, 69.2), (90.4, 53.7), (44.1128.8), (31.8189.4), (170.1, 58.1), (102.2100.6), (91.8182.3), and (176.2182.5). By using the horizontal and vertical coordinates of different centers as the length and width, a prior box can be drawn.

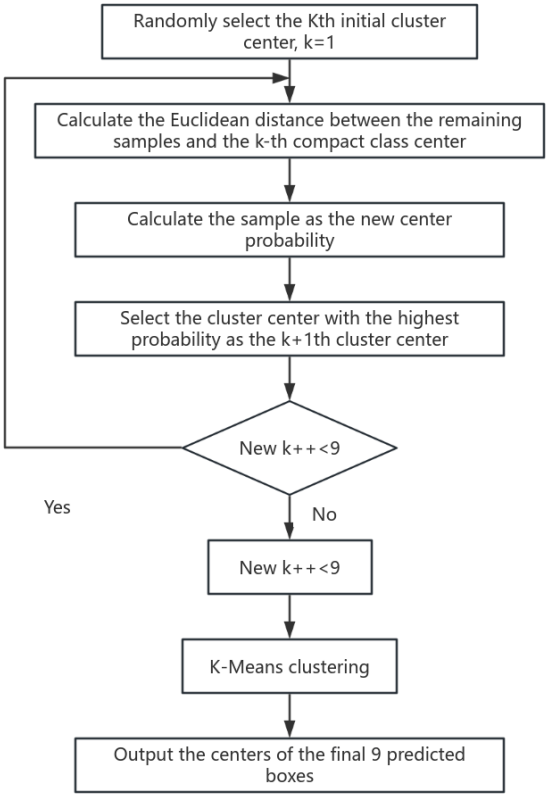

Figure 2: Clustering Process Diagram

3.2 SENet attention mechanism

The attention mechanism can help the network focus on key small target feature information, and this article uses the SENet module. SENet has established connections between features within channels, iteratively updating them based on the weights of different channels to filter out key information on the channels. Its network structure is shown in Figure 3. The input is first compressed and globally average pooled into vectors of C channels. Then, it enters two fully connected layer networks and is normalized using activation functions to obtain the weight coefficients of each channel. Finally, Scale is executed to weight the weights onto the original channels, obtaining the importance of different feature maps. Due to multiple full connections, the SENet module has a larger number of parameters, while CBMA is relatively lighter in weight. This article introduces SENet attention mechanism in the convolution operation of CSPi-X modules in the Backbone and Neck parts of YOLOv5s network.

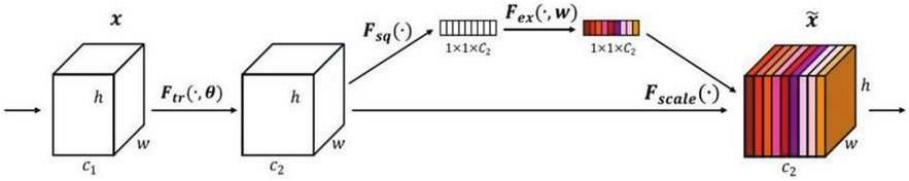

Figure 3: SENet module structure diagram

4. EXPERIMENT AND RESULT ANALYSIS

4.1 Dataset and Experimental Environment

In order to verify the effectiveness of the improved steel surface defect detection algorithm based on YOLOv5 in this article, the experiment used the NEU-DET steel surface defect dataset released by Northeastern University, which has 6 categories and was mixed with the dataset collected by the enterprise for expansion. Firstly, label the self collected dataset with LabelImg, and then name it according to NEU-DET naming conventions, with the suffix

'xml' to maintain consistency with the image names. This article randomly divides the mixed dataset into 8:2 parts, as shown in Table 1.

Table 1: Dataset Classification

Experimental dataset	Number
Total dataset	2100
Various datasets	350
Training set	1680
Test set	420

The experimental environment is Windows 10 operating system, the processor is Inter Core i7-9700k, NVIDIA Geforce GTX 1080 graphics card, and the learning framework is Pytorch version 1.9.

4.2 Evaluation indicators

There are many evaluation indicators for object detection, and this article uses recall rate three indicators, Precision and Mean Average Precision (mAP), are used to evaluate the detection accuracy of the algorithm. The calculation formula for each evaluation indicator is as follows:

$$mAP = \frac{\sum_{i=1}^{M-1} AP_i}{M} \tag{1}$$

In the formula, AP_i represents the average accuracy of each target class, and M represents the number of detection classes. We can use the P-R curve to simultaneously display both accuracy and recall metrics, with AP being:

$$AP = \int_0^1 P(r)dr, r \in (0,1) \tag{2}$$

But

$$R = \frac{TP}{TP+FN} \tag{3}$$

$$P = \frac{TP}{TP+FP} \tag{4}$$

In the formula: TP is the number of correctly detected samples, FP is the number of positive samples that were mistakenly detected, and FN is the number of positive samples that were missed. The recall rate R represents whether the model detection is complete, and the accuracy P evaluates whether the prediction is accurate.

4.3 Analysis of Experimental Results

This article conducts ablation experiments on a mixed dataset using the improved YOLOv5 algorithm. The experiment used YOLOv5 model pre trained on COCO and ImageNet. During training, some important parameters were set as follows: input image size was expanded to 640×640, iteration was 300 rounds, batch size was 64, optimization algorithm was selected as SGD, momentum factor was 0.937, initial learning rate was 0.01, and final learning rate was 0.0001. The cosine annealing method was used to adjust the dynamically adjusted learning rate.

The experimental results are shown in Table 2. This article conducts Experiment 1 to compare the YOLOv5s algorithm's original adaptive computation prediction box and K-Means++ clustering prediction box methods, verifying the effectiveness of the K-Means++ clustering algorithm in multi-scale detection. The comparison of detection performance under the same input conditions is shown in Figure 4. The results indicate that using K-Means++ clustering prediction boxes is more in line with the true size of defect targets and more suitable for multi-scale defect targets.

Original algorithm prediction box K-means++clustering prediction box

Figure 4: Comparison of detection effects of different anchor box calculations

Introducing SENet attention mechanism in the network for Experiment 2, it can be seen that the mAP is 2.1% higher than the original algorithm. It utilizes the channel relationship of features to focus on the location information of the target, which is more suitable for localization tasks. However, it also increases the number of parameters and reduces inference speed. This article combines two strategies to improve the YOLOv5s network for Experiment 3, with a 6.2% increase in mAP. The improved network outperforms the original YOLOv5s network in both P and R. The P-R curve comparison is shown in Figure 5, indicating that various defect APs have been improved after the algorithm improvement, with particularly cracking increasing to 42.0% and 39.7%.

Table 2: Comparison of Detection Accuracy for Different Improvement Strategies

YOLOv5s	K-Means++	SENet	R_P/%	R_P/%	mAP	Speed(fps/s)
√			51.9	79.1	77.1	101
√	√		55.3	79.5	79.2	125
√	√	√	57.9	85.2	83.3	96.5

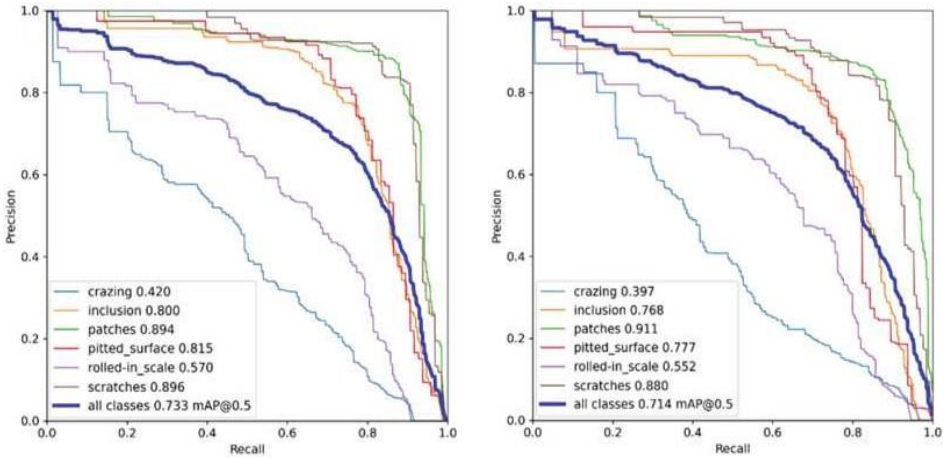

Figure 5: Performance Comparison of P-R Curve between Original YOLOv5 and Improved Algorithm

Through fusion experiments, we can see that the detection accuracy of the fused strategy is as high as 83.3%, thus obtaining the improved YOLOv5 algorithm with the best performance. The algorithm shows a slight decrease in detection speed due to the introduction of attention mechanism. It establishes channel connections between pixels, which increases computational complexity and requires a certain amount of time for learning. Figure 6 shows the performance of the improved algorithm in steel surface defect detection tasks. For different categories, the detected boxes are very close to the manually annotated boxes, indicating that the algorithm has high accuracy and achieves more precise actual detection results.

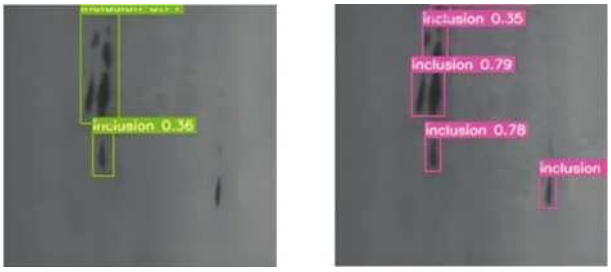

Figure 6: Comparison of detection performance between YOLOv5s and improved algorithm

5. CONCLUSION

The current surface defect detection of steel has poor feature extraction and detection accuracy due to the small target size. This article introduces the SENet module to increase the weight of small targets and improve detection accuracy. Meanwhile, due to the random assignment of initial cluster centers in the original K-Means algorithm, it is easy to converge to a local optimal solution. To solve this problem, the K-Means++ algorithm is used to re cluster the prior box centers, and weighted distances are used to find initial cluster centers that are farther apart from each other. The optimized prior boxes can fit various defects of different sizes, further suitable for surface defects in steel, and help improve detection accuracy. Finally, a self built dataset was used for steel surface defect detection, and the results showed that the improved YOLOv5s algorithm improved detection accuracy by 6.2% by integrating two improvement strategies without increasing too many parameters. However, adding attention mechanisms will increase the training time and storage space of the model. The improved defect detection algorithm has reduced the detection rate and FPS by 4.5%. Although sacrificing computing space for improved detection accuracy, it can still meet the practical needs of industry. This performance improvement makes the algorithm highly applicable in practical industrial production. In the future, more accurate detection can be achieved for targets with large intra class differences, and lightweight models can be further developed to facilitate network deployment on actual mobile devices.

REFERENCES

[1] Zeng, Yuan, et al. "Education investment, social security, and household financial market participation." Finance Research Letters 77 (2025): 107124.
[2] Wang, Hao. "Joint Training of Propensity Model and Prediction Model via Targeted Learning for Recommendation on Data Missing Not at Random." AAAI 2025 Workshop on Artificial Intelligence with Causal Techniques. 2025.
[3] Ding, C.; Wu, C. Self-Supervised Learning for Biomedical Signal Processing: A Systematic Review on ECG and PPG Signals. medRxiv 2024.
[4] D. Restrepo, C. Wu, S.A. Cajas, L.F. Nakayama, L.A. Celi, D.M. López. Multimodal deep learning for low-resource settings: A vector embedding alignment approach for healthcare applications. (2024), 10.1101/2024.06.03.24308401
[5] Xie, Minhui, and Shujian Chen. "InVis: Interactive Neural Visualization System for Human-Centered Data Interpretation." Authorea Preprints (2025).
[6] Zhu, Bingxin. "RAID: Reliability Automation through Intelligent Detection in Large-Scale Ad Systems." (2025).
[7] Zhang, Yuhan. "InfraMLForge: Developer Tooling for Rapid LLM Development and Scalable Deployment." (2025).
[8] Hu, Xiao. "GenPlayAds: Procedural Playable 3D Ad Creation via Generative Model." (2025).
[9] Qin, Haoshen, et al. "Optimizing deep learning models to combat amyotrophic lateral sclerosis (ALS) disease progression." Digital health 11 (2025): 20552076251349719.
[10] Wang, Yang, and Zhejun Zhao. "Advancing Abstract Reasoning in Artificial General Intelligence with a Hybrid Multi-Component Architecture." 2024 4th International Symposium on Artificial Intelligence and Intelligent Manufacturing (AIIM). IEEE, 2024.
[11] Fu, Lei, et al. "Adversarial Prompt Optimization in LLMs: HijackNet's Approach to Robustness and Defense Evasion." 2025 4th International Symposium on Computer Applications and Information Technology (ISCAIT). IEEE, 2025.
[12] Lei, Fu, et al. "Teacher-Student Framework for Short-Context Classification with Domain Adaptation and Data Augmentation." (2025).

[13] Zheng, Haoran, et al. "FinGPT-Agent: An Advanced Framework for Multimodal Research Report Generation with Task-Adaptive Optimization and Hierarchical Attention." (2025).

Author Profile

Qian Zhang (November 1996-), female, master's student, research direction is object detection.