Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 5, No. 01, 2026

Enhanced Deep Learning Framework for Steel Surface Defect Detection under Class Imbalance

Fatih Demir
Lakehead University, Thunder Bay, Canada
Fatih.Demir9923@lakeheadu.ca

Abstract: Steel surface defect detection is a critical task in intelligent manufacturing, where accuracy and real-time performance are essential for ensuring product quality. Although recent deep learning–based object detection methods have achieved promising results, challenges remain in handling complex defect morphologies, scale variations, and class imbalance. To address these issues, this paper proposes an enhanced steel surface defect detection framework based on a lightweight one-stage detection architecture. Specifically, Deformable Convolution v3 (DCNv3) is introduced into the detection head to improve the model's adaptability to geometric deformations and irregular defect patterns. In addition, Focal Loss is employed to alleviate the imbalance between foreground and background samples, enabling the model to focus more effectively on hard-to-detect defects. The proposed method is evaluated on the NEU steel surface defect dataset under a consistent experimental setup. Experimental results demonstrate that the proposed approach achieves notable improvements in detection accuracy while maintaining real-time performance, outperforming the baseline detection framework in terms of mAP_0.5. These results indicate that the proposed framework is effective and practical for industrial surface defect inspection tasks.

Keywords: Steel surface defect detection; deformable convolution; focal loss; object detection

1. Introduction

With the advent of Industry 4.0 era, as an indispensable raw material in the manufacturing industry, the importance of surface defect detection [1] has become increasingly prominent. However, the traditional manual detection method is limited by the experience, fatigue and subjective judgment of the detection personnel, and has some problems such as low efficiency and inconsistent detection results. The machine vision detection [2] method has the problems of high system cost, high requirements for image processing algorithms, and sensitivity to ambient light. It is difficult to meet the needs of modern production for efficient and accurate detection.

In recent years, the rapid development of deep learning technology has brought innovative methods for the detection of steel surface defects. In particular, YOLO (You Only Look Once) series algorithms have made great achievements in the field of object detection by virtue of their fast and accurate characteristics. Shi et al[3]Propose an improved YOLOv5 algorithm that incorporates the attention mechanism module CBAM to enhance the focus on key information, and introduces K-means clustering optimization to improve the detection accuracy of small targets and extreme aspect ratio defects. The improved YOLOv5+CBM algorithm can improve the average detection accuracy and is also suitable for automatic detection of steel surface defects. wen et al[4]proposed an improved YOLOv5 algorithm to build lightweight networks with new modules, introduce space-aware self-attention mechanism and improve Atrous spatial pyramid pool to improve feature extraction and detection speed and reduce model volume reduction. Lu et al[5] proposed SS-YOLO, an enhanced lightweight YOLOv7 model for surface defect detection of steel strips. By replacing the CBS module of the backbone network with MobileNetv3 network, and introducing D-SimSPPF module and non-parametric attention mechanism SimAM, the model size is reduced, the inference speed is accelerated, and the detection accuracy and feature extraction ability are improved. However, these methods still have shortcomings in terms of accuracy and real-time performance. In this paper, based on yolov8n algorithm, combined with the application of DCNv3 technology in the head part, and introduced Focal Loss function, the proposed algorithm achieved significant improvement in real-time, stability and accuracy through experimental verification.

2. YOLOv8 algorithm introduction

YOLOv8 is an object detection algorithm introduced in the YOLO series, which officially debuted in 2022. This version continues the traditional advantages of the YOLO family, including real- time processing power, structural simplicity and efficient operation, while being optimized in all aspects of performance and user friendliness. Like YOLOv5, YOLOv8 also offers five models of different sizes, namely x, l, m, s, and n, where x is the largest model and n is the smallest. This study chooses YOLOv8n[6]as the research basis, and the specific network structure details are shown in Figure 1.

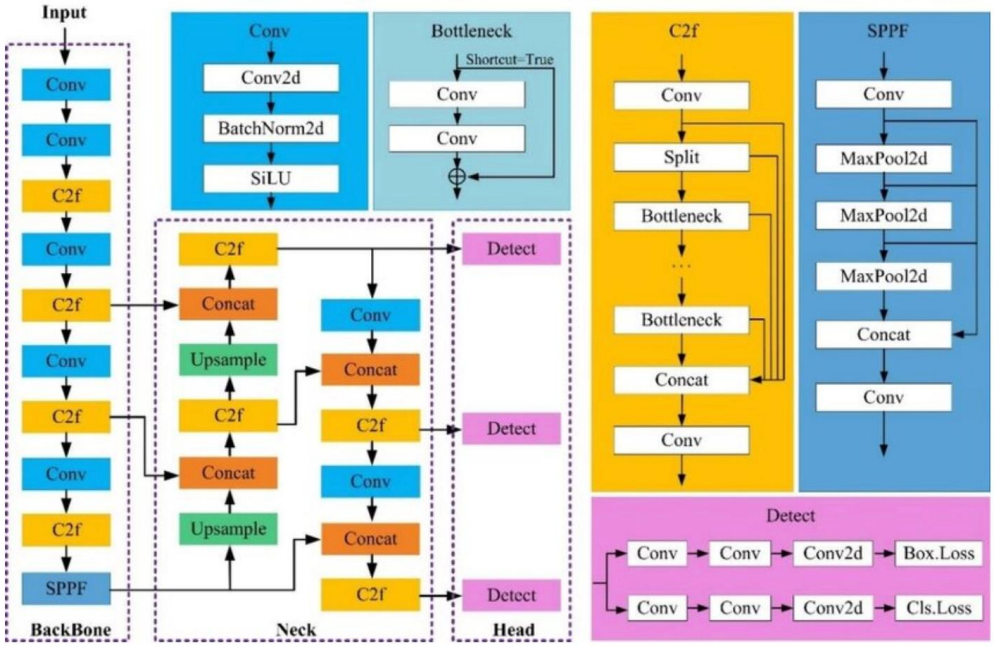

Figure 1. yolov8n Network Architecture

YOLOv8 has demonstrated excellent performance on multiple target detection datasets by introducing novel network structures and training strategies. The algorithm performs better in identifying small, dense and obscured objects. Compared to the previous YOLO series, YOLOv8 replaces the original 6x6 convolution kernel with 3x3 convolution kernel in its Backbone's initial convolution layer; At the same time, the C3 structure was updated to the C2f structure with richer gradient flow [7], and the depth and number of channels of the modules were adjusted according to models of different scales. In the Neck section, the 1x1 convolution layer used for dimensionality reduction was removed and the C3 module was replaced with a C2f structure. In the Head part, compared with YOLOv5[8], YOLOv8 has achieved two major improvements: First, it adopts Decoupled Head structure to separate classification and detection tasks; Second, there is a shift from an Anchor-Based mechanism to an Anchor-Free mechanism. In addition, YOLOv8 abandons traditional IOU matching or unilateral proportional allocation methods in favor of Task-Aligned Assigner for positive and negative sample matching and introduces Distribution Focal Loss (DFL). In the training stage, data enhancement also played an important role. In the last 10 cycles of training, YOLOv8 learned from the practice of YOLOX algorithm and selectively removed the Mosiac enhancement steps to enhance the generalization ability of the model.

3. Model improvement

3.1 DCNv3

DCNv3[9], Deformable Convolutional Networks version 3, is an advanced convolutional neural network (CNN) module designed to enhance feature representation in computer vision tasks.As a new member of the deformable Convolutional network (DCN) family, the core highlight of DCNv3 is the introduction of spatial sampling offsets to enhance the model's adaptability to geometric transformations.The essence of DCNv3 is its deformable convolution layer, which breaks through the limitation of traditional convolution sampling on fixed grid points and allows the convolution kernel to dynamically shift according to the content of the input feature map. This kind of migration is realized by learning the parameters, so that the network can adaptively grasp the geometric changes of the target.

In the implementation of DCNv3, the sampling points of each convolution kernel can be personalized to the content of the feature graph, and this process is learned by a clever subnetwork.This offset learning mechanism greatly improves the flexibility of the network to adapt to targets of various shapes and sizes.In addition, DCNv3 significantly improves the ability to encode target geometry information by using deformable sampling points, which is especially critical for target detection scenarios with significant deformations or complex backgrounds. It is worth noting that DCNv3 achieves performance improvements without significantly increasing the computational burden, maintaining the lightweight model.

$$y(p_0) = \sum_{g=1}^{G} \sum_{k=1}^{K} w_g m_{gk} x_g (p_0 + p_k + \Delta p_{gk})$$

In the field of computer vision, especially in object detection, semantic segmentation and other tasks, class imbalance has been a major challenge in training effective models. The traditional cross entropy loss function is often powerless in the face of the imbalance of positive and negative samples or difficult samples. For this purpose, Focal Loss was proposed. It was an innovative loss function that could dynamically adjust the loss weights in the training process, so

that the model could focus on those samples that were difficult to classify, and pay less attention to those samples that were easy to classify.

Focal Loss [10] was proposed by Lin et al to solve the problem of classification imbalance, especially in object detection. Focal losses are defined as follows:

$$FL = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

Where, pt is the prediction probability of the model for each category, a is the coefficient used to balance the positive and negative sample weights, and r is the parameter for adjusting the weight of the difficult and easy samples.

Focal Loss is centered on two key parameters: the focusing parameter γ and the balance factor α. The effect of the γ parameter is to adjust the model's attention to easy to classify and difficult to classify samples, by increasing the value of γ, the model will pay more attention to those samples that are difficult to distinguish. The α parameter is used to balance the weight of positive and negative samples, which is especially important in tasks such as object detection, because the number of background (negative samples) often far exceeds the number of foreground targets (positive samples), resulting in a model that may be overly biased to predict the background. By adjusting α properly, the model's attention to positive samples can be improved, and the recognition ability of a few categories can be improved.

4. Analysis of experimental results

4.1 Data set introduction

This paper uses a surface defect database published by Northeastern University (NEU), which focuses on hot-rolled steel strips and covers six common surface defect types: rolled oxide (RS), patch (Pa), cracking (Cr), pitted surface (PS), inclusions (In), and scratches (Sc). The database contains 1,800 grayscale images equally distributed for each defect type, with 300 samples each, ensuring a balanced and representative dataset.

4.2 Training parameters and Experimental environment

The training parameters are shown in Table 1

Table 1: Training parameters

Name	Numerical value
Training Picture Size (imgsz)	640×640
batch-size	8
Training iterations (epochs)	200
Initial learning Rate (lr0)	0.01

The experimental environment is shown in Table 2

Table 2: Experimental environment

Name	Related configuration
CPU	12 vCPU Intel(R) Xeon(R) Silver 4214R CPU @
	2.40GHz
Internal memory	12GB
GPU	RTX 3080 Ti
CUDA/CUDNN	11.8
Operating system	Linux Ubuntu 20.04.4 LTS
Python	3.1

4.3 Evaluation index

In this object detection experiment, we use average accuracy (mAP) as the core evaluation index. Specifically, we chose mAP_0.5, which is the average accuracy when the IoU (crossover ratio) threshold is 0.5. mAP_0.5 can effectively reflect the comprehensive performance of the model when detecting different types of targets. By calculating the accuracy of each category under different recall rates and averaging, we get a comprehensive and objective evaluation result. This evaluation standard has high authority and universal applicability in the field of target detection, which helps us to better analyze and compare the performance of different models in the experiment.

4.4 Experimental Results

The training results of the improved yolov8 algorithm on the steel defect data set are shown in Figure 2, and the data are shown in Table 3.

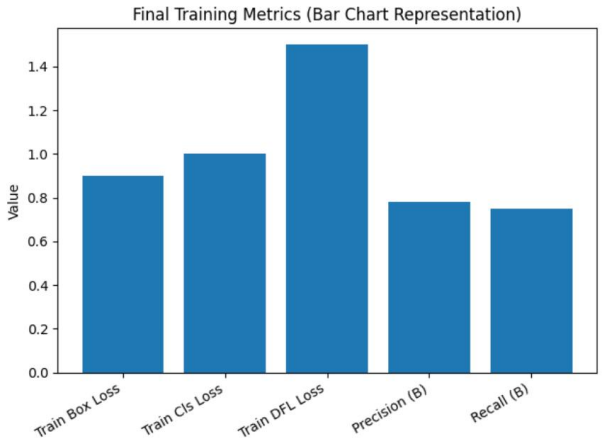

Figure 2. yolov8-DCNv3-Focal training results

Table 3: Experimental Results

Model	mAP_0.5
Yolov8	0.803
Yolov8-DCNv3	0.811
Yolov8-DCNv3-Focal	0.816

It can be analyzed from Figure 3 that box_loss and cls_loss are both high at the beginning, but gradually decrease and tend to be stable with the increase of the number of rounds, and finally drop below 1.0. mAP_0.5 is low at the beginning, and the detection fluctuates, but tends to be stable with the increase of the number of rounds. It can be seen from Figure 4 that the accuracy rate of steel defect detection has reached 80%. It can be seen from Table 3 that mAP_0.5 increased by 1.3% after the introduction of DCNv3 and Focal Loss.

5. Conclusion

Aiming at the lack of accuracy in steel defect detection, the YOLOV8-DCNV3-FOCAL algorithm proposed in this paper was verified by experiments, and the mAP_0.5 value of the algorithm reached 81.6%. Compared with traditional detection methods, this significant improvement not only reflected the advantages of YOLOv8 model in real-time detection speed. By integrating DCNv3, the model's ability to capture complex and variable defects on the steel surface was enhanced, and the classification imbalance was effectively alleviated by Focal Loss, further improving the detection accuracy. However, instability occurs in the early stage of training, and this problem will be studied in the future based on the algorithm in this paper

References

[1] Gao Y, Lv G, Xiao D, et al. Research on steel surface defect classification method based on deep learning[J]. Scientific Reports, 2024, 14(1): 8254.

[2] Tang B, Chen L, Sun W, et al. Review of surface defect detection of steel products based on machine vision[J]. IET Image Processing, 2023, 17(2): 303–322.

[3] Shi J, Yang J, Zhang Y. Research on steel surface defect detection based on YOLOv5 with attention mechanism[J]. Electronics, 2022, 11(22): 3735.

[4] Wen H, Li Y, Wang Y, et al. Usage of an improved YOLOv5 for steel surface defect detection[J]. Materials Testing, 2024, 66(5): 726–735.

[5] Lu J, Yu M M, Liu J. Lightweight strip steel defect detection algorithm based on improved YOLOv7[J]. Scientific Reports, 2024, 14(1): 13267.

[6] Bao J, Li S, Wang G, et al. Improved YOLOv8 network and application in safety helmet detection[C]// Journal of Physics: Conference Series. IOP Publishing, 2023, 2632(1): 012012.

[7] Solawetz F. What is YOLOv8? The ultimate guide[EB/OL]. 2023-01-11 [2023-06-21].

[8] Lu Y, Huang Z C, Jiang Y Q, et al. Lightweight-Detection: The strip steel surface defect identification based on improved YOLOv5[J]. Materials Today Communications, 2024: 109814.

[9] Li H, Zhang Y, Zhang Y, et al. DCNv3: Towards Next Generation Deep Cross Network for CTR Prediction[J]. arXiv preprint arXiv:2407.13349, 2024.

[10] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980–2988.