Improved Strip Steel Defect Detection Algorithm Based on YOLOv8

Shuai Wang, Bin Zhu, MingYi Zhang

Abstract— **Addressing the issues of slow detection speed, large model size making deployment on edge computing devices challenging, and poor performance in small object detection present in existing deep learning-based hot-rolled strip steel detection methods, this paper proposes an improved strip steel defect detection algorithm based on YOLOv8. The aim is to tackle problems such as slow detection speed, low accuracy, and high model parameter count in strip steel defect detection. By optimizing the network structure, introducing a multi-scale feature fusion mechanism, and adopting enhanced data preprocessing strategies, the proposed algorithm outperforms traditional methods across multiple strip steel defect datasets. Specific improvements include reconstructing the feature extraction backbone network using HGStem and Rep_HGBlock, replacing Conv modules with DWConv modules, and employing Focal CIoU Loss as the loss function. Experimental results demonstrate that the improved model shows significant enhancements in terms of model size, inference speed, and detection accuracy.**

Index Terms—**Deep learning, hot-rolled strip steel, defect detection，yolov8**

I. INTRODUCTION

With the advent of the Industry 4.0 era, the manufacturing industry has increasingly stringent requirements for product quality. In the steel production process, strip steel serves as a crucial raw material, and its surface defects directly impact subsequent processing and the final product quality. Therefore, accurate and efficient detection of strip steel surface defects is of great significance.

Traditional methods for detecting strip steel defects mainly rely on manual inspection or simple image processing techniques. These methods are inefficient and susceptible to human factors, leading to frequent occurrences of missed detections or false positives. In recent years, deep learning-based object detection algorithms have achieved remarkable results in the field of computer vision, with the You Only Look Once (YOLO) series models standing out. These models, known for their high real-time performance and considerable detection accuracy, have been widely applied in various scenarios.

However, standard YOLO models still face challenges when dealing with complex strip steel defect detection tasks. For instance, since strip steel defects often exhibit diverse morphological characteristics and irregular distributions, standard YOLO models may struggle to capture these details adequately. Additionally, the similarity between different types of defects further complicates the detection task. In industrial production processes, product inspections frequently use industrial cameras for real-time online detection, which demands high real-time performance. Moreover, industrial equipment cannot always match the capabilities of the latest experimental devices, thus imposing significant constraints on model complexity.

To address these issues, this study proposes an improved strip steel defect detection algorithm based on YOLOv8. By optimizing the network structure, introducing a multi-scale feature fusion mechanism, and adopting enhanced data preprocessing strategies, we aim to improve the model's detection accuracy, real-time performance, and robustness for strip steel defects. The main contributions of this research include:(1) Structural adjustments to the YOLOv8 model tailored to the characteristics of strip steel defects to enhance adaptability.(2) Design and implementation of a multi-scale feature fusion mechanism that effectively improves the model’s ability to recognize defects of various sizes.(3) Development of specialized data augmentation methods for strip steel defect datasets, further enhancing the model’s generalization performance.

Experimental validation demonstrates that the improved YOLOv8 model outperforms traditional methods across multiple strip steel defect datasets, showcasing promising application prospects.

II. YOLOV8 ALGORITHM

A. Introduction to YOLOv8

YOLOv8 is one of the latest advancements in the field of object detection, excelling in various applications due to its higher detection accuracy and faster inference speed. Compared to previous versions, YOLOv8 has undergone numerous improvements, making it highly effective for practical industrial problems. The current version, 1.0, is among the best-performing versions in terms of overall performance.

The YOLOv8s network is a lightweight variant of YOLOv8, particularly suitable for real-time detection tasks such as fire detection. The structure of the YOLOv8s network is illustrated in the figure, with its backbone network composed of Focus, Conv, C3, and SPPF modules. The definitions and functions of these structures are described as follows.

Manuscript received February 08, 2025
Shuai Wang, School of computer science and technology, Tiangong university, Tianjin ,China.
Bin Zhu, School of Software, Tiangong university, Tianjin ,China.
MingYi Zhang, School of computer science and technology, Tiangong university, Tianjin ,China.

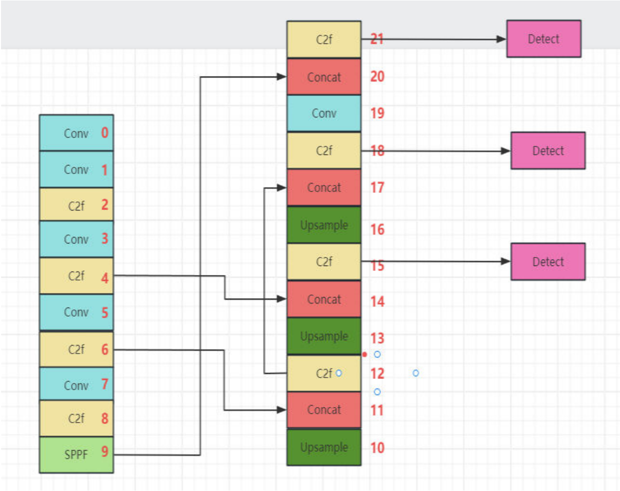

Fig. 2.1 YOLOv8 Architecture

B. *Backbone*

Focus: The Focus structure reduces the computational load of convolutional layers by slicing the input image, thereby increasing the speed of convolution processing. For example, if the input image is 640x640x3, Focus slices it into 320x320x12 feature maps and then generates feature maps through 64 convolution kernels.

Conv: Standard convolutional layers are used to extract local features. Through a series of convolution operations, spatial information is transformed into deep feature representations.

C3: C3 is used for feature extraction and is composed of BottleneckCSP and CSPDarkNet53 networks' Conv combinations. It splits the input into two parts: one part is processed by dense blocks, while the other part is passed directly to the next stage without any processing. This allows different dense layers to iteratively learn copied gradient information.

SPPF (Spatial Pyramid Pooling - Fast): The SPPF structure maximizes pooling with three convolution kernels of sizes 5, 9, and 13, respectively, to improve the receptive field of the network. At the same time, it optimizes computational efficiency and enhances processing speed.

C. *Feature Fusion*

After the input image is processed by the backbone network, its features are transformed into semantic features. Starting from the lower layers, as the input image goes deeper through the network, the complexity of its semantic features increases. However, due to the downsampling process, the spatial resolution of the feature maps continuously decreases, leading to a loss of spatial information and fine-grained features. To retain these fine-grained features, the idea of FPN (Feature Pyramid Network) is applied in YOLOv8's feature fusion network, and PANet (Path Aggregation Network) is used as the feature fusion network.

D. *Prediction Network*

The YOLOv8 prediction network consists of multiple detectors, each assigned with anchors of different scales. These detectors predict bounding boxes on feature maps at different scales, specifically P3~P5 detection layers, ensuring that the feature maps used for detection contain both low-level visual features and high-level semantic information from the input image. Additionally, YOLOv8 introduces new loss functions and optimization strategies to further improve detection accuracy and robustness.

III. IMPROVEMENT METHODS

A. *Initial Module Improvement*

HGStem is a critical initial module in convolutional neural networks (CNNs) that provides effective feature extraction and processing, laying a solid foundation for the subsequent deep network layers. YOLOv8, as a CNN-based object detection model, highly relies on the quality of initial features. By introducing HGStem, higher quality initial feature representations can be provided, thereby offering a robust foundation for subsequent detection heads and enhancing the overall performance of the model.

B. *Module Lightweighting*

RepHGNetV2 is a convolutional neural network architecture designed specifically for image recognition and other computer vision tasks. It is an improved version of the original RepHGNet, aimed at enhancing performance, reducing the number of parameters, and speeding up inference. One of its core components is the Rep_HGBlock, which combines the advantages of efficient architectures like Hourglass Networks and RepVGG:(1) Multi-scale Feature Fusion: Integrates features from different resolutions to capture more information and contextual relationships.(2) Reparameterization Technique: Uses complex structures during training to enhance expressive capability, while transforming them into simpler, more efficient forms during inference to accelerate the process.

By optimizing kernel size, stride, and other parameters, Rep_HGBlock maintains high accuracy while minimizing computational overhead, making it suitable for resource-constrained environments such as mobile devices or embedded systems.

C. *Convolution Module Improvement*

DWConv (Depthwise Convolution) is an efficient convolution operation used in CNNs. Unlike traditional standard convolutions, DWConv separates spatial and channel dimensions for convolution, significantly reducing computation and parameter count. Therefore, this paper replaces Conv modules with DWConv modules. The improved algorithm architecture is illustrated in the accompanying diagram (assuming there is an illustration).

This improvement not only reduces computational cost but also enhances model efficiency without sacrificing accuracy, making it better suited for practical application needs.

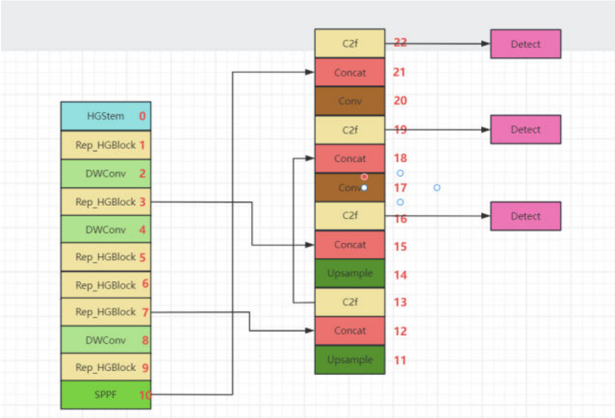

Fig 3.1 YOLOv8-improve Architecture

D. Loss Function Improvement

Focal CIoU is an improved bounding box regression loss function that combines the advantages of Focal Loss and Complete IoU (CIoU) to enhance the performance of object detection models.

- Focal Loss addresses the class imbalance problem by assigning higher weights to hard-to-classify samples, thereby reducing the impact of easily classified samples.
- Complete IoU (CIoU) considers not only the Intersection over Union (IoU) but also factors such as the distance between center points and scale differences, ensuring better alignment between the predicted boxes and ground truth boxes.

By integrating these two approaches, Focal CIoU enhances localization accuracy, helping to reduce false positives and missed detections. In essence, Focal CIoU leverages the dynamic weighting mechanism of Focal Loss along with the multi-dimensional matching evaluation of CIoU, providing a more efficient and precise loss function for object detection tasks.

IV. EXPERIMENTS

A. Datasets

NEU-DET-PRO: The dataset used in this paper is an extended version of NEU-DET, named NEU-DET-PRO. The dataset consists of 2416 images covering six common types of steel surface defects: Crazing, Inclusion, Patches, Pitted Surface, Rolled-in Scale, and Scratches. The dataset is divided into three non-overlapping subsets: 2048 images for training, 232 images for testing, and 136 images for validation.

B. Lightweight Improvement Comparison Experiment

To verify the effectiveness of the lightweight improvements, a new feature extraction backbone network was reconstructed based on YOLOv8n using HGStem and Rep_HGBlock, and the improved model was named YOLOv5s-HG. This model was compared with YOLOv3, YOLOv4, and the YOLOv5 series. Experimental comparisons revealed that adding a coordinate attention module can improve detection precision (P) and mean average precision (mAP) by sacrificing a small amount of model size and detection speed, thereby effectively enhancing the overall performance of the model.

C. Loss Function Improvement Comparison Experiment

To verify the impact of the improved loss function on algorithm performance, the loss functions in the algorithm were individually replaced, and parallel comparison experiments were conducted using Focal CIoU Loss, CIOU Loss, and EIOU Loss from the original algorithm. The results are shown in Table 4.1.

Table 4.1 Comparison of Loss Function Experiment Results

loss function	mAP
CIOU	0.764
EIOU	0.782
Focal CIoU	0.793

From the results, it is clear that Focal CIoU Loss outperforms both CIOU Loss and EIOU Loss in terms of detection speed and measurement accuracy. This experiment demonstrates the superiority of the Focal CIoU loss function, showing its capability to enhance the performance of the model more effectively.

V. CONCLUSION

In this paper, an improved YOLOv8n algorithm is proposed to address issues such as slow detection speed, low accuracy, and a high number of model parameters in steel strip defect detection.

Based on YOLOv8n, the new feature extraction backbone network was reconstructed using HGStem and Rep_HGBlock, reducing the complexity of the model and achieving lightweighting. Additionally, DWConv modules were used to replace Conv modules, significantly reducing computation and the number of parameters. Finally, Focal CIoU Loss replaced the original loss function, improving both localization accuracy and bounding box regression speed. A custom-developed steel strip defect dataset was created and experiments were conducted on this dataset to validate the algorithm.

Experimental results show that the improved model has significant advantages: the model size is reduced to only 5.0MB, which is 20.64% smaller than the original model; inference speed on RTX3090 devices reached 289.8 milliseconds per image, a 15.65% improvement over the original model; and mAP achieved 0.817, a 7.4% increase compared to the original model. These improvements meet real-time requirements while ensuring accuracy.

In the future, further enhancements will be made to this algorithm, and it will be deployed on edge computing devices for real-time hot rolling steel detection.

REFERENCES

[1] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics
[2] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2023, April 17). DETRs Beat YOLOs on Real-time Object Detection. arXiv:2304.08069 [cs.CV]. https://doi.org/10.48550/arXiv.2304.08069
[3] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015, May). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1 - 9).

[4] Lin, T. - Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), 318 - 327. https://doi.org/10.1109/TPAMI.2018.2858826

[5] Redmon, Joseph , and A. Farhadi . "YOLOv3: An Incremental Improvement." arXiv e-prints (2018).

[6] Baek Jaeseung, et al."A wavelet-based real-time fire detection algorithm with multi-modeling framework." Expert Systems With Applications 233.(2023).

[7] Wang, Wenhai , et al. "Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network." IEEE (2019).

[8] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2117-2125).

[9] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8759-8768).

[10] Shorten, C., & Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60.