W. Z. TENG, Y. J. ZHANG, H. G. ZHANG, D. X. GAO

ISSN 0543-5846
METABK 63(3-4) 399-402 (2024)
UDC – UDK 621.771-536.45:669.14:620.191-620.197.6:62.001.57=111

# SURFACE DEFECT DETECTION OF STEEL BASED ON IMPROVED YOLOv7 MODEL

Received – Primljeno: 2024-02-09
Accepted – Prihvaćeno: 2024-04-02
Original Scientific Paper – Izvorni znanstveni rad

In response to the inevitable surface defects in the manufacturing process of hot-rolled steel, this paper proposes an improved steel surface defect detection model based on YOLOv7. In the Extended Efficient Large Aggregation Network (E-ELAN), the model replaces conventional convolution with Omni-Dimensional Dynamic Convolution (ODConv) to enhance the network's sensitivity to feature extraction using a combination of various attention mechanisms. Additionally, the detection head in the head section is replaced with an Efficient Decoupled Detection Head, enhancing the model's capability to classify and locate small defects. The proposed model is tested on the public dataset NEU-DET, achieving a high mAP of 76,5 %. This effectively enhances the model's ability to detect surface defects in steel while maintaining a fast detection speed.

*Keywords:* hot-rolled, steel, surface defect, YOLOv7 model, efficient decoupled detection head

## INTRODUCTION

Steel is an important raw material in the field of mechanical design and manufacturing. Hot-rolled steel refers to the production of steel strips and plates using the method of hot rolling. Due to the high-temperature environment in the production process of hot-rolled steel, surface defects are inevitable. If not detected in a timely manner, it can affect the quality of the product and even cause serious harm to the users of the steel. Therefore, to ensure the safety performance of steel products, it is essential to conduct defect detection on the surface of hot-rolled steel [1].

Traditional steel defect detection mainly relies on human observation. Under long-term, high-intensity work, inspectors are prone to visual fatigue, leading to situations of false positives and false negatives, as well as slow detection speeds. With the continuous development of computer vision technology, using image processing techniques for steel defect detection has become mainstream. However, there are still some shortcomings. Therefore, this paper proposes an improved model based on the YOLOv7 framework for common surface defects in hot-rolled steel. The model is validated on the NEU-DET dataset, and through comparative experiments with other similar algorithms, the paper demonstrates the advancement of the improved model.

## RALATED WORK

Deep learning-based object detection models can be divided into two categories: one is the two-stage object detection model represented by R-CNN, Faster R-CNN, etc., and the other is the one-stage object detection model represented by SSD, YOLO, etc. Object detection models based on deep learning provide a solid theoretical foundation for the defect detection on the surface of hot-rolled steel, and experts and researchers in the field have achieved significant results. He et al. proposed an improved steel defect detection model based on the Faster R-CNN framework and trained it on the Northeastern University open dataset, however, the detection speed is difficult to meet the requirements of industrial applications [2]. Li et al. proposed a shallow feature enhancement network based on YOLOv4, incorporating a feature pyramid network and convolutional block attention mechanism to enhance the extraction capability for steel surface defects [3]. The aforementioned studies have laid a solid foundation for steel defect detection, but there is still room for improvement in terms of accuracy and robustness. In order to meet the real-time and accurate requirements of defect detection on the surface of hot-rolled steel, this paper proposes an improved model based on YOLOv7. After testing on the NEU-DET dataset, the detection accuracy reaches an mAP of 76,5 %, and the detection speed reaches 87,5 frame $\cdot s^{-1}$.

## METHODOLOGY

The YOLOv7 series algorithm is proposed by Alexey Bochkovskiy and has a significant advantage over the previous YOLO series in terms of detection accuracy and speed [4]. The YOLOv7 network model mainly consists of three parts: the backbone network, the neck network, and the head. The backbone network is responsible for

W. Z. Teng, Y. J. Zhang, H. G. Zhang, D. X. Gao, School of Computer Science and Software Engineering, University of Science and Technology Liaoning, China. Corresponding author: Y. J. Zhang (1997zyj@163.com)

extracting features from the input images, the Neck is responsible for merging the extracted features to obtain small, medium, and large-sized features. Finally, the fused features are passed to the detection head, and after detection, the final results are output. The network architecture of YOLOv7 is shown in Figure 1.

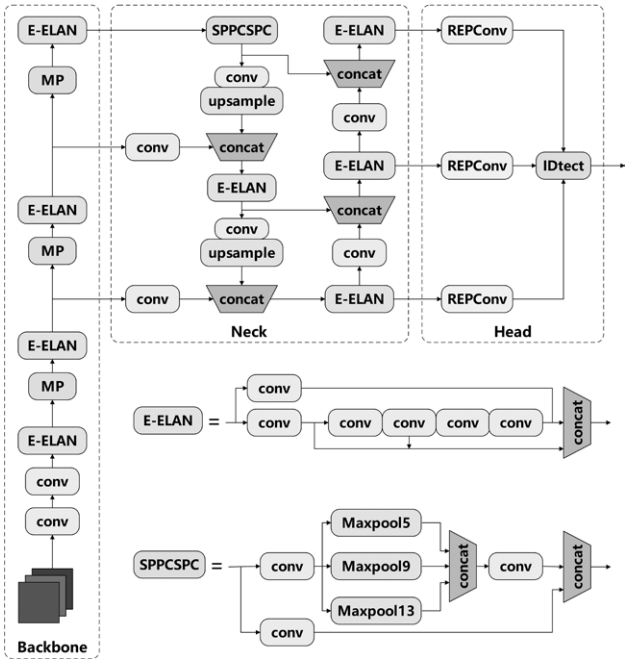


**Figure 1** The network structure of YOLOv7

The backbone network of YOLOv7 consists mainly of convolutional layers, Expanded Efficient Lightweight Aggregation Network (E-ELAN), MPConv and SPPCSPC. The E-ELAN module enhances the learning capacity of the network without disrupting the gradient path. The MPConv module expands the receptive field of the current feature layer and fuses it with information processed by conventional convolution, improving the network's generalization. The SPPCSPC module is applied at the end of the backbone network, introducing a series of convolution operations in parallel pooling to avoid issues like image distortion and addressing challenges in extracting redundant features in convolutional neural networks. In the feature fusion Neck network, YOLOv7 shares a similar structure with the YOLO series, employing the Path Aggregation Feature Pyramid Network (PAFPN) structure and incorporating the E-ELAN module. It aggregates information from different network paths or feature pyramids effectively using an adaptive receptive field to enhance the network's capability in detecting small targets. In the detection head, YOLOv7 utilizes the IDtect detection head for three different target sizes. It incorporates the Reparameterized Convolution (RepConv), introducing learnable parameters in the convolution kernel, making the network adaptive and better able to capture features in the data. This enhances the model's performance and generalization ability.

Based on the improved YOLOv7 network model in this paper, the first enhancement is the replacement of conventional convolution in the E-ELAN module with Omni-Dimensional Dynamic Convolution (ODConv) [5]. This modification utilizes a combination of multiple attention mechanisms to enhance the network's sensitivity in feature extraction, thereby improving its capability to detect small defects. In traditional convolutions, stacking or adding convolutional layers is a common approach, which not only increases computational costs but also hinders the efficiency of defect detection in steel. ODConv employs a multi-dimensional attention strategy, applying attention weighting in parallel along the four dimensions of the convolutional kernel. This dynamic allocation of different weights to different convolutional kernels enhances their adaptive capabilities. Compared to traditional convolutions, full-dimensional convolutions are suitable for features of different sizes and shapes, providing greater flexibility and expressive power. The structure of the Omni-Dimensional Dynamic Convolution module is illustrated in Figure 2.

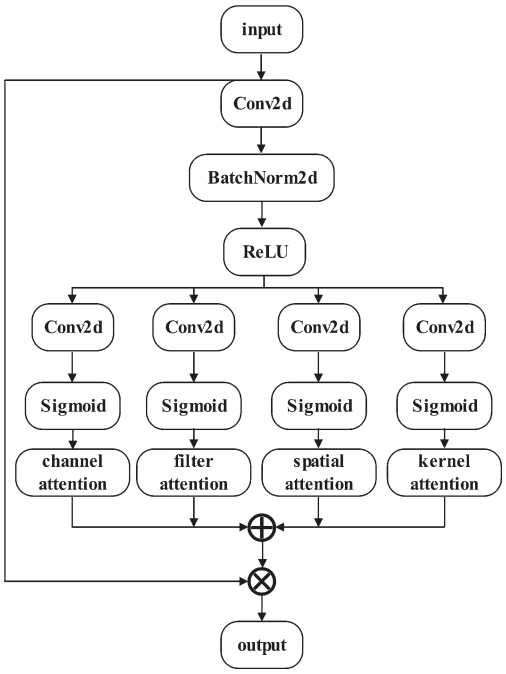


**Figure 2** ODConv Structure.

In the ODConv module, the input feature map undergoes operations such as average pooling, fully connected layers, batch normalization, and activation functions. Based on different attention types, it calculates attention weights for different channels. These attention weights are then multiplied with the feature map to aggregate feature information along four dimensions, yielding the output result after full-dimensional dynamic convolution. The formula for calculating attention weights is shown in Formula 1.

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \cdots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * \chi \quad (1)$$

where $\alpha_{w1}$ denotes assigning different values to $n$ overall convolution kernels, $\alpha_{f1}$, $\alpha_{c1}$ represents assigning different

attention weights to convolution filters for various output and input channels, and $\alpha_{s1}$ signifies assigning different attention weights to parameters at convolutional spatial positions. ⊙ indicates multiplication along different dimensions of the convolutional kernel. Through the collaborative efforts of these four dimensions of attention, fully utilizing the convolutional kernel space, input and output channel information to acquire rich contextual information, it provides assurance for enhancing the detection capability of the network model. Additionally, it reduces runtime through parallel operations.

Furthermore, since surface defect detection in steel is a multi-scale, multi-target detection problem, differences in scale can cause detection models to overlook minor defects, leading to instances of missed detection. Therefore, surface defect detection tasks in steel require more accurate localization information and richer classification information. In this regard, this paper replaces the original detection head with the Efficient Decoupled Detection Head(EDDH) in the head section to predict targets. YOLOv7 initially utilizes a coupled detection head, where classification and localization tasks share the same parameters. This joint processing approach results in mutual interference between classification and regression tasks, thereby impacting detection accuracy. The EDDH handles classification and regression tasks separately, allowing the network to focus more on each individual task and, consequently, improving the model's detection accuracy for small-sized defects. The structure of the Efficient Decoupled Head is depicted in Figure 3.

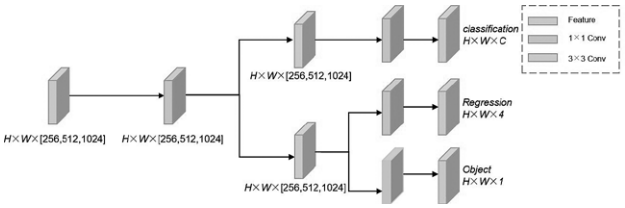


**Figure 3** Efficient decoupled detection head structure.

As shown in Figure 3, the input feature information undergoes channel adjustment through a 1×1 convolution, followed by feeding the feature map into two parallel channels. Each channel comprises a 3×3 convolutional layer for feature extraction. The upper channel, after feature extraction, adjusts the feature channel number to accomplish the classification task. The lower channel, after feature extraction, further branches into two subpaths. One sub-path is responsible for obtaining the parameters of the bounding box (height, width, and center coordinates). The other sub-path is responsible for obtaining the confidence parameter. Compared to traditional decoupled heads, this approach can enhance detection accuracy while making the network more efficient.

## METHOD OF IMPLEMENTATION

This paper utilizes the Northeastern University open dataset (NEU-DET) for training and validation. The dataset consists of 1800 grayscale images of steel surface defects, encompassing six different defect types: Rolled Scale (Rs), Patches (Pa), Scratches (Sc), Cracks (Cr), Inclusions (In), and Pitted-Surface (Ps). In the experiments, the dataset is divided into a training set and a validation set with an 8:2 ratio.

The experiments were conducted on the Windows 10 operating system, utilizing the pytorch1,12 framework, and powered by an NVIDIA RTX 3080 GPU. The batch size was set to 8, and the training process spanned 300 epochs. Stochastic Gradient Descent (SGD) was employed to adjust the network parameters. The initial learning rate was set to 0,01, with a weight decay coefficient of 0,0005. Additionally, a cosine annealing algorithm was applied for learning rate adjustments.

The experiment selected Average Precision (AP), Mean Average Precision (mAP), and Frames Per Second (FPS) as evaluation metrics for the steel surface defect detection results. The formulas for calculating each metric are as follows:

$$AP = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$mAP = \frac{\sum_{n=1}^{Num(class)} AP_{(n)}}{TP + TN + FP + FN} \tag{3}$$

In the formulas, TP represents the number of correctly identified positive samples; TN represents the number of correctly identified negative samples; FP represents the number of negative samples incorrectly identified as positive; FN represents the number of positive samples incorrectly identified as negative.

## EXPERIMENT AND ANALYSIS

To verify the effectiveness of the improved model proposed in this paper, four mainstream object detection models, including Faster R-CNN [6], SSD, YOLOv3 and YOLOv7, were selected for comparative experiments on the NEU-DET dataset. The experimental results are shown in Table 1.

As shown in Table 1, From the analysis of Table 1, it is evident that among various models, YOLOv7 achieves a high detection accuracy of 74,9 % and a detection speed of up to 87,9 frame · $s^{-1}$, making it a rea-

Table 1 **Results of comparative experiments on different algorithms.**

| Category | Faster R-CNN | SSD | YOLOv3 | YOLOv7 | ours |
|---|---|---|---|---|---|
| Cr | 34,2 | 36,5 | 22,0 | 40,2 | 50,9 |
| In | 70,1 | 67,2 | 68,0 | 79,7 | 82,1 |
| Pa | 81,2 | 86,0 | 77,2 | 95,5 | 88,3 |
| Ps | 81,5 | 55,6 | 33,3 | 80,8 | 83,7 |
| Rs | 53,0 | 61,2 | 21,1 | 61,4 | 62,3 |
| Sc | 79,1 | 57,9 | 61,1 | 91,7 | 91,5 |
| mAP | 66,5 | 60,8 | 47,1 | 74,9 | 76,5 |
| FPS | 28,4 | 47,0 | 50,8 | 87,9 | 87,5 |

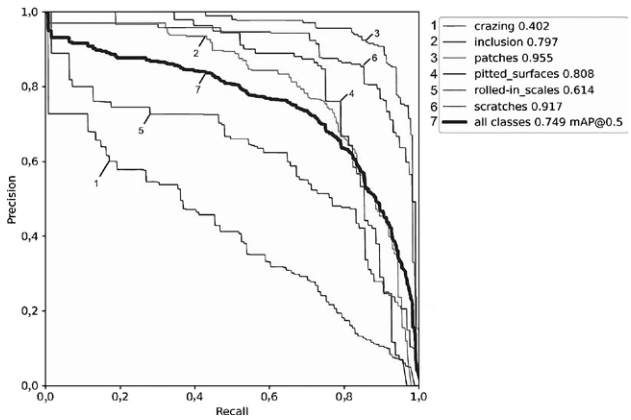


**Figure 4** Original YOLOv7 results.

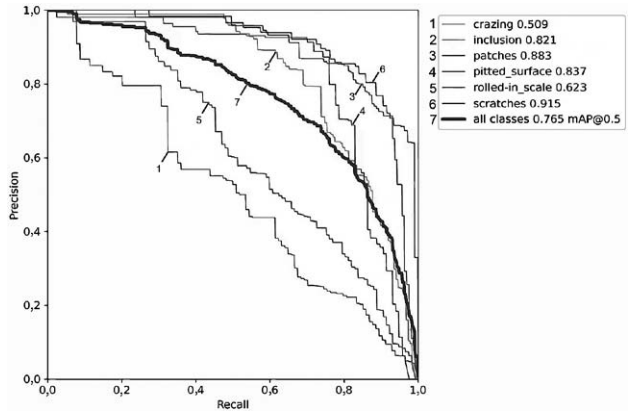


**Figure 5** Improved YOLOv7 results.

sonable choice for the model selected for improvement in this paper. After the improvements made in this paper, the model's detection accuracy reaches 76,5 %. Although there is a slight decrease in the detection accuracy of Patches and Scratches, the overall average accuracy is the highest among all models. In conclusion, the proposed network model demonstrates more accurate recognition of surface defects in steel materials and exhibits good generalization and robustness.

## CONCLUSION

Addressing the issue of surface defects in the manufacturing process of hot-rolled steel, this paper proposes an object detection model based on YOLOv7 that integrates Omni-Dimensional Dynamic Convolution and an Efficient Decoupled Detection Head. Comparative experiments with other models demonstrate that the improved model in this paper exhibits strong capabilities in defect detection, classification, and localization, meeting the high-speed requirements of hot-rolled steel production lines.

## Acknowledgements

This work was supported in part by Key Laboratory of iot Application Technology for Intelligent Construction in Liaoning Province, Project number: 2021JH13 /10200051.

## REFERENCES

[1] Luo Q, Fang X, Liu L, et al. Automated visual defect detection for flat steel surface: A survey[J]. IEEE Transactions on Instrumentation and Measurement, 69(2020)3, 626-644.

[2] He Y, Song K, Meng Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features[J]. IEEE transactions on instrumentation and measurement, 69(2019)4, 1493-1504.

[3] Li M, Wang H, Wan Z. Surface defect detection of steel strips based on improved YOLOv4[J]. Computers and Electrical Engineering, (2022), 102, 108208.

[4] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conferenceon Computer Vision and Pattern Recognition. (2023), 7464-7475.

[5] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arxiv preprint arxiv:2209.07947, 2022.

[6] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, (2015), 28.

**Note:** The responsible translators for English language is J. Wang – University of Science and Technology Liaoning, China