

Defect identification method for steel surfaces based on improved YOLOv5

Wang Shuo¹ Zhang Liaojun² Yin Guojiang³

(¹ College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China)

(² College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

(³ Safety Testing Center of Hydraulic Metal Structure of the Ministry of Water Resources, Hohai University, Nanjing 210098, China)

Abstract: Traditional machine vision detection methods suffer from low accuracy in identifying small-scale defects. To address this, a nondestructive identification method for steel surface defects is proposed based on an enhanced version of the fifth version of the You Only Look Once (YOLOv5) algorithm. In this improved approach, the Res2Block module is incorporated into the backbone of the YOLOv5 algorithm to expand the receptive field and improve computational efficiency. Additionally, the recursive gated convolution structure is fused into the neck of the YOLOv5 algorithm to further enhance the computational performance of the surface defect identification method. To validate the effectiveness of the proposed method, a series of ablation experiments were conducted using different module combinations. These results were then compared with those obtained through other object detection methods. This comparison reveals that the proposed method achieves a mean average precision of 67.8% and an F_1 -score of 86.0% in steel surface defect identification. When compared with the original YOLOv5 algorithm, the proposed method exhibits superior performance, particularly in the identification of small-scale steel surface defects. Furthermore, it also surpasses other object detection methods, such as SSD, YOLOv3, YOLOv5-Lite, and YOLOv8, demonstrating significant improvements in computational accuracy.

Key words: steel; defect detection; convolutional neural network; You Only Look Once (YOLO)

DOI: 10.3969/j.issn.1003-7985.2024.01.006

Steel is a critical material in various fields, such as transportation, construction, and industrial manufacturing^[1]. The surface quality of steel determines the performance and application of the products it forms. How-

ever, owing to the manufacturing process and production environment, steel often exhibits surface defects such as crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches. These defects not only affect the appearance of steel but also significantly damage its quality, leading to decreased corrosion resistance and wear resistance. Undertaking surface defect detection can help identify different types of defects, allowing for appropriate measures to maintain product quality and reliability. This process is of immense importance for improving product quality, ensuring product safety, and facilitating remanufacturing and defect repair.

The primary objective of surface defect detection in steel plates is to accurately categorize, locate, and determine defects. Historically, this task relied heavily on manual feature extraction and machine learning methods. However, recent years have seen notable advancements in defect detection through the adoption of deep learning methods, particularly convolutional neural networks and their derivatives^[2-4]. In the domain of steel surface defect detection, numerous target detection models have been proposed and extensively used^[5-7]. Two-stage target detection models employ selective search and anchor box generation algorithms to identify target candidate regions, subsequently conducting position regression and classification on these regions^[8-9]. Despite their superior training accuracy, two-stage target detection models tend to be more intricate and less efficient in terms of training steps. By contrast, single-stage object detection models excel in real-time processing. They provide rapid and efficient detection through a streamlined, end-to-end approach. Representative algorithms of such models include You Only Look Once (YOLO)^[10], which has found widespread application in object detection^[11-12].

Surface defect detection on steel plates presents a unique challenge owing to large variations, particularly when identifying smaller defects. To address this, Zhao et al.^[13] proposed an RDD model based on YOLOv5. The backbone of this model mainly consists of Res2Net blocks designed to enlarge the receptive field and extract features at different scales. Additionally, a dual-feature pyramid network was designed to enhance the neck and generate

Received 2023-09-18, **Revised** 2023-12-13.

Biographies: Wang Shuo (1992—), female, doctor; Zhang Liaojun (corresponding author), male, doctor, professor, ljzhang@hhu.edu.cn.

Foundation items: The Natural Science Foundation of Jiangsu Province (No. BK20230956), the Jiangsu Funding Program for Excellent Postdoctoral Talents (No. 2022ZB188), the Transportation Technology Plan Project of Jiangsu Province (No. 2020QD28).

Citation: Wang Shuo, Zhang Liaojun, Yin Guojiang. Defect identification method for steel surfaces based on improved YOLOv5[J]. Journal of Southeast University (English Edition), 2024, 40(1): 49 – 57. DOI: 10.3969/j.issn.1003-7985.2024.01.006.

rich representations. A decoupled head was used to separate the regression and classification tasks, resulting in higher detection accuracy. Guo et al.^[14] proposed the MSFT-YOLO model, incorporating TRANS modules based on the transformer design into the backbone and detection head to combine features with global information. The bidirectional feature pyramid network (BiFPN) was used to fuse information at different scales. Data augmentation and multistep training methods were introduced to further improve performance and detection speed. Liao et al.^[15] proposed an improved YOLOv5 method for recognizing surface defects on Si_3N_4 ceramic bearing balls. Coordination attention was applied to the backbone, and the defect feature information was fused at different scales into a weighted BiFPN. All these studies have contributed significantly to improving methods for identifying steel surface defects. However, despite these extensive research efforts, room for improvement remains, particularly in small defect detection.

To enhance the accuracy of detecting small defects on steel surfaces, this study proposes an innovative refinement to the deep learning object detection model. Our enhancement strategy involves incorporating the Res2Net to strengthen the backbone, enabling finer receptive fields at various granularities. We strategically employ the Recursive Gated Convolution ($g^n\text{Conv}$) module in the neck to augment the model's spatial interaction capabilities. Fur-

thermore, we integrate the use of the complete intersection over union (CIoU) loss, offering a more comprehensive evaluation metric that considers not only the spatial overlap of bounding boxes but also their shapes and aspect ratios. This research holds practical value for steel surface defect detection and recognition, offering valuable insights into addressing the challenges posed by varied defect scales.

1 Improved YOLOv5 Steel Surface Defect Detection Framework

1.1 Improved YOLOv5 network

In object detection applications, striking a balance between real-time detection and accuracy is crucial. Current research has adopted the YOLOv5 algorithm for steel defect identification, addressing this need. However, when dealing with steel defects of similar categories and significant variations in scale, improving the YOLOv5 algorithm can further enhance detection accuracy and efficiency.

The proposed YOLOv5 module is shown in Fig. 1. The YOLOv5 algorithm consists of three main components: the backbone, neck, and head. The backbone primarily incorporates CBS (Conv + BN + SiLU), Res2Block, and SPPF (spatial pyramid pooling-fast) modules to extract image feature information. The neck is responsible for fusing the extracted features from different

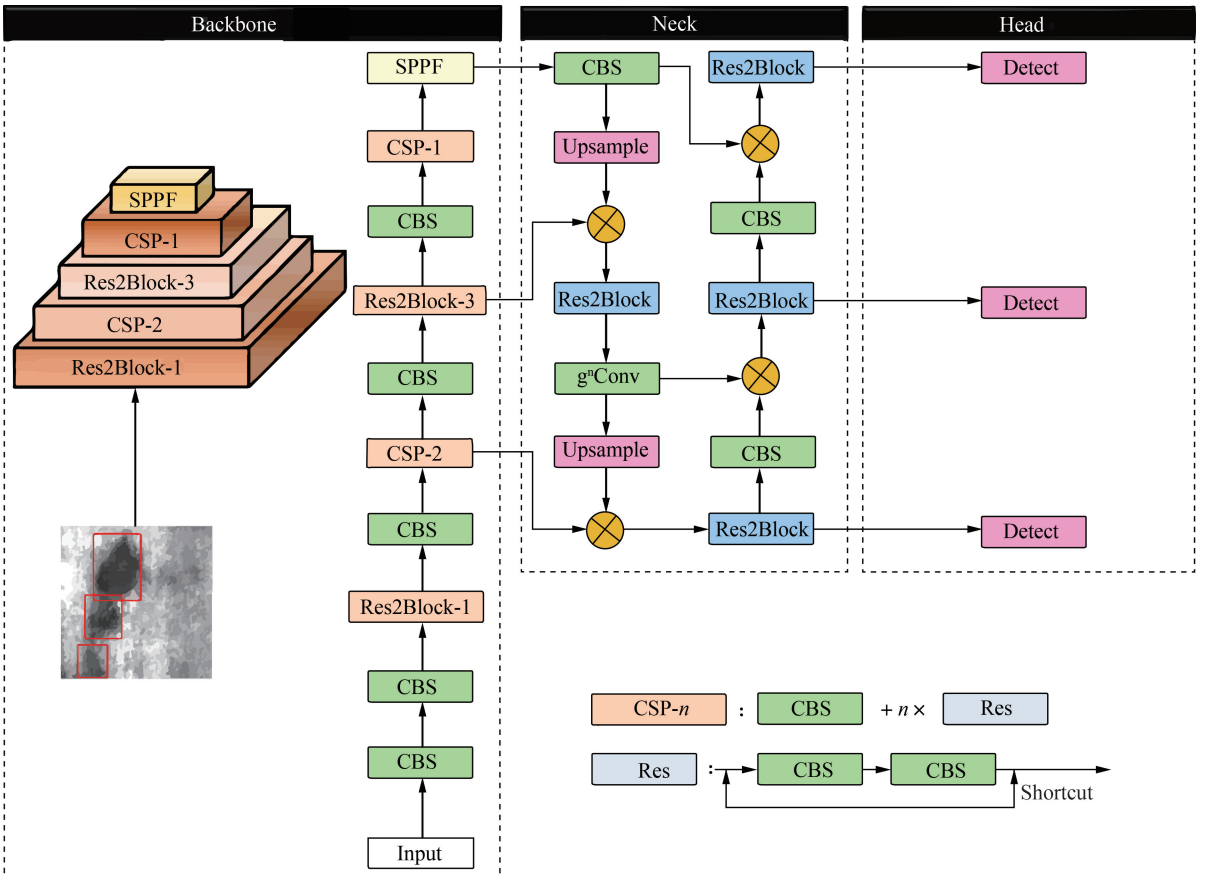


Fig. 1 Schematic of the improved YOLOv5 module with Res2Net in the backbone and $g^n\text{Conv}$ in the neck

layers. In our proposed module, we introduce $g^n\text{Conv}$ to the neck. The head uses three types of loss functions to compute target classification loss, target localization loss, and confidence loss, thereby enhancing network detection accuracy through nonmaximum suppression (NMS). The model accepts a default input image size of 640×640 RGB images, and the final output format is $3 \times (5 + n_{\text{cls}})$, where n_{cls} represents the number of object detection classes.

1.2 Res2Net added to the backbone

To improve the feature extraction capability, we introduce the Res2Net^[16] module to further extract features from the downsampled feature maps transmitted through the skip connection layer.

Fig. 2 shows the structure of ResNet and Res2Net. As shown in Fig. 2(a), ResNet^[17] divides the original network features into two parts using convolutional operations. Both structural blocks have the same input features, with one block remaining untouched by convolution and the other undergoing the operation. The feature that did not undergo convolution is then merged with the output feature of the convolutional operation.

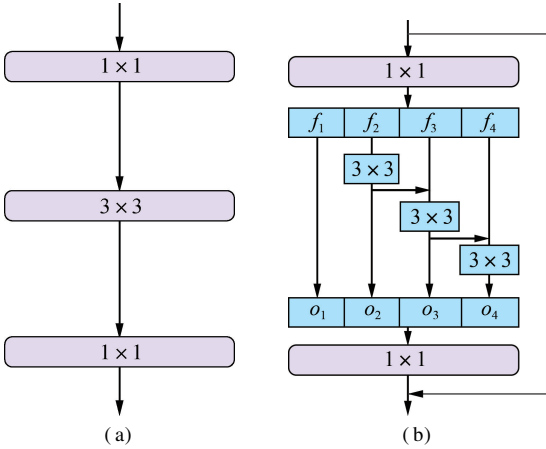


Fig. 2 Structure of ResNet and Res2Net. (a) ResNet; (b) Res2Net

Res2Net improves upon the traditional ResNet by constructing hierarchical residual connections within a residual block, thereby expanding the receptive field of each network layer. As shown in Fig. 2(b), the Res2Net network structure performs grouping first, followed by residual operations. The process involves grouping the input features, with one group of filters performing convolutional operations to extract the features of the input information. The obtained features and another group of features prepared for input are then combined and inputted to the next filter. This process repeats until all input features are fully processed. Upon completion, the operation ends, and the feature maps are connected. These connected feature maps pass through a 1×1 filter to fuse all features. During the feature propagation process, the

input features can be transformed in any path. When passing through a 3×3 filter that is the same as the previous convolution, the receptive field expands owing to the convolutional operation. The calculation for this operation is represented as

$$o_i = \begin{cases} f_i & i = 1 \\ K_i(f_i + o_{i-1}) & 1 < i < n \end{cases} \quad (1)$$

where $K_i(\cdot)$ denotes the convolutional operation performed on the feature. Starting from the calculation of the second group of features, each $K_i(\cdot)$ calculation takes as input the concatenation of the residuals o_{i-1} from the previous group and the features f_i of the current group. Ultimately, all the multiscale features o_i obtained are concatenated and input into the next convolutional layer, resulting in the output of the Res2Net module.

Res2Net replaces the original filter with a set of smaller filter banks in the convolutional part, which expands the receptive field of the model and efficiently achieves multiscale feature extraction. To obtain more fine-grained image features in surface defect detection of strip steel, Res2Net is adopted in the current research to improve the YOLOv5 structure.

1.3 $g^n\text{Conv}$ added to the neck

$g^n\text{Conv}$, proposed by Rao et al.^[18], utilizes gated convolution to enable spatial interactions and operates recursively to facilitate interactions between pixel features of arbitrary input and context features of arbitrary order. Fig. 3 illustrates three recursive operations of this gated convolution. In the diagram, C represents the number of channels, Proj represents the linear projection operation, DWConv stands for depth-wise separable convolution, and Mul represents element-wise multiplication.

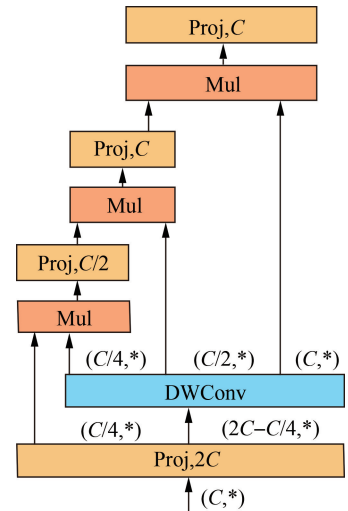


Fig. 3 $g^n\text{Conv}$ structure

Let $x \in \mathbf{R}^{H \times W \times 2C}$ represent the input feature. The gate convolution operates on this input feature, adjusting its channel numbers through linear projection. Subsequent-

ly, the output feature is divided into two parts, p_0 and q_0 , as shown in the following equation:

$$[p_0^{HW \times C/4}, q_0^{HW \times (2C - C/4)}] = \varphi_{in}(x) \in \mathbf{R}^{HW \times 2C} \quad (2)$$

where φ_{in} is the linear projection layer for executing channel mixing. Note $p_1^{(i,c)} = \sum_{j \in \Omega_i} w_{i-j}^c q_0^{(j,c)} p_0^{(i,c)}$. Ω_i represents the local window centered at i , while w denotes the convolutional weights of depth-wise convolution f . Therefore, Eq. (2) explicitly introduces the interaction between adjacent features $p_0^{(i)}$ and $q_0^{(j)}$.

After processing with deep separable convolution, the features q'_0 with the same number of channels as p_0 are obtained. An element-wise multiplication between p_0 and q'_0 is then performed as shown in the following equation:

$$p_1 = f(q_0) \odot p_0 \in \mathbf{R}^{HW \times C} \quad (3)$$

where f represents the depth-wise convolution.

After integrating the number of channels through linear projection operations, the first-order spatial interaction of features is achieved. The output of the gated convolution, $y = gconv(x)$, can be written as

$$y = \varphi_{out}(p_1) \in \mathbf{R}^{HW \times C} \quad (4)$$

where φ_{out} is the linear projection layer that performs channel mixing.

To further enhance the model capacity, we introduce high-order information interaction. This strategy helps preserve the information extracted by the convolutional layers to a great extent. First, φ_{in} is used to obtain a set of projected features p_0 and $\{q_k\}_{k=0}^{n-1}$.

$$[p_0^{HW \times C_0}, q_0^{HW \times C_0}, \dots, q_{n-1}^{HW \times C_{n-1}}] = \varphi_{in}(x) \in \mathbf{R}^{HW \times (C_0 + \sum_{0 \leq k \leq n-1} C_k)} \quad (5)$$

Then, we execute gated convolution recursively through the following equation:

$$p_{k+1} = \frac{f_k(q_k) \odot g_k(p_k)}{\alpha} \quad k = 0, 1, \dots, n-1 \quad (6)$$

Here, the output is scaled by $1/\alpha$ to stabilize the training process. The sets $\{f_k\}$ and $\{g_k\}$ represent depth-wise convolution layers and dimension matching in different orders, respectively.

$$g_k = \begin{cases} \text{Identity} & k = 0 \\ \text{Linear}(C_{k-1}, C_k) & 1 \leq k \leq n-1 \end{cases} \quad (7)$$

Finally, the output of the last recursive step q_n is fed into the projection layer φ_{out} to obtain the result of g^n Conv. g^n Conv takes a feature map with a channel size of C as input. Following the first convolution layer, this channel count is doubled. The output of the first convolution layer is divided into two parts: the first part is directed to the subsequent layer, while the second part undergoes depth-wise separable convolution, generating three parts as input for the remaining three layers. In the recursive process of g^n Conv, there is a progressive in-

crease in the channel count, scaling from smaller to larger sizes. Thanks to its proficiency in feature extraction across diverse granularity levels and its ability to concurrently reduce computational costs, g^n Conv has been successfully applied in steel defect detection^[19] and railway panoramic segmentation^[20].

1.4 Nonmaximum suppression method

In object detection, YOLOv5s employs an NMS method to eliminate redundant candidate bounding boxes. The NMS mechanism in YOLOv5s ranks candidate boxes according to confidence scores, using intersection over union (IoU) as the evaluation metric for selection. IoU, which assesses the accuracy of object localization, is depicted in Fig. 4 (a) and calculated by the following equation:

$$L_{IoU} = \frac{A \cap B}{A \cup B} \quad (8)$$

where L_{IoU} is a measure of the overlap between two bounding boxes.

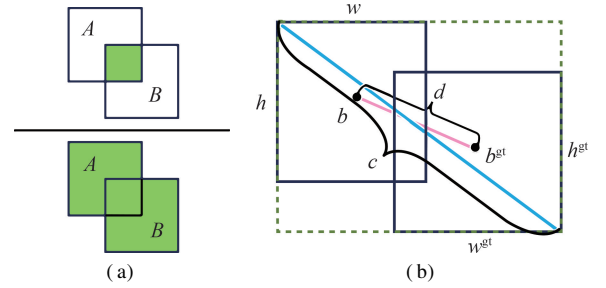


Fig. 4 Schematic diagram of IoU and CIoU. (a) IoU; (b) CIoU

IoU may not precisely compute the loss function for bounding box regression, particularly in scenarios involving overlapping or nested boxes, leading to slow convergence and less accurate results. To address this shortfall, the CIoU loss^[21] has been introduced. It demonstrates faster convergence and superior performance compared with the original IoU. This loss function optimizes the overlapping area, center point distance, and aspect ratio, thereby ensuring more stable regression for target boxes. Fig. 4 (b) provides a schematic representation of the CIoU, with calculation equations detailed in the following:

$$L_{CIoU} = 1 - L_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (9)$$

$$\alpha = \frac{v}{1 - L_{IoU} + v} \quad (10)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (11)$$

where $\rho^2(b, b^{gt})$ is the Euclidean distance between the center point of the prediction frame and the real frame, α is a trade-off parameter, v represents the width-to-height ratio measurement function, while w and h are the height and width of the prediction box, respectively. Similarly,

w^{gt} and h^{gt} are the height and width of the real box, respectively. Lastly, c signifies the diagonal distance that can simultaneously contain the minimum closure area of both the prediction box and the real box.

The CIoU loss function considers the overlapping area, center point distance, and aspect ratio. Directly minimizing the distance between the detection box and the annotated box's center points enhances the reasonability and effectiveness of results obtained from NMS. This leads to more accurate model predictions. Consequently, the CIoU loss function is adopted in the current research.

1.5 Evaluation of performance metrics

1) Precision (P) represents the ratio of accurately predicted defective samples to the total predicted defective samples. The equation for calculating P is the following:

$$P = \frac{T_p}{T_p + F_p} \quad (12)$$

where T_p represents the true positive, which indicates the scenario where both the actual detection and the sample identify as a defect. Conversely, F_p represents the false positive, which means that the actual detection identifies a defect, but the sample is not labeled as a defect.

2) Recall (R) refers to the ratio of correctly predicted positive examples to the total number of actual positive examples in the dataset. The calculation equation for R is the following:

$$R = \frac{T_p}{T_p + F_N} \quad (13)$$

where F_N represents the false negative, which means that the actual detection is not a defect, but the sample is labeled as a defect.

3) For a specific target type, the area under the P - R curve is called the average precision (AP), representing the model's testing accuracy. By calculating the mean of

the APs across all target types, we derive the mean average precision (mAP), which is the primary performance metric for evaluating a models' detection accuracy across the entire detection data set.

4) F_1 -score is the harmonic mean of precision and recall. It ranges from 0 to 1, with 1 representing the model's best output and 0 representing the worst. The equation for F_1 -score is the following:

$$F_1 = \frac{2PR}{P + R} \quad (14)$$

5) IoU represents the overlap ratio between the predicted bounding box and the ground truth bounding box labeled on the original image. A perfect prediction result, where the predicted bounding box perfectly matches the ground truth, would yield an IoU ratio of 1.

2 Experimental Results and Discussion

2.1 Data source and image augmentation

The Northeastern University (NEU) surface defect data set is a publicly accessible resource that contains 1 800 defect images, each with a resolution of 200×200 ^[7]. Each defect image comes with annotated bounding box coordinates and its corresponding category label, marked against either a single or complex background. These bounding boxes are considered ground truth boxes, totaling nearly 3 900 in the data set. In the experiments, 1 200 images were randomly selected to train or fine-tune the detection network, while the remaining images were used to verify the detection performance.

Owing to the insufficient volume of original data to meet the training requirements of the deep learning network, three offline image augmentation techniques, namely ShiftScaleRotate, Crop, and Contrast, were employed. The offline augmentation results are illustrated in Fig. 5. Following augmentation, the total number of im-

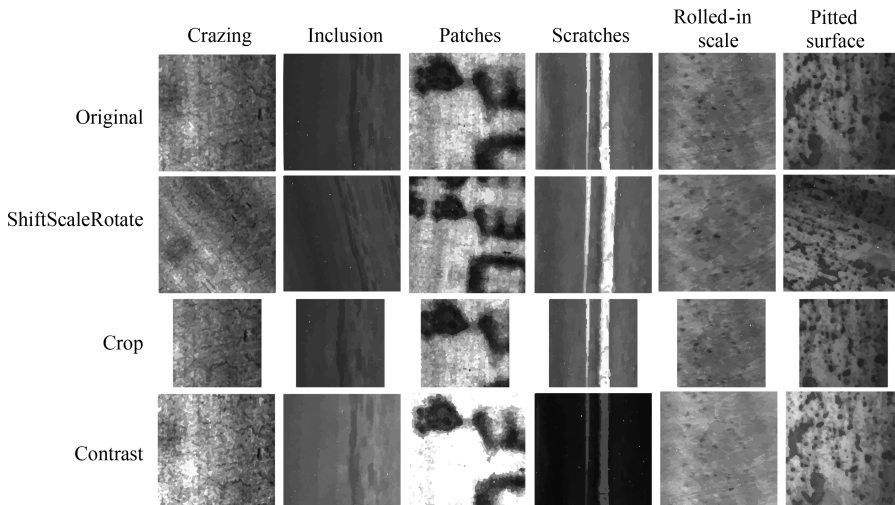


Fig. 5 Different surface defect types and offline image augmentation results

ages increased to 7 200, with 6 480 images allocated for the training set and the remaining 720 images for the test set.

Mosaic, an effective online data augmentation method, was adopted in the training process of YOLOv5. This augmentation strategy, inspired by the CutMix^[22] method, is shown in Fig. 6. First, four images are ran-

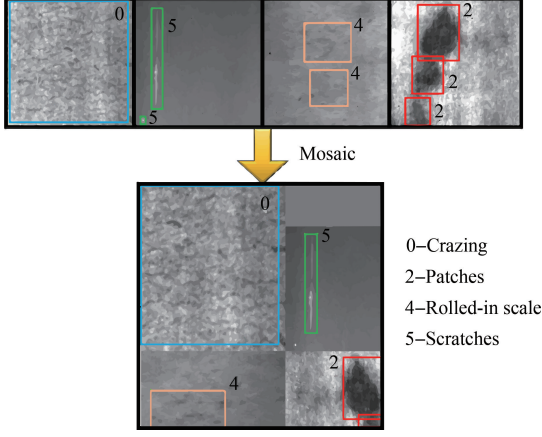
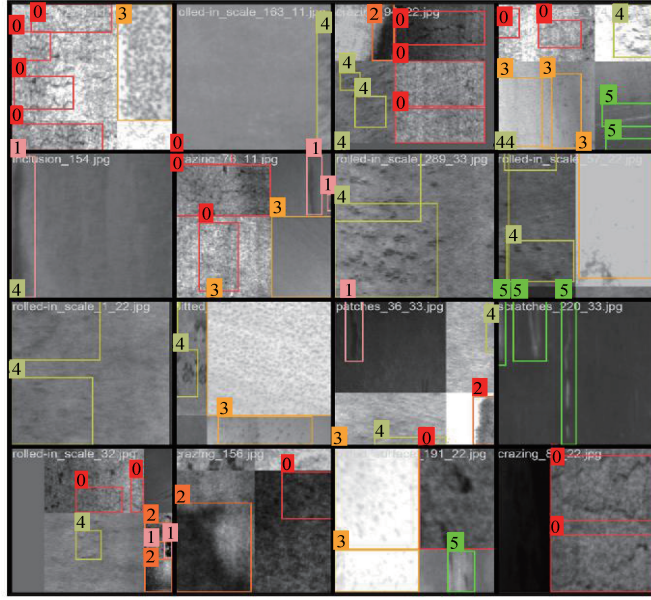


Fig. 6 Data augmentation effect diagram using the mosaic method



0-Crazing; 1-Inclusion; 2-Patches; 3-Pitted surface; 4-Rolled-in scale; 5-Scratches

Fig. 7 Augmentation effect of partial image data during training

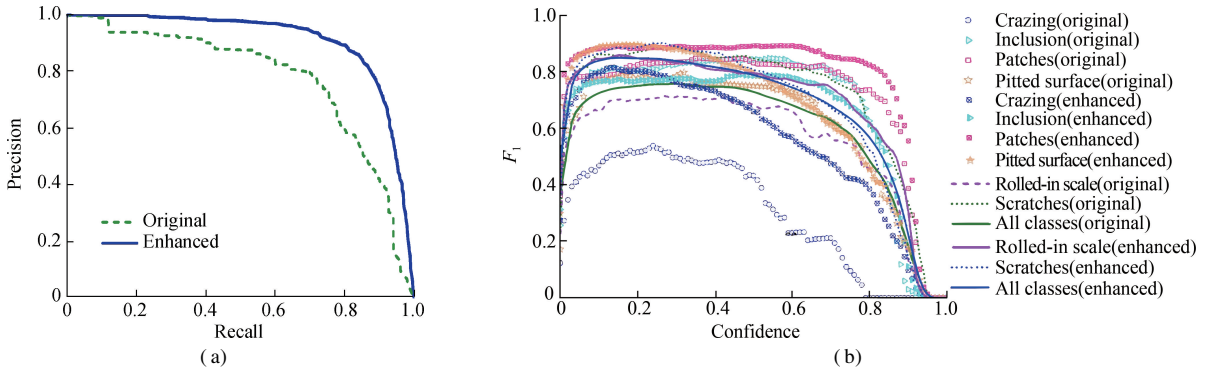


Fig. 8 Comparison of detection results based on original and online enhanced images. (a) P - R curve; (b) F_1 -confidence curve

domly selected from the labeled defect area data set. These chosen images then undergo preliminary data augmentation operations such as random cropping, flipping, scaling, and color space transformation. The enhanced images are then placed on a gray background in the sequence of top left, top right, bottom left, and bottom right. Ultimately, the four images are captured and concatenated into a new image using the matrix method. This process enriches the image background compared to the original data, contributing to an improved batch size.

The online data augmentation effect on partial image data during the training process is shown in Fig. 7. The augmented P - R and F_1 -confidence curves are depicted in Fig. 8. Obviously, the P - R value and F_1 -score of the enhanced results consistently surpass those of the original results. Consequently, the combination of offline and online image augmentation significantly enhances the accuracy of detection results, providing a foundation for subsequent studies on the structure of the detection network.

2.2 Steel surface defect detection using the improved YOLOv5 model

The improved YOLOv5 model is employed for object detection, leveraging the augmented data. First, an ablation study is conducted using various block combinations within the YOLOv5 architecture. Table 1 shows the ablation results. The study reveals that incorporating Res2Block in the backbone and g^n Conv in the neck yields the highest values for P , F_1 -score, and mAP. When compared to the original YOLOv5 architecture, the proposed method, with similar GFLOPs, shows improvements of 5.08%, 1.30%, and 3.35% for P , F_1 -score, and mAP, respectively. In terms of GFLOPs, adding Res2Net to the backbone facilitates a more efficient study by leveraging various residual structures. Meanwhile, including g^n Conv in the neck introduces a computational workload aimed at refining the model's accuracy. The strategic combination of Res2Block in the backbone and g^n Conv in the neck is particularly advantageous. This combination optimally balances the efficient multiscale feature integration provided by g^n Conv with the computational refinement introduced by Res2Block. Importantly, this pairing achieves superior performance without significantly increasing the computational workload.

Table 1 Ablation experiment results

Number	Backbone add-in	Neck add-in	P /%	R /%	F_1 -score/%	mAP/%	GFLOPs
1	YOLOv5		84.7	85.4	84.9	65.6	15.8
2	Res2Block		85.9	85.0	85.2	63.8	15.0
3		g^n Conv	86.5	85.1	85.7	66.6	16.9
4	Res2Block	g^n Conv	89.0	83.3	86.0	67.8	15.9

Fig. 9 provides a detailed comparison of the detection results between the original YOLOv5 algorithm and the

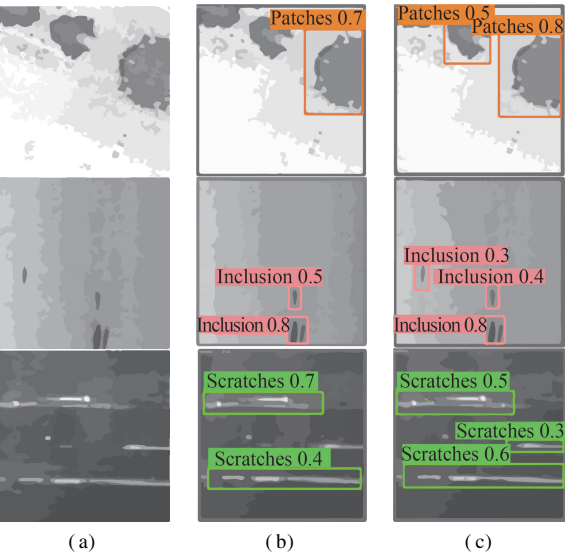


Fig. 9 Comparison of detection results using the original YOLOv5 and the proposed method. (a) Original; (b) YOLOv5; (c) Proposed method

proposed method. Evidently, the proposed method surpasses the original YOLOv5 algorithm, particularly in small-scale defect detection. The combined effect of Res2Net and g^n Conv is instrumental in achieving these advancements. Res2Net aids in expanding the model's receptive field by increasing output features representing scale. In addition, g^n Conv enhances the interaction between model features and the surrounding space, reducing information loss and enhancing feature extraction ability.

Fig. 10 shows the P -confidence curve and the confusion matrix. The P -confidence curve delineates the nuanced changes in precision across diverse decision thresholds. In this graph, the horizontal axis corresponds to the decision threshold, whereas the vertical axis captures the precision values. As shown in Fig. 10 (a), the model achieves an accuracy of 100% when the decision threshold is set at 0.9. This implies that, under this particular decision threshold, the proposed YOLOv5 model can accurately predict all types of defects. The primary diagonal elements in the confusion matrix, presented in Fig. 10 (b), are significantly larger than the non-diagonal elements. This indicates that the model can accurately identify different defect types. Fig. 11 shows the prediction results using the proposed method with labeled boxes. These results show that the proposed method performs well in detecting various defect types in steel plates.

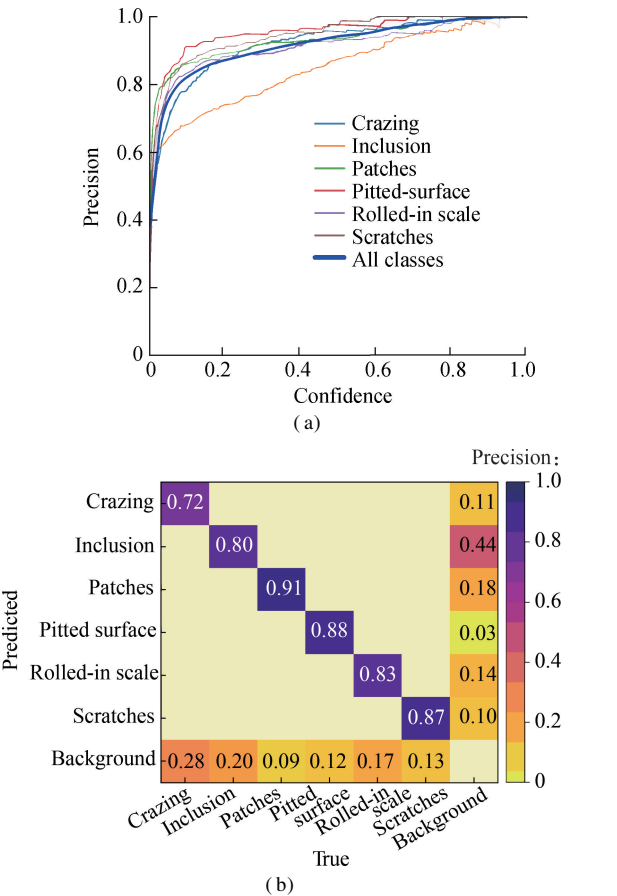


Fig. 10 Prediction performance using the improved YOLOv5 model. (a) P -confidence curve; (b) Confusion matrix

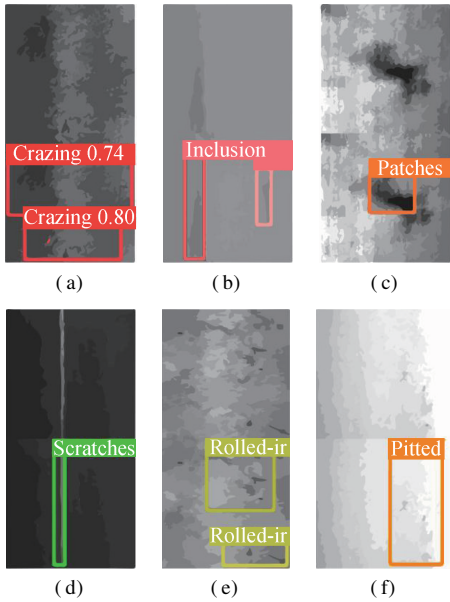


Fig. 11 Detected steel defects using the improved YOLOv5 model. (a) Crazing; (b) Inclusion; (c) Patches; (d) Scratches; (e) Rolled-in scale; (f) Pitted surface

Table 2 Results compared with SOTA defect detectors on the NEU-DET dataset

Model	Backbone	AP/%						mAP/%
		Crazing	Inclusion	Patches	Scratches	Rolled-in scale	Pitted surface	
SSD	VGG-16	46.7	43.5	62.6	37.4	53.2	61.4	50.8
YOLOv3	DarkNet-53	29.7	40.3	62.7	21.0	47.8	38.7	40.0
YOLOv5-Lite	CSPDarknet-53	18.0	41.9	60.3	40.1	36.9	43.8	40.2
YOLOv8s	CSPDarkNet53	66.3	67.2	80.5	81.6	70.7	76.7	73.8
YOLOv8n	CSPDarkNet53	51.1	58.4	74.1	71.1	58.8	70.7	64.0
Ours	CSPDarknet-53 + Res2Block	60.2	59.9	77.3	68.3	67.2	73.8	67.8

3 Conclusions

- 1) A steel surface defect detection method is proposed based on the Res2Net- g^n Conv YOLOv5. This innovative approach enhances the YOLO backbone with Res2Net, improves the neck of YOLO using g^n Conv, and incorporates the CIoU loss function.
- 2) The experimental results demonstrate the superior performance of our proposed method, achieving an mAP of 67.8% and an F_1 -score of 86.0% in steel surface defect identification. With nearly equivalent GFLOPs, it exhibits improvements of 5.08%, 1.30%, and 3.35% for P , F_1 -score, and mAP, respectively, compared to the original YOLOv5 architecture.
- 3) The effectiveness of our proposed method is further validated through a comprehensive comparative study in which several SOTA detection methods serve as benchmarks. The mAP exhibits notable improvements of 33.46%, 69.50%, 68.66%, and 5.94%, respectively, when compared with the results obtained using the SSD, YOLOv3, YOLOv5-Lite, and YOLOv8n methods.
- 4) Looking forward, we plan to explore the integration of attention mechanisms with lightweight frameworks. This initiative aims to address the challenges of precision and efficiency in detecting small-scale defects on steel surfaces. Our goal is to further enhance the multiscale perspective while concurrently improving the com-

To further establish the superior efficacy of our proposed method in surface defect detection, we conducted a comprehensive comparison study. Several state-of-the-art (SOTA) detection methods, including SSD, YOLOv3, YOLOv5-Lite, and YOLOv8, were used as benchmarks, as shown in Table 2. Among these, YOLOv8s exhibits better performance with an mAP of 73.8%, compared to the 67.8% achieved by our proposed method. However, the computational demand, expressed in GFLOPs, is significantly higher at 28.4, compared to the 15.9 of our YOLOv5 model. This discrepancy results in a noticeably larger computational workload. Therefore, we chose YOLOv8n for further comparative analysis. Our proposed method demonstrates superior performance in terms of mAP when compared with SSD and other well-established detectors. Specifically, it shows substantial improvements of 33.46%, 69.50%, 68.66%, and 5.94% when juxtaposed with the results obtained using the SSD, YOLOv3, YOLOv5-Lite, and YOLOv8n, respectively. This significant enhancement underscores the ability of our proposed method to meet the accuracy requirements inherent in surface defect detection tasks more effectively.

putational efficiency.

References

[1] Hu X J, Yang J, Jiang F L, et al. Steel surface defect detection based on self-supervised contrastive representation learning with matching metric[J]. *Applied Soft Computing*, 2023, **145**: 110578. DOI: 10.1016/j.asoc.2023.110578.

[2] Zhang S Y, Zhang Q J, Gu J F, et al. Visual inspection of steel surface defects based on domain adaptation and adaptive convolutional neural network [J]. *Mechanical Systems and Signal Processing*, 2021, **153**: 107541. DOI: 10.1016/j.ymssp.2020.107541.

[3] Roy A M, Bhaduri J. Dense SPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism [J]. *Advanced Engineering Informatics*, 2023, **56**: 102007. DOI: 10.1016/j.aei.2023.102007.

[4] Ni Y H, Lu H, Ji C, et al. Comparative analysis on bridge corrosion damage detection based on semantic segmentation[J]. *Journal of Southeast University (Natural Science Edition)*, 2023, **53**(2): 201 – 209. DOI: 10.3969/j.issn.1001-0505.2023.02.003. (in Chinese)

[5] Gao Y P, Gao L, Li X Y. A hierarchical training-convolutional neural network with feature alignment for steel surface defect recognition[J]. *Robotics and Computer-Integrated Manufacturing*, 2023, **81**: 102507. DOI: 10.1016/j.rcim.2022.102507.

[6] Xing J J, Jia M P. A convolutional neural network-based

method for workpiece surface defect detection[J]. *Measurement*, 2021, **176**: 109185. DOI: 10.1016/j.measurement.2021.109185.

[7] Liu R Q, Huang M, Gao Z M, et al. MSC-DNet: An efficient detector with multi-scale context for defect detection on strip steel surface[J]. *Measurement*, 2023, **209**: 112467. DOI: 10.1016/j.measurement.2023.112467.

[8] Zhao W D, Chen F, Huang H C, et al. A new steel defect detection algorithm based on deep learning[J]. *Computational Intelligence and Neuroscience*, 2021, **2021**: 5592878. DOI: 10.1155/2021/5592878.

[9] Wang S, Xia X J, Ye L Q, et al. Automatic detection and classification of steel surface defect using deep convolutional neural networks[J]. *Metals*, 2021, **11**(3): 388. DOI: 10.3390/met11030388.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, 2016: 779 – 788. DOI: 10.1109/CVPR.2016.91.

[11] Yin Z W, Shao J Y, Zhang N. YOLO-DAW: Object detection model based on dual attention mechanism within windows[J]. *Journal of Southeast University (Natural Science Edition)*, 2023, **53**(4): 718 – 724. DOI: 10.3969/j.issn.1001-0505.2023.04.019. (in Chinese)

[12] Yuan T, Zhao X, Liu R, et al. Speed prediction model at urban intersections considering traffic participants[J]. *Journal of Southeast University (Natural Science Edition)*, 2023, **53**(2): 326 – 333. DOI: 10.3969/j.issn.1001-0505.2023.02.016. (in Chinese)

[13] Zhao C, Shu X, Yan X, et al. RDD-YOLO: A modified YOLO for detection of steel surface defects[J]. *Measurement*, 2023, **214**: 112776. DOI: 10.1016/j.measurement.2023.112776.

[14] Guo Z X, Wang C S, Yang G, et al. MSFT-YOLO: Improved YOLOv5 based on transformer for detecting defects of steel surface[J]. *Sensors*, 2022, **22**(9): 3467. DOI: 10.3390/s22093467.

[15] Liao D H, Cui Z H, Zhu Z X, et al. A nondestructive recognition and classification method for detecting surface defects of Si₃N₄ bearing balls based on an optimized convolutional neural network[J]. *Optical Materials*, 2023, **136**: 113401. DOI: 10.1016/j.optmat.2022.113401.

[16] Gao S H, Cheng M M, Zhao K, et al. Res2Net: A new multi-scale backbone architecture[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, **43**(2): 652 – 662. DOI: 10.1109/TPAMI.2019.2938758.

[17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, 2016: 770 – 778. DOI: 10.1109/CVPR.2016.90.

[18] Rao Y M, Zhao W L, Tang Y S, et al. HorNet: Efficient high-order spatial interactions with recursive gated convolutions [EB/OL]. (2022-07-28) [2023-03-18]. <http://arxiv.org/abs/2207.14284>. pdf.

[19] Zhou X, Hao W J, Bian C G, et al. Detection method for welding defects of YOLOv5 steel pipe based on gⁿConv and GAM[J]. *Microelectronics & Computer*, 2023, **40**(9): 29 – 37. DOI: 10.19304/J. ISSN1000-7180.2022.0778. (in Chinese)

[20] Chen Y, Zhou F C, Zhang J J, et al. Railway panoramic segmentation based on recursive gating enhancement and pyramid prediction[J/OL]. (2023-10-07) [2023-11-11]. *Journal of Beijing University of Aeronautics and Astronautics*. <https://bhxb.buaa.edu.cn/bhzk/en/article/doi/10.13700/j.bh.1001-5965.2023.0492>. DOI: 10.13700/j.bh.1001-5965.2023.0492. (in Chinese)

[21] Zheng Z H, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, **34**(7): 12993 – 13000. DOI: 10.1609/aaai.v34i07.6999.

[22] Yun S, Han D, Chun S, et al. CutMix: Regularization strategy to train strong classifiers with localizable features [C]//2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, South Korea, 2019: 6022 – 6031. DOI: 10.1109/ICCV.2019.00612.

基于改进 YOLOv5 的钢材表面缺陷识别方法

王 硕¹ 张燎军² 尹国江³

(¹河海大学土木与交通学院,南京 210098)

(²河海大学水利水电学院,南京 210098)

(³河海大学水利部水工金属结构安全检测中心,南京 210098)

摘要: 由于传统的机器视觉检测方法在小尺度钢材表面缺陷识别中存在检测精度较差的问题,提出了一种基于改进 YOLOv5 算法的钢材表面缺陷无损识别方法. 将 Res2Block 模块应用于 YOLOv5 算法的骨干,在扩大感受野的同时提高计算效率;在 YOLOv5 算法的颈部融合 gⁿConv 结构,以提高表面缺陷识别方法的计算性能. 为验证所提方法的有效性,进行了不同模块组合的消融试验,并与其他目标检测方法进行了对比. 结果表明:所提方法在钢材表面缺陷识别中实现了 67.8% 的 mAP 和 86.0% 的 F₁ 值;与原始 YOLOv5 算法相比,所提方法在小尺度钢材表面缺陷识别方面表现更为优越;与其他目标检测方法如 SSD、YOLOv3、YOLOv5-Lite、YOLOv8 相比,所提方法的计算精度有明显的提高.

关键词: 钢材; 缺陷检测; 卷积神经网络; YOLO

中图分类号: TP391.41; TP18; TG142