scientific reports

Check for updates

OPEN

RSTD-YOLOv7: a steel surface defect detection based on improved YOLOv7

Hongru Song[1,2]✉

Steel surface defect detection is one of the important applications of object detection technology in industry, which can accurately detect surface defects and improve the quality of products. To address the issues of low detection accuracy caused by less area, small scale and similarity between defects and background of steel surface defects. We proposes a RSTD-YOLOv7 method based on YOLOv7 for steel surface defect detection. First, the RFBVGG module and SimAM attention mechanism are integrated into the YOLOv7 backbone network to expand the receptive field, reduce the loss of texture information, and enhance the target feature extraction ability of the model. Second, the STRVGG module, constructed using the Swin Transformer, is incorporated into the neck network. This enhancement improves the extraction ability to capture deep information concealed within the feature maps, reduces feature loss, and improve the ability of feature detection. Then, an improved DSDH detector head is employed to elevate the model's detection precision and network convergence speed. Finally, comparative experiments are conducted on the NEU-DET and GC10-DET datasets. The results show that our proposed method attains the highest detection accuracy, achieving an mAP of 79.3% and 73.2% respectively, compared with the original YOLOv7 model, the mAP increased by 15.9% and 9.6% respectively, the parameters were reduced by 11.3 M and 11.5 M, respectively, the FPS increased by 15.7% and 11.5%, respectively. These results show that our proposed model excels in detection accuracy and speed, exhibiting remarkable generalization capabilities.

Keywords Surface defect detection, Feature extraction, YOLOv7, Swin transformer

Steel is an important raw material used in manufacturing, aerospace, defense, and industrial equipment. The actual production process is inevitably affected by many technical factors, resulting in various defects and damage[1]. Surface defects such as crazing, inclusions, patches, pitted surfaces, and rolled-in-scale defects can reduce the strength of steel to different degrees, shorten the service life of steel workpieces, and affect quality. Therefore, the detection of surface defects in steel is highly significant[2]. With the continuous development and application of computer vision, deep learning and other new technologies, have been widely used in steel surface defect detection, and steel surface defect detection has become an important research topic in the field of engineering technology.

The traditional detection method for steel surface defects relies mainly on manual observation and detection, which has high error and missed detection rates, low detection efficiency, and high costs. Therefore, image recognition and detection technology based on deep learning has become the main method of surface defect object detection. The object detection algorithms can be structurally divided into one-stage detection algorithms and two-stage detection algorithms. The representative algorithms of one-stage detection are the Single Shot Neural Networks (SSD)[3] and You Only Look Once (YOLO)[4–9]. One-stage detection algorithms do not require a Region Proposal Network (RPN) stage and directly obtain detection results. This method has fast detection speed and can respond to the real-time detection requirements of modern industrial production lines. Two-stage representative algorithms include the Regon Convolution Neural Networks(R-CNN)[10]. Faster Regon Convolution Neural Networks (Faster R-CNN)[11] and Mask R-CNN[12]. The two-stage detection algorithm first generates candidate regions via a Region Proposal Network (RPN), then makes a classification prediction for the target, and finally, accurately determines the target position. This method has the advantages of accurate target positioning and higher detection accuracy, but the disadvantage is that the speed is slow and cannot meet real-time requirements. Most of the steel surface defect areas account for less of the original image area, the defect scale is small, and the edge feature information is not obvious. It is difficult to obtain differentiated defect features, which leads to inaccurate detection and positioning of the target defect, resulting in a detection effect

[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou 31002, China. [2]College of Electrical Engineering, Tongling University, Tongling 244061, China. ✉email: hongrusong@tlu.edu.cn

that is not obvious and a low detection accuracy. To better solve the existing problems, this article proposes the RSTD-YOLOv7 model based on improved YOLOv7[9]. This model is designed in the following aspects: (1) We propose introducing the RFBVGG and SAM-SPFCSPC modules into the backbone network to preserve more texture information in the image, enhance the feature extraction ability of the model, and further improve the accuracy of object detection by fusing features with the neck network. (2) We propose introducing the STRVGG module, which is designed by combining the Swin Transformer and RepVGG in the neck network, this module can help better capture and utilize information in the feature map, and improve the detection accuracy of defect targets with unclear feature information. (3) We adopt an improved DSDH decouped head to better solve the conflict between classification and regression tasks, thus further improving the detection accuracy and speed of the model. (4) We conducted fusion experiments on the proposed model on the NEU-DET and GC10-DET datasets, and the experimental results revealed that the RSTD-YOLOv7 model achieved good results in defect detection with a small improvement area and unclear feature information, which can further meet the requirements of enterprises for surface defect detection accuracy. The other organizational structure of this article is as follows: Sect. 2 introduces relevant literature on surface defect detection methods in related work. Section 3 describes in detail the new architecture design method of the proposed RSTD-YOLOv7 model. Section 4, we conduct ablation experiments on the proposed new model to verify its effectiveness. Section 5 summarizes the entire text.

Related work

Many scholars have extensively researched the problems of surface defect detection in steel. Li et al.[13] proposed an improved YOLOv4 algorithm, which includes a Convolution Block Attention Module (CBAM) and Receptive Field Block (RFB) for strip surface defect detection, and improved the detection accuracy of defects with unclear edge feature information, but lacked the ability to detect and analyze small area defects. Wu et al.[14] proposed a surface defect detection method called SDD-YOLO, which is based on YOLOv5s. In this method, the convolution GhostNet Hybrid Module (CGH) and multiconvolution feature fusion block (MCFF) are designed to improve the efficiency of feature extraction, and carafe interpolation is used to replace bilinear interpolation upsampling to improve the extraction of image features and improve the detection accuracy. However, the improvement in defect detection is less pronounced when the edge feature information is not obvious. However, the detection effect for defects with unclear edge feature information is not obvious. Guo et al.[15] proposed the MSFT-YOLO model, which is a detection method in which a Transformer module is added to the YOLOv5 backbone and detection head. This method combines feature extraction with global information to expand the receptive field and improves the detection performance for defects that are easily confused and have small scales. However, it lacks analysis for defects with unclear edge feature information. Zhao et al.[16] proposed a steel surface defect detection model based on improved YOLOv5, which uses Res2Net blocks, Dual Feature Pyramid Networks (DFPN), and decoupled heads to improve detection accuracy. This method has an obvious effect on defect detection with unclear edge feature information, but there is still a problem of unclear improvement effects on defect detection with small areas and small defect sizes. Peng et al.[17] used YOLOv5s as the benchmark model, which was combined with the Swin Transformer structure and the improved model of the Global Attention Mechanism (GAM) to effectively improve the detection accuracy. This method improves the ability to detect defects with smaller areas, but there is a problem of not significantly improving the detection of defects with unclear edge feature information. Zhang et al.[18] proposed an improved YOLOv5 fusion attention mechanism surface defect detection algorithm, which improves the feature extraction ability of the model by introducing dual spatial and channel attention mechanisms in the backbone network. Then, deformable processing is added to the receptive field of the detector head to expand the receptive field of the network and improve the accuracy of network detection. The algorithm lacks analysis of the defect of unclear edge feature information. Su et al.[19] designed a lightweight YOLOv5s strip surface defect detection method, which uses RepGhost with multibranch structure to replace some of the convolutional layers in the backbone network, enhances the ability of the backbone to extract feature information, and utilizes the GG-FPN network to improve the accuracy and speed of the model. However, this method lacks analysis for defects such as small scale and unclear edge information. Cui et al.[20] proposed combining the improved MCB module of CBAM and replacing the four head ASFF with YOLOv8 for the steel surface defect detection algorithm, and achieved good detection accuracy, but this method lacked the specific analysis of defects with small scale and unclear edge information. Xu et al.[21] proposed an efficient detector based on multi-scale information extraction(MSI-YOLO), which uses YOLOv8s as the baseline model. The model uses the channel prior convolution attention mechanism and the lightweight feature fusion network(Slim-neck) to improve the detection accuracy and speed. Lu et al.[22] proposed an improved SS-YOLO surface defect detection method based on the YOLOv7 model, which utilizes a lightweight MobileNetv3 network and D-SimSPPF module to enhance the accuracy and speed of defect detection. Ma et al.[23] proposed a lightweight method based on improved YOLOv8, using GhostNet as the backbone network and Multi Path Coordinate Attention (MPCA) mechanism to reduce the number of model parameters, computational complexity, and enhance feature extraction capabilities, thereby achieving a balance between lightweight and detection accuracy of the model.

The above studies have improved the detection accuracy of steel surface defects to some extent, but problems of low detection accuracy caused by small scale defects and unclear edge feature information still exist, and further research is lacking.

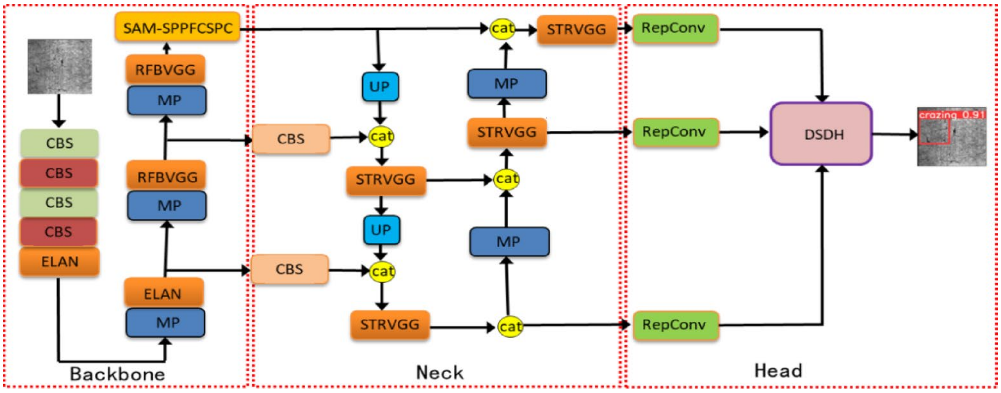


Fig. 1. The architecture of RSTD-YOLOv7.

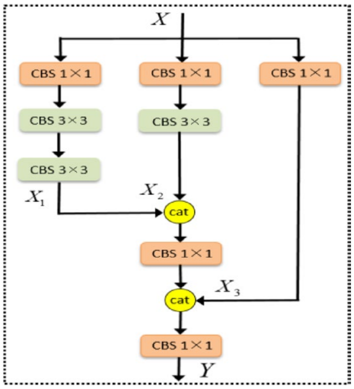


Fig. 2. The architecture of the RFBVGG.

Methodology

Overall structure

To improve the detection accuracy of defects with small scale and unclear edge feature information, we propose the RSTD-YOLOv7 model. The overall architecture is shown in Fig. 1. The main process of the model improvement is as follows: (1) The constructed RFBVGG and SAM-SPPFCSPC modules are introduced into the backbone network, which can retain more texture information in the image, reduce feature loss, enhance the feature extraction ability of the target, and further improve the target positioning accuracy and model detection accuracy. (2) The STRVGG module is introduced to better capture and utilize the semantic information in the feature map, thus improving the detection accuracy. (3) By introducing the DSDH detector, the contradiction between the classification task and regression task can be better solved, and the detection accuracy of the model can be further improved.

RFBVGG module

The RepVGG[24,25] network adopts multibranch structure with convolution kernels of different sizes, which helps the model learn more feature information. Introducing RFB into this network can enhance the feature extraction ability of the network, improve the detection ability of the model for objects of different scales, and achieve faster running speed while improving the detection accuracy. This article introduces RepVGG and RFB network structures for design and proposes a new structure called RFBVGG. The main function of this module is to preserve more texture information and enhance the feature extraction ability of the target. The structure of the RFBVGG module is shown in Fig. 2. The module is composed of three branches. One branch is first compressed by a 1 × 1 convolution to reduce the parameters required for the subsequent 3 × 3 convolution, and then, the features are extracted by two 3 × 3 convolutions, The other branch is first compressed by a 1 × 1 convolution and then extracted by a 3 × 3 convolution, and the two branches are fused. After 1 × 1 convolution, the fusion result is fused with the third 1 × 1 convolution branch to further extract image features. Finally, the fused feature results use a 1 × 1 convolution to adjust the number of channels and output to the final target prediction.The mathematical expressions of the RFBVGG module can be written as follows:

$$X_1 = f_{conv}^{3\times 3}\left\{f_{conv}^{3\times 3}\left[f_{conv}^{1\times 1}(X)\right]\right\} \tag{1}$$

$$X_2 = f_{conv}^{3\times3}\left\{f_{conv}^{1\times1}\left(X\right)\right\} \tag{2}$$

$$X_3 = Cat\left(X_1, X_2\right) \tag{3}$$

$$Y = f_{conv}^{1\times1}\left\{Cat\left[f_{conv}^{1\times1}\left(X\right), f_{conv}^{1\times1}\left(X_3\right)\right]\right\} \tag{4}$$

.

where $f_{conv}^{1\times1}$ and $f_{conv}^{3\times3}$ represent the standard convolution operations with kernel sizes of 1 × 1 and 3 × 3, respectively. Cat(·) is the feature map concatenation operation. X denotes the proceeding input features, Y denotes the proceeding output features.

STRVGG module

The Swin Transformer[26] module is an improved algorithm of the deep learning algorithm Transformer[27]. It not only maintains the displacement invariance and phased resolution reduction characteristics of convolutional neural networks, but also improves the detection speed of the network. Therefore, the introduction of the Swin Transformer module into the network can enhance the feature network's ability to extract deep feature information from images. To better solve the problems of small surface defect features and unclear edge feature information in steel, which makes it difficult to extract defect features, we design the STRVGG module, which is composed of the RepVGG module combined with the Swin Transformer, among which Swin Transformer module is abbreviated as STRBlock. The module structure is shown in Fig. 3. The module is divided into three parts: the left part uses a 1 × 1 convolution and the STRBlock module to fuse the channel information, which is added and fused with the output results in the middle. In the middle part, 1 × 1 convolution and 3 × 3 convolution are used to fuse the channel information, which is added and fused with the output results on the left and right sides. The right part uses a 1 × 1 convolution and two 3 × 3 convolutions to form the fusion channel information, which is added and fused with the output results on the left and right sides. This module is applied in the head layer to replace the ELEN-H module. By using its sliding window attention mechanism and dynamic partition method, it can better capture and use the semantic information in the feature map, to reduce the loss of features and improve the accuracy of model detection. The mathematical expressions of the STRVGG module can be written as follows:

$$X_1{}' = f_{conv}^{3\times3}\left\{f_{conv}^{3\times3}\left[f_{conv}^{1\times1}\left(X'\right)\right]\right\} \tag{5}$$

$$X_2{}' = f_{conv}^{3\times3}\left\{f_{conv}^{1\times1}\left(X'\right)\right\} \tag{6}$$

$$X_3{}' = S\left\{f_{conv}^{1\times1}\left(X'\right)\right\} \tag{7}$$

$$Y' = f_{conv}^{1\times1}\left\{X_1{}' \oplus X_2{}' \oplus X_3{}'\right\}. \tag{8}$$

where $f_{conv}^{1\times1}$and $f_{conv}^{3\times3}$represent the standard convolution operations with kernel sizes of 1 × 1 and 3 × 3, respectively.$\oplus$ represents the elementwise addition operation of the feature map. S represents the output feature map of the STRBlock module.

SAM-SPPFCSPC module

Because the 3D parameter free SimAM[28] module is based on the theory of visual neuroscience, the design of the energy function is optimized. An energy function has been defined for each neuron, which can effectively generate realistic 3D weights, help the model learn image features without introducing additional parameters or increasing computational complexity, enhance feature extraction capabilities, and thereby improve the detection accuracy of the model. To enhance the ability to extract edge feature information from steel surface defects with small scales and less edge feature information, we introduce the SimAM attention mechanism module

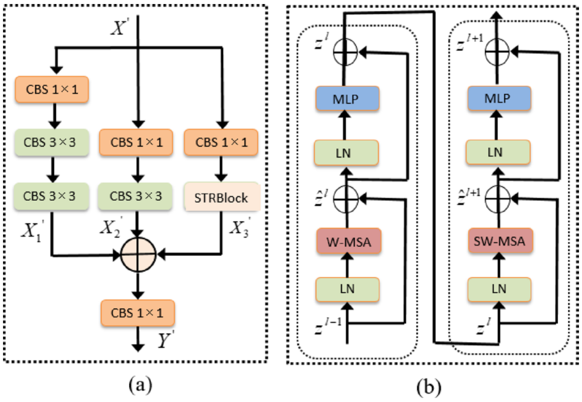


Fig. 3. The architecture of the STRVGG. (**a**) the TRVGG module, and (**b**) the STRBlock module.

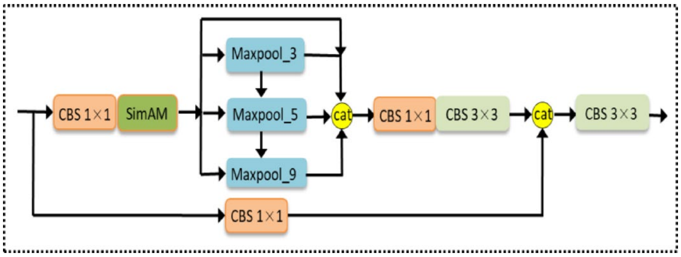


Fig. 4. The architecture of the SAM-SPPFCSPC.

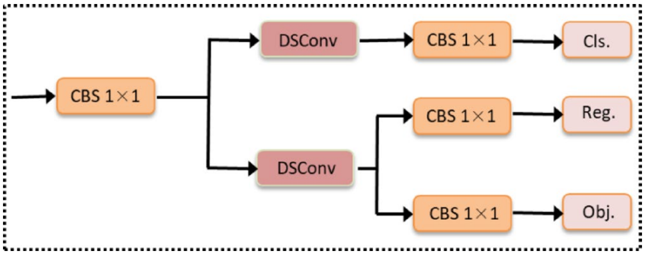


Fig. 5. The architecture of the DSDH detector head.

in the improved SPPFCSPC module to form a new module named SAM-SPPFCSPC. The structure of SAM-SPPFCSPC is shown in Fig. 4, and the specific structure is as follows: (1) The two CBS layers in front of the pooling layer are cut to reduce the filtering of small target edge information by the convolution layer, and only one layer of the CBS smoothing feature is used to reduce number of calculations and the number of parameters of the module. (2) Introduce the attention mechanism SimAM before the pooling layer to increase attention to the target area, reduce effective feature loss, and suppress confusing feature expression. (3) The setting of the three maximum pooling layers is changed from (5, 5, 5) to (3, 5, 9), so that the receptive field of the expansion of the small pooling core is matched with the scale of the small targets, which is conducive to extracting more small target features and further improving the accuracy of small target detection.

DSDH detector head

In the target detection task, the header is used to predict the classification and location information of the target. This paper adopts a new DSDH detection head module combined with a decoupled head and the DSConv convolution method to further improve the detection accuracy of the model and the convergence speed of the network. The DSDH detector head is shown in Fig. 5, which is called the DSDH detector head module. The module first reduces the dimension through a 1 × 1 convolution to reduce the number of parameters, and then connects two branches composed of a DSConv convolution block and a 1 × 1 convolution block to realize classification and positioning respectively. Because of the lightweight characteristics of DSConv convolution, it can reduce the number of calculations and parameters while maintaining high detection accuracy, and improve the speed and efficiency of the target detection model.

Experiments and analysis

Datasets

This paper uses the NEU-DET[29] and GC10-DET[30] datasets to verify the effectiveness of RSTD-YOLOv7. The NEU-DET dataset contains 1800 grayscale images with a resolution of 200 × 200 collected in the production environment, consisting of six types of surface defects, each of which consists of 300 images, including cracks (cr), inclusions (in), patches (p.a.), pitted_surface (ps), rolled in_scale (rs) and scratches (sc). The dataset is randomly partitioned into training, test and validation sets using a ratio of 8:1:1. This division yields a training set of 1440 images, a test set of 180 images, and a validation set of 180 images. Typical defect classification images are shown in Fig. 6, the red box in the figure represents the defective part. Because the area of the cr and rs defects is small, and the edge feature information of ps and sc is not obvious, it is difficult to extract these types of defect features, which leads to difficulty in target defect detection and positioning, making the detection accuracy low and the detection effect not obvious.

The GC10-DET dataset contains a total of 2294 grayscale images with a resolution of 2048 × 1000 collected in the production environment. It is composed of 10 types of surface defects, including punching (pu), welding_line (wl), crescent_gap (cg), water_spot (ws), oil_spot (os), silk_spot (ss), inclusion (in), rolled_pit (rp), crease (cr), and waist_folding (wf) defects. The dataset is randomly partitioned into training, test and validation sets using a ratio of 8:1:1. This division yields a training set of 1834 images, a test set of 230 images, and a validation set of 230 images. Typical defect classification images are shown in Fig. 7, the red box in the figure represents the defective part. It can be seen from the picture characteristics that the areas of the pu, os, in and rp defects are

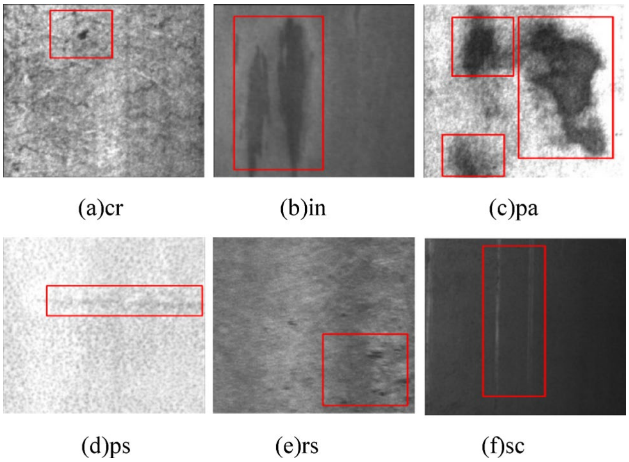


Fig. 6. Examples of different defects in the NEU-DET dataset.

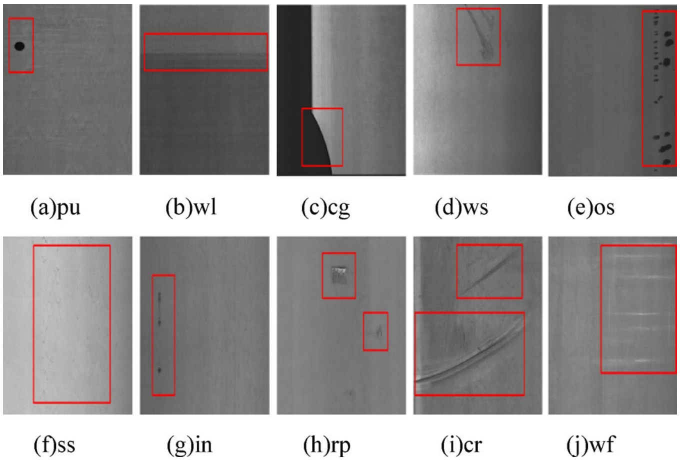


Fig. 7. Examples of different defects in the GC10-DET dataset.

small, and the feature information of the ss and wf defects is not obvious. It will be difficult to extract these types of defect features, which will make it difficult to detect and locate the target defects and make the defect detection effect not obvious, resulting in low overall detection accuracy.

Evaluation index

In this paper, recall (R), precision (P), average precision (AP) and average precision across categories (mAP) are used as evaluation indicators. Where recall represents the ratio of the correctly detected image to all test images of each defect category. Precision represents the average detection accuracy of each defect category; AP represents the prediction accuracy of each individual category; and mAP is the average detection accuracy of all categories. Specifically, as shown in formulas (9) to (12), true potential (TP) represents the number of positive samples predicted by the model and actually positive samples; false positive (FP) indicates the number of predicted positive samples and actual negative samples; and false negative (FN) indicates the number of samples predicted to be negative but actually positive. The total number of samples is N.

$$P=\frac{TP}{TP+FP} \tag{9}$$

$$R=\frac{TP}{TP+FN} \tag{10}$$

$$AP=\int_{0}^{1}P\left(R\right)dR \tag{11}$$

$$mAP=\frac{\sum_{i=1}^{N}AP_{i}}{N}.s \tag{12}$$

Methods	AP(%)						mAP(%)	*P*(%)	*R*(%)	FPS
	cr	in	*p*.a.	ps	rs	sc				
YOLOv7	18.8	71.2	87.2	87.5	35.8	79.9	63.4	56.7	63.6	35.7
mothod 1	38.5	71.9	87.0	84.0	52.6	62.0	66.0	58.9	61.9	40.5
mothod 2	33.9	74.1	90.5	78.3	52.5	88.7	69.7	68.5	71.2	39.2
mothod 3	34.1	73.8	91.3	81.4	53.6	92.9	71.2	66.7	65.5	38.5
mothod 4	49.6	76.3	**92.9**	91.1	65	92.6	77.9	74.4	**72.8**	**42.8**
RSTD-YOLOv7	**51.2**	**78.1**	92.7	**92.4**	**67.2**	**94**	**79.3**	**75.1**	71.5	41.3

Table 1. Detection values on the NEU-DET dataset. Significant values are in bold.

Methods	AP(%)										mAP(%)	*P*(%)	*R*(%)	FPS
	pu	wl	cg	ws	os	ss	in	rp	cr	wf				
YOLOv7	93.8	73.2	96.2	82.8	47.9	58.5	27.4	8.39	66.2	81.9	63.6	67.2	61.3	34.8
mothod 1	93.5	74.6	96.1	83.1	50.2	58.2	**26.8**	25.4	67.8	85.1	66.1	63.5	62.2	35.4
mothod 2	94.6	82.3	**96.7**	84.2	56.7	58.5	25.9	29.4	66.9	86.3	68.2	**68.2**	65.3	35.1
mothod 3	95.8	94.4	96.3	85.3	64.5	58.9	26.4	30.2	67.7	88.6	70.8	70.1	68.1	34.6
mothod 4	97.6	96.3	93.8	**87.1**	67.6	61.5	26.7	32.4	**70.3**	88.7	72.2	65.4	**72.6**	36.9
RSTD-YOLOv7	**98.2**	**96.4**	94.2	84.6	**68.1**	**63.4**	26.1	**41.3**	70.2	**89.4**	**73.2**	67.4	71.8	**38.8**

Table 2. Detection values on the GC10-DET dataset. Significant values are in bold.

Ablation experiment

We use Pytorch deep learning framework to train and test our proposed model. The experimental environment is used with the following configuration: AMD 5800 × 8-core processor with CPU frequency of 3.80 GHz, and Nvidia GeForce RTX3060 GPU with 12 GB memory. All models were trained for 200 epochs using the Adam optimizer with an initial learning rate of 0.0001 and a BatchSize set to 16. To ensure the fairness of the experiments, all models were subject to the same experimental setup.

In this section, ablation experiments are conducted to verify the effectiveness of the proposed improved algorithm. Therefore, in the ablation experiment, the performance of the improved RFBVGG, SAM-SPPFCSPC, STRVGG and decoupled head modules was verified via experiments on the NEU-DET and GC10-DET datasets. To facilitate research and analysis, the improved models are named as follows.(1)YOLOv7with the RFBVGG module is named mothod 1. (2)YOLOv7with the RFBVGG module and SAM-SPPFCSPC module is named mothod 2. (3)YOLOv7 with the RFBVGG module and SAM-SPPFCSPC module is named mothod 3. (4) YOLOv7 with the RFBVGG module SAM-SPPFCSPC module and DSDH is named mothod 4. (5)YOLOv7 with the RFBVGG module SAM-SPPFCSPC module, STRVGG and DSDH is named RSTD-YOLOv7.

NEU-DET experiment

In order to verify the effectiveness of the improved model proposed in this paper, the experimental results on NEU-DET data set are shown in Table 1. It can be seen from the experimental values in Table 1 that the mAP, P and FPS of mothod 1 with RFBVGG module are improved by 2.6%, 2.2% and 4.8 respectively. The improved model obviously improves the detection accuracy of cr and rs defects, and the AP value increases by 19.7% and 16.8% respectively. Compared with the mAP, P and R of mothod 1, the mothod 2 model has improved 3.7%, 9.6% and 9.3% respectively, and further improved the detection accuracy of defects in in, p.a. and sc. Compared with mothod 2, the mAP of RPS-YOLOv7 is increased by 1.5%, and the AP value of ps and sc defect detection is increased by 3.1% and 4.2% respectively. Compared with mothod 3, the mAP of mothod 4 is improved by 6.7%, and the FPS is improved by 3.3%, so that the AP values of cr, rs and ps defects are all improved, which shows that the introduction of decoupling head eliminates the negative influence of coupling head on regression and classification tasks and improves the detection accuracy and speed. The mAP of RSTD-YOLOv7 reached 79.3%, which was 1.4% higher than that of mothod 4, and the AP values of cr, rs, ps and sc defects were further improved.

By comparing the experimental data of the improved model, it can be seen that the RSTD-YOLOv7 model has the best performance in defect detection. Compared with the original YOLOv7, the mAP is increased from 63.4 to 79.3%, the R value is increased from 56.7 to 75.1%, the P value is increased from 63.6 to 71.5%, and the FPS is increased from 35.7 to 41.3, which greatly improves the AP values of cr, ps, rs and sc defects. It is verified that the improved model proposed in this paper has obviously improved the detection effect of surface defects with small area and unclear feature information on NEU-DET dataset.

GC10-DET experiment

To further verify the effectiveness of the RSTD-YOLOv7 model for detecting defects with small areas and unclear edge feature information, we conducted ablation experiments on the GC10-DET dataset. The experimental values are shown in Table 2. The mAP value of YOLOv7 is 63.6%, and the mAP values of the improved mothod

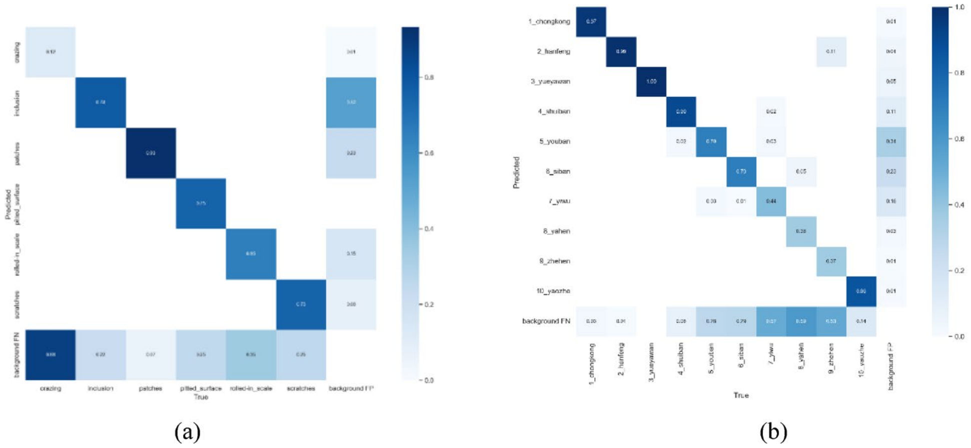


Fig. 8. Confusion matrices produced by our algorithm on different datasets. (**a**) Confusion matrix obtained on the NEU-DET dataset. (**b**) Confusion matrix obtained on the GC10-DET dataset.

Methods	AP(%)						mAP(%)	Params(M)	FPS
	cr	in	*p*.a.	ps	rs	sc			
Zhang's method[18]	**87.5**	75.4	76.2	81.7	65.3	72.2	76.4	-	-
LAACNet[31]	47.1	**88.4**	**92.8**	76.5	59.3	92.2	76.0	-	-
YOLOv5l[8]	45.9	77.2	91.1	87.1	56.3	93.0	75.1	24.6	37.4
YOLOv7[9]	18.8	71.2	87.2	87.5	35.8	79.9	63.4	27.8	35.7
YOLOv8n	46.0	80.9	91.4	89.1	64.6	91.4	77.2	**7.9**	**57.6**
YOLOv9s	54.2	81.0	91.1	90.5	66.1	92.3	79.2	8.3	51.2
YOLOv10s	46.7	76.9	89.6	89.3	60.3	92.4	75.9	9.3	47.8
RSTD-YOLOv7	51.2	78.1	92.7	**92.4**	**67.2**	**94.0**	**79.3**	16.5	41.3

Table 3. Test results of different methods on the NEU-DET dataset. Significant values are in bold.

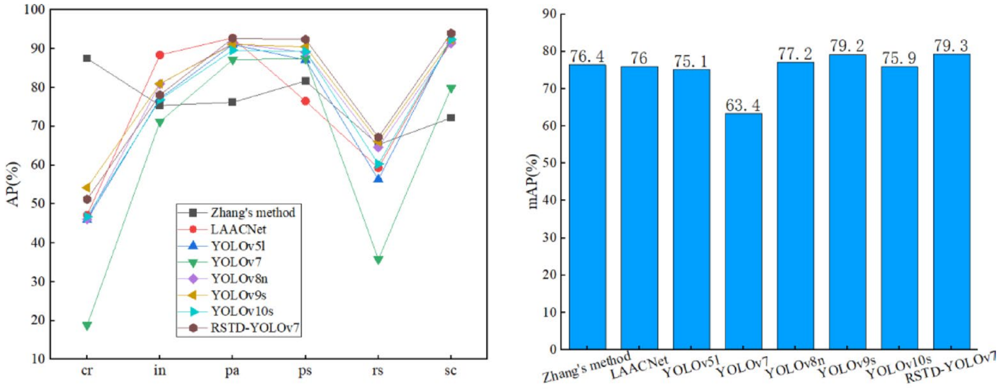
1, mothod 2, mothod 3, mothod 4 and RSTD-YOLOv7 models yield valuse of 66.1%, 68.2%, 70.8%, 72.2%and 73.2%, respectively, which are 2.5%, 4.6%, 7.2%, 8.6% and 9.6% higher than those of the original YOLOv7 model. The mAP value of the improved model is gradually increased. Among the improved models, the AP value of defect detection with small area of pu, os and rp and no obvious characteristic information of ss and wf is improved, and the RSTD-YOLOv7 model has the highest detection accuracy for pu, wl, os, ss, rp and wf defects. Compared with the original YOLOv7, the RSTD-YOLOv7 model improved mAP from 63.6 to 73.2%, P value from 67.2 to 67.4%, R value from 61.3 to 71.8%, the FPS from 34.8 to 38.8, and AP values of pu, wl, os, ss, rp and wf increased by 4.4%, 23.2%, 20.2%, 4.9%, 32.9% and 7.9%, respectively.

The comparison of the experimental results in Table 2, it is shows that the proposed RSTD-YOLOv7 model improves the detection accuracy of defects with less area and less edge feature information on the GC10-DET dataset, and the detection performance has been greatly improved, which further verifies the effectiveness of the improved RSTD-YOLOv7 model.

Comparison with state-of-the-art algorithms

To verify the advancement of the method proposed in this paper, we conducted a comparative experiment with other mainstream detection methods, including Zhang's method[18] LAANet[31] YOLOv5l, YOLOv7, YOLOv8n, YOLOv9s, and YOLOv10s. In order to compare the performance of several algorithms, various evaluation indicators were used in the experiment, including mAP for multi-class defects, FPS and parameters (Params). The confusion matrix generated by the method proposed in this paper on NEU-DET and GC10-DET datasets is shown in Fig. 8. Each element in the confusion matrix reflects the correspondence between the actual category and the predicted category. An analysis of the results obtained for each defect category, reveals that the algorithm proposed in this article result in almost no false detections. This shows that our model has good performance in detection tasks.

The test results of the proposed method on NEU-DET dataset are shown in Table 3. According to the comparison results of Table 3, the RSTD-YOLOv7 method proposed in this paper performs best in terms of detection accuracy, and has the best detection effect on defects with small defect area rs and ps and sc, which edge feature information is not obvious. Compared with the literature 18 and 31 model, mAP increased by 2.9% and 3.3% respectively. Compared with YOLO5l and YOLOv7, mAP and FPS have been improved, mAP has increased by 4.2% and 15.9%, FPS has increased by 3.9 and 5.6, and parameters has decreased by 8.1 M and



(a) (b)

Fig. 9. Comparison of (**a**) various types of AP (%) for six kinds of typical surfaces detected by different methods, and (**b**) detection mAP (%) by different methods.

11.3 M, respectively. Compared with the benchmark models YOLOv8n, YOLOv9s and YOLOv10s, the mAP obtained is the best, but there are also problems that the parameters param and FPS need to be further optimized.

In order to more intuitively reflect the detection results of several methods on the NEU-DET datasets. As shown in Fig. 9 is the comparison of the detection values, and in Fig. 10 is the comparison of the detection effect.

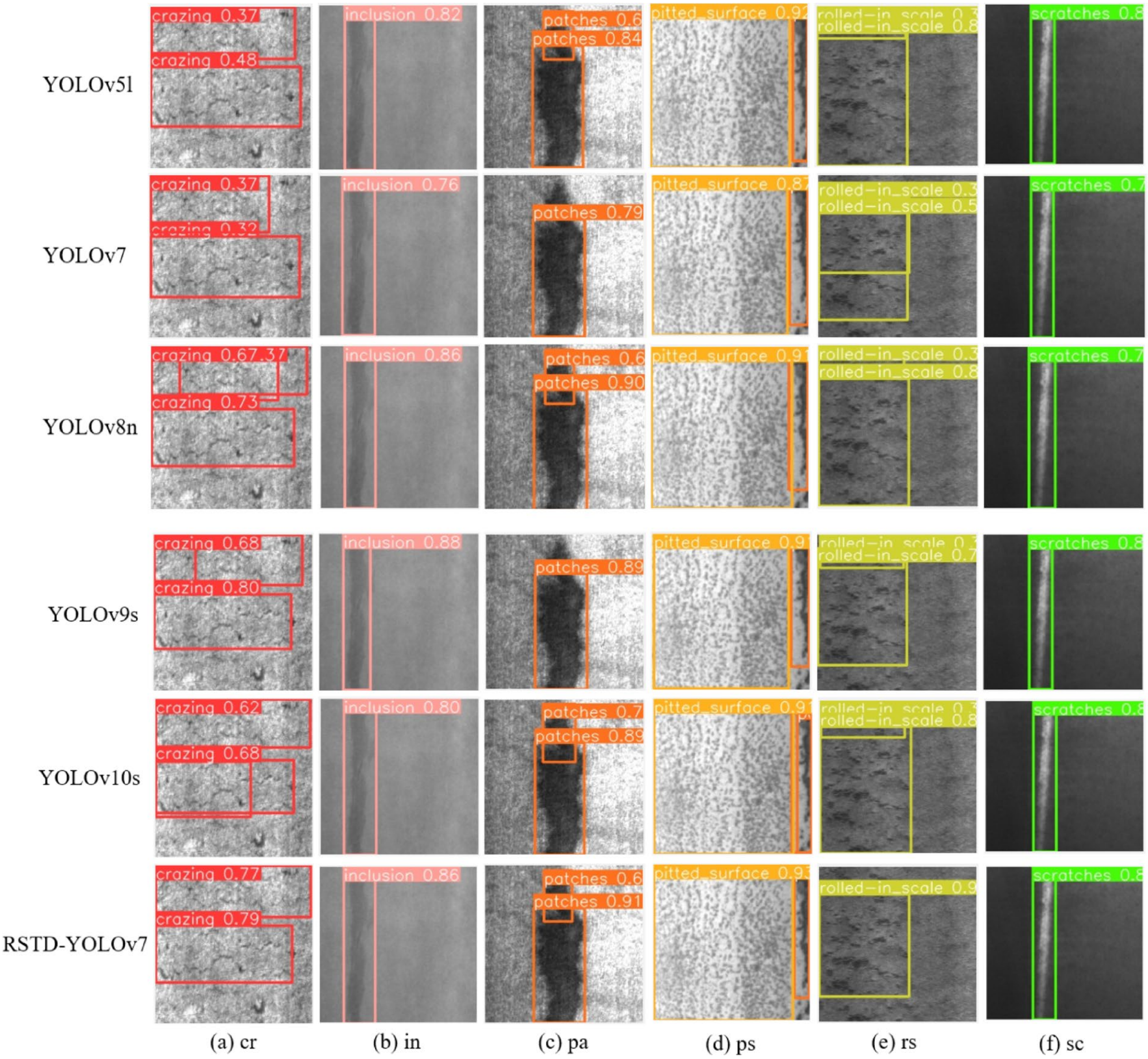
The results of the detection and comparison of the proposed method on the GC10-DET dataset are shown in Table 4. From the comparison results in Table 4, it can be seen that the proposed method mAP is the best, achieving an increase of about 1.7–9.6%, especially for wl, rp with small defect area, and wf with unobvious feature information. Other defects have also achieved good detection results, which further verifies the effectiveness of our method.

The experimental comparison results show that the proposed method not only significantly improves the detection accuracy and speed, but also reduces the number of parameters, and shows better detection effect for defects with small area, small scale and unclear edge feature information, which proves its advancement in surface defect detection. As shown in Fig. 11 is the comparison of the detection values, and in Fig. 12 is the comparison of the detection effect on the GC10-DET datasets.

Conclusions

Aiming at the problems of less area of steel surface defects, small defect size, and unclear edge feature information, which lead to inaccurate detection and positioning, making the detection effect not obvious and the detection accuracy low. In this paper, a surface defect detection method based on RSTD-YOLOv7 is proposed. First, RFBVGG module and SAM-SPPFCSPC module are introduced into the backbone network to expand the receptive field and enhance the feature extraction ability of defect object, so as to improve the detection accuracy of the network. Second, the introduction of STRVGG module can better capture the feature information in the feature map, and improve the feature detection ability of the model. Finally, the DSDH detector head is used to solve the conflict between regression and classification tasks to obtain higher detection accuracy and speed.

The experimental results show that the map of RSTD-YOLOv7 model on NEU-DET and GC10-DET datasets is 79.3% and 73.2%, respectively, which is 15.9% and 9.6% higher than that of YOLOv7 model, the parameters are reduced by 11.3 M and 11.5 M respectively, and the FPS is increased by 15.7% and 11.5% respectively. Especially, the detection accuracy of defects with small area, small defect scale and unclear edge feature information has achieved good results. Compared with the lightweight YOLO model, RSTD-YOLOv7 model has achieved good results in mAP, but it needs further optimization in reducing the amount of parameters and improving the detection speed. In the future research, we will carry out lightweight optimization design for the model, and design the model to reduce the amount of parameters and improve the speed, so as to better adapt to the real-time application in the industrial scene.

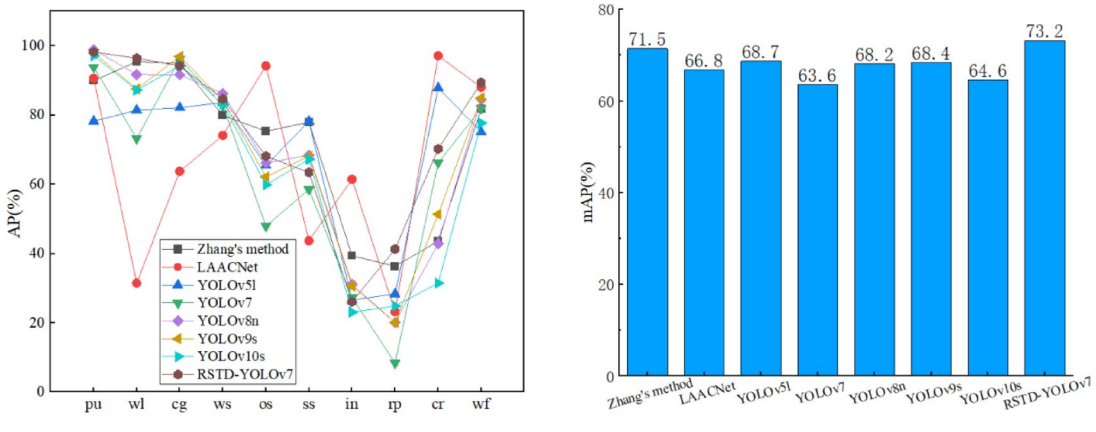


(a) cr (b) in (c) pa (d) ps (e) rs (f) sc

Fig. 10. The detection effects of different methods were compared on NEU-DET dataset.

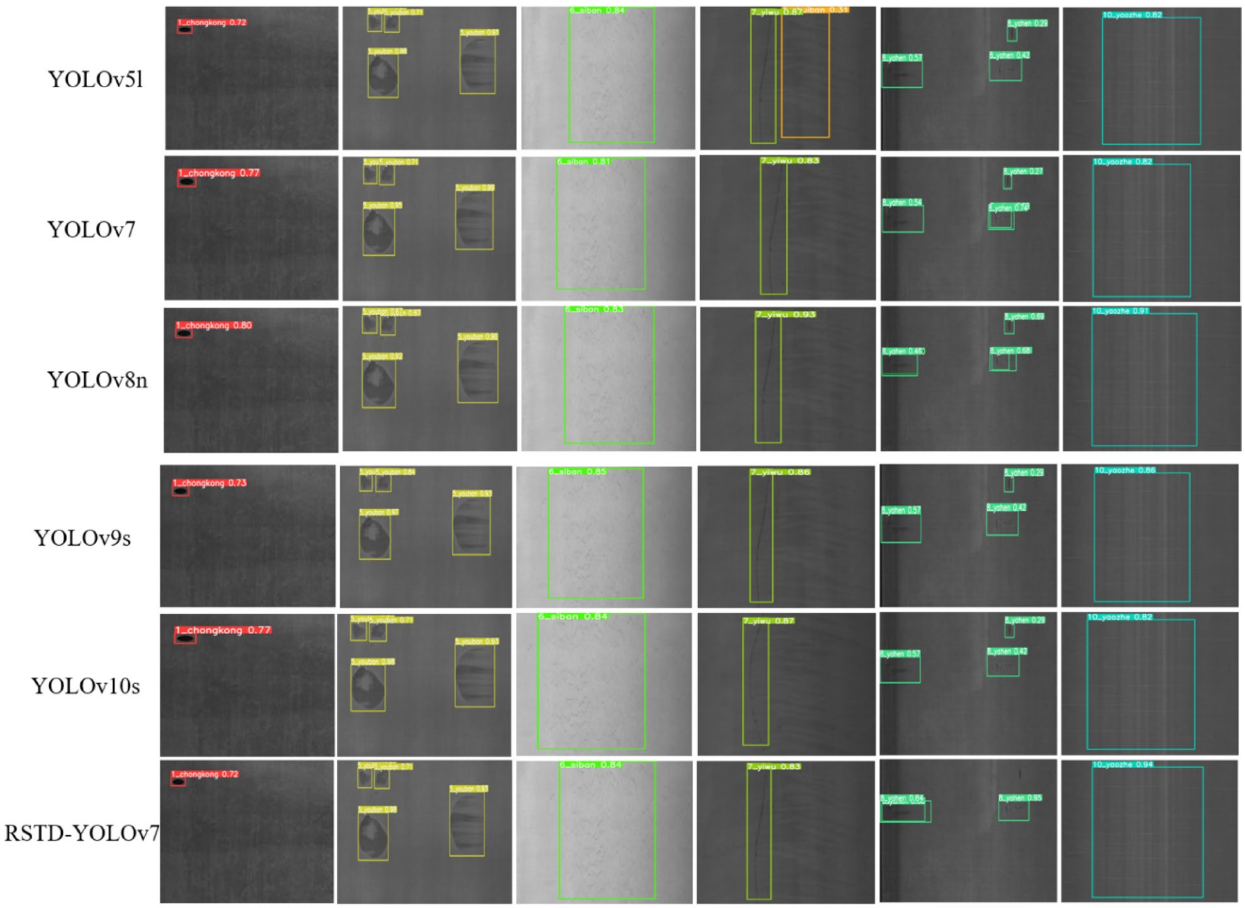
Methods	AP(%)										mAP(%)	Params(M)	FPS
	pu	wl	cg	ws	os	ss	in	rp	cr	wf			
Zhang's method[18]	89.9	95.4	94.9	79.9	75.3	77.9	39.4	36.4	43.6	81.8	71.5	–	–
LAANet[31]	90.6	31.5	63.7	74.1	**94.2**	43.7	**61.4**	23.2	**97.1**	88.0	66.8	–	–
YOLOv5l[8]	78.2	81.4	82.1	83.6	65.4	**78.2**	26.6	28.4	87.8	75.1	68.7	24.8	35.7
YOLOv7[9]	93.8	73.2	96.2	82.8	47.9	58.5	27.4	8.4	66.2	81.9	63.6	28.9	34.8
YOLOv8n	**98.9**	91.7	91.6	**86.2**	66.1	68.4	31.2	20.1	42.9	84.5	68.2	**6.8**	**58.2**
YOLOv9s	98.0	87.5	**96.9**	84.3	62.1	68.1	30.7	20.1	51.3	84.9	68.4	8.2	49.9
YOLOv10s	97.1	87.2	94.2	83.1	59.8	67.2	23.1	24.9	31.5	77.7	64.6	8.9	46.5
RSTD-YOLOv7	98.2	**96.4**	94.2	84.6	68.1	63.4	26.1	**41.3**	70.2	**89.4**	**73.2**	17.4	38.8

Table 4. Test results of different methods on the GC10-DET dataset. Significant values are in bold.



(a) (b)

Fig. 11. Comparison of (**a**) various types of AP (%) for ten kinds of typical surfaces detected by different methods, and (**b**) detection mAP (%) by different methods.



(a) pu (b) os (c) ss (d) in (e) rp (f) wf

Fig. 12. The detection effects of different methods were compared on GC10-DET dataset.

Data availability

The data that support the findings of this study are openly available in [NEU-DET] at http://faculty.neu.edu.cn/songkc/en/zhym/263264/list/index.htm, and [GC10-DET] at https://github.com/lvxiaomin2019/GC10-DET-Metallic-Surface-Defect-Datasets. All data generated or analysed during this study are included in this published article.

Received: 7 April 2025; Accepted: 29 May 2025
Published online: 04 June 2025

References

1. Xu, Y., Li, D., Xie, Q., Wu, Q. & Wang, J. Automatic defect detection and segmentation of tunnel surface using modified mask R-CNN. *Measurement* **178** (4), 109316. https://doi.org/10.1016/j.measurement.2021.109316 (2021).
2. Cheng, X. & Yu, J. RetinaNet with difference channel attention and adaptively Spatial feature fusion for steel surface defect detection. *IEEE Trans. Instrum. Meas.* **70**, 1–11. https://doi.org/10.1109/TIM.2020.3040485 (2021).
3. Liu, W. et al. SSD: single shot multiBox detector. Preprint at http://arxiv.org/pdf/1512.02325v5 (2016).
4. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 779–788 (IEEE, 2016).
5. Redmon, J. & Farhadi, A. Y. O. L. O. Better, Faster, Stronger. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 6517–6525 (IEEE, 2017).
6. Redmon, J. & Farhadi, A.Yolov3: an incremental improvement. Preprint at https://doi.org/10.48550/arXiv.1804.02767 (2018).
7. Bochkovskiy, A., Wang, C. Y. & Liao, H. Y. M. Yolov4: Optimal speed and accuracy of object detection. Preprint at https://doi.org/10.48550/arXiv.2004.10934 (2020).
8. Glenn, J. YOLOv5 release v5.0. https://github.com/ultralytics/yolov5/releases/tag/v5.0 (2021).
9. 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. Preprint at https://doi.org/10.48550/arXiv.2207.02696 (2022).
10. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* 580–587 (IEEE, 2014).
11. Ren, S., He, K., Girshick, R. & Sun, J. Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2017).
12. He, K., Gkioxari, G., Dollár, P. & Girshick, R. *Mask R-C-N-N, IEEE International Conference on Computer Vision (ICCV)* 2980–2988 (IEEE, 2017).
13. Li, M. J., Wang, H. & Wan, Z. B. Surface defect detection of steel strips based on improved YOLOv4. *Comput. Electr. Eng.* **102** (6), 108208. https://doi.org/10.1016/j.compeleceng.2022.108208 (2022).
14. Wu, Y. Y., Chen, R., Li, Z., Ye, M. & &Dai, M. S. D. D. Y. O. L. O. A lightweight, High-Generalization methodology for Real-Time detection of strip surface defects. *Metals* **14** (6), 650. https://doi.org/10.3390/met14060650 (2024).
15. Guo, Z., Wang, C., Yang, G., Huang, Z. & Li, G. M. S. F. T. Y. O. L. O. Improved YOLOv5 based on transformer for detecting defects of steel surface. *Sensors* **22** (9), 3467. https://doi.org/10.3390/s22093467 (2022).
16. Zhao, C., Shu, X., Yan, X., Zuo, X. & Zhu, F. RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement* **214**, 112776. https://doi.org/10.1016/j.measurement.2023.112776 (2023).
17. Peng, Y. F., Yuan, X. L., Chen, Y. K. & Zhao, T. Improved YOLOv5s strip surface defect detection method. *Mech. Sci. Technol. Aerosp. Eng.* https://doi.org/10.13433/j.cnki.1003-8728.20230254 (2023).
18. Zhang, C., Bai, X. & Zhang, X. Metal plate surface defect Dtection method based on attention mechanism. *Comput. Integr. Manuf. Syst.* **03** (2023).
19. Su, Y. Y. & He, Y. P. Lightweight strip surface detection method based on improved YOLOv5s. *J. Optoelectron. Laser*. **35**. https://doi.org/10.16136/j.joel.2024.07.0163 (2024).
20. Cui, K. B. & Jiao, J. Y. Steel surface defect detection algorithm based on MCB-FAH-YOLOv8. *J. Graph.* **45**, 112–125 (2024).
21. Xu, W., Zhang, Y., Jiang, X., Lian, J. & Lin, Y. An efficient steel defect detection model based on multi-scale information extraction. *Robotic Intell. Autom.* **44**, 817–829. https://doi.org/10.1108/RIA-03-2024-0065 (2024).
22. Lu, J. B., Yu, M. M. & Liu, J. Y. Lightweight strip steel defect detection algorithm based on improved YOLOv7. *Sci. Rep.* **14**, 13267. https://doi.org/10.1038/s41598-024-64080-x (2024).
23. Ma, S., Zhao, X., Wan, L. & Zhang, Y. A lightweight algorithm for steel surface defect detection using improved YOLOv8. *Sci. Rep.* **15**, 8966. https://doi.org/10.1038/s41598-025-93469-5 (2025).
24. Goyal, A., Bochkovskiy, A., Deng, J. & Koltun, V. Non-deep networks. Preprint at http://arxiv.org/arxiv:2110.07641v1 (2021).
25. Ding, X. et al. Making VGG-style ConvNets Great Again. Preprint at http://arxiv.org/pdf/2101.03697v3 (2021).
26. Liu, Z., Lin, Y. T., Cao, Y. & Hu, H. Swin Transformer: hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* 9992–10002 (IEEE, 2021).
27. Dosovitskiy, A. & Kolesnikov, B. E. Y. E. R. L. A. An image is worth 16×16 words: Transformers for image recognition at scale. Preprint at http://arxiv.org/pdf/2010.11929v2 (2020).
28. Yang, L., Zhang, R. Y., Li, L., Xie, X. & SimAM A simple, parameter-free attention module for convolutional neural networks. *JInt Conf. Mac Learn.* **139**, 11863–11874 (2021).
29. Song, K. C. & &Yan, Y. H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **285**, 858–864. https://doi.org/10.1016/j.apsusc.2013.09.002 (2013).
30. Lv, X. M., Duan, F. & Jiang, J. J. Deep metallic surface defect detection: the new benchmark and detection network. *Sensors* **20** (1562). https://doi.org/10.3390/s20061562 (2020).
31. Lv, Z. L. et al. Lightweight adaptive activation Convolution network based defect detection on polished metal surfaces. *Eng. Appl. Artif. Intell.* **133**, 108482. https://doi.org/10.1016/j.engappai.2024.108482 (2024).

Author contributions

Conceptualization, editing, methodology and formal analysis, H.S.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.