IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 January 2024, accepted 22 February 2024, date of publication 29 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3371584

RESEARCH ARTICLE

Steel Surface Defect Detection Using Improved Deep Learning Algorithm: ECA-SimSPPF-SIoU-Yolov5

FEI REN[1], (Student Member, IEEE), JIAJIE FEI[2], HONGSHENG LI[2], AND BONIFACIO T. DOMA JR.[1]

[1]School of Information Technology, Mapua University, Manila 999005, Philippines
[2]School of Automation, Nanjing Institute of Technology, Nanjing 211167, China

Corresponding authors: Hongsheng Li (zdhxlhs@njit.edu.cn) and Bonifacio T. Doma Jr. (btdoma@mapua.edu.ph)

This research was funded by the Office of Directed Research for Innovation and Value Enhancement (DRIVE) of Mapua University.

ABSTRACT Steel surface defect detection is an indispensable part of industrial production and processing processes. It helps to reduce production costs, ensure product quality, improve production safety and compliance, and maintain sustainability and competitiveness. To address the low detection accuracy of traditional methods, this paper developed and investigated an improved algorithm based on YOLO for steel surface defect detection: ECA(Efficient Channel Attention) -SimSPPF (Simplified Spatial Pooling - Fast) -SIoU (Scylla Intersection over Union) -Yolov5. First, deformable convolutions were used to replace some conventional convolutions in the model, which expanded the receptive field and improved detection accuracy. Additionally, Efficient Channel Attention was integrated into the model to improve the weight of important information. Then, the SimSPPF was employed in place of the SPP module in the model, reducing computational complexity. Finally, the SIoU loss function was utilized to handle bounding box regression more effectively. The paper conducted different ablation experiments, and the improved ECA-SimSPPF-SIoU-Yolov5 algorithm demonstrated superior detection performance. Using the NEU-DET dataset, the mAP reached 78.8%, which was a 7.1% improvement higher than the original model, while the Recall reached 76.4% and improved by 3.7% compared to the original model. The improved model showed significant improvements in terms of mAP and Recall. Furthermore, the paper conducted multiple comparative experiments, comparing the model with other attention mechanisms and loss functions. The results demonstrated that the improved ECA-SimSPPF-SIoU-Yolov5 algorithm achieved good detection results in terms of mAP and Recall. In the third comparative experiment, the model was compared with YOLOv5 model with different network depths and the latest Yolov8 model, and the improved model also achieved good detection accuracy.

INDEX TERMS Defect detection, deep learning, object detection, convolutional neural network, defect classification.

I. INTRODUCTION

Steel, as a raw material, is highly applied in multiple fields such as aerospace, automotive, and chemical industries. With the development of production technology, the quality of steel has become increasingly important. Surface defect detection of steel is an essential part of steel quality inspection [1]. There are different types of defects on the surface of steel, most of which are caused by multiple factors such as environment and equipment when producing and processing, as shown in Figure 1. The images in Figure 1 are from the NEU-DET dataset [2] created by Northeastern University. The service life and strength of steel will be influenced by these defects, thereby impacting the quality and sales

The associate editor coordinating the review of this manuscript and approving it for publication was Sawyer Duane Campbell.

of subsequent products. Therefore, it is very important to timely and effectively complete the detection of defects on the surface of steel.

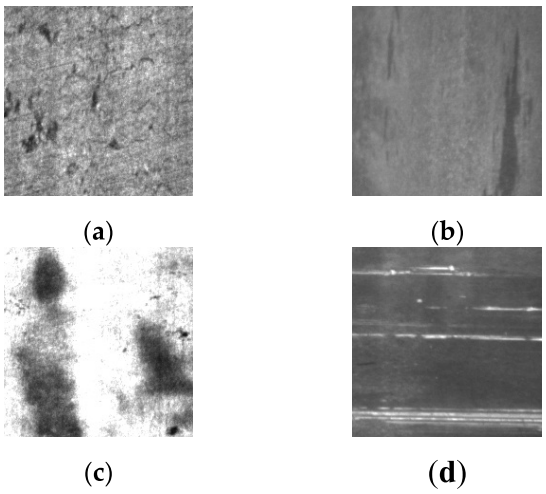

FIGURE 1. Steel surface defect (a) crazing; (b) inclusion; (c) patches; (d) scratches.

The traditional method for detecting surface defects on steel is the stroboscopic method, which is one of the manual inspection methods. However, this method has significant drawbacks such as low efficiency, low reliability, noticeable false negatives, and high labor requirements [3]. Due to the rapid growth of artificial intelligence in recent years, neural network has been utilized to various detection fields, including steel defect detection. The models that use deep learning to complete defect detection are divided into two categories: two-stage detection and one-stage detection. When using the two-stage detection, features are first extracted, followed by the generation of region proposals, and finally, the defect detection is performed. The R-CNN (Regional Convolutional Neural Network) series [4] are the representative algorithms of the two-stage detection. On the other hand, one-stage detection eliminates the generation of region proposals and directly performs detection after feature extraction. The most popular YOLO (You Only Look Once) series algorithms [5] currently are representative algorithms for single stage detection. SSD (Single Shot MultiBox Detector) algorithms [6] also belong to a category of single stage order taking. The former tends to achieve better detection accuracy, while the latter offers faster detection speed. The surface defect detection of steel materials involves various task requirements, such as object detection, instance segmentation, or semantic segmentation. U-Net is commonly employed for semantic segmentation tasks due to its proficiency in learning high-resolution semantic features from input images. When the dataset comprises images with relatively small defect sizes or dense defect distributions, U-Net may demonstrate enhanced effectiveness. However, in the context of the dataset in this study, where defects are larger and sparsely distributed, the YOLO algorithm is deemed more suitable. YOLOv8, the latest object detection model released by Ultralytics, incorporated numerous advanced features and technologies. However, given its recent release, it had not undergone extensive market validation or practical application testing. In contrast, YOLOv5 was a mature and thoroughly validated object detection algorithm that had gained widespread recognition and adoption in both academia and industry. Its stability and reliability had been proven in numerous practical projects. Hence, the paper chose to build upon the foundation of YOLOv5.

To solve the task of detecting surface defects in steel, a modified algorithm based on YOLOv5 was investigated in the study with the aim of improving the average precision of detection tasks. The main work of this study is as follows:

(1) Improvement of the detection accuracy of irregular defects. Some convolutions in the YOLOv5 model were replaced with deformable convolutions. This increases the receptive field during sampling, allowing for a detection that was closer to the actual shape of irregular defects and thus improving the detection accuracy.

(2) By incorporating the ECA [7] (Efficient Channel Attention) attention mechanism into the model, we increased the weight of important information channels, thus improving detection accuracy. Moreover, ECA introduces only a small number of parameters to the model.

(3) The spatial pyramid pooling in the model was replaced with the SimSPPF (Simplified Spatial Pooling - Fast) structure, reducing computational complexity. Because SimSPPF, with the replaced activation function, achieved better results in testing detection speed.

(4) The SIoU (Scylla Intersection over Union) substituted for the CIoU (Complete Intersection over Union) in the original model, leading to better inference results and improved detection accuracy.

II. RELATED WORK

A. APPLICATION OF YOLOv5 IN OBJECT DETECTION

When it comes to computer vision, object detection will never be forgotten because it is currently one of the most challenging issues. Completing the classification of objects is its main task, and on this basis, further positioning tasks can be completed. Current research efforts primarily focus on object classification and localization, which are collectively referred to as detection tasks. Two-stage detection and one-stage detection have their own advantages and disadvantages. However, from the perspective of practical applications, one-stage detection can greatly improve detection speed while sacrificing a small portion of detection accuracy, making it more promising for the future. YOLOv5 is one of the most widely used algorithms in this field, including various detection scenarios such as agricultural disease detection, industrial part inspection, assisted medical diagnostics, pedestrian detection, vehicle detection, and more.

Liu et al. [8] utilized YOLOv5 combined with SimAM attention mechanism and added a layer of small object

detection in the layer of Neck to detect tassels in corn, achieving an mAP (mean average precision) of 44.7%, which was a 2.1% improvement over the original algorithm. However, further improvements in detection accuracy are still necessary for real-world applications.

Ma et al. [9] replaced the convolutional layers in YOLOv5 and applied it to fire and smoke detection scenarios, achieving a detection accuracy of 87.6%. However, since it was applied to fire and smoke detection, further improvements in detection accuracy are required for practical real-life applications.

Li et al. [10] completed the detection of typical satellite components using YOLOv5 and achieved an mAP of 95.8%, while reducing the model size by 66%. This provided a possibility for practical applications in this domain.

B. OBJECT DETECTION IN STEEL DEFECT SCENARIOS

In recent years, researchers have continuously achieved results in the research direction of object detection technology. With the increasing demand in industrial settings, many researchers have applied object detection techniques to defect detection in industrial environments, including steel surface defects. Early works mostly relied on image processing techniques to detect defects. They used image detection techniques to identify the presence of defects and then further highlighted the defects using gradient algorithms or region-growing algorithms. However, with the advancement of deep learning, subsequent works have predominantly utilized deep learning for defect detection.

Yu et al. [11] employed the anchor-free FCOS (Fully Convolutional One-Stage) detection framework and incorporated a channel attention mechanism. They used the FPN (Feature Pyramid Network) in place of BFFN (Bidirectional Feature Fusion Network). Experimental results showed an mAP of 76.68%, which was a 4.43% improvement over the original algorithm. This approach achieved good detection results, but it had a drawback of having a large number of parameters, which might impact its industrial applicability.

Feyza Selamet et al. [12] added SFS(Shape From Shading) to the Faster R-CNN model to address the influence of environmental factors such as lighting on detection. The experimental mAP reached 83%, meeting the detection requirements effectively. Although this research achieved good detection accuracy, it still suffered from the common issue of two-stage detection such as low detection speed.

Yang et al. [13] incorporated the CBAM (Convolutional Block Attention Module) into YOLOv5 and preprocessed images using a filtering algorithm. They achieved an mAP of 82.7%. However, this experiment had the drawback of insufficient generalization performance, with lower detection accuracy for two types of defects: cracks and scratches. Further improvements are required in this regard.

C. ATTENTION MECHANISM

Generally speaking, the larger the total number of parameters contained in a model, the better its expressive power will be. However, as the total number of parameters in a model increases to a limit, it can lead to information overload, causing computers to struggle in processing the information within images, which subsequently affects the subsequent detection results. Attention mechanisms have been proposed to address such issues. The main task of attention mechanisms is to identify important information, increase the weight of relevant information, at the same time decrease the weight of irrelevant information. They can help alleviate information overload to some extent and improve detection accuracy and speed, especially when computational resources are limited.

Attention mechanisms such as SENet (Squeeze-and-Excitation Networks) [14], CBAM (Convolutional Block Attention Module) [15], ECA (Efficient Channel Attention), CA (Coordinate Attention) [16], NAM (Normalization-based Attention Module) [17], and others have been widely applied in neural network models to enhance the performance of various detection tasks. Chen et al. [18] incorporated SENet into YOLOv3 and applied it to nut detection. He et al. [19] combined the ECA module with the ResNet model, and Wang et al. [20] combined multiscale transformer and CBAM, both applied to remote sensing image detection. Dou et al. [21] added the NAM module to YOLOv5, resulting in a 17% in-crease in computational speed. Xuan et al. [22] added the CA module to YOLOX and applied it to defect detection in printed circuit boards (PCB), achieving improved detection performance.

D. TYPES OF CONVOLUTIONS

Convolutions, as the fundamental building blocks in neural networks, are widely used and modifying the convolutions in a model can have a significant impact. Many researchers have focused on the study of convolutions and have proposed different types of convolutions, such as group convolution [23], depthwise separable convolution [24], and dilated convolution [25]. These convolutions have their own advantages and disadvantages, and researchers from various fields have extensively applied them in the field of detection.

Group convolution can lead to a decrease in the total number of parameters in the model, thereby achieving the goal of reducing computational costs. Chen et al. [26] utilized the advantage of group convolution to achieve aerial image segmentation. Depth-wise separable convolution significantly reduces the number of convolution parameters but may have an impact on detection accuracy. Training on GPUs with depthwise separable convolution is generally slower. Yan et al. [27] leveraged the characteristics of depthwise separable convolution to build a lightweight face recognition model and improving real-time detection. Dilated convolution expands the receptive field while reducing computational complexity. Compared to the face residual model, the total number of parameters in this model has been reduced by about 45%. However, it can lead to insufficient continuity of feature information in the image during sampling. Yang et al. [28] combined dilated convolution

into a neural network for image classification, achieving a classification accuracy of 93.5% and saving half of the training time.

III. METHODS

The improved network architecture, ECA-SimSPPF-SIoU-YOLOv5, is depicted in Figure 2. Several convolutions in the backbone and neck have been replaced with deformable convolutions. The SPP structure in the backbone has been replaced with the SimSPPF structure, and two additional layers of SimSPPF structure have been added. An ECA attention mechanism has been incorporated into the backbone, and the loss function has been replaced with SIoU.

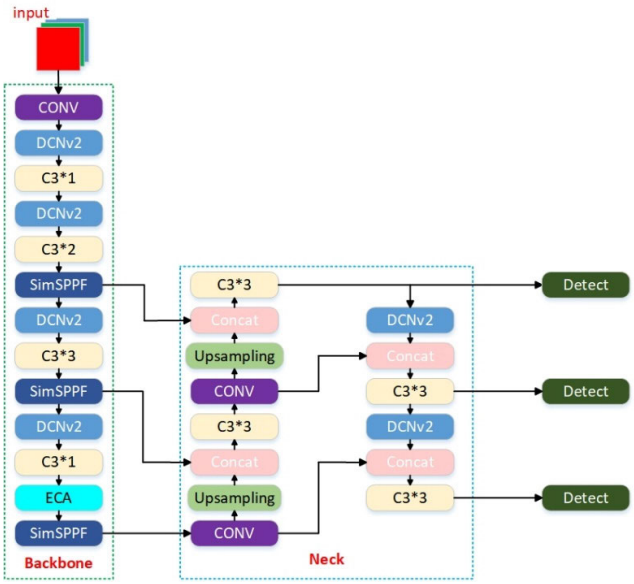

FIGURE 2. Improved YOLOv5 framework.

A. SimSPPF

The SPP (Spatial Pyramid Pooling) module was proposed by He et al. in 2015 [29] to address the problem of image distortion and significantly improve the speed of generating candidate boxes, thereby saving computational costs. Building upon this, Glenn Jocher, the author of YOLOv5, introduced the SPPF (Spatial Pyramid Pooling - Fast) module. The SPPF module transforms the separate max-pooling operations in the SPP module into sequential operations, leading to a decrease in the total number of parameters in the model, thereby achieving the goal of reducing computational costs. However, the authors utilized the SimSPPF (Simplified Spatial Pyramid Pooling - Fast) module in this paper. Its structure is shown in Figure 3. The difference between SimSPPF and SPPF lies in the activation function used within the module. While SPPF employs the SiLU activation function, SimSPPF utilizes the ReLU activation function. Using SimSPPF yields better detection speed compared to SPPF.

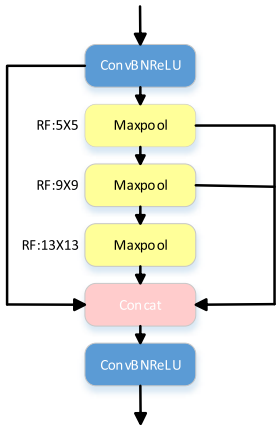

FIGURE 3. SimSPPF structure.

B. ECA ATTENTION MECHANISM

The SE attention mechanism is a method that allows neural networks to automatically determine the importance of feature channels and assign different weights accordingly, aiming to achieve better detection accuracy. In this paper, the ECA attention mechanism was utilized, which is an improved version of the SE attention mechanism. Its structure is depicted in Figure 4.

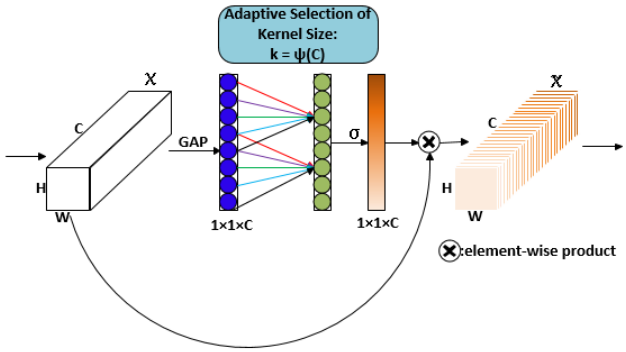

FIGURE 4. ECA structure.

In SENet, it is necessary to consider the relationships among all channels to achieve weight distribution. However, this approach also incurs significant computational costs. On the other hand, ECA overcomes this issue by replacing the fully connected layer with a convolutional layer, significantly reduced the total parameter count of the model after adding attention mechanisms. Comparatively, the ECA attention mechanism can achieve good detection results with only a small increase in parameters.

The size of the convolution kernel in Figure 4 is obtained through an adaptive function, which is defined as follows:

$$k = \left| \frac{log_2 c}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{1}$$

where k represents the size of the obtained convolution kernel, c denotes the size of the number of channels, γ and b are set to 2 and 1, respectively. The notation $||_{odd}$ indicates

that the resulting convolution kernel size can only be an odd number.

C. DEFORMABLE CONVOLUTION

When using regular convolutions for image sampling, the same receptive field is applied to extract features from objects of different sizes and irregular shapes. This feature extraction method is clearly not ideal. In this paper, deformable convolution [30] was employed, which could address some of the limitations of regular convolutions. The convolution operation of deformable convolution is depicted in Figure 5. The left side of Figure 5 illustrates the regular convolution operation, while the right side demonstrates the deformable convolution. From the convolution operation, it is evident that deformable convolution can obtain a larger receptive field compared to regular convolution. The range of receptive field during sampling can be changed as the size of the detected target changes, making the feature extraction process more aligned with the actual detection objects.

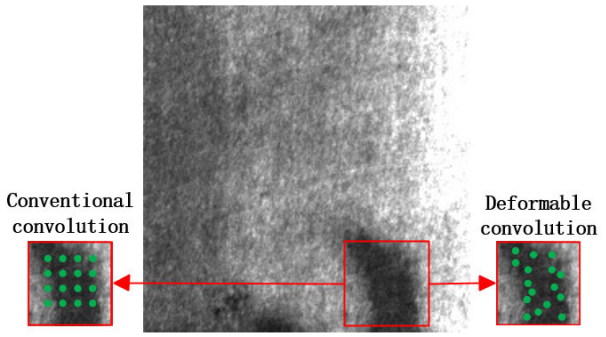

FIGURE 5. Deformable convolution operation.

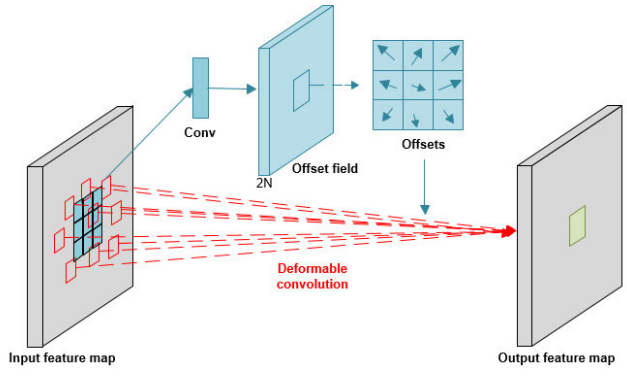

FIGURE 6. Deformable convolution structure.

The structure of deformable convolution is depicted in Figure 6. The input feature map is first passed through a convolutional layer to obtain an offset field, where N represents the size of the convolutional kernel, resulting in N offsets. Since the image is a 2D image, there are a total of 2N offsets, which correspond to the displacement values. These displacement values are then incorporated into the computation of regular convolution. The goal of expanding the receptive field can be achieved, thereby ultimately completing the task of matching the size of the detection target. The formula for incorporating the displacement values into regular convolution is as follows:

$$y(p_0) = \sum_{p_n \in R} w(p_n) * x(p_0 + p_n + \Delta p_n) \tag{2}$$

where $y(p_0)$ represents a point on the output feature map, where p_0 is the center point of the convolutional kernel, p_n represents the offset of regular convolution, $R \in \{(-1,-1), (-1,0), \ldots, (0,1), (1,1)\}$, $w(\cdot)$ denotes the weight of the corresponding offset point, and Δp_n is the offset amount for deformable convolution. However, since the generated offsets are usually floating-point numbers, bilinear interpolation is utilized to obtain the corresponding pixel values after the offset.

Due to the range of the receptive field exceeded the scope of the detected objects in the initial version (v1) of deformable convolution, which aimed to expand the receptive field. As a result, the extracted features extended beyond the detection range. To address this issue, a second version (v2) was proposed. In this version, a penalty mechanism was introduced to penalize the weights that extend beyond the detection range. This mechanism helps reduce such occurrences. The deformable convolution used in this paper specifically referred to the v2 version.

D. SIOU LOSS FUNCTION

Loss functions can be used to determine the similarity between predicted bounding boxes and ground truth boxes in object detection tasks. The higher the similarity between these two boxes, the better the detection result. The detection performance of the entire model will be better by properly utilizing loss functions. Several loss functions have been proposed to address the bounding box regression problem, such as IoU(Intersection over Union), GIoU(Generalized-IoU), DIoU(Distance-IoU), and CIoU(Complete-IoU). In YOLOv5, the CIoU loss function was chosen. However, these loss functions only consider aspects such as the overlapping area, distance between centers, and aspect ratios, without considering the mismatch in orientation between the two. This paper used SIoU instead of CIoU in the original model. It redefines the penalty mechanism and considers the angle between the predicted and ground truth boxes, which effectively improves the detection accuracy.

IV. RESULTS

A. DATASET

The NEU-DET dataset [31], created by Northeastern University, was used in this experiment. This dataset consists of 6 classes of defects: crazing (Cr), inclusion (In), patches (Pa), pitted surface (Ps), rolled-in scale (Rs), and scratches (Sc). There are 300 images in every types of defect, resulting in a

total of around 1800 images with approximately 4200 defects. The images have a resolution of 200 × 200 pixels.

B. HYPERPARAMETER OPTIMIZATION

The hyperparameter settings of the network can impact the detection results of the model. After multiple repeated experiments and comparisons, a set of optimal model training parameters was selected. Initial learning rate was 0.01, cyclic learning rate was 0.2, and the number of training epochs was 200 epochs. The SGD optimizer was used, and the weight decay was set to 0.0005.

C. PERFORMANCE METRICS

Using accuracy alone is not sufficient to effectively measure the detection performance of the model. Therefore, evaluation primarily focused on mAP (mean Average Precision), recall, and the total number of model parameters to assess the detection performance of the model in this experiment. mAP represents the average precision across all defect classes, providing an overall measure of detection accuracy.

D. ABLATION EXPERIMENTS

In order to validate the effectiveness of each improvement module in this experiment, several ablation experiments were conducted using the NEU-DET dataset. Each improvement module was evaluated using a controlled variable approach.

The specific results and experimental model of the ablation experiment are shown in Table 1. The second row represents the model without any improvement modules. Here, "DCNv2" refers to deformable convolution.

TABLE 1. Ablation experiment.

SimSPPF	ECA	DCNv2	SIoU	mAP	Recall	parameters
				0.717	0.727	7067395
√				0.729	0.679	**6863747**
	√			0.766	0.754	7067398
		√		0.76	0.687	7277509
			√	0.734	0.727	7067395
√	√	√	√	**0.788**	**0.764**	7073864

From Table 1, it can be observed that each individual improvement module, when added to the original model, can moderately increase mAP. The inclusion of SimSPPF can also reduce model parameters by nearly 3%. When all four improvements were added to the original model, it is evident that the mAP improved by approximately 7%, reaching 78.8%, with a 3.7% increase in recall, reaching 76.4%. These results indicated that the improved model exhibits good detection performance. The obvious comparison of three key data in the ablation experiment is shown in Figure 7.

In Figure 7, first Y-axis from left represents the magnitude of mAP values, and different improvement methods are represented by black line graphs. Second Y-axis from left

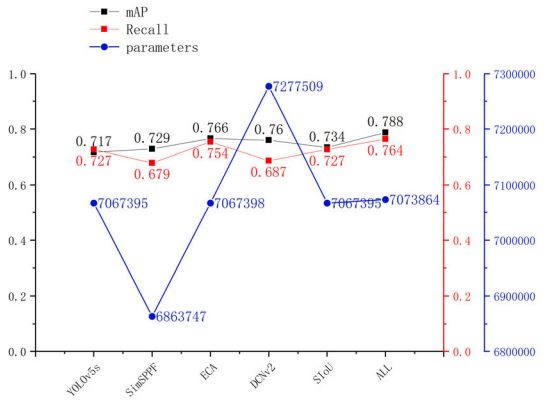

FIGURE 7. Ablation experiment.

represents the size of parameters in blue. Third Y-axis from left represents the magnitude of recall values in red. A blue line graph is used to compare the parameter sizes of models improved by different methods, while a red line graph is used to compare the recall values of models improved by different methods. It is evident that each improvement can increase the model's mAP, and when all improvements were added, the highest mAP of 78.8% and recall of 76.4% were achieved. This is significantly better than models with only one type of improvement. Additionally, the model only increased the parameter size by less than 1% compared to the YOLOv5s model.

E. COMPARATIVE EXPERIMENTS

In order to validate the effectiveness of the detection method in this experiment, certain improvement modules were replaced. For example, while keeping other improvement modules unchanged, the type of attention mechanism in the model was changed. The ECA attention mechanism was replaced with other types of attention mechanisms such as SA, SE, and NAM, among others. These comparative experiments were conducted to verify the advantages of the improved methods selected in this experiment. The experiments conducted in this study focused on attention mechanisms and loss functions. The specific results of the experiment can be well reflected in Tables 2 and 3.

From the various data in Table 2, it is easy to conclude that using different attention mechanisms only causes an increase or decrease of less than 1% in the total parameter quantity of the model. Among the five different attention mechanisms, the ECA mechanism achieves the best AP among three types of defects: inclusion, pitted surface and scratches. It also achieved the best mAP and Recall results in models that use different types of attention.

From the various data in Table 3, it is easy to conclude that compared to the other three different loss functions, the proposed algorithm achieved the best AP between two types of defects: patches and pitted surface. It also achieved the best mAP and Recall performance among all the models with different loss functions.

TABLE 2. Comparative experiments of attention.

Scheme	Crazing	Inclusion	Patches	Pitted Surface	Rolled-in Scale	Scratches	mAP	Recall	Parameters
CA-SimSPPF-SIoU-Yolov5	**0.63**	0.834	0.951	0.857	0.408	0.862	0.757	0.731	7099509
NAM-SimSPPF-SIoU-Yolov5	0.542	0.818	0.941	0.859	0.542	0.888	0.765	0.701	7074885
CBAM-SimSPPF-SIoU-Yolov5	0.543	0.79	0.934	0.819	**0.554**	0.912	0.759	0.727	7107272
SE-SimSPPF-SIoU-Yolov5	0.529	0.82	**0.953**	0.832	0.511	0.896	0.757	0.742	7106629
SA-SimSPPF-SIoU-Yolov5	0.601	0.801	0.935	0.876	0.543	0.905	0.777	0.715	7074053
ECA-SimSPPF-SIoU-Yolov5（ours）	0.591	**0.837**	0.949	**0.891**	0.535	**0.922**	**0.788**	**0.764**	**7073864**

TABLE 3. Comparative experiments of loss functions.

Scheme	Crazing	Inclusion	Patches	Pitted Surface	Rolled-in Scale	Scratches	mAP	Recall	Parameters
ECA-SimSPPF-CIoU-Yolov5	0.558	0.834	0.944	0.875	0.531	**0.932**	0.779	0.743	7073864
ECA-SimSPPF-DIoU-Yolov5	0.594	0.819	0.941	0.878	**0.56**	0.879	0.778	0.758	7073864
ECA-SimSPPF-GIoU-Yolov5	**0.598**	**0.871**	0.928	0.864	0.459	0.888	0.768	0.737	7073864
ECA-SimSPPF-SIoU-Yolov5（ours）	0.591	0.837	**0.949**	**0.891**	0.535	0.922	**0.788**	**0.764**	**7073864**

Based on the information presented in Tables 1-3, it was evident that the introduced enhancements in our model had yielded favorable outcomes. The incorporation of deformable convolution allowed the model to closely align with the shape of defects during the sampling process, facilitating the extraction of superior features. Additionally, the inclusion of the ECA attention mechanism effectively amplified the significance of crucial information in the detection process, thereby enhancing detection accuracy. Furthermore, the paper had refined SPP into SimSPPF to reduce the computational complexity associated with using this module. Lastly, by substituting the SIoU loss function, the paper had taken angles into account, effectively addressing the issue of bounding box regression. The obvious comparison of three key data on attention mechanisms and loss functions are illustrated in Figure 8 and Figure 9.

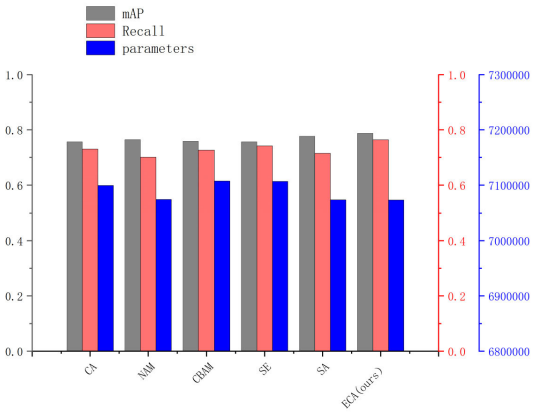

FIGURE 8. Comparative experiments of attention mechanisms.

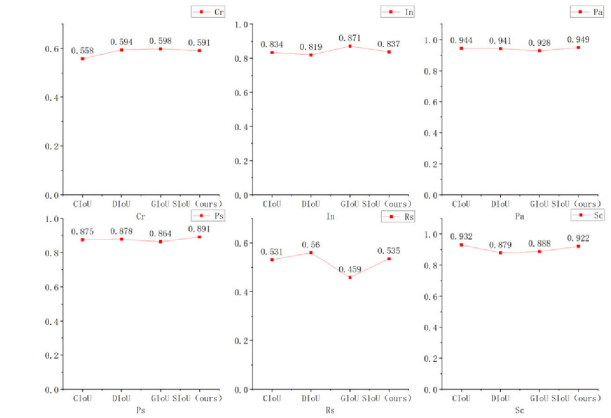

FIGURE 9. Comparative experiments of loss functions.

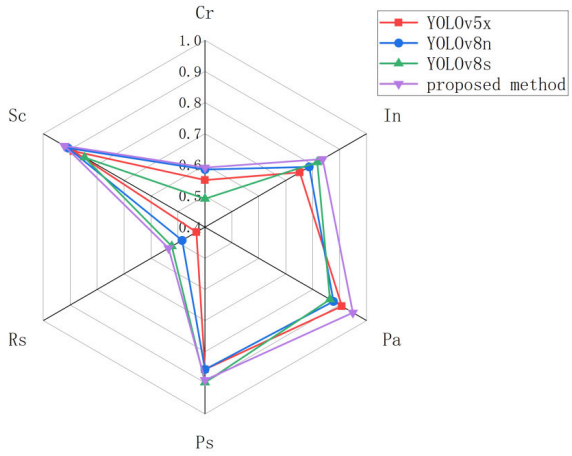

FIGURE 10. Other comparative experiments.

Figure 8 presents a comparison of different attention mechanisms. First Y-axis from left represents mAP, while second Y-axis from left represents the parameter size in blue. Third Y-axis from left represents Recall in red. The gray bars in the graph represent the magnitude of mAP, the blue bars represent the parameter size, and the red bars represent the Recall. From the graph, it becomes more visually apparent to observe the differences in mAP under different attention mechanisms. It can be observed that the

TABLE 4. Other comparative experiments.

Scheme	Crazing	Inclusion	Patches	Pitted Surface	Rolled-in Scale	Scratches	mAP	Recall	Parameters
YOLOv5x	0.551	0.752	0.908	0.857	0.432	0.888	0.731	**0.766**	87232339
YOLOv8n	0.585	0.787	0.877	0.857	0.485	0.908	0.75	0.683	**3006818**
YOLOv8s	0.491	0.817	0.864	**0.899**	0.523	0.847	0.74	0.718	11127906
ECA-SimSPPF-SIoU-Yolov5	**0.591**	**0.837**	**0.949**	0.891	**0.535**	**0.922**	**0.788**	0.764	7073864

proposed ECA-SimSPPF-SIoU-Yolov5 achieved the highest mAP of 78.8%, surpassing the second-ranked SA-SimSPPF-SIoU-Yolov5 with an mAP of 77.7%. It also achieved the highest Recall of 76.4% with the least number of parameters, which was only 7,073,864 in this comparison group.

Figure 9 illustrates the comparison of different loss functions for detecting different defects. It is evident that the proposed method demonstrated good detection results for various defects. The detection accuracy of patches reached the highest of 94.9%, and the detection accuracy of pitted surfaces reached the highest of 89.1%.

In order to validate the effectiveness of the improved ECA-SimSPPF-SIoU-Yolov5 model, comparative experiments were conducted with other models from the YOLO series, such as different network depths of YOLOv5 models and some models from the YOLOv8 series. The specific results of these comparative experiments can be well reflected in Table 4, demonstrating the effectiveness of this experiment.

From Table 4, it is easy to conclude that the proposed method showed the best detection effect for five out of six defects, excluding the ''rolled-in scale'' defect. It achieved the highest detection accuracy of 59.1% for ''crazing,'' 83.7% for ''inclusion,'' 94.9% for ''patches,'' 53.5% for ''rolled-in scale,'' and 92.2% for ''scratches.'' The AP for all five defect classes reached the highest value, and the mAP also reached the highest value. Figure 10 presents a radar chart comparing the proposed method with other approaches.

In the radar chart, the red color represents YOLOv5x, the blue color represents YOLOv8n, the green color represents YOLOv8s, and the purple color represents the proposed method. YOLOv8 is the latest algorithm introduced in the YOLO series. From the radar chart, it is evident that the ECA-SimSPPF-SIoU-Yolov5 model proposed in this paper outperformed the latest generation YOLOv8 in terms of detection accuracy. Compared to YOLOv8n, the proposed model achieved higher detection accuracy while reducing the overall parameter size.

V. CONCLUSION

This study improved upon the YOLOv5 model by making several modifications. First, the convolutional layers were enhanced by replacing conventional convolutions with deformable convolutions, which increased the receptive field and extracted larger features. Second, an ECA mechanism was introduced to assign higher weights to important feature information. Third, the pooling module was replaced to reduce computational complexity. Lastly, the loss function was replaced to better address the bounding box regression problem.

The algorithm proposed in this article was validated through the highly recognized NEU-DET dataset, and the experimental results demonstrated that the ECA-SimSPPF-SIoU-Yolov5 algorithm achieved good detection accuracy. It yielded an mAP of 78.8%, which was a 7.1% improvement over the original YOLOv5s model. It also outperformed other comparative experiments, achieving a Recall of 76.4%, which was a 3.7% improvement over YOLOv5s model without any improvements added. The total parameter size only increased by a small amount.

However, all the deep learning models proposed in this study still exhibit relatively low detection accuracy for the ''crazing'' and ''rolled-in scale'' defects. The images of these defects reveal that the exhibited characteristics of these two types of defects are less pronounced compared to other defects. Consequently, directly extracting these two types of defects from the background is not an easy task, leading to a lower detection accuracy for these specific defects. In future work, we plan to address this issue by increasing training data for these two types of defects, employing data augmentation techniques, and optimizing the model.

REFERENCES

[1] Y. Liu, J. Wang, H. Yu, F. Li, L. Yu, and C. Zhang, ''Surface defect detection of steel products based on improved YOLOV5,'' in *Proc. 41st Chin. Control Conf. (CCC)*, Anhui, China, Jul. 2022, pp. 5794–5799.

[2] K. Song, S. Hu, and Y. Yan, ''Automatic recognition of surface defects on hot-rolled steel strip using scattering convolution network,'' *J. Comput. Inf. Syst.*, vol. 10, no. 7, pp. 3049–3055, 2014.

[3] Y. Wang, H. Wang, and Z. Xin, ''Efficient detection model of steel strip surface defects based on YOLO-V7,'' *IEEE Access*, vol. 10, pp. 133936–133944, 2022, doi: 10.1109/ACCESS.2022.3230894.

[4] B. Hu and J. Wang, ''Detection of PCB surface defects with improved faster-RCNN and feature pyramid network,'' *IEEE Access*, vol. 8, pp. 108335–108345, 2020, doi: 10.1109/ACCESS.2020.3001349.

[5] S. Ikechukwu and E. Akin, ''High performance network for detection of surface defects on hot-rolled steel strips based on an optimized YOLO V3,'' in *Proc. 9th Int. Conf. Electr. Electron. Eng. (ICEEE)*, Alanya, Turkey, Mar. 2022, pp. 1–6.

[6] S. Zhai, D. Shang, S. Wang, and S. Dong, ''DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion,'' *IEEE Access*, vol. 8, pp. 24344–24357, 2020, doi: 10.1109/ACCESS.2020.2971026.

[7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, ''ECA-Net: Efficient channel attention for deep convolutional neural networks,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 11531–11539.

[8] W. Liu, K. Quijano, and M. M. Crawford, ''YOLOV5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOV5 based on transfer learning,'' *IEEE J. Sel. Topics Appl. Earth Obser. Remote Sens.*, vol. 15, pp. 8085–8094, 2022, doi: 10.1109/JSTARS.2022.3206399.

[9] J. Ma, Z. Zhang, W. Xiao, X. Zhang, and S. Xiao, ''Flame and smoke detection algorithm based on ODConvBS-YOLOV5S,'' *IEEE Access*, vol. 11, pp. 34005–34014, 2023, doi: 10.1109/ACCESS.2023.3263479.

[10] C. Li, G. Zhao, D. Gu, and Z. Wang, "Improved lightweight YOLOV5 using attention mechanism for satellite components recognition," *IEEE Sensors J.*, vol. 23, no. 1, pp. 514–526, Jan. 2023, doi: 10.1109/JSEN.2022.3222868.
[11] J. Yu, X. Cheng, and Q. Li, "Surface defect detection of steel strips based on anchor-free network with channel attention and bidirectional feature fusion," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2021.3136183.
[12] F. Selamet, S. Cakar, and M. Kotan, "Automatic detection and classification of defective areas on metal parts by using adaptive fusion of faster R-CNN and shape from shading," *IEEE Access*, vol. 10, pp. 126030–126038, 2022, doi: 10.1109/ACCESS.2022.3224037.
[13] N. Yang and W. Guo, "Application of improved YOLOv5 model for strip surface defect detection," in *Proc. Global Rel. Prognostics Health Manag. (PHM Yantai)*, Yantai, China, 2022, pp. 1–5.
[14] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.
[15] S. Woo, J. Park, and J. Lee, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[16] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 13708–13717.
[17] Y. Liu, Z. Shao, Y. Teng, and N. Hoffmann, "NAM: Normalization-based attention module," 2021, *arXiv:2111.12419*.
[18] Q. Chen, L. Liu, R. Han, J. Qian, and D. Qi, "Image identification method on high speed railway contact network based on YOLO V3 and SENet," in *Proc. Chin. Control Conf. (CCC)*, Guangzhou, China, Jul. 2019, pp. 8772–8777.
[19] Y. He, S. Zhou, and X. Quan, "Remote sensing image scene classification based on ECA attention mechanism convolutional neural network," in *Proc. IEEE 4th Int. Conf. Civil Aviation Saf. Inf. Technol. (ICCASIT)*, Dali, Dali, China, Oct. 2022, pp. 1265–1269.
[20] W. Wang, X. Tan, P. Zhang, and X. Wang, "A CBAM based multiscale transformer fusion approach for remote sensing image change detection," *IEEE J. Sel. Topics Appl. Earth Obser. Remote Sens.*, vol. 15, pp. 6817–6825, 2022, doi: 10.1109/JSTARS.2022.3198517.
[21] X. Dou, T. Wang, and S. Shao, "A lightweight YOLOV5 model integrating GhostNet and attention mechanism," in *Proc. 4th Int. Conf. Comput. Vis. Image Deep Learn. (CVIDL)*, May 2023, pp. 348–352.
[22] W. Xuan, G. Jian-She, H. Bo-Jie, W. Zong-Shan, D. Hong-Wei, and W. Jie, "A lightweight modified YOLOX network using coordinate attention mechanism for PCB surface defect detection," *IEEE Sensors J.*, vol. 22, no. 21, pp. 20910–20920, Nov. 2022, doi: 10.1109/JSEN.2022.3208580.
[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1–9.
[24] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
[25] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.
[26] K. Chen, K. Fu, X. Gao, M. Yan, W. Zhang, Y. Zhang, and X. Sun, "Effective fusion of multi-modal data with group convolutions for semantic segmentation of aerial imagery," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Aug. 2019, pp. 3911–3914.
[27] W. Yan, T. Liu, S. Liu, Y. Geng, and Z. Sun, "A lightweight face recognition method based on depthwise separable convolution and triplet loss," in *Proc. 39th Chin. Control Conf. (CCC)*, Shenyang, China, Jul. 2020, pp. 7570–7575.
[28] J. Yang and J. Jiang, "Dilated-CBAM: An efficient attention network with dilated convolution," in *Proc. IEEE Int. Conf. Unmanned Syst. (ICUS)*, Oct. 2021, pp. 11–15.
[29] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015, doi: 10.1109/TPAMI.2015.2389824.
[30] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.
[31] K. Song and Y. Yan, "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," *Appl. Surf. Sci.*, vol. 285, pp. 858–864, Nov. 2013, doi: 10.1016/J.APSUSC.2013.09.002.

FEI REN (Student Member, IEEE) was born in China, in 1994. He received the bachelor's degree in mechanical design, manufacturing, and automation, in 2016, and the master's degree in mechanical engineering (machine vision and artificial intelligence), in 2021. He is currently pursuing the Ph.D. degree in computer science with Mapua University, Philippines.

He has published more than ten domestic and international papers, published and authorized nearly 30 patents, won more than ten national and provincial competition awards, and published one monograph. His research interests include machine vision, machine learning, artificial intelligence, natural language processing, and OCR technology.

Mr. Ren served as an off campus tutor for graduate students in Nanjing's universities, a member of Baidu PaddlePaddle Doctor's Association, and a student member of the Institute of Electrical and Electronics Engineers.

JIAJIE FEI received the bachelor's degree in measurement and control technology and instrumentation from Nanjing Institute of Technology, in 2022, where he is currently pursuing the master's degree in mechanical engineering.

His research interests include machine vision, defect detection, and artificial intelligence.

HONGSHENG LI was born in August 1966. He received the Ph.D. degree in engineering from the Department of Control Science and Engineering, Southeast University.

He is currently a Professor with the School of Automation, Nanjing Institute of Technology. He has led and completed more than ten important projects and multiple industry university research engineering research projects. He has published over 70 papers in important academic journals, such as IEEE TCST, *Journal of Mechanical Engineering, Pattern Recognition and Artificial Intelligence, Information and Control*; and authorized four invention patents. His research interests include robot control, machine vision, CNC technology, and intelligent control.

BONIFACIO T. DOMA JR. received the Ph.D. degree in chemical engineering from the University of the Philippines Diliman. He is a professor of chemical engineering and business analytics. His current research is in the area of engineering applications of artificial intelligence and machine learning.