Defect Detection Algorithm for Body-in-white Weld Spots Based on Improved YOLOv5

LIU Si-yuan, LI Yi-xian, WANG Cong-hui, HU Zheng-yi, PEI Kai, LI Hong-lin

***Abstract*—Spot welding is a crucial process in automotive manufacturing. This paper proposes a method for detecting appearance defects in body-in-white weld spots based on an improved YOLOv5. First, the PConv convolution is combined with conventional convolution to propose the C3-Faster module, aimed at reducing redundant computation and memory access. Second, the WIoUv3 loss function is introduced, employing Wise-IoU dynamic non-monotonic FM (static focusing mechanism) to evaluate anchor box quality via outlier degree and using a gradient gain allocation strategy to reduce harmful gradients from low-quality examples. Experiments conducted with images collected from an automotive chassis production site indicate that the proposed model achieves higher mAP and faster detection speed compared to the YOLOv5, with an mAP of 0.974. The proposed algorithm demonstrates superior accuracy and speed compared to ten other commonly used algorithms on our dataset.**

***Index Terms*—Weld Spot Defect Identification, Deep Learning, YOLOv5, Image Processing**

I. Introduction

IN automotive manufacturing, the welding quality of weld spots, nuts, and bolts is fundamental to ensuring the proper installation of other components. Achieving real-time feedback on welding defects during production is a critical issue in improving product quality for the automotive industry. Currently, most manufacturers rely on visual inspection by personnel to detect spot welding defects[1]. This method's recognition accuracy is highly influenced by the inspectors' proficiency and skill level, with variations in their working conditions also leading to inconsistent results, making it challenging to achieve objective and accurate weld spot defect identification[2]. In contrast, highly efficient and accurate spot welding defect detection technology can significantly enhance production automation, reduce costs, ensure product quality more effectively, and improve a company's competitiveness[3]. Therefore, developing an efficient and accurate defect detection algorithm for automotive spot welding has significant theoretical value and practical importance.

Ultrasonic testing is widely used in spot welding defect detection, but its drawbacks are also significant and cannot be ignored[4]. When the metal sheet is thin, it is difficult to distinguish echo signals from different interfaces, requiring higher frequency transducers. Additionally, indentations on the weld spot can affect ultrasound propagation, leading to inaccurate parameter estimation[5]. In the application of ultrasonic testing, integration with automotive automated welding production lines is challenging, requiring significant modifications to the production line, which results in high costs[6]. In contrast, deep learning-based spot welding defect detection can significantly improve the efficiency and accuracy of feature extraction, and it has a higher adaptability, making it suitable for defect detection in various environments[7]. For different detection requirements, targeted improvements can be made by optimizing the network and other methods. For example, J. Sun[8] used ImageNet 2012 to pre-train the network to address the issue of a small sample size in metal surface defect detection. After pre-training on a general dataset, the weights were transferred to initialize the surface defect detection network. Y. P. Gao[9] proposed a semi-supervised learning method to address the small sample issue by improving CNN performance using pseudo labels, which requires fewer labeled samples. Z. Q. Lin[10] optimized the YOLOv5 model by enhancing the upsampling operator with the CARAFE operator, improving the accuracy of cylindrical coating lithium battery defect detection. Z. X. Zhang[11] introduced a Nonlinear Spatial Pyramid Pooling Fast (NSPPF) module and constructed a combination of CoordConv and SK attention modules, enhancing YOLOv5's performance in road damage detection. Y. Z. Fu[12] added an attention module and modified the loss function of YOLOv5 to detect DTP defects in door trim panels.

To meet the real-time requirements in surface defect detection applications, models or model parameters can be pruned. J. Lei[13] proposed an end-to-end screen defect detection model, incorporating a defect detection network based on a merge-split strategy (MSDDN) to handle various sizes and shapes of defect image patches. Y. F. Pan[14] designed an FPGA architecture for a one-dimensional Fourier reconstruction defect segmentation algorithm with dual-task parallelism and two-pixel parallelism, applying it to texture surface defect segmentation, which effectively reduced detection time.

In summary, body-in-white weld spots are characterized by their large quantity, small size, and indistinct defect features[15], with defective weld spots being relatively rare in actual production, making it challenging to collect

Manuscript received July 11, 2024; revised January 16, 2025.

LIU Si-yuan is an associate professor at Jilin University, Changchun, Changchun 130000, China. (email: liusiy@jlu.edu.cn).

LI Yi-xian is a postgraduate student of Jilin University, Changchun, Changchun 130000, China. (email: lidx23@mails.jlu.edu.cn).

WANG Cong-hui is a professor at Jilin University, Changchun, Changchun 130000, China. (corresponding author to provide phone: 15543009880; email: conghui@jlu.edu.cn).

HU Zheng-yi is a professor at Changchun Automotive Industry Institute, Changchun, Changchun 130000, China. (e-mail: huzhengyi@rossum-robot.com).

PEI Kai is a postgraduate student of Jilin University, Changchun, Changchun 130000, China. (email: peikai22@mails.jlu.edu.cn).

LI Hong-lin is a postgraduate student of Jilin University, Changchun, Changchun 130000, China. (email: honglinl23@mails.jlu.edu.cn).

sufficient samples. Current deep learning model improvement methods have certain limitations in body-in-white weld spot detection. Firstly, existing YOLO models are limited by receptive field variations and multi-scale features, resulting in suboptimal performance in weld spot detection scenarios[16]. Secondly, incorporating lightweight networks into object detection models can significantly reduce the computational requirements and enhance real-time performance. However, this often leads to insufficient feature extraction capability, compromising detection accuracy. To meet the requirements of weld spot appearance quality inspection in industrial environments, this paper optimizes and improves the latest version of the YOLOv5 model. The model combines PConv convolution[17] with conventional convolution to propose the C3-Faster module, replacing the original C3 module in the network to reduce redundant computation and memory access. The WIoUv3 loss function[18] is introduced, which employs a dynamic non-monotonic FM (static focusing mechanism) called Wise-IoU, using outlier degree instead of IoU to evaluate anchor box quality, and applies a gradient gain allocation strategy to reduce harmful gradients from low-quality examples, ultimately achieving accurate and fast weld spot detection.

II. YOLOv5 Model and Algorithm Improvements

Based on the appearance characteristics of body-in-white weld spots, including their large quantity and the significant size variation between near and distant weld spots, this paper selects the YOLOv5s model for optimization and improvement. The PConv convolution is combined with traditional convolution to reduce computational costs, and the WIoUv3 loss function is introduced to mitigate harmful gradients from low-quality examples, enhancing the model's detection capability for body-in-white weld spots.

A. YOLOv5 Model

YOLOv5 is a multi-scale object detection model trained on the COCO dataset[19]. As shown in Figure 1, its network structure consists of three main components: Backbone, Neck, and Head. The input stage processes images with Mosaic data augmentation, adaptive image scaling, and adaptive anchor calculation to enhance training speed and accuracy. The Backbone module is responsible for extracting general feature representations and is typically composed of high-performing classifiers[20]. The Neck network, located between the Backbone and Head, employs SPP (Spatial Pyramid Pooling) and FPN+PAN modules to enhance the diversity and robustness of features. The Head network is primarily used to predict the type and location of objects, using GIoU_Loss as the loss function for bounding boxes, and applying NMS (Non-Maximum Suppression) to remove redundant predictions.

B. C3-Faster Lightweight Network Structure Design

Traditional convolutional neural networks improve detection performance by deepening the network layers, increasing the number of channels, or enhancing image resolution[21]. However, overly deep networks may suffer from vanishing gradients and reduced accuracy gains, while excessively wide networks struggle to extract rich semantic information at deeper levels. High image resolution can also introduce additional training difficulties[22]. Complex network structures lead to higher computational complexity and longer running times. To balance lightweight design and network accuracy, and to create a cost-effective, faster network with reduced computational complexity, we propose the C3-Faster module inspired by the FasterNet network[17]. This module replaces the original C3 module in the YOLOv5 network, as shown in Figure 2.

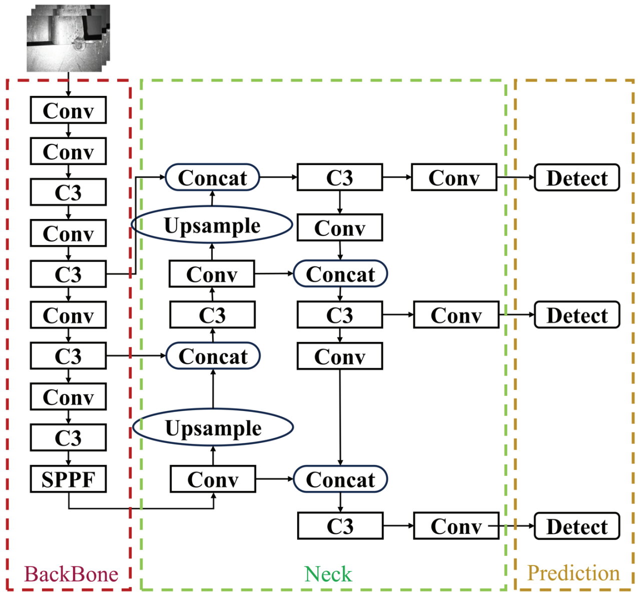


Fig. 1. YOLOv5 Network Architecture Diagram

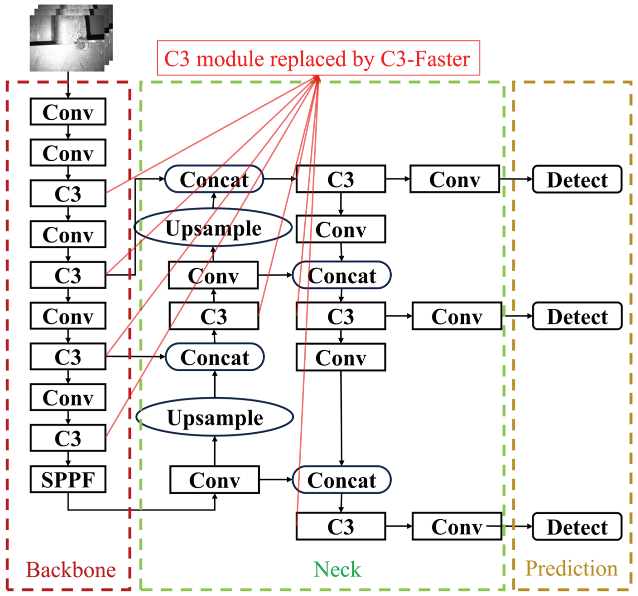


Fig. 2. Replacing the original YOLOv5 network's C3 module with the C3-Faster module

The overall architecture of the FasterNet network has four hierarchical stages, each with a stack of FasterNet blocks, preceded by an embedding or merging layer. The last three layers are used for feature classification. In each FasterNet block, the key lightweight improvement is the use of PConv instead of traditional convolution. PConv is a local convolution used for more efficient extraction of spatial features. Compared to traditional convolution, PConv only utilizes part of the channels for feature extraction, while retaining the remaining channels for subsequent PWConv layers. This approach allows PConv to reduce both redundant computations and memory access, thereby improving the operating speed of the neural network. After modifying YOLOv5's C3 structure with PConv convolution, its lightweight performance is significantly better compared to the BottleNeck structure.

The BottleNeck structure consists of two 1×1 convolution layers with an embedded 3×3 convolution layer in between[23]. The 1×1 convolution layers are responsible for reducing and then increasing the dimensions, effectively keeping the dimensions unchanged, making the 3×3 convolution layer the bottleneck with reduced intermediate dimensions between input and output. In contrast, the C3-Faster module replaces the BottleNeck structure with the FasterNet Block. The FasterNet Block comprises a 3×3 PConv layer and two 1×1 convolution layers, as illustrated in Figure 3.

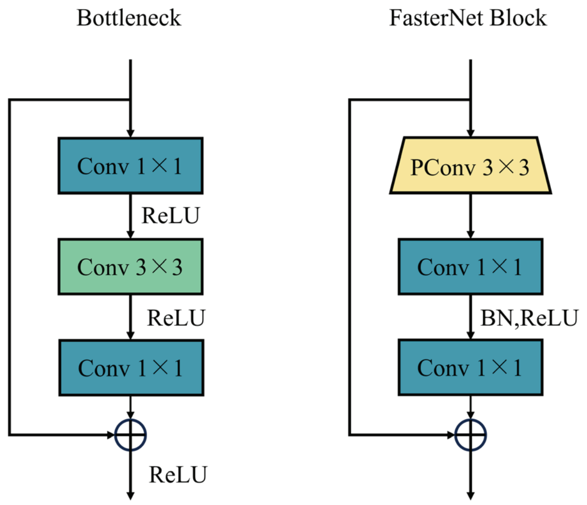


Fig. 3. The structure diagram of the BottleNeck and FasterNet Block

The C3 module divides the input from the previous layer into two branches. One branch goes through convolution, normalization, and activation functions before being passed to the BottleNeck structure, while the other branch only undergoes convolution, normalization, and activation. The two branches are then concatenated and processed again with convolution, normalization, and activation[23]. The C3-Faster replaces the original BottleNeck structure with the FasterNet Block, as shown in Figure 4.

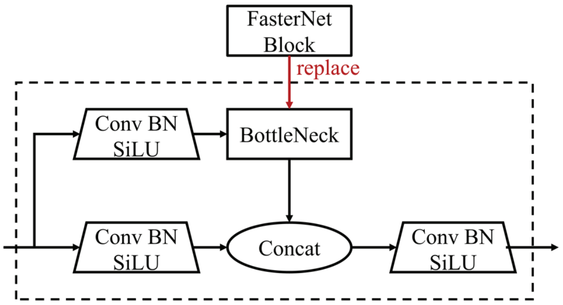


Fig. 4. Replacing the BottleNeck structure with the FasterNet Block

C. Using Wise-IoU v3 to Optimize the Loss Function

The built-in loss function of YOLOv5 is GIoU, a variant based on IoU. The calculation of IoU is as follows:

$$IoU = \frac{D_1 \cap D_2}{D_1 \cup D_2} \tag{1}$$

$$\mathrm{L}_{IoU} = 1 - IoU \tag{2}$$

In the equation, IoU represents the intersection over union of the ground truth box and the predicted box, with D1 being the area of the ground truth box and D2 being the area of the predicted box. GIoU compensates for the issue when IoU becomes zero due to no overlap between the ground truth box and the predicted box, which prevents gradient descent and model optimization[24]. However, GIoU is still based on area metrics. To accelerate convergence and stabilize the regression process, YOLOv5 introduces DIoU, which transforms the area issue into a distance problem[25]. DIoU accelerates network convergence by minimizing the normalized distance between the centroids of the two bounding boxes.

These methods assume that examples in the training data are of high quality and focus on enhancing the fitting ability of the bounding box regression (BBR) loss. However, the weld point images collected from factories contain a large number of distant weld points. These weld points are too small in pixels and do not contain sufficient features to determine whether they are defective. Such low-quality examples are prevalent in weld point datasets, and blindly enhancing BBR on these low-quality examples would jeopardize classification performance. To address this issue, Z. Tong[18] proposed an IoU-based loss function that includes a dynamic non-monotonic FM named Wise-IoU (WIoU). The dynamic non-monotonic FM uses outlier degree instead of IoU to evaluate the quality of anchor boxes and provides a wise gradient gain allocation strategy.

When the predicted box aligns well with the target box, a good loss function should reduce the penalty of geometric factors. Less intervention during training can lead to better generalization ability for the model. Based on this, distance attention was constructed, resulting in WIoU v1, which includes two layers of attention mechanisms:

$$\mathrm{L}_{WIoUv1} = \mathrm{R}_{WIoU} \mathrm{L}_{IoU} \tag{3}$$

$$\mathrm{R}_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \tag{4}$$

In the equation, W_g and H_g are the width and height of the smallest enclosing box formed by the predicted box and the target box. The asterisk (*) indicates that W_g and H_g are detached from the computation graph, preventing R_{WIoU} from generating gradients that hinder convergence.

To ensure that BBR focuses on anchor boxes of regular quality, a small gradient gain is assigned to excessively large and small outlier degrees. A non-monotonic focusing factor, β is constructed and applied to WIoU v1：

$$\mathrm{L}_{WIoUv3} = r \mathrm{L}_{WIoUv1} \tag{5}$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \tag{6}$$

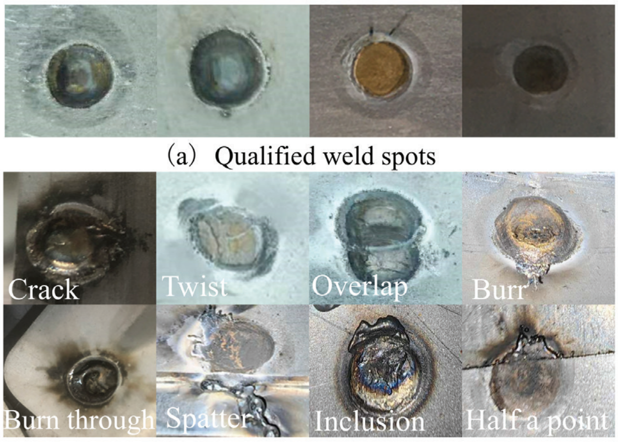
In the equation, β represents the outlier degree of the anchor box, expressed as the ratio of L_{IoU} to $\overline{\mathrm{L}_{IoU}}$, denoted as $\beta = \frac{\overline{\mathrm{L}_{IoU}}}{\mathrm{L}_{IoU}} \in [0, +\infty)$, L_{IoU} is the exponentially weighted moving average of the momentum m:

$$m = 1 - \sqrt[tn]{0.05} \qquad (7)$$

In the equation, t represents the number of epochs and n represents the number of batches. This setup ensures that after training for t epochs, L_{IoU} approaches the actual value. When the outlier degree of the anchor box satisfies *β=C*(where *C* is a constant value)， the anchor box will receive the highest gradient gain. Since both L_{IoU} and the quality division standard of the anchor box are dynamic, the gradient gain allocation strategy of WIoU v3 can adjust in real-time according to the training conditions[18].

III. Dataset Preparation

Weld spot images were captured using industrial cameras from the body-in-white spot welding workshop in an automotive manufacturing plant. According to the weld spot classification standard, different types of qualified weld spots and those with appearance defects are shown in Figure 5.

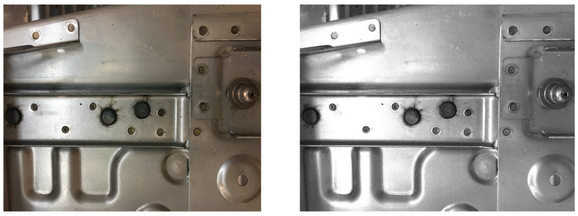


（a） Qualified weld spots

（b） Defective weld spots

Fig. 5. Examples of Weld Spots with Acceptable and Defective Appearances

Grayscale processing was applied to weld spot images, converting RGB color information into a single-channel grayscale value containing only luminance[26]. Grayscale processing reduces the dimensionality of image data, thereby decreasing the computational load and complexity of subsequent image processing. At the same time, it retains the shape features of the targets in the images without affecting object recognition, as shown in Figure 6.



（a）Original image （b）Grayscale image

Fig. 6. The original image and the grayscale processed comparison image

Annotation was performed to obtain the dataset required for subsequent experiments. The commonly used open-source visualization data annotation tool, labelimg, was employed to complete the image annotation, obtaining information on the types of targets contained in the images and the precise locations of the target areas. These data are saved in txt files corresponding to the weld spot images for use in model training and detection. A total of 205 weld spot images were annotated, comprising 2342 weld spots. Among these, there were 1704 good spots, 172 half a point spots, 96 overlap welding spots, 352 twist spots, 10 burr spots, and 8 burn through spots. The maximum pixel size of the weld spots was 392×521, while the minimum pixel size was 19×22. During annotation, only the distinction between appearance-qualified and appearance-defective weld spots was made, resulting in a binary classification training sample. The annotated training samples were divided into training and validation sets in a ratio of 7:3.

IV. EXPERIMENT

Experiments were conducted on the body-in-white dataset using the proposed improved YOLOv5 algorithm, and the detection results were compared with those of the original network to validate the proposed algorithm.

A. Experimental Setup and Parameter Settings

The experimental environment includes the Windows 10 operating system, an Intel Core i9 13900k processor (CPU), an NVIDIA GEFORCE RTX4090 graphics card (GPU) with 24GB of VRAM, 16GB of system RAM, the PyTorch deep learning framework, and the PyCharm development platform. The

TABLE I
Training Parameter Settings Table

Parameter Name	Parameter Value
Leanrning_rate	0.01
Weight decay	0.0005
Warmup_epochs	3.0
Warmup_bias_lr	0.1
Batch_size	16
Epochs	200

B. Evaluation Metrics

In machine vision detection, mAP (mean Average Precision) is generally used as the evaluation metric. First, the PR curves (Precision-Recall curves) for each category are plotted. AP (Average Precision) is the area under the PR curve, and mAP is the average of the AP values for all categories.

Precision refers to the proportion of actual correct targets among all targets predicted as correct by the network. The calculation formula is as follows:

$$Precision = \frac{TP}{TP + FP} \qquad (8)$$

In the formula, TP（True Positives）represents the number of correctly predicted targets among the predicted true targets, and FP（False Positives） represents the number of correctly predicted targets among the predicted false targets.

Recall refers to the proportion of actual correct targets predicted as correct by the network among all actual targets. The calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN} \qquad (9)$$

In the formula, FN（False Negatives）represents the number of incorrectly predicted targets among the actual false targets.

The formula for calculating AP is as follows：

$$\mathrm{AP} = \int_0^1 P(R)dR \qquad (10)$$

In the formula, P represents Precision, and R represents Recall.

Additionally, GFLOPS (Giga Floating-point Operations Per Second) represents the number of floating-point operations performed per second in billions, while FPS (Frames Per Second) refers to the number of forward and backward propagations a neural network model can perform per second. GFLOPS is used to measure the model's computational complexity, and FPS is used to evaluate the detection speed of the model.

C. Experimental Results

Model improvements were made based on YOLOv5s. The improved model underwent 200 iterations, with its mAP and Loss curves shown in Figure 7. From the training plots, it can be observed that as the number of iterations increases, the Loss curve converges smoothly, and the learning efficiency gradually saturates. The mAP value increases rapidly and eventually stabilizes, reaching 0.974 at the end of training. This indicates that the improved model performs well on the weld spot dataset.

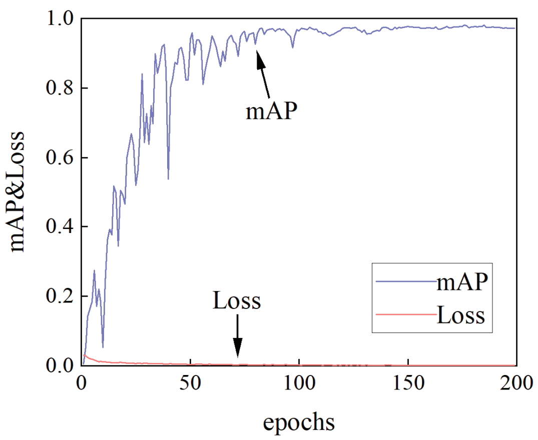
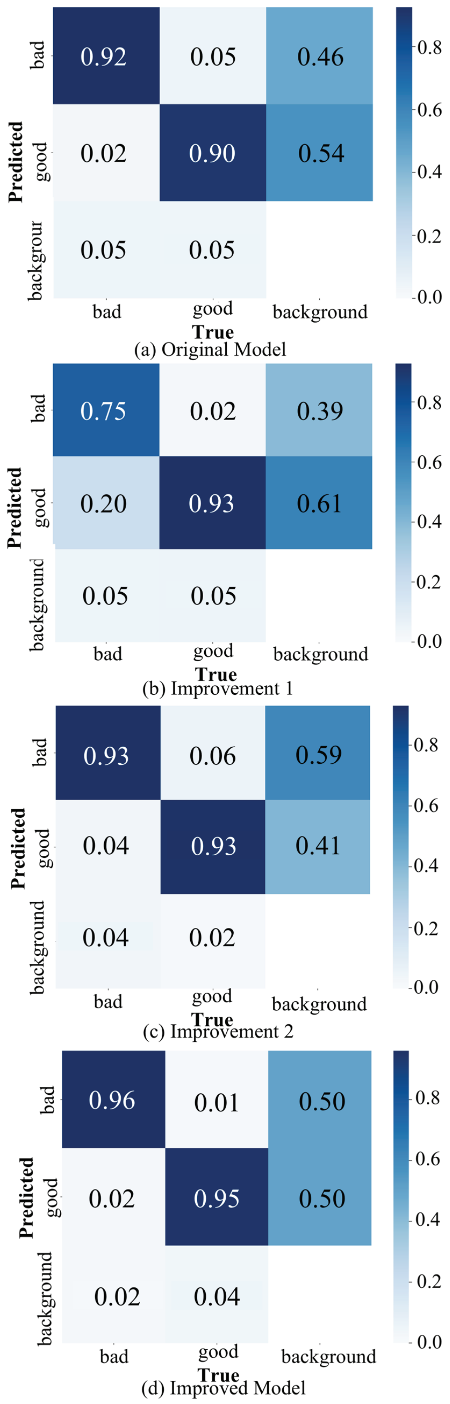


Fig. 7. mAP curve and Loss curve of the improved YOLOv5 model after 200 epochs

To verify the effectiveness of the improved network, ablation experiments were conducted by training the weld spot dataset on the network, and the results are shown in Table 2. From Table 2, it can be seen that using C3-Faster for model lightweight optimization results in a decrease in mAP, which is a common phenomenon, especially when pursuing a lightweight network structure. This is due to the reduction in model complexity, which can lead to a decline in performance. Using WIoUv3 to improve the original model can increase the mAP, and interestingly, using WIoUv3 on the lightweight C3-Faster model yields a greater mAP improvement, indicating good compatibility between the WIoUv3 loss function and the C3-Faster lightweight network.

To further analyze the practical significance of improvements in different parts of the model, confusion matrices for the ablation experiments were plotted, as shown in Figure 8. By comparing the confusion matrices of the different improved models, changes in the number of correctly classified samples for each category can be observed. When using the WIoUv3 improvement, the true positives (TP) for both defective and qualified weld spots increased, indicating enhanced recognition ability for both categories. The confusion matrices also reveal which classes are more likely to be confused by the model. In the model improved with both C3-Faster and WIoUv3, false positives (FP) and false negatives (FN) are significantly reduced compared to other models, indicating enhanced generalization ability of the model, which can more accurately distinguish between different categories with fewer false alarms or missed detections.



(a) Original Model

(b) Improvement 1

(c) Improvement 2

(d) Improved Model

Fig. 8. Confusion Matrix for Ablation Experiment

In summary, the lightweight optimization brought by C3-Faster led to a reduction in mAP, but incorporating WIoUv3 not only compensated for the performance loss but also provided improvements in certain aspects. From the confusion matrices, it can be seen that the improved model exhibits better recognition capabilities across different categories, with fewer false detections and missed detections, resulting in more stable performance. These findings suggest that the WIoUv3 loss function is well-suited to the C3-Faster lightweight network, effectively enhancing model performance. Furthermore, the improved model reduced the parameter count by 77.7% and increased detection speed by 9.0% compared to the original model, as shown in Table 3.

TABLE III
COMPARISON OF DETECTION SPEED

Model	Parameters/GPLOPs	FPS
Original Model	16.6	271
Improved Model	3.7	329

To further verify the effectiveness of the improvements, an additional 70 weld spot images were collected from an automotive chassis production site for comparison, resulting in a total of 628 weld spots detected. As shown in Figure 9, while the original YOLOv5s network could correctly classify most weld spots, it suffered from a considerable number of missed detections due to feature loss during convolution and optimization issues in the loss function. The introduction of C3-Faster significantly reduced detection speed but did not lead to more missed detections, though the number of false positives increased slightly. The model with the WIoUv3 introduction significantly reduced the number of falsely detected weld spots, particularly excelling in reducing missed detections. The final network, combining all improvements, enhanced recognition accuracy while reducing recognition time.

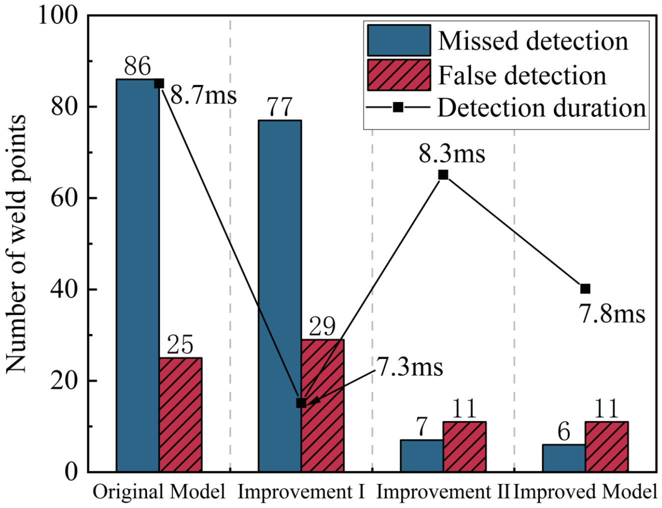


Fig. 9. The number of missed and false detected weld spots during actual detection, along with the average time taken to detect a weld spot image

WIoUv3 dynamically adjusts weights so that targets that are difficult to detect contribute more to the loss, which causes the model to focus more on these targets that are easily missed during training, thereby improving recall and reducing missed detections. Furthermore, in combination with C3-Faster, the enhanced detection speed was maintained without sacrificing detection accuracy, even resulting in a slight improvement. Figure 11 illustrates the improvement brought by the enhanced model. The two weld point images were detected using the original and improved models. In the left image, missed and false detections are observed with the original model, while the right image shows a false detection at the lower weld point, which is correctly detected by the improved model.

During the detection phase, the model's task is to perform forward propagation on new input images and generate predictions. The C3 module functions the same way during the inference phase as it does during the training phase, mainly for extracting deep features of the input image and feature fusion. Therefore, improvements to the C3 module also optimized the detection time for weld spots. The inference speed statistics are shown in Table 4, where the original YOLOv5s model took 721 ms in total, while the improved model took 651 ms. Specifically, "pre-process" refers to the time for data preprocessing, "inference" refers to the time for model inference, and "NMS" refers to the time for non-maximum suppression. These results were obtained under an input data size of (1, 3, 640, 640), where "1" represents the batch size, "3" represents the number of channels (RGB image), and "640" represents the image width and height. The data shows that the main contribution of the improved model to reducing detection time lies in reducing inference time, indicating that the improved model reduces forward propagation computational overhead by reducing the number of parameters.

TABLE IV
EXECUTION TIME OF EACH PART OF WELD SPOT DETECTION

Model	pre-process per image	inference per image	NMS per image	Total time
Original Model	0.8 ms	8.7 ms	0.8 ms	721 ms
Improvement 1	0.8 ms	7.3 ms	0.7 ms	616 ms
Improvement 2	0.6 ms	8.3 ms	0.8 ms	679 ms
Improved Model	0.7 ms	7.8 ms	0.8 ms	651 ms

We conducted comparative experiments using a self-made dataset to further analyze the detection performance of the proposed algorithm. The Improved Model was compared with other popular object detection techniques. Each

TABLE II
RESULTS TABLE OF ABLATION EXPERIMENTS

Method	C3-Faster	WIoU v3	Precision	Recall	mAP
Original Model	×	×	0.844	0.972	0.939
Improvement 1	✓	×	0.892	0.963	0.927
Improvement 2	×	✓	0.913	0.989	0.956
Improved Model	✓	✓	0.951	0.984	0.974

detection technique was tested using the same data from the training and testing sets under consistent parameter settings. The proposed algorithm demonstrated superior performance across all aspects on our dataset, as shown in Table 5. Our method outperformed the SSD model by 35.4% mAP and the Faster R-CNN model by 24.9% mAP, while still maintaining training and testing sets under consistent parameter settings. The proposed algorithm demonstrated superior performance fewer parameters and higher computational speed. The improved YOLOv5 model showed better accuracy and speed compared to all versions of YOLOv5, and its performance was enhanced in all aspects when compared with other classic and latest versions of YOLO.

We conducted experiments using common improvement methods on the original YOLOv5s network, including improving the network with Squeeze-and-Excitation Networks[27], improving the network with Convolutional Block Attention Module[28], improving the loss function with Normalized Wasserstein Distance[29], improving the loss function with variFocalLoss[30], and finally improving the loss function with WIoUv3. A total of five sets of ablation experiments were conducted, and the comparison results are shown in Figure 10. After 200 epochs, the mAP value obtained from our improvements was significantly higher than those from the other methods, validating the effectiveness of our approach.

TABLE V
COMPARISON OF ALGORITHM PERFORMANCE

Model	mAP	GFLOPS	FPS
SSD	0.620	15.3	263
Faster R-CNN	0.725	31.2	183
YOLOv5s	0.939	16.4	271
YOLOv5m	0.942	29.7	192
YOLOv5l	0.944	46.8	165
YOLOv5x	0.949	85.0	124
YOLOv3	0.763	17.7	201
YOLOv7	0.857	72.1	130
YOLOv8m	0.922	34.3	174
YOLOv8n	0.876	8.6	306
Improved Model	0.974	3.7	329

V CONCLUSION

This paper proposes a body-in-white weld spot appearance defect detection algorithm based on YOLOv5. First, the PConv convolution is combined with traditional convolution operations to create the C3-Faster module, which replaces the original C3 module in the network, reducing redundant computation and memory access, thereby enabling the model to more effectively extract spatial features. Second, the WIoUv3 loss function is introduced, incorporating a dynamic non-monotonic FM (static focusing mechanism) named Wise-IoU that evaluates anchor box quality based on outlier degree instead of IoU and applies a gradient gain allocation strategy to reduce harmful gradients from low-quality examples. Finally, ablation experiments reveal that while the C3-Faster module successfully reduces the model size, it slightly decreases accuracy. However, when combined with the WIoUv3 loss function, accuracy improves more significantly than embedding WIoUv3 alone in the original model. The improved network was also compared with several commonly used networks in terms of parameter size and detection speed, demonstrating the superiority of the lightweight design. Further, experiments comparing several common improvement methods to the addition of the WIoUv3 loss function revealed the effectiveness of WIoUv3 in enhancing detection accuracy for the body-in-white weld spot dataset.

The proposed algorithm demonstrated superior performance on the body-in-white weld spots dataset. In terms of detection accuracy, the mAP was on average 12.03% higher than that of ten other commonly used models, 3.5% higher than the original YOLOv5s model, and 9.8% higher than the latest YOLOv8 model. Regarding computational speed, both parameter size and detection speed were better than those of the ten models, with the parameter count reduced by 77.7% and detection speed increased by 9.0% compared to the original model.

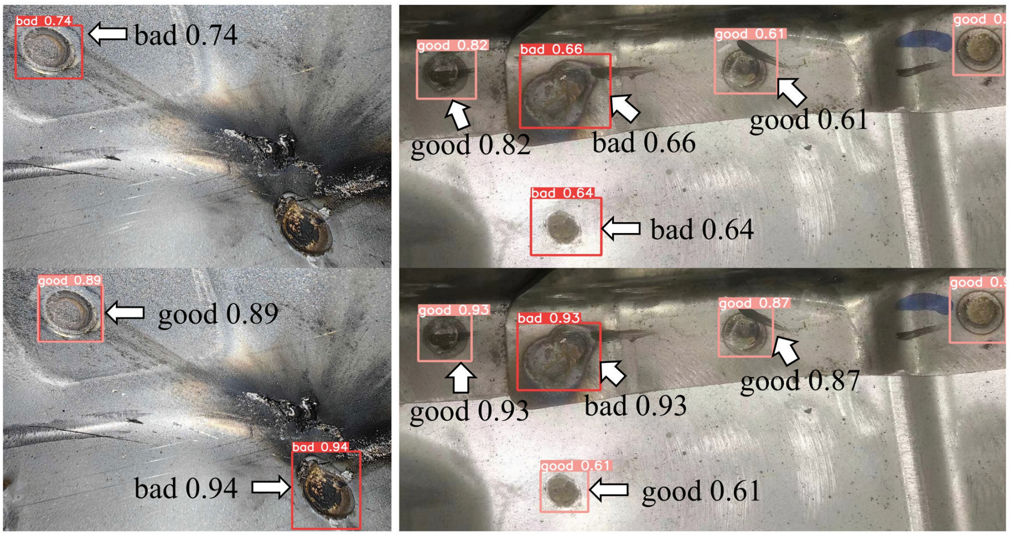


Fig. 10. Comparison of detection results between the original network and the improved network using images collected in the factory (The original network detection results are presented above, and the improved network detection results are presented below)

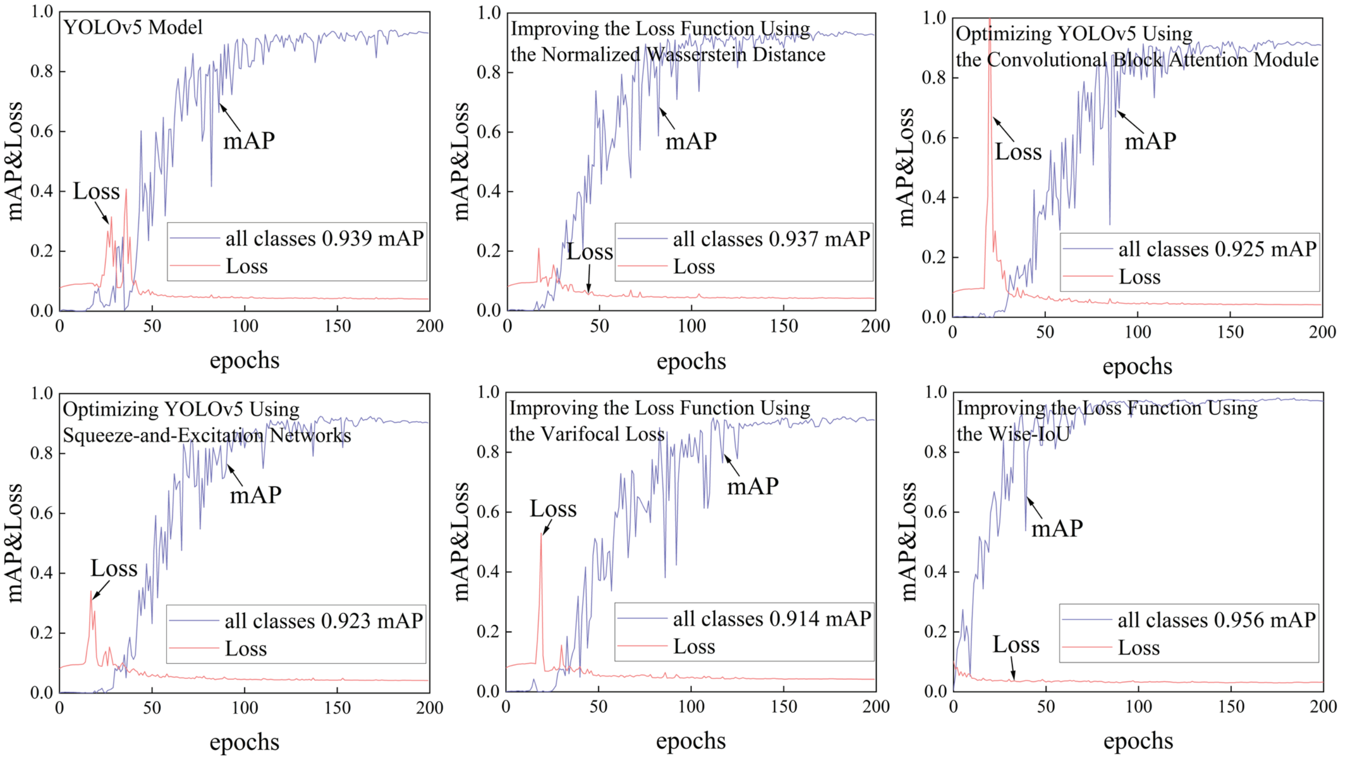


Fig. 11. Comparison Chart of mAP with Five Improvement Methods Added to YOLOv5 Network

REFERENCES

[1] P. LIU, “Detection method of spot welding based on multi-information fusion and fractal,” Chinese Journal of Mechanical Engineering (English Edition), pp. 76, 2008.

[2] D. Bakavos, P. B. Prangnell, “Mechanisms of joint and microstructure formation in high power ultrasonic spot welding 6111 aluminium automotive sheet,” Materials Science and Engineering a-Structural Materials Properties Microstructure and Processing, vol. 527, no. 23, pp. 6320-6334, 2010.

[3] X. K. Wang, S. Y. Guan, L. Hua, B. Wang, and X. M. He, “Classification of spot-welded joint strength using ultrasonic signal time-frequency features and PSO-SVM method,” Ultrasonics, vol. 91, pp. 161-169, 2019.

[4] S. Liu, D. Erdahl, I. C. Ume, A. Achari, and J. Gamalski, “A novel approach for flip chip solder joint quality inspection: Laser ultrasound and interferometric system,” Ieee Transactions on Components and Packaging Technologies, vol. 24, no. 4, pp. 616-624, 2001.

[5] I. N. Ermolov, “Progress in the theory of ultrasonic flaw detection. Problems and prospects,” Russian Journal of Nondestructive Testing, vol. 40, no. 10, pp. 655-678, 2004.

[6] M. Pouranvari, and S. P. H. Marashi, “Critical review of automotive steels spot welding: process, structure and properties,” Science and Technology of Welding and Joining, vol. 18, no. 5, pp. 361-403, 2013.

[7] W. Dai, D. Y. Li, D. Tang, Q. Jiang, D. Wang, H. M. Wang, and Y. H. Peng, “Deep learning assisted vision inspection of resistance spot welds,” Journal of Manufacturing Processes, vol. 62, pp. 262-274, 2021.

[8] J. Sun, P. Wang, Y. K. Luo, and W. Li, “Surface Defects Detection Based on Adaptive Multiscale Image Collection and Convolutional Neural Networks,” Ieee Transactions on Instrumentation and Measurement, vol. 68, no. 12, pp. 4787-4797, 2019.

[9] Y. P. Gao, L. Gao, X. Y. Li, and X. G. Yan, “A semi-supervised convolutional neural network-based method for steel surface defect recognition,” Robotics and Computer-Integrated Manufacturing, vol. 61, 2020.

[10] Z. Q. Lin, L. J. Zhu, J. Y. Zhang, Y. H. Zhang, and X. D. Liu, “Research on Improving YOLOv5s Algorithm for Defect Detection in Cylindrical Coated Lithium-ion Batteries,” Engineering Letters, vol. 32, no. 7, pp. 1521-1528, 2024.

[11] Z. X. Zhang, W. H. Cui, Y. Tao, and T. W. Shi, “Road Damage Detection Algorithm Based on Multi-scale Feature Extraction,” Engineering Letters, vol. 32, no. 1, pp. 151-159, 2024.

[12] Y. Z. Fu, L. Qiu, X. Kong, and H. F. Xu, “Deep Learning-Based Online Surface Defect Detection Method for Door Trim Panel,” Engineering Letters, vol. 32, no. 5, pp. 939-948, 2024.

[13] J. Lei, X. Gao, Z. L. Feng, H. M. Qiu, and M. L. Song, “Scale insensitive and focus driven mobile screen defect detection in industry,” Neurocomputing, vol. 294, pp. 72-81, 2018.

[14] Y. F. Pan, R. S. Lu, and T. D. Zhang, “FPGA-accelerated textured surface defect segmentation based on complete period Fourier reconstruction,” Journal of Real-Time Image Processing, vol. 17, no. 5, pp. 1659-1673, 2020.

[15] Z. Mo, L. Chen, and W. You, “Identification and Detection of Automotive Door Panel Solder Joints based on YOLO,” Chinese Control And Decision Conference (CCDC), pp. 5956-5960, 2019.

[16] L. Li, G. Shi, and T. Jiang, “Fish detection method based on improved YOLOv5,” Aquacultural Engineering, vol. 31, no. 5, pp. 2513-2530, 2023.

[17] J. Chen, S.-h. Kao, H. He, W. Zhuo, S. Wen, C.-H. Lee, and S.-H. G. Chan, “Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks,” IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp. 12021-12031, 2023.

[18] Z. Tong, Y. Chen, Z. Xu, and R. Yu, “Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism,” Computer Vision and Pattern Recognition (cs.CV), 2023.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[20] X. K. Zhu, S. C. Lyu, X. Wang, Q. Zhao, and I. C. Soc, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," IEEE International Conference on Computer Vision Workshops. pp. 2778-2788, 2021.

[21] W. Rawat, and Z. Wang, “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,” Neural Computation, pp. 2352-2449, 2017.

[22] B. Yan, P. Fan, X. Y. Lei, Z. J. Liu, and F. Z. Yang, “A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5,” Remote Sensing, vol. 13, no. 9, 2021.

[23] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” Computer Vision and Pattern Recognition (cs.CV), 2020.

[24] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, and I. C. Soc, "Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression," IEEE Conference on Computer Vision and Pattern Recognition. pp. 658-666, 2019.

[25] Z. H. Zheng, P. Wang, W. Liu, J. Z. Li, R. G. Ye, D. W. Ren, and I. Assoc Advancement Artificial, "Distance-IoU Loss: Faster and Better

Learning for Bounding Box Regression," AAAI Conference on Artificial Intelligence. pp. 12993-13000, 2020.

[26] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” International Journal of Robotics Research, vol. 32, no. 11, pp. 1231-1237, 2013.

[27] J. Hu, L. Shen, S. Albanie, G. Sun, and E. H. Wu, “Squeeze-and-Excitation Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011-2023, 2020.

[28] H. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," Lecture Notes in Computer Science. pp. 3-19, 2018.

[29] C. Xu, J. W. Wang, W. Yang, H. Yu, L. Yu, and G. S. Xia, “Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark,” Isprs Journal of Photogrammetry and Remote Sensing, vol. 190, pp. 79-93, 2022.

[30] H. Y. Zhang, Y. Wang, F. Dayoub, N. Sunderhauf, and S. O. C. Ieee Comp, "VarifocalNet: An IoU-aware Dense Object Detector," IEEE Conference on Computer Vision and Pattern Recognition. pp. 8510-8519, 2021.