

# YOLO-WP: A Lightweight and Efficient Algorithm for Small-Target Detection in Weld Seams of Small-Diameter Stainless Steel Pipes

Huaishu Hou, Yukun Sun\*, Chaofei Jiao

School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai, 201418, China

**Abstract**—To address the low detection efficiency and high computational resource demands of current welded pipe defect detection algorithms for small target defects, this paper proposes the YOLO-WP algorithm based on YOLOv5s. The improvements of YOLO-WP are mainly reflected in the following aspects: First, an innovative GhostFusion architecture is introduced in the backbone network. By replacing the C3 modules with C2f modules and integrating the Ghost CBS module inspired by Ghost convolution, cross-stage feature fusion is achieved, significantly enhancing computational efficiency and feature representation for small target defects. Second, the Slim-Neck lightweight design based on GSConv is employed in the neck to further optimize the network structure and reduce the number of parameters. Additionally, the SimAM lightweight attention mechanism is incorporated to improve the network's ability to extract defect features, and the Focal-EIoU loss is utilized to optimize CIoU loss, thereby enhancing small object detection and accelerating loss convergence. The experimental results show that the AP(D1) and mAP@0.5 of the YOLO-WP model are improved by 5.3% and 3%, respectively, over the original model. In addition, the number of model parameters and FLOPs are reduced by 40% and 45%, respectively, achieving a good balance between performance and efficiency. We evaluated the performance of YOLO-WP using other datasets and showed that YOLO-WP exhibits excellent applicability. Compared to existing mainstream detection algorithms, YOLO-WP is more advanced. The YOLO-WP model significantly enhances production quality in industrial defect detection, laying the foundation for building compact, high-performance embedded weld pipe surface defect detection systems.

**Keywords**—Welded pipe; lightweight model; defect detection; deep learning; feature extraction; attention mechanism

## I. INTRODUCTION

Small-diameter stainless steel welded pipes are widely used across various fields, including oil and gas transportation, chemical production, medical equipment, and automotive components [1]. The widespread use is due to their excellent welding performance, lower manufacturing cost compared to seamless pipes, small diameter, lightweight design, high strength characteristics [2], and superior corrosion resistance [3-4]. The welding of small-diameter stainless steel pipes primarily employs the Gas Tungsten Arc Welding (GTAW) technique. In practical production, two common defects are observed on the weld surface. The first is weld voids, which can result from misalignment or omission during the welding process [5]. The second is welding porosity or pits, typically caused by tungsten electrode contamination or grinding wheel damage [6]. These

defects can significantly impact the performance of the pipes. Therefore, it is essential to perform real-time inspection of the weld seams of small-diameter stainless steel welded pipes on the production line. This approach enables the timely detection and correction of defects, preventing defective products from proceeding to subsequent offline inspection stages and avoiding unnecessary quality control costs.

Current methods for online inspection of weld surface defects in stainless steel pipes mainly include manual inspection, X-ray inspection, eddy current inspection, and ultrasonic inspection [7]. Manual inspection is considered inefficient [8]. X-ray and ultrasonic inspections require high operational skills from personnel [9]. Additionally, eddy current inspection signals are vulnerable to interference from external factors [10]. Therefore, there is an urgent need for a detection method that is efficient, easy to operate, and highly resistant to interference. Machine vision, a non-destructive testing method, can fulfill this requirement. Within recent years, with the progress of machine vision and deep learning technologies, numerous defect detection methodologies leveraging these approaches have been extensively applied in various inspection contexts, including food and agriculture, electronics fabrication, metal materials, the semiconductor sector and healthcare [11-15]. Nevertheless, machine vision-based inspection methods for detecting weld surface defects on small-diameter stainless steel welded pipes are still not widely used.

Employing machine vision for inspection not only eliminates human subjectivity but also enables quantitative defect descriptions, thereby minimizing variability in the results. This innovation enhances detection efficiency and accuracy, fostering the advancement of industrial automation. Currently, mainstream detection algorithms can be categorized into two types [16]: single-stage algorithms and two-stage algorithms. Advanced single-stage object detection algorithms include the YOLO series, DETR, SSD and CenterNet. Advanced two-stage object detection algorithms include Faster R-CNN, Mask R-CNN, Libra R-CNN and HTC. Yang [17] applied the YOLOv5 algorithm to the welded pipe defect detection and achieved excellent results. Compared to the representative two-stage object detection algorithm Faster R-CNN, YOLOv5 demonstrates superior precision and detection speed. Therefore, different algorithms should be selected and modified for different application scenarios to target specific tasks. There are many existing strategies for algorithm improvement that focus on optimizing model components to achieve desired outcomes. Zhou et al. [18]

introduced a novel model in the YOLOv5 algorithm that combines the advantages of the CSPPlayer module with a global attention enhancement mechanism, improving accuracy for metal material detection. However, this approach increases model complexity and demands higher computational resources. Zhao et al. [19] enhanced the Faster R-CNN algorithm by replacing some traditional convolutional networks with deformable convolutional networks to improve the detection capability for small-size defects on steel strips. However, this approach leads to a significant increase in model parameters, making deployment more challenging and less suitable for direct application in industrial settings. Shao et al. [20] proposed the TD-Net network for detecting tiny defects in industrial products, addressing the limitations of current image-based defect detection methods in identifying small and irregularly shaped defects. However, the detection speed has decreased. Ji et al. [21] introduced the Yolo-tla algorithm, which integrates the C3CrossConv module into the YOLOv5 backbone, effectively reducing computational demand and parameter count, thus making the model more lightweight. However, the detection accuracy has slightly decreased. Han et al. [22] proposed the DFW-YOLO algorithm based on YOLOv5, which automatically calls defect indications, resolves redundant defect feature maps, and incorporates the FasterNet backbone to enhance the model's feature extraction capability. However, the detection accuracy for small target defects has decreased. Zhou et al. [23] proposed the SKS-YOLO algorithm, using EfficientNetv2 as the backbone, which significantly reduces computation and accelerates training speed while maintaining accuracy, and employs the Simplified Intersection over Union (SIoU) loss function to improve the model's capability in locating and detecting surface defects on steel plates. However, the ability to extract small target or high-frequency features has decreased in certain scenarios. Yuan et al. [24] presented the YOLO-HMC algorithm, which uses the HorNet network (MCBAM) as its backbone and incorporates an improved multi-convolution block attention module to enhance feature extraction capabilities. However, the model requires more computational resources. Despite these studies achieving breakthroughs in specific scenarios, existing algorithms still face challenges in balancing the detection accuracy of small targets, model lightweighting, and real-time performance.

However, the existing methods for detecting surface defects in small-diameter stainless steel welded pipes still face these challenges. Firstly, the irregular sizes of weld hole defects and the extremely small shapes of welding porosities further complicate the detection process. Secondly, the high computational resources required by deep learning models pose limitations on their application in online detection within actual production environments. To address these challenges, the online detection of surface defects in small-diameter stainless steel welded pipes demands models with high speed, accuracy, and ease of deployment. To this end, we developed an enhanced model called YOLO-WP. This study focuses on the following key areas:

1) *Network structure optimization*: By optimizing the deep learning network structure, we enhance the detection accuracy for small target defects and improve overall detection performance. This is achieved by removing redundant structures and modules within the network to design a lightweight model,

thereby increasing efficiency in resource-constrained environments while maintaining or even improving model performance.

2) *Incorporation of attention mechanism*: By introducing attention mechanisms, we improve detection accuracy and enhance the ability to detect small targets. Attention mechanisms allow the model to focus on critical features and regions, thereby improving the detection of small and complex defects.

3) *Loss function improvement*: By optimizing the loss function and tailoring it to the characteristics of the dataset, we further enhance the model's localization accuracy and classification performance for small target defects. This leads to improved overall detection performance.

## II. ALGORITHM DESCRIPTION

### A. Baseline YOLOv5s

The YOLO series of algorithms has evolved through several versions, with YOLOv5 being widely used in the field of industrial real-time detection caused by its excellent detection accuracy and outstanding detection speed. Innovations such as the CSPDarknet53 backbone, Feature Pyramid Network (FPN), adaptive anchor box computation, and advanced data augmentation techniques have enhanced the model's performance and flexibility. YOLOv5 offers multiple versions of the model, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, to accommodate different computational needs and scenarios [25]. Among these, the YOLOv5s model features the smallest network depth and width, and the fastest detection speed, making it relatively well-suited for the requirements of industrial online detection. Accordingly, this paper selects YOLOv5s as the base model.

### B. The Overview of YOLO-WP

Although YOLOv5s demonstrates advantages in speed, accuracy, and terminal applications, it still faces certain limitations in practical use, such as suboptimal localization accuracy, and high computational resource demands [26]. In particular, it often suffers from missed detections and inaccurate target localization when detecting small-scale defects. Hence, this paper proposes an online detection model named YOLO-WP, specifically designed for detecting weld seam surface defects in small-diameter stainless steel pipes. The design aims of the YOLO-WP model have two aspects: first, to improve operational efficiency in resource-constrained environments, and second, improve its capability in handling small-sized defects. Although the optimization and validation of this model primarily target the detection of weld seam surface defects in small-diameter stainless steel pipes, its architectural design and improvement strategies are equally applicable to other fields requiring efficient object detection. The structure of YOLO-WP model is illustrated in Fig. 1. The targeted improvements include:

1) *In the backbone network*, the paper proposes the GhostFusion architecture, which enhances feature expression through multiscale cross-stage fusion while maintaining efficient computation. In the neck network section, a lightweight Slim-Neck network, based on GSConv [27], is referenced to reduce network parameters and enhance computational resource efficiency.

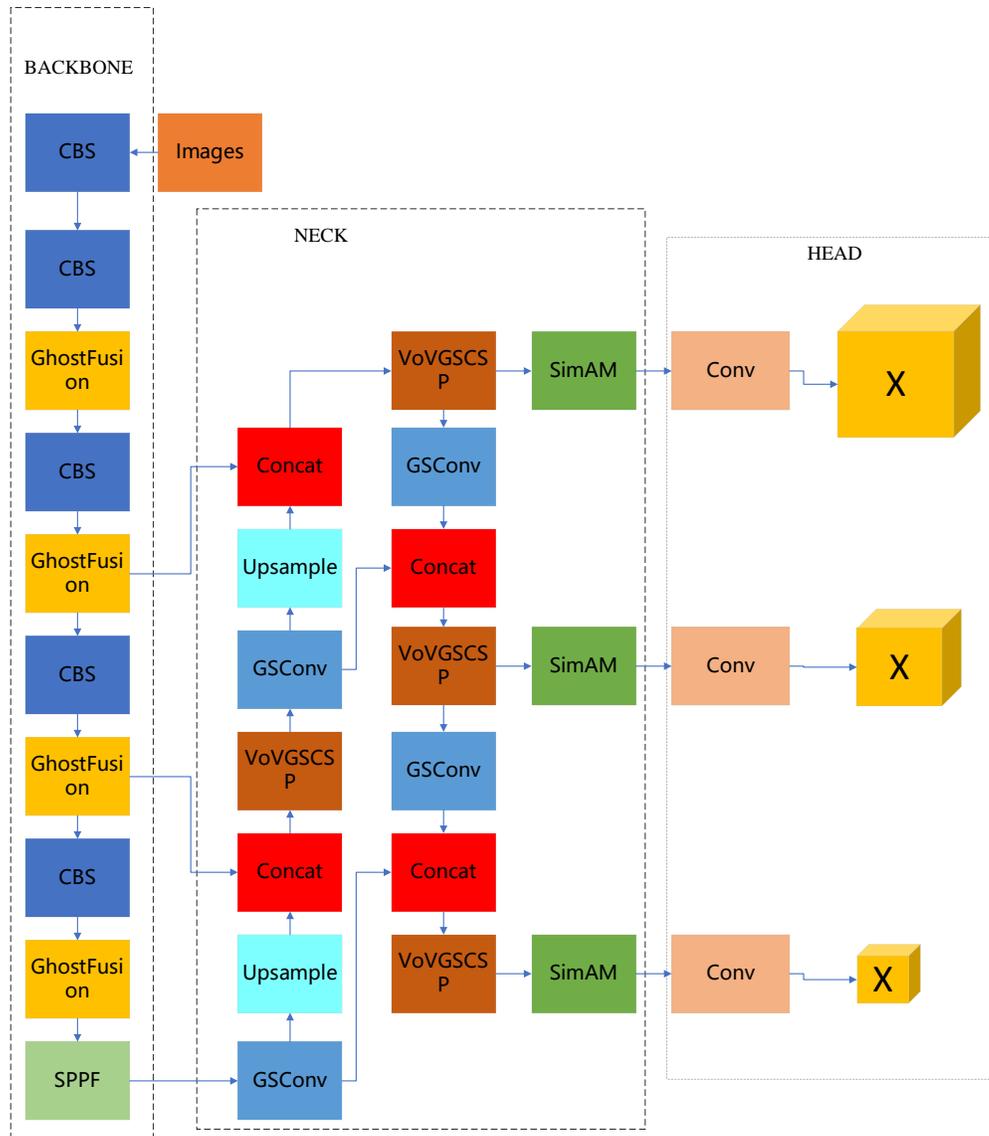


Fig. 1. Schematic model of YOLO-WP.

2) Adding the lightweight SimAM attention mechanism [28] to the neck network helps the model focus on key areas of the image, improves the fusion of feature maps across different scales, and enhances detection accuracy.

3) The Focal-EIOU loss [29] is utilized to substitute the original CIOU loss, aiming to enhance the detection of small-scale defects, address sample imbalance, and improve robustness on small and noisy datasets.

### III. STRUCTURE OF KEY IMPROVEMENT COMPONENTS

#### A. GhostFusion Architecture

In the design of the backbone network, the innovative GhostFusion architecture is proposed. The construction process involves replacing the C3 modules with C2f modules and optimizing the CBS module in the C2f module by borrowing the core idea of Ghost convolution, and innovatively proposing the

Ghost CBS module, as depicted in Fig. 2. This design significantly enhances computational efficiency and feature representation capability through cross-stage fusion, achieving resource savings and improving detection performance for small target defects.

The core idea of Ghost Conv is to decompose conventional convolution operations into two stages: the main convolution stage and the ghost convolution stage. In the main convolution stage,  $1 \times 1$  convolution kernels are used to extract condensed features, while in the ghost convolution stage, cheap  $5 \times 5$  convolution kernels generate the remaining feature maps. The complete feature layer is formed by concatenating these two parts [30]. The detailed operation process is shown in Fig. 3. Additionally, the C2f module splits the input data into two parts through a Split operation. One part is directly retained, while the other part is processed through multiple BottleNeck structures to achieve multi-scale feature fusion [31].

The design greatly improves the computational efficiency and feature representation capability through cross-stage fusion, realizes resource saving, and improves the detection performance of small target defects. The core of the GhostFusion architecture lies in cross-stage feature fusion. By combining the Ghost CBS module with the C2f module, not only the computational efficiency is greatly improved, but also the semantic information of the features is significantly enhanced through the dynamic fusion of multi-scale features. This fusion mechanism enables the model to handle different sizes of receptive fields simultaneously, thus extracting more comprehensive feature information in complex scenes. The GhostFusion architecture performs well in the small target defect detection task. Through the synergy of the optimized Ghost CBS module and the C2f module, the model is able to efficiently extract features of small targets and further enhance the expressiveness of these features through the cross-stage fusion mechanism. This design not only improves the detection accuracy, but also ensures that the model can maintain efficient operation in resource-constrained scenarios. The GhostFusion architecture is designed with resource conservation and performance balance in mind. By introducing the efficient features of Ghost convolution, the

number of parameters and computation amount of the model can be significantly reduced, and at the same time, through the cross-stage fusion mechanism, the feature expression ability of the model is further enhanced. This design not only improves the operational efficiency of the model, but also ensures its high performance in complex tasks.

*B. Slim-Neck Structure Based on GSConv Module*

To elevate the model’s detection speed and computational efficiency, this paper adopts a Slim-Neck structure based on the lightweight convolutional GSConv module [32]. Compared to traditional convolution operations, GSConv reduces the number of model parameters and computational complexity load by dividing the input features into multiple groups and independently performing convolution operations and depthwise separable convolutions on each group. GSConv introduces a mechanism that focuses on important feature channels, thus improving the model’s feature extraction capability. The core principle of GSConv is to combine the characteristics of depthwise separable convolutions (DSC) and standard convolutions (SC) to achieve efficient feature map fusion and information flow. The structure of GSConv is shown in Fig. 4.

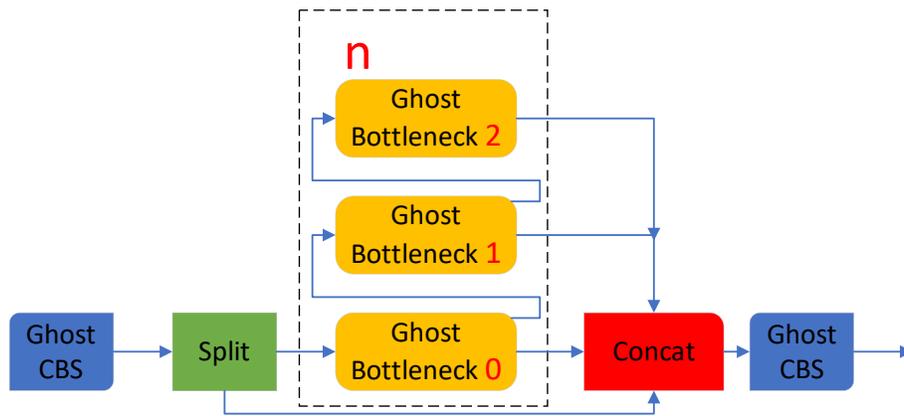


Fig. 2. Schematic diagram of the GhostFusion architecture.

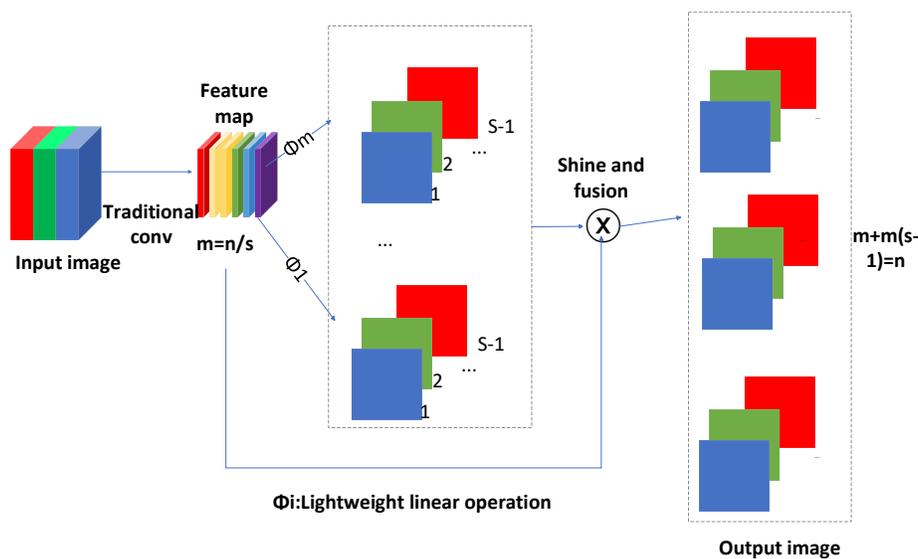


Fig. 3. GhostConv module.

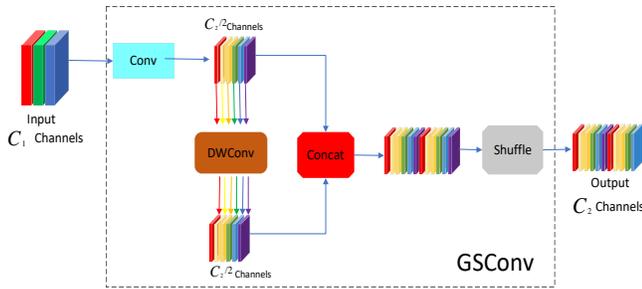


Fig. 4. The structure of GSConv.

First, GSConv inputs a downsampled standard convolution, followed by a depthwise convolution (DWConv), which concatenates the depthwise standard convolution (SC) and the depthwise separable convolution (DSC). Subsequently, a shuffle operation is applied to align the DSC output with the SC output, preserving channel and semantic information in the feature map.

The Slim-Neck lightweight neck network structure can effectively fuse and enhance features while maintaining the model's detection performance, even as it reduces computational load and the number of parameters.

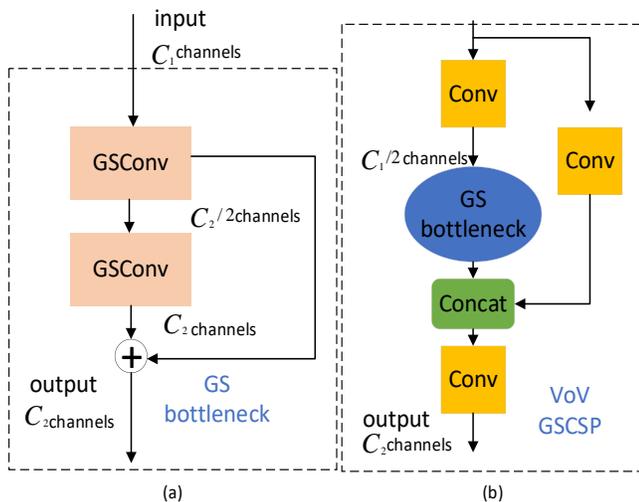


Fig. 5. Slim-Neck structure: (a) GSbottleneck module; (b) VoV-GSCSP module.

The basic building block of Slim-Neck is called VoV-GSCSP, which can replace the CSP layers comprised of standard convolutions. Among its components is the GSbottleneck, which uses GSConv as its building block, as shown in Fig. 5(a). The VoV-GSCSP module uses a one-shot aggregation method, shown in Fig. 5(b), to improve the model's target detection across different sizes by merging multi-scale feature maps, while reducing computational load and complexity.

### C. SimAM Attention Mechanism

To enhance the model's representation of key features and improve its detection performance, this paper introduces an attention mechanism into the neck network of the model, thereby improving the Neck section and boosting the model's robustness. The attention mechanism typically includes channel atten-

tion and spatial attention. The parameter-free attention mechanism SimAM adopted in this paper combines both, as shown in Fig. 6. Adjacent pixels in an image usually have strong similarities, while distant pixels have weaker similarities. SimAM generates attention weights by calculating the similarity between each pixel and its neighbors in the feature map, thus inferring three-dimensional attention weights for the feature map. This effectively integrates channel and spatial attention, significantly improving the model's detection performance [33].

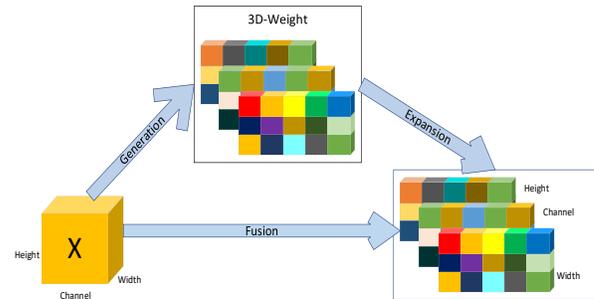


Fig. 6. The architecture of the SimAM attention mechanism module.

SimAM is grounded in the theory of visual neuroscience, where neurons with more information tend to exhibit more prominent activity compared to their adjacent neurons. In the task of surface defect detection for small-diameter stainless steel welded tubes, these neurons typically extract key features and should be assigned higher weights. This paper introduces the SimAM attention mechanism into the Neck network of the YOLOv5s model to optimize feature fusion and enhancement, while balancing network width, depth, and detection speed, thereby improving the accuracy of surface defect detection without increasing the network parameters. As shown in Eq. (1) ~ (4), SimAM evaluates neurons using an energy function for linear separability, where  $t$  represents the target neuron,  $x$  represents the adjacent neurons, and  $\lambda$  is a hyperparameter. The lower the energy  $e_t^*$ , the higher the distinguishability and importance of the neuron. As Eq. (4) shows, neurons are weighted based on their importance using  $\frac{1}{e_t^*}$ . SimAM assesses the importance of features using the energy function, providing higher interpretability and without introducing additional learnable parameters.

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

$$\hat{\mu} = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i \quad (2)$$

$$\hat{\sigma}^2 = \sum_{i=1}^{M-1} (x_i - \hat{\mu})^2 \quad (3)$$

$$\bar{X} = \text{sigmoid} \left( \frac{1}{E} \right) \odot X \quad (4)$$

### D. Improvement of the Loss Function

The loss function determines the degree of agreement between the true values and the predicted values, and its performance largely reflects the model's effectiveness. In the YOLO algorithm, there are three types of loss functions: classification loss, confidence loss, and localization loss. Among these, the localization loss represents the error between the predicted bounding box and the ground truth bounding box. This paper

conducts research and improvement on the localization loss function. The original localization loss function of YOLOv5s is CIOU loss, which is centered around the concept of calculating the positional alignment error between the ground truth bounding box and the predicted bounding box based on the size of the IoU (Intersection over Union). The calculation process is expressed as:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (5)$$

Where  $b$  and  $b^{gt}$  represent the centroids of the prediction frame and the true frame respectively,  $\rho$  represents the computation of the Euclidean distance between the two centroids, so  $\rho^2(b, b^{gt})$  is the distance between the centroid of the prediction frame and the centroid of the defective true bounding box, and  $c$  represents the diagonal distance of the smallest closed region that can contain both the prediction frame and the true frame.  $\alpha v$  denotes the aspect ratio between the prediction frame and the true bounding box.

From Eq. (5), it is evident that despite CIOU loss function considering the overlap area, distance between center points, and aspect ratio of the regression bounding boxes, there are still some issues. Specifically, it adjusts based solely on the aspect ratio without considering the specific values of width and height. Additionally, the gradients of width and height have opposite signs, preventing simultaneous increase or decrease, leading to potentially amplifying the width or height during optimization when both the width and height of the anchor box are greater than the defect to be detected. Therefore, this paper selects the EIOU loss function, which modifies the aspect ratio adjustment in the CIOU loss function to specific width and height regression, enabling the model to converge faster and achieve higher accuracy, hence improving the detection efficiency for small target defects. The formula is as follows:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (6)$$

Where  $w, h$  represents the width and height of the predicted box,  $w^{gt}, h^{gt}$  represent the width and height of the real box,  $C_w$  and  $C_h$  are the width and height of the smallest outer box that covers both boxes.

In this paper, variations in viewing angles and lighting may impact dataset quality during defect identification. In addition, the irregular appearance of defects complicates precise manual labelling, resulting in imperfect alignment between the aiming frame and defects. These factors cause dramatic fluctuations in loss value when training on low-quality samples, which can severely affect model performance. The goal of the proposed Focal II is to address the imbalance between high- and low-quality samples. Balance problem, and combined with EIOU loss to form Focal-EIOU loss.

$$L_{Focal-EIOU} = IOU^\gamma L_{EIOU} \quad (7)$$

Where  $\gamma$  is a constant, with a verified  $\gamma$  value of 0.5 giving the best results [34].

## IV. EXPERIMENTAL PROCEDURE DESIGN

### A. Experimental Platform and Parameter Design

During the model experiments, to maintain consistency with the comparison models and ensure the comparability of the experimental results, the SGD optimizer was used with an initial learning rate of 0.01, a momentum of 0.935, and a weight decay coefficient of 0.0005. The experiment was conducted with a batch size of 16 across 100 epochs. The experimental environment configuration is shown in Table I, and all experiments in this paper were conducted using this configuration.

TABLE I. EXPERIMENTAL ENVIRONMENT CONFIGURATION

Category	Configuration
CPU	Intel® Core™ i5-12490F Processor
GPU	NVIDIA GeForce RTX 3070ti 8G
RAM	32G
Operation System	Windows 10
Framework	PyTorch 2.0.0
Programming environment	Python 3.9
CUDA	11.8

### B. Experimental Datasets

In this paper, we selected stainless steel welded pipes with defects, produced in actual manufacturing, with diameters ranging from 7mm to 9mm as samples. The defects studied in this paper are shown in Fig. 7. Due to the high visual similarity between welding sand holes and welding pores on these small-diameter stainless steel welded pipes, we categorized them as a single class.

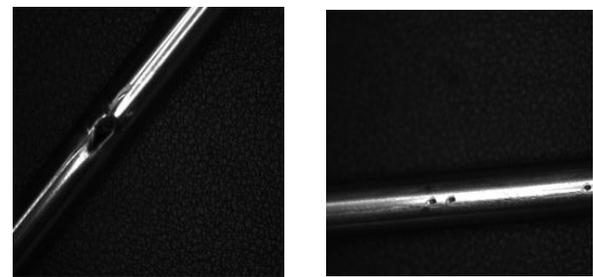


Fig. 7. The two types of defects studied in this paper.

In the absence of publicly available datasets dedicated to this field, this study established an experimental platform for image acquisition. An industrial matrix camera (model MV-CS004-10GM) was employed, precisely positioned above the welded pipe and aligned with the center of the annular aperture to ensure consistency and high quality in image acquisition. The smooth surface of the pipe reflects light centrally, resulting in brighter tones in the image, whereas weld void defects scatter the light due to their surface characteristics, creating darker areas that clearly outline the defects. Additionally, this study utilized an adjustable-brightness annular LED aperture as the light source and integrated a slider motor to accurately adjust the distance between the welded pipe and the camera, enabling rapid and precise focusing for pipes of different specifications and enhancing the flexibility and adaptability of the system. The specific setup is shown in Fig. 8.

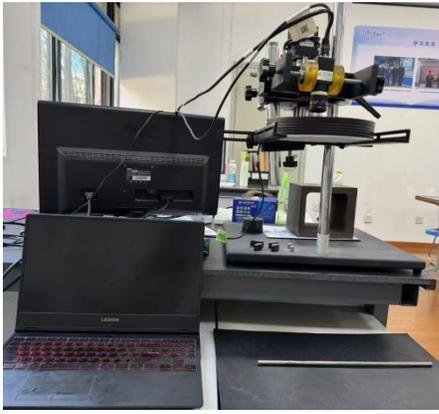


Fig. 8. Image acquisition platform.

To demonstrate the improved model's generalization capability, this study captured images of two defects types under varying lighting conditions, angles, distances, and focal lengths. For each type of defect, 2,000 images were selected to form the training set and 600 images for the validation set, with all images having a resolution of  $640 \times 640$  pixels. Subsequently, the images were annotated, and the number of defect instances in both the training and validation sets was statistically analyzed. Detailed data are presented in Table II.

TABLE II. DEFECT DATA LABELING STATISTICS

Defect Type	Dataset Labels	
	Training Dataset	Testing Dataset
D: weld hole	2866	784
D1: weld porosity and weld blister	2995	813

## V. EXPERIMENTAL RESULTS AND ANALYSES

### A. Algorithm Evaluation Metrics

Accuracy P, mean accuracy mAP, recall R, parameters, model complexity (FLOPs) and FPS were used to evaluate the performance of the model.

$$P = \frac{TP}{TP+FP} \quad (8)$$

$$R = \frac{TP}{TP+FN} \quad (9)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{C} \quad (10)$$

Where TP is the number of correctly detected defects; FP is the number of incorrectly detected defects; FN is the number of undetected defects; AP denotes the accuracy of the detection; the value of mAP is obtained by averaging all the category APs; and C is the total number of detected categories.

When larger mAP and P values indicate higher detection accuracy, smaller parameters and FLOPs reflect a more light-weight model, and higher FPS reflects the faster algorithm detection speed.

### B. YOLO-WP Ablation Experiments

To validate the effectiveness of the proposed improvements, this study conducted ablation experiments using YOLOv5s as the baseline model, incrementally adding improvement modules across six sets of experiments, as shown in the Table III. Compared to the first experimental group, the second group, after incorporating the GhostFusion efficient feature fusion network structure, saw an improvement of 3.2% in AP(D1) and 1.6% in mAP@0.5. Additionally, the number of parameters and FLOPs decreased by 22.2% and 33.5%, respectively, which not only enhanced computational efficiency but also improved the detection capability for small target defects. Compared to the second group, the third group introduced the Slim-Neck architecture based on the lightweight convolutional GSCConv module into the neck network. This resulted in a further reduction of 27.1% in parameters and 17.4% in FLOPs while maintaining the model's detection performance. Compared to the third set, the fourth set introduced the lightweight SimAM attention mechanism in the neck network. Although FPS decreased by 2.2, there was no increase in parameters and FLOPs. Meanwhile, AP(D), AP(D1), and mAP@0.5 improved by 0.5%, 1.2%, and 0.8%, respectively. In the fifth set, the Focal-EIOU loss was used to optimize CIOU loss, further improving the model's localization accuracy and convergence speed compared to the fifth set. AP(D), AP(D1), and mAP@0.5 increased by 0.5%, 0.8%, and 0.7%, respectively, with mAP@0.5 reaching 96.6%. The ablation experiments from the second to the fifth set demonstrate that each improvement method effectively optimizes the model. Compared to the baseline model YOLOv5s in the first set, the improved YOLO-WP model (sixth set) achieved a 5.3% increase in AP(D1) and a 3% increase in mAP@0.5, while reducing parameters and FLOPs by 40% and 45%, respectively. The YOLO-WP model efficiently detects small target defects with higher performance and lower computational cost, thereby enhancing overall detection accuracy. Although FPS decreased by 2.6, it still meets the requirements for online detection of surface defects in small-diameter stainless steel welded pipe seams.

TABLE III. ABLATION EXPERIMENTS

Group Model		AP(%)		P(%)	mAP@0.5 (%)	Parameters(106)	FLOPs (G)	FPS (F/S)
		D	DI					
1	YOLOv5s(Baseline)	97.1	90.3	93.3	93.7	7.2	15.5	105.3
2	YOLOv5s+GhostFusion	97.3	93.2	94.1	95.2	5.9	10.3	104.9
3	YOLOv5s+GhostFusion+Slim-Neck	97.1	93.1	93.9	95.1	4.3	8.5	103.7
4	YOLOv5s+GhostFusion+Slim-Neck+SimAM	97.6	94.3	94.5	95.9	4.3	8.5	101.5
5	YOLOv5s+GhostFu- sion+SlimNeck+SimAM+Focal-Elou	98.1	95.1	94.7	96.6	4.3	8.5	102.7

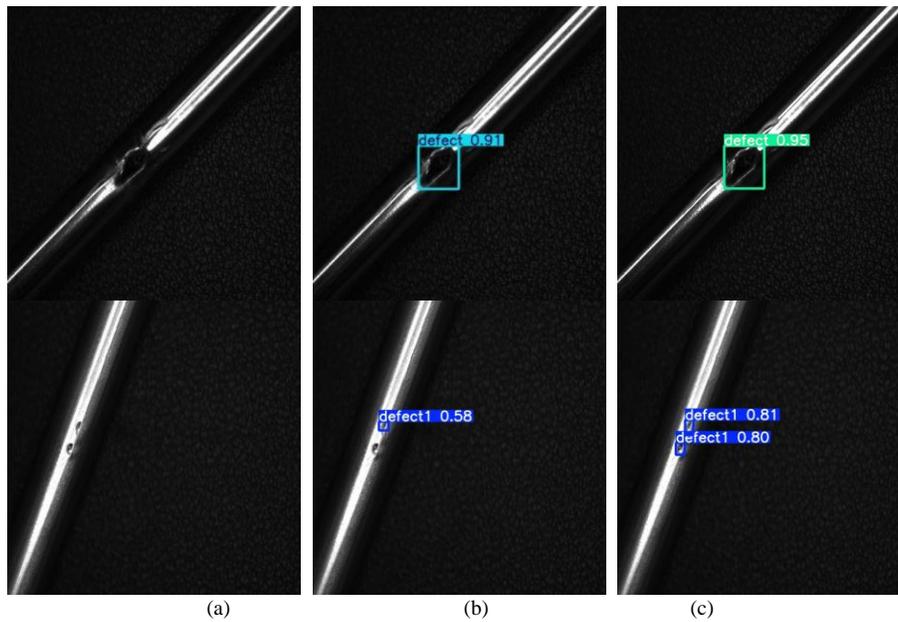


Fig. 9. Validation results for YOLO-WP and YOLOv5s. (a) Original image; (b) YOLOv5s; (c) YOLO-WP.

To offer a clearer illustration of the detection performance of the enhanced model, two images with two types of defects were randomly selected from the test set to evaluate the YOLO-WP model. This paper provides a visual comparison of the detection outcomes of the YOLO-WP model and the YOLOv5s model on images of small-diameter stainless steel welded pipes. As shown in Fig. 9, the YOLOv5s algorithm model exhibited missed detections and low accuracy in detecting small target defects. In contrast, the improved YOLO-WP algorithm model accurately located the weld seam surface defects within the images, demonstrating superior overall detection accuracy compared to the YOLOv5s algorithm model. This effectively addresses the issues of low detection efficiency and missed detections of small target on small-diameter stainless steel pipes.

### C. Comparative Experiments with Multiple Datasets

To validate the generalization capability and robustness of the YOLO-WP model and to ensure its effectiveness across a wide range of applications, this paper designs experiments to test YOLOv5s and YOLO-WP on the MT Defects Dataset and NEU-DET Dataset. The MT Defects Dataset, used for magnetic tile surface defect detection, consists of 1,344 images that encompass five types of defects: pores, cracks, wear, fractures, and uneven surfaces. The NEU-DET dataset, utilized for detecting surface defects on hot-rolled steel strips, comprises 1,800 images that encompass six defect types: rolled scale, cracks, patches, pitted surfaces, inclusions, and scratches. Both datasets are divided into training and validation sets at a ratio of 7:3. The experimental results are shown in Tables IV and V. Table IV shows that the YOLO-WP model's mAP@0.5 increased by 0.5% compared to YOLOv5s. Additionally, the model's parameters and FLOPs are reduced by 38% and 43% respectively, while the detection speed is nearly unaffected. These results indicate that the YOLO-WP model achieves higher accuracy in detecting magnetic tile surface defects, demonstrates greater computational efficiency, and is easier to

deploy. As shown in Table V, the YOLO-WP model's mAP@0.5 increased by 0.9% compared to YOLOv5s. Moreover, the model's parameters and FLOPs were reduced by 36% and 43%, respectively, while the detection speed remained almost unchanged. These findings further demonstrate that the YOLO-WP model outperforms YOLOv5s in detecting surface defects in the hot-rolled steel strip dataset. The experimental results from both the MT Defects Dataset and the NEU-DET Dataset confirm that YOLO-WP surpasses YOLOv5s in terms of accuracy, model complexity, and computational efficiency. Overall, these experiments suggest that YOLO-WP offers greater practical value and stronger robustness for industrial applications compared to YOLOv5s.

TABLE IV. EXPERIMENTAL RESULTS OF MT DEFECTS DATASET

Model	mAP@0.5(%)	Parameters(106)	FLOPs(G)	FPS(F/S)
YOLOv5s	89.1	7.3	15.4	87
YOLO-WP	89.6	4.5	8.7	86

TABLE V. EXPERIMENTAL RESULTS OF NEU-DET ON THE DATASET

Model	mAP@0.5(%)	Parameters(106)	FLOPs(G)	FPS(F/S)
YOLOv5s	85.3	7.3	15.4	84
YOLO-WP	86.1	4.6	8.7	82

### D. Comparison of Frontier Models

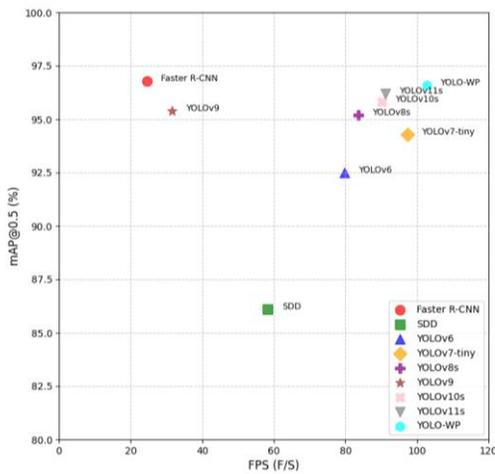
To more effectively assess the performance of the enhanced model introduced in this paper, we carried out comparative experiments using various object detection algorithms on a dataset for detecting surface defects in small-diameter stainless steel welded pipe seams. These algorithms include Faster R-CNN, SSD, YOLOv6, YOLOv7-tiny [35], YOLOv8s, YOLOv9, YOLOv10s [36] and YOLOv11s [37], along with our algorithm

(YOLO-WP). As shown in the Table VI, YOLO-WP achieved an mAP@0.5 of 96.6%, with 4.3 million parameters, 8.5 GFLOPs, and a speed of 102.7 FPS. The experimental results indicate that, compared to other algorithms, the YOLO-WP model has the smallest number of parameters and FLOPs, as well as the fastest detection speed. Although YOLO-WP's mAP@0.5 is 0.2% lower than that of Faster R-CNN, it significantly reduces the number of parameters and FLOPs.

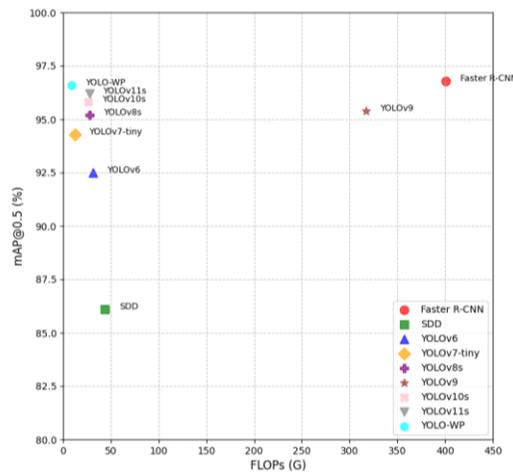
To better demonstrate the balanced advantages of the YOLO-WP model regarding detection accuracy, model complexity, and detection speed, this paper presents scatter plots illustrating the relationships among these factors. As shown in Fig. 10(a), the YOLO-WP model is positioned in the upper right corner of the two-dimensional coordinate system, reflecting its ability to balance speed and detection accuracy. In Fig. 10(b), the YOLO-WP model is positioned in the upper left corner of the two-dimensional coordinate system, indicating its capability to balance computational efficiency and detection accuracy.

TABLE VI. COMPARISON OF FRONTIER MODELS

Model	mAP@0.5(%)	Parameters(10 <sup>6</sup> )	FLOPs(G)	FPS(F/S)
Faster R-CNN	96.8	135.4	400.8	24.5
SDD	86.1	63.5	43.7	58.3
YOLOv6	92.5	15.6	31.3	79.7
YOLOv7-tiny	94.3	6.3	12.7	97.4
YOLOv8s	95.2	10.3	27.5	83.6
YOLOv9	95.4	70.5	317.2	31.5
YOLOv10s	95.8	11.3	26.2	90.1
YOLOv11s	96.2	13.4	27.3	91.2
YOLO-WP(Ours)	96.6	4.3	8.5	102.7



(a) Scatter plot of the relationship between mAP@0.5 and FPS.



(b) Scatter plot of the relationship between mAP@0.5 and FLOPs.

Fig. 10. Relationship scatter plot.

Based on the comparative experimental data and analysis presented above, the improved algorithm proposed in this paper achieves a balance between accuracy and lightweight design, efficiently utilizing computational resources to reach an optimal balance between model accuracy and training weights. This further underscores the superiority of the YOLO-WP algorithm. By significantly reducing the number of parameters and FLOPs, the YOLO-WP algorithm lowers hardware requirements. Consequently, this improvement meets the demands for online detection of weld seam surface defects in small-diameter stainless steel pipes.

## VI. CONCLUSIONS

To address the gap in detecting surface defects on small-diameter stainless steel pipe weld seams, and to overcome the limitations of YOLOv5s on terminal devices, which arise from insufficient computational power and poor detection capabilities for small object defects, this paper proposes the significantly improved YOLO-WP algorithm. By introducing the innovative GhostFusion architecture, Slim-Neck lightweight design, SimAM lightweight attention mechanism, and Focal-EIou

loss function optimization, the YOLO-WP model achieves a 5.3% and 3% increase in AP(D1) and mAP@0.5, respectively, compared to the original model. Additionally, the number of model parameters and FLOPs are reduced by 40% and 45%, respectively, significantly enhancing the efficiency of small target detection and the model's applicability in resource-constrained environments. Experimental results show that the YOLO-WP model achieves high detection accuracy, low complexity, minimal computational requirements, and rapid detection speeds. This model has demonstrated robustness across different datasets. Compared to other models, YOLO-WP exhibits strong competitiveness, improving production quality and reducing costs, thereby making it suitable for industrial online inspection.

However, this study still has certain limitations. Although we collected and trained common defect data during the online production process, which to some extent reduced the burden of subsequent offline detection, we have not yet achieved comprehensive coverage of surface defects in small-diameter stainless steel pipe weld seams. Moreover, YOLO-WP still faces challenges in dealing with more complex defect types, such as

occluded or extremely small targets. In future research, we plan to introduce more types of defects for study and verify the capability of our algorithm model in offline detection of surface defects in stainless steel pipe weld seams. We will also specifically improve a set of algorithm models suitable for offline detection to achieve collaborative work between offline and online detection, thereby fully applying visual inspection to all quality control processes.

This study provides a novel solution for the detection of surface defects in small-diameter stainless steel pipe weld seams. By optimizing the network architecture, incorporating lightweight design, and improving the loss function, the model's detection performance in resource-constrained environments has been significantly enhanced. These improvements not only offer an efficient and accurate detection tool for industrial online inspection but also provide valuable references for future research in related fields.

#### VII. AUTHORS' CONTRIBUTION

Conceptualization, Yukun Sun and Huaishu Hou; methodology, Yukun Sun and Chaofei Jiao; validation, Yukun Sun; investigation, Huaishu Hou; data curation, Huaishu Hou and Yukun Sun. writing— original draft preparation, Yukun Sun; writing—review and editing, Huaishu Hou. All authors have read and agreed to the published version of the manuscript.

#### FUNDING

None

#### DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

#### CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

#### REFERENCES

- [1] Bettahar K, Bouabdallah M, Badji R, et al. Microstructure and mechanical behavior in dissimilar 13Cr/2205 stainless steel welded pipes. *Materials & Design*, 2015, 85: 221-229.
- [2] Kumar M V, Balasubramanian V. Microstructure and tensile properties of friction welded SUS 304HCu austenitic stainless steel tubes. *International Journal of Pressure Vessels and Piping*, 2014, 113: 25-31.
- [3] Ressa J, Monrrabal G, Díaz A, et al. Microbiologically influenced corrosion of welded AISI 304 stainless steel pipe in well water. *Engineering Failure Analysis*, 2020, 116: 104734.
- [4] Baddoo N R. Stainless steel in construction: A review of research, applications, challenges and opportunities. *Journal of constructional steel research*, 2008, 64(11): 1199-1206.
- [5] Wang Y, Guo Z, Bai X, et al. Effect of weld defects on the mechanical properties of stainless-steel weldments on large cruise ship. *Ocean Engineering*, 2021, 235: 109385.
- [6] Ressa J, Monrrabal G, Díaz A, et al. Microbiologically influenced corrosion of welded AISI 304 stainless steel pipe in well water. *Engineering Failure Analysis*, 2020, 116: 104734.
- [7] Lee J K, Bae D S, Lee S P, et al. Evaluation on defect in the weld of stainless steel materials using nondestructive technique. *Fusion Engineering and Design*, 2014, 89(7-8): 1739-1745.
- [8] Ge J, Zhu Z, He D, et al. A vision-based algorithm for seam detection in a PAW process for large-diameter stainless steel pipes. *The international journal of advanced manufacturing technology*, 2005, 26: 1006-1011.
- [9] Gök D A. Destructive and non-destructive testings of 304 austenitic stainless steel produced by investment casting method. *Nondestructive Testing and Evaluation*, 2024: 1-13.
- [10] Peng J, Xu Z, Chen H, et al. Detection of brazing defects in stainless steel core plate using the first peak value of pulsed eddy current testing signals. *Construction and Building Materials*, 2023, 408: 133636.
- [11] Huang C Y, Hong J H, Huang E. Developing a machine vision inspection system for electronics failure analysis. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2019, 9(9): 1912-1925.
- [12] Ahmad H M, Rahimi A. Deep learning methods for object detection in smart manufacturing: A survey. *Journal of Manufacturing Systems*, 2022, 64: 181-196.
- [13] Badgajar C M, Poulouse A, Gan H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, 2024, 223: 109090.
- [14] Usamentiaga R, Lema D G, Pedrayes O D, et al. Automated surface defect detection in metals: a comparative review of object detection and semantic segmentation using deep learning. *IEEE Transactions on Industry Applications*, 2022, 58(3): 4203-4213.
- [15] Li Z, Dong M, Wen S, et al. CLU-CNNs: Object detection for medical images. *Neurocomputing*, 2019, 350: 53-59.
- [16] Demetriou D, Mavromatidis P, Robert P M, et al. Real-time construction demolition waste detection using state-of-the-art deep learning methods; single-stage vs two-stage detectors. *Waste Management*, 2023, 167: 194-203.
- [17] Yang D, Cui Y, Yu Z, et al. Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence*, 2021, 35(15): 1237-1249.
- [18] Zhou C, Lu Z, Lv Z, et al. Metal surface defect detection based on improved YOLOv5. *Scientific Reports*, 2023, 13(1): 20803.
- [19] Zhao W, Chen F, Huang H, et al. A new steel defect detection algorithm based on deep learning. *Computational Intelligence and Neuroscience*, 2021, 2021(1): 5592878.
- [20] Shao R, Zhou M, Li M, et al. TD-Net: tiny defect detection network for industrial products. *Complex & Intelligent Systems*, 2024: 1-12.
- [21] Ji C L, Yu T, Gao P, et al. Yolo-tla: An Efficient and Lightweight Small Object Detection Model based on YOLOv5. *Journal of Real-Time Image Processing*, 2024, 21(4): 141.
- [22] Han Z, Li S, Chen X, et al. DFW-YOLO: YOLOv5-based algorithm using phased array ultrasonic testing for weld defect recognition. *Nondestructive Testing and Evaluation*, 2024: 1-24.
- [23] Zhou S, Ao S, Yang Z, et al. Surface Defect Detection of Steel Plate Based on SKS-YOLO. *IEEE Access*, 2024.
- [24] Yuan M, Zhou Y, Ren X, et al. YOLO-HMC: An improved method for PCB surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [25] Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments. *Procedia computer science*, 2022, 199: 1066-1073.
- [26] Liu H, Sun F, Gu J, et al. Sf-yolov5: A lightweight small object detection algorithm based on improved feature fusion mode[J]. *Sensors*, 2022, 22(15): 5817.
- [27] Wang L, Zhang Y, Lin Y, et al. Ship Detection Algorithm Based on YOLOv5 Network Improved with Lightweight Convolution and Attention Mechanism[J]. *Algorithms*, 2023, 16(12): 534.
- [28] Shang J, Wang J, Liu S, et al. Small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. *Electronics*, 2023, 12(11): 2434.
- [29] Huang B, Liu J, Liu X, et al. Improved YOLOv5 Network for Steel Surface Defect Detection[J]. *Metals*, 2023, 13(8): 1439.
- [30] Han K, Wang Y, Xu C, et al. GhostNets on heterogeneous devices via cheap operations. *International Journal of Computer Vision*, 2022, 130(4): 1050-1069.

- [31] Wang Z, Zhou D, Guo C, et al. Yolo-global: a real-time target detector for mineral particles. *Journal of Real-Time Image Processing*, 2024, 21(3): 1-13.
- [32] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. *Journal of Real-Time Image Processing*, 2024, 21(3): 62.
- [33] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR, 2021: 11863-11874.
- [34] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 2022, 506: 146-157.
- [35] Hu S, Zhao F, Lu H, et al. Improving YOLOv7-tiny for infrared and visible light image object detection on drones. *Remote Sensing*, 2023, 15(13): 3214.
- [36] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.
- [37] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.