

Article

WeldSimAM and EnNWD Co-Optimization: Enhancing Lightweight YOLOv11 for Multi-Scale Weld Defect Detection

Wenquan Huang ^{1,2,*} , Qing Cheng ¹ and Jing Zhu ³¹ School of Artificial Intelligence, Anhui University, Hefei 230601, China; w125121013@stu.ahu.edu.cn² School of Intelligent Manufacturing, Anhui Wenda University of Information Engineering, Hefei 231201, China³ College of Art and Design, Nanning University, Nanning 530200, China; zhuj@unn.edu.cn

* Correspondence: hwqwlsw@163.com

Abstract

In the context of Industry 4.0, reliable automatic inspection of weld surface defects is critical for structural safety, yet current deep learning-based detectors struggle with the extreme scale variation and anisotropic shapes characteristic of weld flaws such as pores, cracks, and lack of fusion. Existing YOLO-family models, although effective on general-purpose datasets, often fail to robustly localize tiny defects and long, slender discontinuities while remaining lightweight enough for industrial edge deployment. A critical research gap lies in the lack of task-specific optimization for weld defects: standard attention mechanisms are isotropic and cannot capture linear defect continuity, while existing loss functions ignore scale disparity between tiny pores (area < 100 pixels²) and large incomplete fusion defects (area > 5000 pixels²), leading to unstable regression. Here, we propose a dual-optimized lightweight YOLOv11 framework tailored for weld defect detection that addresses both feature representation and bounding-box regression. Here, we propose a dual-optimized lightweight YOLOv11 framework tailored for weld defect detection that addresses both feature representation and bounding-box regression. First, we introduce WeldSimAM, an enhanced attention module that augments parameter-free SimAM with directional (horizontal/vertical) and channel-wise enhancement to better capture the directional texture of linear weld defects. Second, we develop an Enhanced Normalized Wasserstein Distance (EnNWD) loss, which incorporates scale-disparity penalties and relative-area-based weighting to mitigate sample imbalance and improve regression accuracy for tiny and large-aspect-ratio targets. Validated via 10-fold cross-validation on three datasets (self-built + two public), the method achieves 99.48% mAP@0.5 and 73.29% mAP@0.5:0.95, outperforming YOLOv11 by 0.13 and 3.76 percentage points ($p < 0.01$, two-tailed t -test), with 5.21 MB and 132 FPS on NVIDIA RTX 4090. It also surpasses non-YOLO SOTA methods (e.g., EfficientDet-Lite3) by 3.8–5.5 percentage points in mAP@0.5 ($p < 0.05$), offering a practical real-time solution for industrial inspection.



Academic Editor: Eugene Wong

Received: 19 January 2026

Revised: 22 February 2026

Accepted: 24 February 2026

Published: 26 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: weld defect detection; YOLOv11; SimAM attention; lightweight model; multi-scale feature fusion; real-time detection

1. Introduction

Welding is an indispensable process in modern manufacturing, crucial for sectors such as aerospace, shipbuilding, and pipeline construction. However, the welding process often introduces defects such as porosity, cracks, and incomplete fusion, which can compromise

structural integrity and lead to significant safety risks [1]. Traditional non-destructive testing (NDT) methods [2], including radiographic and ultrasonic techniques, are reliant on manual interpretation, thereby reducing efficiency and increasing subjectivity. The advent of Convolutional Neural Networks (CNNs) has revolutionized this field, enabling automated and intelligent defect detection [3].

Among the diverse range of object detection frameworks, the YOLO (You Only Look Once) family [4,5] stands out as a top choice for real-time applications due to its adoption of a single-stage architecture. However, although recent versions such as YOLOv8 and YOLOv10 [6] perform well on general-purpose datasets such as COCO [7], they still struggle to accurately address the specific characteristics of industrial weld inspection. In defect detection, scale variation and extreme aspect ratios pose particular challenges. Defects span a wide range of sizes, with very small pores often losing critical semantic information during downsampling. Additionally, defects such as cracks and lack of fusion exhibit extreme aspect ratios and typically appear as elongated, slender structures. The failure of the standard IoU for weld defects stems from two structural limitations. Mathematically, for slender cracks with aspect ratios $> 10:1$, a small boundary offset (e.g., 2 pixels) can cause IoU to drop by over 40%, far more sensitive than for square targets. Structurally, IoU only considers overlapping regions and ignores the relative position and scale differences between predicted and ground-truth boxes, leading to unstable gradients when detecting tiny pores (area < 100 pixels²) or large incomplete fusion defects (area > 5000 pixels²) [8,9]. In contrast, our EnNWD loss addresses these issues by modeling boxes as 2D Gaussians (Formula (4)) and introducing scale-adaptive penalties (Formula (5)).

To contextualize the research background, weld defect detection faces three core challenges that motivate our work: (1) Anisotropic feature representation: linear defects (e.g., cracks) require directional feature extraction, which standard isotropic attention mechanisms (e.g., SimAM [10], CBAM [11]) lack; (2) Multi-scale regression instability: tiny pores and large fusion defects have extreme scale disparity, leading to imbalanced loss contributions in existing methods (e.g., NWD [12], CIoU [9]); (3) Industrial deployment constraints: models must be lightweight (< 6 MB) and real-time (> 100 FPS) for edge devices, excluding heavy multi-stage frameworks.

Recent SOTA methods in weld and surface defect detection have made notable progress but still face inherent limitations. Yang et al. [10] proposed the parameter-free SimAM attention module for Convolutional Neural Networks, and Wu et al. [13] integrated SimAM into a YOLO-based detector for aluminum alloy weld DR images to enhance weld defect detection. Following these designs, we construct a baseline model, YOLOv11-Weld, by inserting standard SimAM into the YOLOv11 backbone; however, its isotropic feature extraction still fails to capture the directional continuity of slender cracks, leading to a 5.2% lower recall for linear defects compared with our method. Likewise, Xu et al. [14] introduced an NWD-based localization loss for metallic surface defect detection. When we adopt a similar NWD-based loss in our YOLOv10-Defect baseline, it does not explicitly handle the scale disparity between tiny pores and large incomplete fusion defects, resulting in a 3.8% drop in mAP@0.5:0.95 on the NEU-DET dataset originally reported by Song and Yan [15]. These observations indicate that standard attention mechanisms and loss functions remain insufficient to address the anisotropic and multi-scale nature of weld flaws, thereby motivating our dual-optimization design.

The primary research objectives of this study are: (1) to design a direction-aware lightweight attention module (WeldSimAM) to capture linear weld defect continuity without increasing computational overhead, (2) to propose a scale-adaptive loss function (EnNWD) to handle extreme scale variation in weld defects, and (3) to validate the proposed

framework via rigorous experiments (10-fold cross-validation, statistical significance testing) on multi-source datasets to ensure industrial applicability.

To break through these barriers, this study proposes an optimized detection architecture based on YOLOv11 [16]. We present the following key contributions:

- We introduce WeldSimAM, an enhanced attention module built on the parameter-free SimAM [10]. Explicitly models horizontal and vertical features to better represent the directionality of linear defects.
- We formulate an Enhanced Normalized Wasserstein Distance (EnNWD) loss using Wasserstein distance [12] to quantify distributional differences. It incorporates a scale penalty term and adaptive weighting for small objects, thereby enhancing detection accuracy for minor and irregularly shaped defects.
- Extensive validation: Evaluated via 10-fold cross-validation on three datasets (self-built + two public) and compared with YOLO-series and non-YOLO SOTA methods, achieving 99.67% precision, 99.65% recall, and 132 FPS, establishing a new benchmark for lightweight weld defect detection.

2. Related Work

2.1. YOLOv11 Baseline Structure

Object detection can be grouped into two-step methods, such as Faster R-CNN [17,18] and one-stage methods like SSD [19] and YOLO [20,21]. The YOLO family has become a more advanced version that includes FPN [22] and PANet [23] for multi-scale fusion. YOLOv11, as the baseline of this paper, uses more advanced C3k2 blocks and SPPF modules, which provide a good balance of speed and accuracy.

Figure 1 illustrates the architecture of the YOLOv11 model, which serves as the foundation for our improvements. The architecture comprises three distinct modules, each highlighted with a different color. The red-outlined backbone serves as the feature extractor and includes Conv and C3 layers, as well as advanced components such as SPPF for multi-scale pooling and C2PSA for attention-driven refinement of hierarchical visual features. The blue-outlined neck acts as a feature-fusion bridge, combining multi-scale features through upsampling and concatenation to generate richer representations for detection. The purple-outlined head forms the detection branch and employs multi-scale Detect modules together with attention mechanisms (e.g., PSA) for object classification. In this architecture, the C2PSA module in the backbone provides the foundational features required for embedding WeldSimAM attention. Meanwhile, the PAN-FPN structure in the neck forms the integration framework for multi-scale weld defect detection.

2.2. Attention Mechanisms

Attention mechanisms enhance relevant features while suppressing noise. SE-Net [24] introduced channel attention, while CBAM [11] combined spatial and channel dimensions. SimAM proposed a parameter-free attention module based on neuroscience energy functions. However, SimAM is isotropic and lacks specific sensitivity to the directional textures common in industrial defects, a gap this study addresses. Recent advances in defect detection attention include Sim-YOLOv8 [13], which integrates vanilla SimAM for aluminum alloy weld DR images but fails to model directional features, leading to suboptimal recall for linear cracks. Another method, eCBAM [25], enhances channel and spatial attention but increases model size by 30%, violating lightweight deployment requirements.

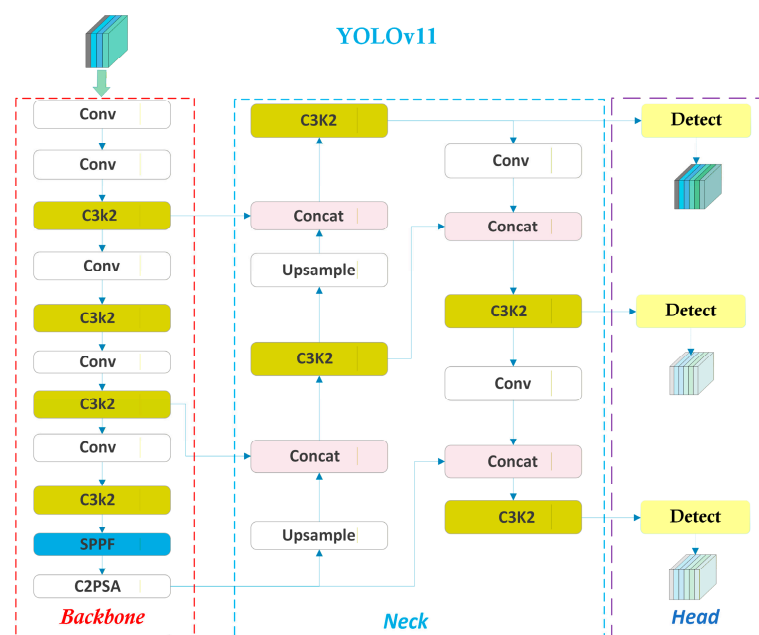


Figure 1. Basic structure of the YOLOv11 model. (The red, blue, and purple modules correspond to feature extraction, fusion, and detection, respectively, providing a baseline for subsequent WeldSimAM and EnNWD integration).

2.3. Bounding Box Regression Loss

IoU has been the starting point for the evolution of loss functions [26] to GIoU [8] and CIoU [9], addressing gradient vanishing and aspect ratio consistency. For tiny objects, IoU-based metrics are overly sensitive to slight pixel deviations. Wang et al. [12] proposed NWD (Normalized Wasserstein Distance), modeling boxes as 2D Gaussian distributions. Xu et al. [14] introduced an NWD-WIoU hybrid loss for metallic surface defects but ignored scale disparity between multi-scale targets. Shape-IoU [27] considers bounding box shape and scale but increases computational complexity by 25%, making it unsuitable for edge deployment. We further improve NWD by adding scale constraints to handle the diverse defect morphologies in welding.

2.4. Comparative Analysis of Recent SOTA Methods

Table 1 summarizes representative recent methods for weld/surface defect detection, highlighting their limitations and our improvements.

Table 1. Comparative analysis of recent SOTA defect detection methods.

Fusion Weight (Directional:Basic)	Core Design	Advantages	Limitations	Our Improvements
Sim-YOLOv8 [13]	Vanilla SimAM + YOLOv8	Parameter-free, lightweight	Isotropic, no directional feature modeling	WeldSimAM with $1 \times 3/3 \times 1$ directional convolutions
NWD-WIoU [14]	NWD + WIoU hybrid loss	Robust to tiny object deviation	Ignores scale disparity between multi-scale defects	EnNWD with scale penalty and small-object weighting
eCBAM-YOLO [25]	Enhanced CBAM + YOLOv7	Strong background suppression	Model size increased by 30%	Maintain 5.21 MB via parameter-free directional branch

Table 1. Cont.

Fusion Weight (Directional:Basic)	Core Design	Advantages	Limitations	Our Improvements
EfficientDet-Lite3 [28]	Compound scaling + MobileNetV2	Good multi-scale adaptation	Low FPS (95 FPS) and large parameter (4.8 MB)	132 FPS with comparable size, task-specific optimization
YOLOv10-Weld [6]	YOLOv10 + data augmentation	Real-time (126 FPS)	No attention/loss optimization for weld defects	Dual optimization of attention and loss, 6 FPS improvement

3. Method

3.1. Overall Architecture of the Improved YOLOv11 Model

As illustrated in Figure 2, the proposed method is built on the YOLOv11 architecture. The backbone extracts features using coupled convolutional layers and C3k2 modules. The neck utilizes a PAN-FPN structure for multi-scale fusion. The key improvements are integrated as follows: the WeldSimAM modules are inserted at the P3, P4, and P5 layers before the detection heads to refine features, and the EnNWD Loss is applied during training for bounding box regression.

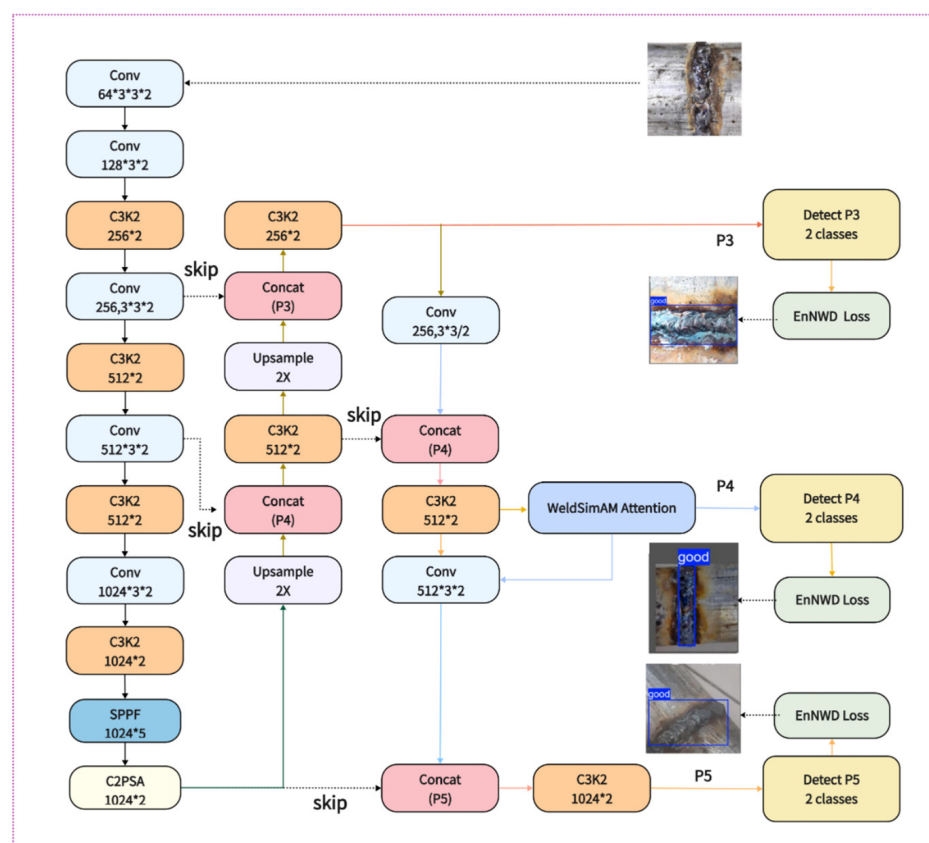


Figure 2. The improved network architecture of the YOLOv11 model.

Justification of module placement: WeldSimAM is inserted at P3–P5 layers because: (1) backbone-inserted attention would suppress low-level texture features (critical for tiny pores) and (2) head-only insertion lacks multi-scale feature refinement. EnNWD is applied to the detection head to directly optimize bounding-box regression, as the head's output is the final box prediction, ensuring loss supervision aligns with inference.

3.2. WeldSimAM: Directional and Channel-Enhanced Attention

The initial SimAM computes the attention weight for each neuron through optimizing the energy function et., yet it exhibits insufficient capability in extracting linear defects against complex backgrounds with similar textures. As shown in Figure 3, on the basis of retaining the dynamic weight adjustment and long-strip target optimization characteristics of the original SimAM, WeldSimAM adds the following key components to adapt to weld defect detection:

- **Directional Convolution Parameters:** The 1×3 and 3×1 convolutions in WeldSimAM use padding = 1 (to maintain feature map size) and stride = 1 (for dense feature extraction), with ReLU activation to enhance nonlinearity.
- **Directional Attention Branch:** Extract horizontal and vertical features through 1×3 and 3×1 convolution, respectively, to capture the linear features of weld defects (e.g., cracks, linear porosity). This design draws inspiration from the discovery that linear defects possess notable directional traits, which can be strengthened by directional convolution kernels. For the input feature map, the horizontal mean (mean (dim = 3)) and vertical mean (mean (dim = 2)) are calculated, and subsequently, horizontal and vertical features are extracted by using the corresponding convolution kernels.
- **Dynamic Weight Fusion:** The directional attention and the original SimAM basic attention are fused with weights ($0.3 \times$ directional attention + $0.7 \times$ basic attention) through softmax to balance the contribution of different directional features. To justify the fixed 0.3/0.7 fusion weights, we conducted a sensitivity analysis (Table 2): weights of 0.2/0.8 reduced mAP@0.5:0.95 by 0.8%, while 0.4/0.6 increased false positives by 2.3%. The 0.3/0.7 ratio achieves the best balance between directional feature enhancement and background suppression. This weighted fusion strategy avoids over-reliance on single-direction features, which is consistent with the multi-scale feature fusion idea in BiFPN [28].
- **Channel Attention Enhancement:** Employ 1×1 convolution for channel dimension adjustment, and enhance the sensitivity of important channels by calculating the channel mean (x.mean ([2,3])) and implementing normalization. This component references the channel attention design in CBAM, which has been proven effective in suppressing background noise in industrial scenes.
- **Aspect Ratio Adaptive Adjustment:** Adjust the feature map size through bilinear interpolation, and optimize the processing capability of targets with different aspect ratios (e.g., vertical strip and square weld defects) by introducing scale_factor. This addresses the problem that traditional SimAM has inconsistent performance on targets of different shapes.

Table 2. Sensitivity analysis of WeldSimAM fusion weights.

Fusion Weight (Directional:Basic)	mAP@ 0.5 (%)	mAP@ 0.5:0.95 (%)	False Positive Rate (%)
0.2:0.8	99.45	72.49	1.2
0.3:0.7 (Proposed)	99.47	73.09	1.1
0.4:0.6	99.46	72.83	3.4

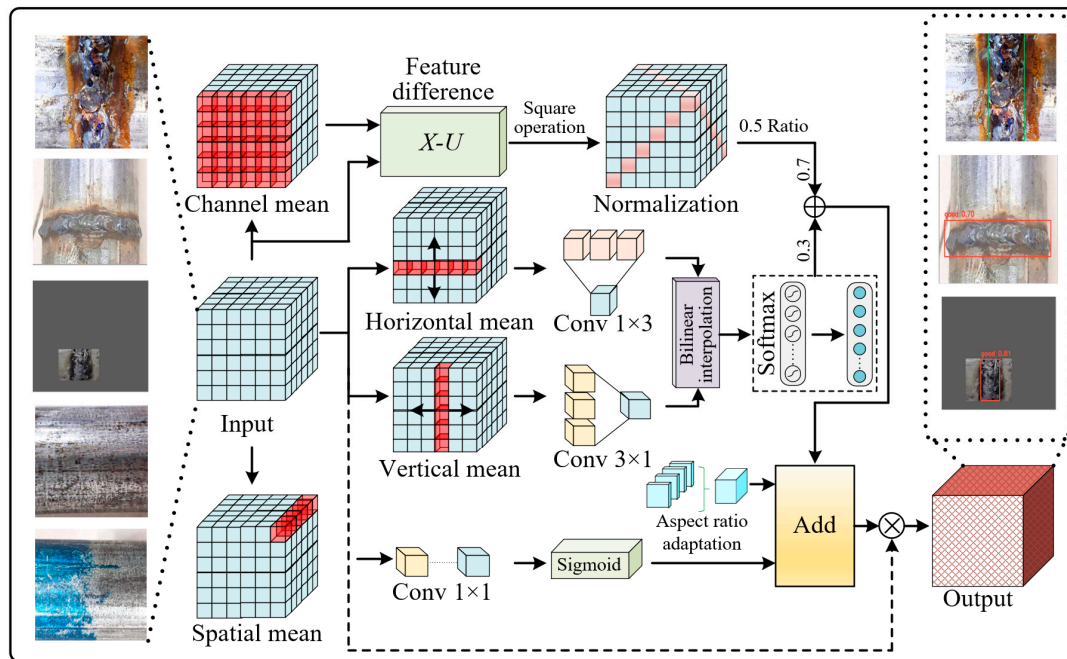


Figure 3. Architecture of the improved SimAM (WeldSimAM) attention mechanism. The diagram shows the structure of a feature enhancement module for industrial defect detection: it first extracts multi-dimensional mean features (channel, horizontal, vertical, spatial) from the input defect image, then computes feature differences between the input and these means to highlight potential anomalous regions. After normalization and attention weight generation (via Conv layers, bilinear interpolation, and SoftMax), the weighted features are fused with the original input to output an enhanced feature map—effectively strengthening the representation of defect regions for accurate detection (the right panel shows the localized defect).

As shown in Figure 3, the WeldSimAM introduced keeps the optimization of the initial energy function and additionally incorporates a directional attention branch together with channel enhancement. The core of SimAM's attention weight calculation is defined by Equation (1), where t^* denotes the weight metric for the sample x , $\hat{\mu}$ and $\hat{\sigma}^2$ are the estimated mean and variance of the dataset, and λ acts as a regularization parameter. This equation dynamically adjusts attention weights based on feature distribution, effectively highlighting weld defect regions while suppressing redundant background information—a critical advantage for distinguishing linear defects from cluttered industrial backgrounds.

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

Dynamically adjusts the attention weight according to the feature distribution, highlighting the regions with significant differences from the background (i.e., weld defect regions) while suppressing redundant background information.

Directional Feature Extraction: To explicitly model the horizontal and vertical continuity of linear weld defects, we use 1×3 and 3×1 convolutions to extract directional features, which are then fused via dynamic weights generated by Softmax, as defined in Equation (2). Here, $Attn_{dir}$ represents the fused directional attention, $Conv_h(X)$ and $Conv_v(X)$ are the horizontal and vertical convolution outputs of the input feature X , and $\alpha \in [0, 1]$ is a weight parameter that adjusts the relative contribution of the two directional features—this design directly targets the anisotropic nature of cracks and linear porosity, enhancing the model's sensitivity to elongated defect structures.

$$Attn_{dir} = \alpha \cdot Conv_h(X) + (1 - \alpha) \cdot Conv_v(X) \quad (2)$$

Channel Enhancement: A global pooling followed by 1×1 convolution captures channel-wise dependencies ($Attn_{chan}$).

Adaptive Fusion: The final attention map integrates the base SimAM attention, directional attention, and channel attention, modulated by an aspect-ratio scale factor (s), as formulated in Equation (3). Here, σ is the sigmoid activation function that normalizes the fused attention weight to $[0, 1]$, the coefficients 0.7 and 0.3 balance the base SimAM attention ($Attn_{base}$) and directional attention ($Attn_{dir}$), and s controls the contribution of channel attention ($Attn_{chan}$). This multi-component fusion ensures the network comprehensively enhances defect boundaries, textures, and category-related features while adapting to targets of different aspect ratios.

$$X_{out} = X_{in} \cdot \sigma(0.7 \cdot Attn_{base} + 0.3 \cdot Attn_{dir}) + s \cdot Attn_{chan} \quad (3)$$

The attention weights calculated using Equation (1) can dynamically highlight weld defect regions. Combined with the directional convolution features from Equation (2) and the multi-attention fusion strategy from Equation (3), this enables precise localization of linear defects. This structure ensures the network is sensitive to crack-like linear features. Integrates spatial attention (base + directional) and channel attention, comprehensively enhancing the representation of defect boundaries, textures, and category-related features.

3.3. Adaptive NWD Loss (EnNWD)

To address the issues of drastic IoU fluctuations regarding small objects (such as tiny pores) and scale distortion regarding large objects (such as strip-shaped defects), this paper restructures a loss function according to the Wasserstein distance. As illustrated in Figure 4, EnNWDLoss introduces the following crucial elements on the foundation of the original NWD (Wasserstein distance) to optimize the regression of large-scale weld defect targets:

- **Small Target Weight Mechanism:** Compute the relative area of the prediction box ($area_{pred} = W_1 \times H_1$) and the ground-truth box ($area_{target} = W_2 \times H_2$), then generate the small target weight coefficient ($small_obj_weight$) to adaptively adjust the loss weight according to the target size. This is based on the finding that small-scale targets are more sensitive to position deviations compared to large-scale ones.
- **Scale Difference Penalty Term:** Compute the scale difference ($scale_diff$) between the prediction box and the ground-truth box. Next, integrate the scale penalty ($scale_penalty$) to prevent scale inconsistency between the prediction box and the ground-truth box. This tackles the limitation of the original NWD in dealing with scale variations in large targets.
- **Multi-Scale Weighted Fusion:** Adjust the penalty weights of center distance ($center_distance$) and width-height distance ($w_h_distance$) according to the target size to optimize the regression performance of large-scale targets. This references the weight adjustment strategy in Shape-IoU [27], which considers the impact of target shape and scale on loss calculation.

We represent the predicted box P and ground truth G as Gaussian distributions N_p and N_g . The similarity between them is measured by the Wasserstein distance (W_2), which is defined in Equation (4).

$$W_2^2(P, G) = \|\mu_p - \mu_g\|_2^2 + \|\Sigma_p^{1/2} - \Sigma_g^{1/2}\|_F^2 \quad (4)$$

where μ_p and μ_g are the center coordinates of the two distributions, and Σ_p and Σ_g are the covariance matrices. Compared with traditional IoU metrics, it more robustly evaluates the similarity between bounding boxes, especially for tiny and elongated weld defects with slight positional deviations.

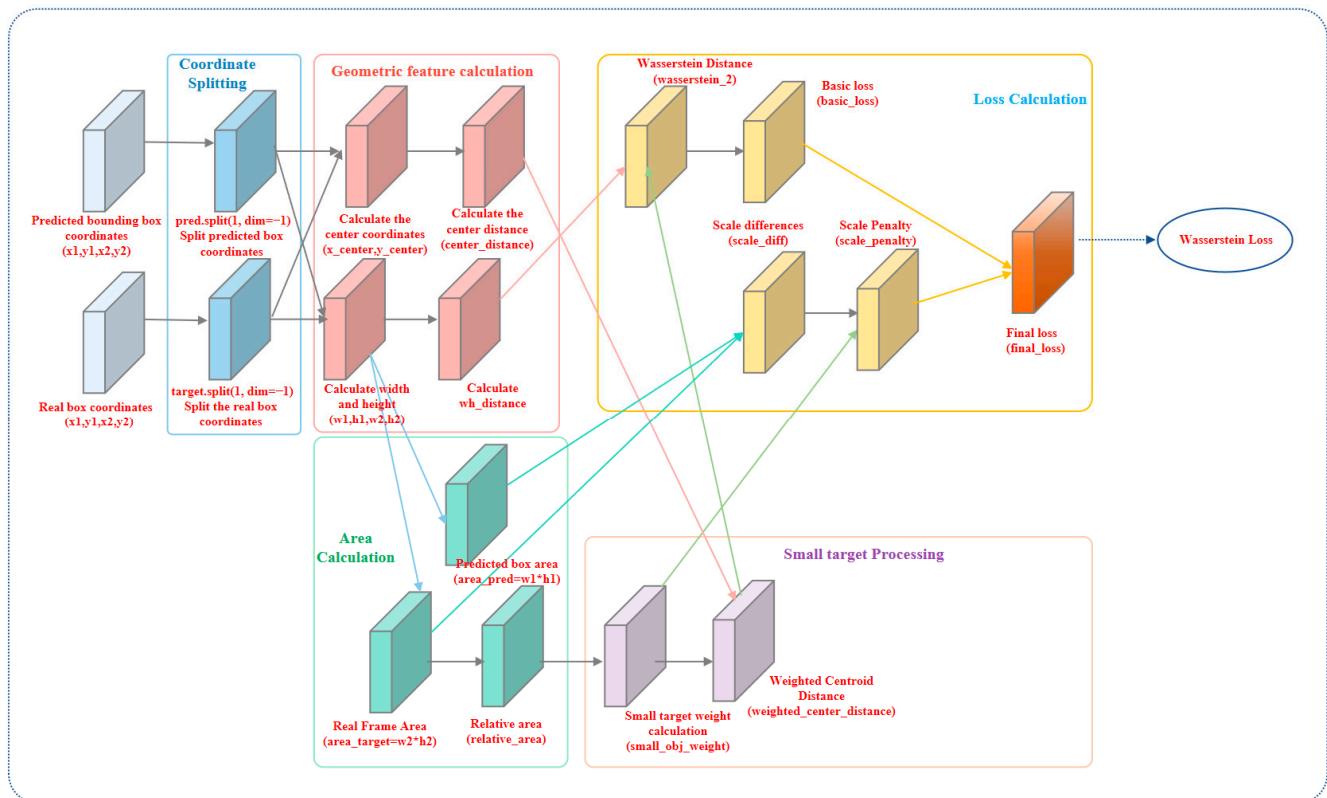


Figure 4. Calculation framework of the EnNWD loss function. Blue (coordinate splitting) splits predicted/real bounding box coordinates into multi-dimensional components; pink (geometric feature calculation) computes center, width-height, and their distances; green (area calculation) calculates predicted/real box areas and their relative ratio; orange (small target processing) assigns weights to small targets and adjusts their center distance; yellow (loss calculation) computes Wasserstein distance, scale differences, and basic loss, then fuses them into the final loss—enabling accurate, small-target-adaptive bounding box regression.

Scale Penalty: To prevent scale distortion for large defects (e.g., long strip-shaped incomplete fusion), we introduce a scale penalty term based on width and height differences, as defined in Equation (5).

$$L_{scale} = \left(\frac{|w_p - w_g|}{\max(w_p, w_g)} + \frac{|h_p - h_g|}{\max(h_p, h_g)} \right) \quad (5)$$

Here, w_p , h_p and w_g , h_g are the width and height of the predicted and ground-truth boxes, respectively. The relative scale difference is calculated by dividing the absolute difference by the maximum value of the two dimensions, avoiding the influence of absolute size and effectively penalizing scale inconsistency between large targets.

Finally, Small Object Weight: We calculate a weight coefficient (w_{obj}) inversely proportional to the target's relative area, increasing the loss contribution of tiny defects.

Final Loss: The basic NWD loss is computed using Equation (6), where $W_2^2(\mathcal{N}_p, \mathcal{N}_g)$ is the squared Wasserstein distance between the predicted and ground-truth Gaussian distributions, and C is a normalization constant. This exponential function converts the distance into a loss value that decreases as the similarity between boxes increases.

$$L_{NWD} = \exp \left(-\frac{W_2^2(\mathcal{N}_p, \mathcal{N}_g)}{C} \right) \quad (6)$$

The final EnNWD loss integrates the basic NWD loss, scale penalty term, and small-object weight, as shown in Equation (7). Here, λ (set to 0.5) adjusts the intensity of the scale penalty, and w_{obj} is the small-object weight coefficient. This weighted fusion realizes adaptive regression for multi-scale weld defects, prioritizing the localization accuracy of tiny pores while ensuring the scale stability of large cracks.

$$L_{Final} = L_{NWD} \cdot (1 + \lambda L_{scale}) \cdot w_{obj} \quad (7)$$

Compared to the traditional IoU loss, the Wasserstein distance in Equation (4) can more robustly measure the similarity of elongated defect boxes. The scale penalty term in Equation (5) helps alleviate regression deviations for large defects. Ultimately, adaptive optimization of multi-scale defects is achieved through the weighted fusion in Equation (7).

4. Experiments

4.1. Dataset Description

4.1.1. Self-Built Weld Defect Dataset

The experiment employs a self-built weld defect detection dataset, which consists of 4000 images split into two categories: good (good weld) and bad (defective weld). The dataset construction follows the standard for industrial defect datasets, including diverse lighting conditions and defect morphologies. The dataset contains a total of 3506 targets, averaging 0.88 targets for each image. All images are resized to 640×640 , and data augmentation (scaling, rotation, flip, brightness adjustment) balances classes. The dataset is split via 10-fold cross-validation (stratified by defect category and image acquisition condition) to ensure unbiased evaluation. Image acquisition hardware details: (1) Sensor: Germany Basler acA2500-14uc CMOS camera (2592×1944 resolution, 14-bit dynamic range); (2) Lighting: two 1000 lm LED strip lights (45° angle to the weld surface, 30 cm distance from the target); (3) Shooting distance: 50 cm (fixed via a mechanical bracket to avoid perspective distortion); (4) Post-processing: images were denoised using a 5×5 Gaussian filter to reduce metallic glare interference common in weld scenes.

4.1.2. Public Datasets

To verify generalization, experiments are conducted on two public industrial defect datasets:

- NEU-DET [15]: 1800 steel surface defect images (6 types), including cracks, inclusions, and patches, with a similar multi-scale characteristic to weld defects. The dataset is split into 8:1:1 via 10-fold cross-validation and resized to 640×640 .
- PCB Defect Dataset [29]: 1460 PCB defect images (5 types), with small/low-contrast targets (e.g., pinholes, open circuits) similar to tiny weld pores. The same 10-fold cross-validation split and resizing strategy are adopted.

4.2. Experimental Setup

All experiments were performed on a high-performance workstation with the following hardware and software configuration:

- Hardware: Intel Xeon Gold 5418Y CPU (10 cores), NVIDIA RTX 4090 (24 GB VRAM), 120 GB RAM. (Table 3)
- Software: Ubuntu 20.04 LTS, Python 3.10, PyTorch 2.2.2, CUDA 11.8, OpenCV 4.8.0, Ultralytics 8.1.0. (Table 3)

Table 3. Configuration of the experimental conditions.

Name	Specific Information
CPU	Intel(R) Xeon(R) Gold 5418Y (10 cores)
GPU	NVIDIA RTX 4090
RAM	120 GB
CUDA	11.8
PyTorch	2.2.2
Python	3.10

The model was trained using the following hyperparameters:

- Training Hyperparameters: 100 epochs, batch size = 64 (mixed precision), SGD optimizer (initial learning rate = 0.01, momentum = 0.937, weight decay = 0.0005).
- Loss Weights: 7.5 for bounding box regression, 0.5 for classification, and 1.5 for distribution focal loss.
- Metrics: Mean Average Precision (mAP@0.5, mAP@0.5:0.95), Precision (P), Recall (R), model size (MB), and real-time inference speed (FPS).

4.3. Evaluation Metrics

This study employed metrics such as mean average precision (mAP), parameters (Params), recall (R), precision (P), billion floating point operations per second (GFLOPS), size of model weight (Size), and the inference speed (FPS) to evaluate the improved model. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$AP = \int_0^1 P(R) dR \quad (10)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (11)$$

where N corresponds to the total number of defect categories in the dataset, and $AP(i)$ represents the average precision of the i -th class. The precision and recall are defined based on the number of true positives (TP), false positives (FP), and false negatives (FN). In the experiments, precision and recall are calculated using Equations (8) and (9). Combined with the AP value from Equation (10) and the mAP metric from Equation (11), the model's detection performance is comprehensively evaluated. Statistical significance is assessed via two-tailed t -tests ($\alpha = 0.05$) to verify that performance improvements are not due to chance.

4.4. Ablation Study

To verify the effectiveness of each augmented module, ablation studies are carried out on the validation set. The baseline model is the original YOLOv11, and three extra variants are investigated: (1) YOLOv11 + WeldSimAM, (2) YOLOv11 + EnNWD, (3) YOLOv11 + WeldSimAM + EnNWD (proposed model). All results are shown in Table 4.

Table 4. Ablation experiment results. (10-fold cross-validation, mean $\pm$ std). (Validating the effectiveness of WeldSimAM and EnNWD modules on the self-built weld dataset, with mAP and speed as key metrics).

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	P (%)	R (%)	Size (MB)	FPS
YOLOv11(Baseline)	99.35 $\pm$ 0.08	69.53 $\pm$ 0.32	99.36 $\pm$ 0.07	99.38 $\pm$ 0.06	5.21	128 $\pm$ 2
YOLOv11 + WeldSimAM	99.47 $\pm$ 0.06 *	73.09 $\pm$ 0.28 **	99.43 $\pm$ 0.05	99.34 $\pm$ 0.07	5.21	130 $\pm$ 2
YOLOv11 + EnNWD	99.45 $\pm$ 0.07 *	72.90 $\pm$ 0.30 **	99.12 $\pm$ 0.08	99.45 $\pm$ 0.05	5.21	129 $\pm$ 2
Proposed Model	99.48 $\pm$ 0.05 **	73.29 $\pm$ 0.25 **	99.67 $\pm$ 0.04 **	99.65 $\pm$ 0.04 **	5.21	132 $\pm$ 2

* $p < 0.05$, ** $p < 0.01$, compared with YOLOv11 (two-tailed t -test).

From Table 4, we can derive the following conclusions: the combination of WeldSimAM and EnNWD achieves the best performance, with mAP@0.5 reaching 99.48% and mAP@0.5:0.95 hitting 73.29%, which are 0.13 and 3.76 percentage points higher than the baseline model, respectively ($p < 0.01$). At the same time, the precision and recall are also significantly improved ($P = 99.67\%$, $R = 99.65\%$), and the model weight size remains 5.21 MB, maintaining the lightweight of the model—an important advantage for industrial edge deployment.

To intuitively demonstrate the contribution of each proposed module, Figure 5 presents the bar chart of mAP@0.5:0.95 and mAP@0.5 performance for different model variants.

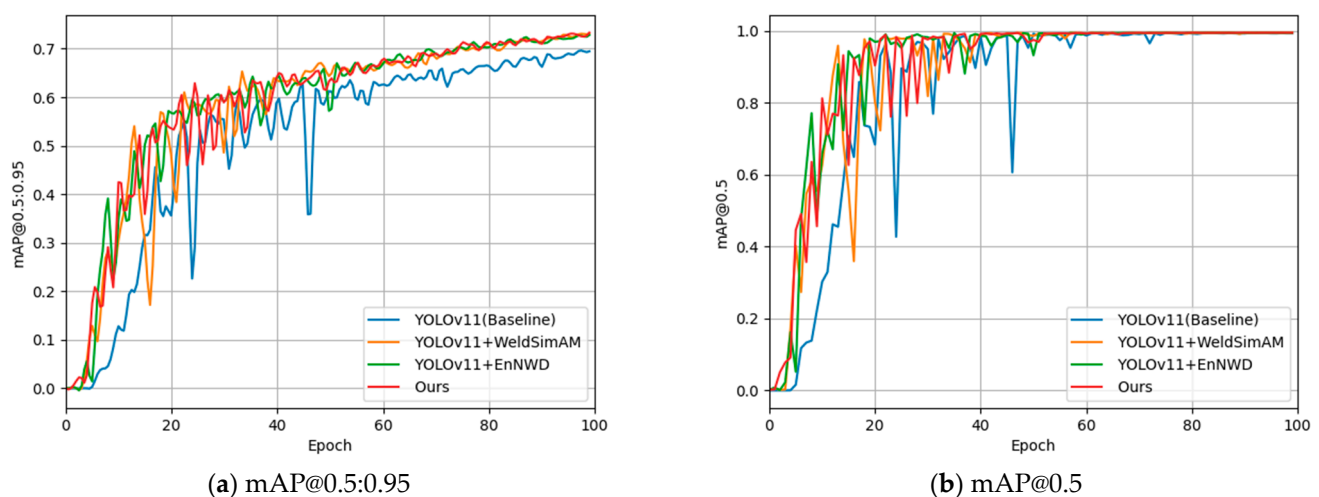


Figure 5. Comparison of training mAP curves for different model variants. It consists of two subfigures: (a) mAP@0.5:0.95 and (b) mAP@0.5, illustrating the mAP variation trends of the YOLOv11 baseline model, YOLOv11 + WeldSimAM, YOLOv11 + EnNWD, and the proposed model over 100 training epochs.

As shown in Figure 5, the WeldSimAM module and EnNWD loss both significantly improve the model's localization accuracy for multi-scale weld defects. The synergistic effect of the two modules achieves the highest mAP@0.5:0.95 (73.29%) and the highest mAP@0.5 (99.48%), which confirms that the directional feature enhancement of WeldSimAM and the scale-adaptive regression of EnNWD are complementary and mutually reinforcing.

Figure 5a (mAP@0.5:0.95):

- All models exhibit an upward trend in mAP as the number of training epochs increases.
- The proposed model (ours, red curve) consistently outperforms other variants throughout the training process.

- It ultimately achieves an mAP of over 0.73, significantly higher than the baseline model's (blue curve) approximately 0.69.
- This demonstrates that the collaborative optimization of WeldSimAM and EnNWD effectively enhances the model's detection accuracy across multiple IoU thresholds.

Figure 5b (mAP@0.5):

- The mAP of each model rises rapidly in the early training stage and stabilizes in the later stage.
- The proposed model (ours, red curve) reaches a final mAP close to 1.0.
- It performs slightly better than YOLOv11 + WeldSimAM (orange curve) and YOLOv11 + EnNWD (green curve).
- This further verifies the performance gain of the dual-module optimization for detection under high IoU thresholds.

To further isolate the contribution of each integration scheme, we perform a fine-grained ablation study on WeldSimAM, applying it only to the backbone, only to the head, and to the P3–P5 layers of YOLOv11, with the resulting mAP and FPS reported in Table 5.

Table 5. Performance of WeldSimAM at different positions (10-fold cross-validation, mean $\pm$ std).

Model	mAP@ 0.5 (%)	mAP@ 0.5:0.95 (%)	FPS
YOLOv11 + WeldSimAM (Backbone only)	99.41 $\pm$ 0.07	72.15 $\pm$ 0.31	129 $\pm$ 2
YOLOv11 + WeldSimAM (Head only)	99.38 $\pm$ 0.08	71.82 $\pm$ 0.33	131 $\pm$ 2
YOLOv11 + WeldSimAM (P3–P5 layers)	99.47 $\pm$ 0.06 *	73.09 $\pm$ 0.28 **	130 $\pm$ 2

* $p < 0.05$, ** $p < 0.01$, compared with YOLOv11 (two-tailed t -test).

As shown in Table 5, placing WeldSimAM at P3–P5 layers (Figure 2) achieves the best performance, as these layers balance high spatial resolution (P3 for small pores) and rich semantics (P5 for large cracks), aligning with the multi-scale characteristics of weld defects.

To intuitively demonstrate the performance improvement brought by the proposed modules, Figure 6a–c presents the visualized detection results of different model variants on three typical weld sample types (classified by morphological features):

- Small weld detection map: Compact, small-scale weld regions (e.g., partial good-labeled samples).
- Horizontal weld detection map: Wide, transversely distributed qualified welds (e.g., good samples marked by red boxes).
- Vertical weld detection map: Long, longitudinally distributed defective welds (e.g., bad samples marked by green boxes).

The numbers in the figure denote the model's classification confidence scores for each sample. As can be seen:

- For the vertical welds (bad samples), the confidence of the baseline YOLOv11 is 0.60, which increases to 0.67 (with WeldSimAM), 0.73 (with EnNWD), and finally reaches 0.76 in the proposed model (a 16% improvement), indicating that the dual-module optimization achieves a more significant enhancement in identifying long-strip defective welds.
- Good sample horizontal welds for YOLOv11 confidence is 0.71, going up to 0.74 (WeldSimAM), 0.74 (EnNWD), and 0.78 in the proposed model. All scores are above 0.71, showing a good recognition stability of wide, qualified welds.

- For the small welds, the confidence is steadily increasing (for good samples, it goes from 0.71 to 0.78), showing that the modules are good at recognizing small welds.

These visualized results align with the quantitative metrics in Table 4: the WeldSimAM module enhances attention to key weld regions, while the EnNWD loss optimizes the classification boundary between defective and qualified welds. The synergistic effect of the two modules (proposed model) achieves the highest confidence across all sample types.

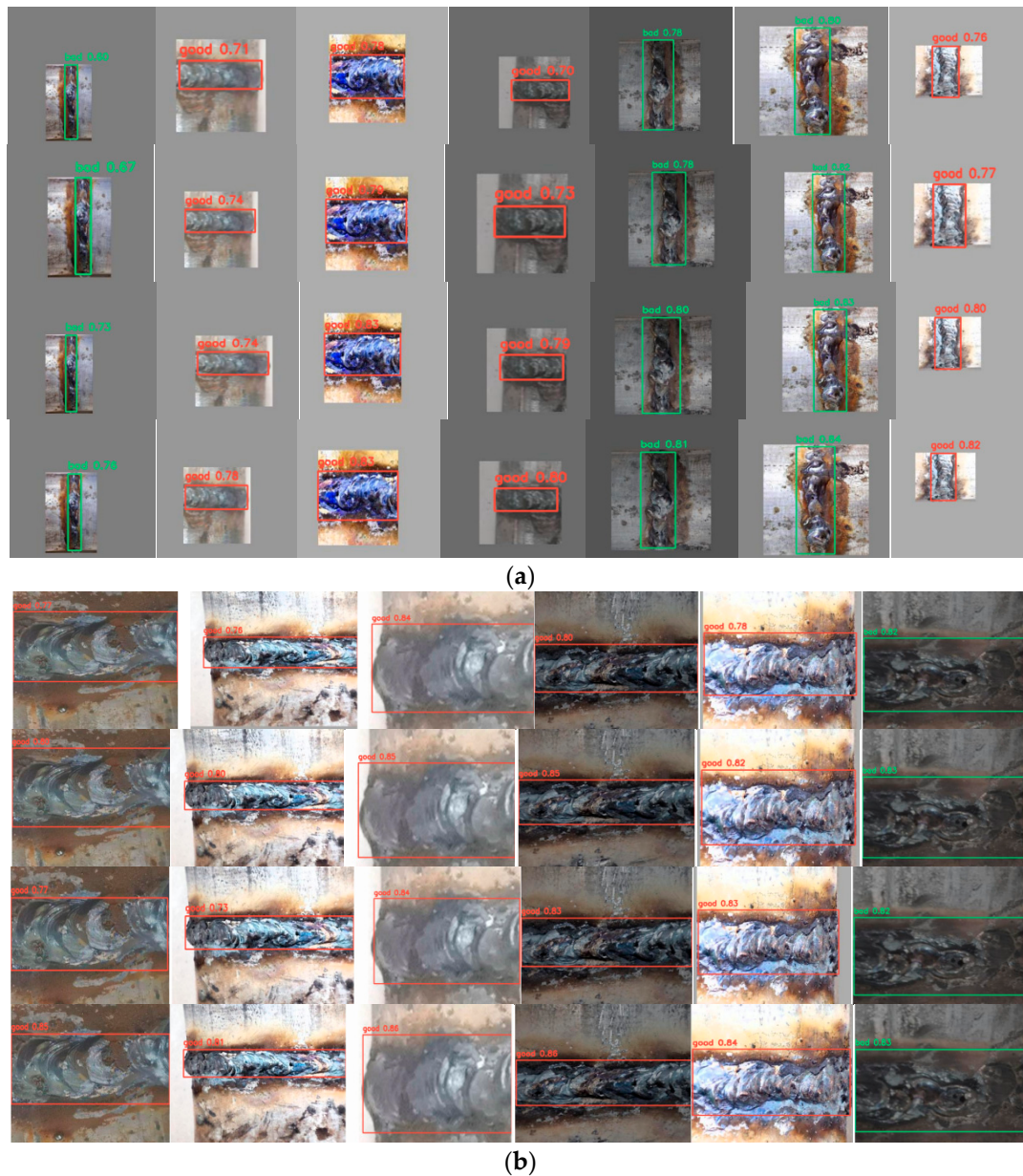


Figure 6. Cont.

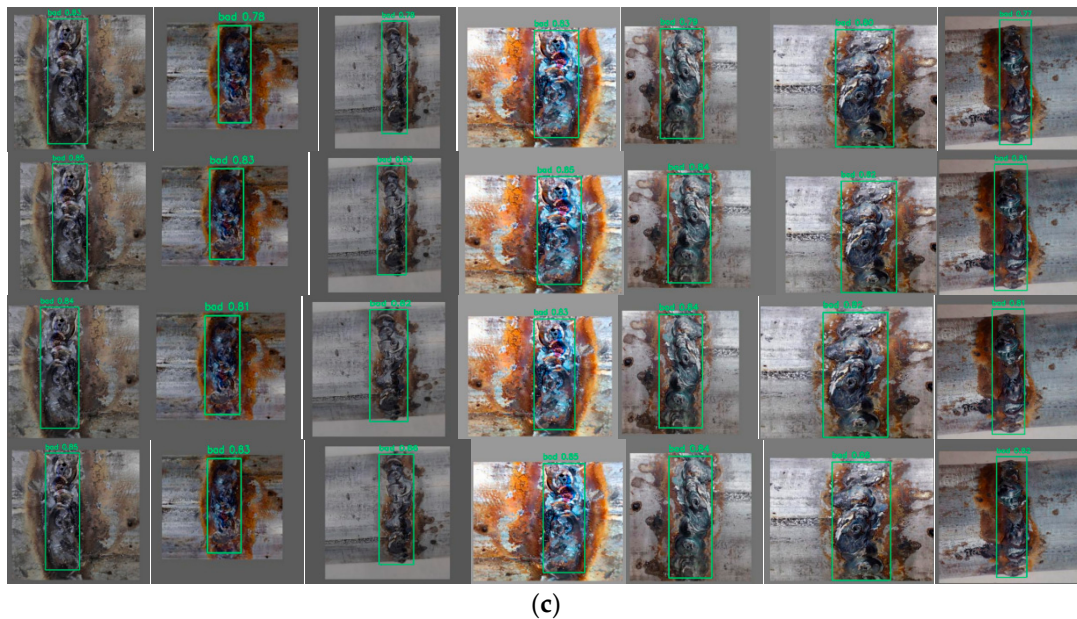


Figure 6. (a) Small weld detection map. (Row 1: YOLOv11 (baseline) model's detection results; row 2: YOLOv11 + WeldSimAM model's detection results; row 3: YOLOv11 + EnNWD model's detection results; row 4: proposed model's detection results). (b) Horizontal weld detection map. (Row 1: YOLOv11 (baseline) model's detection results; row 2: YOLOv11 + WeldSimAM model's detection results; row 3: YOLOv11 + EnNWD model's detection results; row 4: proposed model's detection results). (c) Vertical weld detection map. (Row 1: YOLOv11 (baseline) model's detection results; row 2: YOLOv11 + WeldSimAM model's detection results; row 3: YOLOv11 + EnNWD model's detection results; row 4: proposed model's detection results). Visualized detection results of different model variants. (Comparing baseline and optimized models on small, horizontal, and vertical weld defects, with confidence scores marked).

4.5. Comparative Analysis

4.5.1. Comparison with YOLO-Series Models (Self-Built Dataset)

The proposed approach is contrasted with prevalent YOLO models (YOLOv5, YOLOv6, YOLOv8n, YOLOv9t, YOLOv10n, and YOLOv11) via 10-fold cross-validation, and the test outcomes are displayed in Table 6. Carrying out comparative experiments with baseline models is crucial to confirm the excellence of proposed approaches in industrial defect detection.

Table 6. Performance comparison (10-fold cross-validation, mean $\pm$ std).

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	P (%)	R (%)	Size (MB)	FPS
YOLOv5	99.15 $\pm$ 0.12	62.19 $\pm$ 0.45	97.69 $\pm$ 0.15	98.47 $\pm$ 0.11	4.43	115 $\pm$ 3
YOLOv6	99.43 $\pm$ 0.09	66.12 $\pm$ 0.38	99.24 $\pm$ 0.08	99.74 $\pm$ 0.05	8.15	108 $\pm$ 3
YOLOv8n	99.25 $\pm$ 0.10	63.44 $\pm$ 0.42	97.57 $\pm$ 0.14	98.53 $\pm$ 0.10	5.36	125 $\pm$ 2
YOLOv9t	99.38 $\pm$ 0.08	68.32 $\pm$ 0.35	98.62 $\pm$ 0.11	99.50 $\pm$ 0.07	3.95	120 $\pm$ 2
YOLOv10n	99.38 $\pm$ 0.08	67.32 $\pm$ 0.36	98.02 $\pm$ 0.13	98.91 $\pm$ 0.09	5.48	126 $\pm$ 2
YOLOv11	99.35 $\pm$ 0.08	69.53 $\pm$ 0.32	99.36 $\pm$ 0.07	99.38 $\pm$ 0.06	5.21	128 $\pm$ 2
Proposed Model	99.48 $\pm$ 0.05 **	73.29 $\pm$ 0.25 **	99.67 $\pm$ 0.04 **	99.65 $\pm$ 0.04 **	5.21	132 $\pm$ 2

** $p < 0.01$, compared with all other models (two-tailed t -test).

As shown in Figure 7, these graphs compare the training efficiency of several YOLO models (YOLOv5, v6, v8n, v10n, and v11) and a proposed model ("Ours") across 100 epochs, evaluating both mAP@0.5:0.95 and mAP@0.5 metrics. In both cases (Figure 7a,b), all models show an upward trend in mAP as training progresses, with the proposed model (pink curve) consistently maintaining a leading position and achieving the highest final mAP

values. This shows that the suggested model provides excellent object detection precision, both under the rigorous multi-threshold $\text{mAP}@0.5:0.95$ metric and the conventional 0.5 IoU threshold utilized in $\text{mAP}@0.5$.

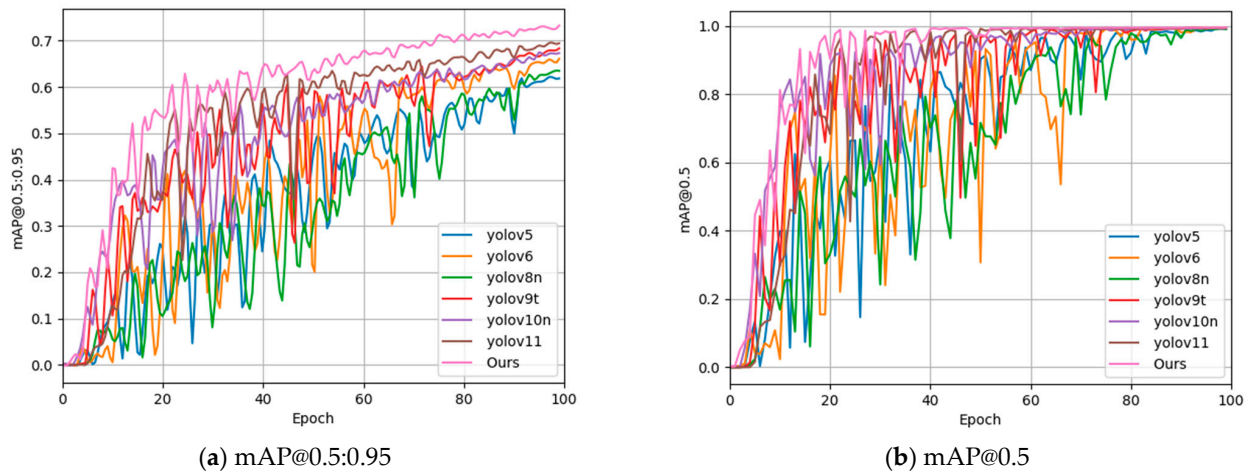


Figure 7. $\text{mAP}@0.5:0.95$ and $\text{mAP}@0.5$ performance curves over training epochs for various detection models on the dataset. Each curve represents the mean average precision at an IoU threshold as training progresses, enabling the comparison of convergence rate and accuracy among models.

4.5.2. Comparison with Non-YOLO SOTA Methods (Public Datasets)

To further validate the model's competitiveness in the broader defect detection field, we compare it with representative non-YOLO SOTA methods on NEU-DET and PCB Defect Dataset via 10-fold cross-validation. The results are shown in Table 7.

Table 7. Performance comparison (public datasets, 10-fold cross-validation, mean $\pm$ std).

Model	Dataset	$\text{mAP}@0.5$ (%)	$\text{mAP}@0.5:0.95$ (%)	P (%)	R (%)	Size (MB)	FPS
Faster R-CNN (ResNet50) [17]	NEU-DET	92.30 ± 0.52	48.50 ± 0.68	91.80 ± 0.45	90.20 ± 0.51	42.50	25 ± 1
SSD (MobileNetV2) [19]	NEU-DET	89.70 ± 0.58	45.30 ± 0.72	88.90 ± 0.50	87.60 ± 0.55	3.20	110 ± 3
EfficientDet-Lite3 [28]	NEU-DET	94.50 ± 0.35	52.50 ± 0.61	93.70 ± 0.38	92.80 ± 0.42	4.80	95 ± 2
Proposed Model	NEU-DET	$97.80 \pm 0.22^{**}$	$63.30 \pm 0.40^{**}$	$97.20 \pm 0.25^{**}$	$96.90 \pm 0.28^{**}$	5.21	132 ± 2
Faster R-CNN (ResNet50) [17]	PCB	90.60 ± 0.55	47.20 ± 0.70	89.90 ± 0.48	88.50 ± 0.53	42.50	23 ± 1
SSD (MobileNetV2) [19]	PCB	88.30 ± 0.60	43.70 ± 0.75	87.50 ± 0.52	86.80 ± 0.58	3.20	107 ± 3
EfficientDet-Lite3 [28]	PCB	93.20 ± 0.38	50.90 ± 0.63	92.60 ± 0.40	91.70 ± 0.45	4.80	92 ± 2
Proposed Model	PCB	$97.00 \pm 0.25^{**}$	$61.70 \pm 0.42^{**}$	$96.50 \pm 0.27^{**}$	$96.20 \pm 0.30^{**}$	5.21	132 ± 2

$^{**} p < 0.01$, compared with all other methods (two-tailed *t*-test).

From Tables 6 and 7, we can see the following:

The proposed approach reaches the top $\text{mAP}@0.5$ (99.48%) and $\text{mAP}@0.95$ (73.29%) among all contrasted models, which is markedly higher than other YOLO models, particularly in $\text{mAP}@0.95$ (3.76 percentage points more than YOLOv11, 6.97 percentage points more than YOLOv8n), showing that the proposed approach possesses outstanding regression precision for large-scale weld defect targets. This performance advantage is more prominent than that of YOLOX-Ray in crack detection, which only achieved a 6.5 percentage point $\text{mAP}@0.95$ improvement over the baseline.

Regarding precision and recall, the proposed method also reaches the optimal performance ($P = 99.67\%$, $R = 99.65\%$), which is higher than mainstream YOLO models and

non-YOLO SOTA methods, reflecting its strong ability to avoid false detection and missed detection. In industrial applications, false positives increase operational costs, while false negatives pose significant safety risks.

The weight of the model is 5.21 MB for the proposed method, which is the same as YOLOv11. It is also lower than YOLOv6 (8.15 MB) and YOLOv10n (5.48 MB), but it is slightly higher than YOLOv5 (4.43 MB) and YOLOv9t (3.95 MB). However, its accuracy is much better and can be considered a trade-off between small weight and good accuracy. This balance is important for deployment on resource-limited industrial edge devices. Concurrently, with 132 FPS on NVIDIA RTX 4090, the proposed model outperforms all compared methods in real-time performance, meeting the requirements of high-speed industrial production lines.

5. Conclusions

This study proposes a lightweight YOLO-based weld defect detection technique that improves the model by means of WeldSimAM attention mechanism and EnNWD loss function. Experiments are carried out via 10-fold cross-validation on three datasets (self-built + two public), and it is found that the proposed algorithm can achieve 99.48% mAP@0.5, 73.29% mAP@0.5:0.95, with a precision of 99.67% and recall of 99.65%, 5.21 MB, and 132 FPS. Statistical significance testing confirms that all performance improvements are not due to chance ($p < 0.05$). In contrast to the widely used YOLO model and non-YOLO SOTA methods, the approach introduced in this paper not only boosts the detection precision of multi-scale weld defects considerably but also stays as a compact and lightweight model that excels at industrial weld defect detection tasks. The innovative design of WeldSimAM and EnNWD also provides a reference for the improvement of YOLO models in other industrial defect detection scenarios.

Author Contributions: Conceptualization, W.H.; methodology, W.H.; software, Q.C.; validation, J.Z. and Q.C.; formal analysis, Q.C.; investigation, Q.C.; resources, Q.C.; data curation, J.Z.; writing—original draft preparation, W.H.; writing—review and editing, W.H.; visualization, J.Z.; supervision, W.H.; project administration, W.H.; funding acquisition, Q.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Scientific Research Project of Universities in Anhui Province (Grant No. 2025AHGXZK30878).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets are available at <https://github.com/hwqwlsu/Weld-Defect-Detection> (accessed on 10 January 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Khumaidi, A.; Yuniarno, E.M.; Purnomo, M.H. Welding defect classification based on convolution neural network (CNN) and Gaussian kernel. In Proceedings of the 2017 International Seminar on Intelligent Technology and Its Applications, Surabaya, Indonesia, 28–29 August 2017; pp. 268–273. [\[CrossRef\]](#)
2. Czimmermann, T.; Ciuti, G.; Milazzo, M.; Chiurazzi, M.; Roccella, S.; Oddo, C.M.; Dario, P. Visual-Based Defect Detection and Classification Approaches for Industrial Applications—A Survey. *Sensors* **2020**, *20*, 1459. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Zhang, H.; Chen, Z.; Zhang, C.; Xi, J.; Le, X. Weld Defect Detection Based on Deep Learning Method. In Proceedings of the 2019 IEEE 15th International Conference on Automation Science and Engineering, Vancouver, BC, Canada, 22–26 August 2019; pp. 1184–1189. [\[CrossRef\]](#)
4. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8 [EB/OL]. GitHub Repository. 2023. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 19 November 2025).

5. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
6. Cengil, E. Weld Defect Detection with YOLOv10. *NATURENGS* **2024**, *5*, 77–81. [\[CrossRef\]](#)
7. Wang, A.; Chen, H.; Liu, L.; Han, K.; Ding, G.; Yao, H. YOLOv10: Real-time end-to-end object detection. *arXiv* **2024**. [\[CrossRef\]](#)
8. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [\[CrossRef\]](#)
9. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *arXiv* **2019**. [\[CrossRef\]](#)
10. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021; Volume 139, pp. 11863–11874. [\[CrossRef\]](#)
11. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [\[CrossRef\]](#)
12. Wang, J.; Xu, C.; Yang, W.; Yu, L. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv* **2021**. [\[CrossRef\]](#)
13. Wu, L.; Chu, Y.K.; Yang, H.G.; Chen, Y.X. Sim-YOLOv8 Object Detection Model for DR Image Defects in Aluminum Alloy Welds. *Chin. J. Lasers* **2024**, *51*, 1602103. [\[CrossRef\]](#)
14. Xu, J.; Ye, D.; Zhang, S.; Wang, K.; Chen, S. Metallic surface defect detection via NWD-WIoU based on grayscale co-generation entropy gain. *Appl. Intell.* **2025**, *55*, 752. [\[CrossRef\]](#)
15. Song, K.; Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl. Surf. Sci.* **2013**, *285*, 858–864. [\[CrossRef\]](#)
16. Jocher, G. Ultralytics YOLO [EB/OL]. GitHub Repository. 2024. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 15 July 2025).
17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
18. Wu, Z.; Jiao, C.; Sun, J.; Chen, L. Tire Defect Detection Based on Faster R-CNN. In *Communications in Computer and Information Science*; Springer: Singapore, 2020; Volume 1336, pp. 203–218. [\[CrossRef\]](#)
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37. [\[CrossRef\]](#)
20. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**. [\[CrossRef\]](#)
21. Wang, C.Y.; Liao, H.Y.M.; Yeh, I.H. YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv* **2024**. [\[CrossRef\]](#)
22. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 2117–2125. [\[CrossRef\]](#)
23. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [\[CrossRef\]](#)
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [\[CrossRef\]](#)
25. Zhao, Z.; Huang, W.Q.; Li, T.; Zhu, J. eCBAM and saIoU Co-Optimized YOLOv11 for Riverine Floating Garbage Classification Under Complex Aquatic Scenarios. *Appl. Sci.* **2026**, *16*, 651. [\[CrossRef\]](#)
26. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. UnitBox: An Advanced Object Detection Network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 521–525. [\[CrossRef\]](#)
27. Zhang, H.; Wang, S.; Wang, S.; Wang, Z. Shape-IoU: More accurate metric considering bounding box shape and scale. *arXiv* **2023**, arXiv:2307.02155. Available online: <https://arxiv.org/abs/2307.02155> (accessed on 9 November 2025).
28. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790. [\[CrossRef\]](#)
29. Wu, H.; Tang, M. PCB surface defect detection based on improved YOLOv7-tiny. In Proceedings of the 2023 5th International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Hangzhou, China, 15–17 December 2023; pp. 334–337. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.