Research on Engineering Structures & Materials
www.jresm.org
Research Group

Research Article

IYOLOv7-tiny-Weld: A lightweight model for ship laser welding defect detection on complex curved surfaces

Tong Chen [*,a]

School of Intelligent Manufacturing, Wenzhou Polytechnic, Wenzhou, China

Article Info

Article History:

Received 15 July 2026

Accepted 25 Aug 2026

Keywords:

Laser welding;
Defect detection;
IYOLOv7-tiny-Weld;
Shipbuilding;
Decoupled Self-Attention (DSA);
Multi-scale feature fusion;
Real-time monitoring

Abstract

Laser welding is a key joining process for large complex curved ship components, where precise real-time weld defect detection is critical to hull structural safety. However, current lightweight detection models struggle to simultaneously achieve high accuracy for tiny defects and high inference speed under industrial conditions, due to challenges such as low-contrast imaging, varying surface reflections from curved geometries, and the multi-scale nature of defect features. To address these issues, this paper proposes an improved YOLOv7-tiny model named IYOLOv7-tiny-Weld (Improved YOLOv7-tiny for Weld Defect Detection). A Decoupled Self-Attention (DSA) module is embedded to enhance local defect feature extraction, and Group Shuffle Convolution (GSConv) is adopted to offset the extra computational overhead introduced by DSA. The model further incorporates a Weighted Bidirectional Feature Pyramid Network (BiFPN) and Content-Aware Reassembly of Features (CARAFE) to achieve superior multi-scale feature fusion and reconstruction. In addition, Decoupled Detection Heads (DDH) are utilized to mitigate the conflicts between classification and regression tasks. Experiments are conducted on a self-built ship laser welding dataset containing 3,200 images, as well as on the public WELD dataset. On the self-built ship laser welding dataset, the model attains a median detection confidence of 89.75% and a median IoU of 82.16%. The model achieves AP values of 92.56%, 89.14%, and 85.43% for porosity, undercut, and lack of fusion, respectively. Under low illumination and long-distance conditions, the precision drops by only 4.17% and 3.41%, outperforming other compared models. Compared with the baseline YOLOv7-tiny, the proposed model achieves a 12.62 percentage point improvement in mAP@0.5 on the public WELD dataset. The modified YOLOv7-tiny achieves simultaneous advances in accuracy, inference speed, and environmental robustness, offering an efficient lightweight solution for real-time online ship laser welding quality monitoring.

1. Introduction

Laser welding offers concentrated energy and minimal thermal deformation, making it the mainstream advanced process for welding and forming large curved components in shipbuilding [1]. However, the welding process is susceptible to disturbances from process parameters, workpiece conditions, and environmental factors, which readily induce defects such as porosity, cracks, lack of fusion, and undercuts. These defects not only degrade the mechanical properties of welded joints but also directly affect the overall structural strength and service safety of ship components, necessitating the development of efficient and accurate defect detection methods [2]. In complex curved surface welding, curvature variations alter the laser incident angle and focal position, resulting in uneven heat input and unstable defect morphology. During image acquisition, varying reflection angles and shooting distances further introduce non-uniform illumination, scale

[*]Corresponding author: wdjxtt@sina.com
[a]https://orcid.org/0009-0009-9624-0036
DOI: http://dx.doi.org/10.17515/resm2026-1883vk0715rs

variation, blur, and geometric distortion, making defect detection significantly more challenging than on flat weld surfaces. Traditional ship weld inspection primarily relies on manual visual inspection, ultrasonic testing, and radiographic testing, which generally suffer from low efficiency, strong subjectivity, and the inability to perform online real-time screening, making them difficult to adapt to the quality inspection demands of intelligent shipbuilding production lines [3]. With the continuous advancement of deep learning-based vision detection technologies, real-time online inspection requires a detector that balances speed and accuracy. The YOLO (You Only Look Once) single-stage detection algorithm, offering fast detection speed, high recognition accuracy, and strong automation capabilities, provides an effective technical solution for online intelligent inspection of ship laser welding defects [4].

In weld defect detection, scholars conducted systematic studies around acoustic feature analysis and visual feature extraction. Yusof M F et al. based on pulsed laser welding acoustic signals, used a wavelet denoising method integrating Z-score threshold. Their constructed classification model reached 95% accuracy for weld defect recognition [5]. Qiu Y et al. targeted the laser-arc hybrid welding process. They designed a hybrid pooling attention module and embedded it into a Convolutional Neural Network (CNN). The recognition accuracy reached 95.02% on a test set of four types of defects [6]. Li H et al. developed a CNN detection algorithm based on a molten pool visual sensing system. They predicted defects by analyzing the morphology of molten pool dark spots and semi-solidified regions, and the accuracy exceeded 95% [7]. Zhang M et al. pointed out that deep learning could effectively make up for the shortcomings of traditional methods. However, it still faced interdisciplinary challenges of deep integration between algorithm design and welding process [8]. These studies proved that deep learning could detect weld defects well. However, lightweight detectors such as YOLOv7-tiny still have limitations in curved-surface weld defect detection. Standard convolutions mainly extract local features with relatively fixed receptive fields, making it difficult to model long-range dependencies in irregular weld morphologies. Meanwhile, repeated down sampling and insufficient multi-scale feature fusion may weaken small defect textures and boundary details, especially for pores, cracks, and incomplete penetration with large scale variations.

Under this background, the single-stage detection algorithm represented by YOLOv7 has become the mainstream path of industrial visual detection due to its excellent balance of precision and speed. For weld defect detection, Xu X et al. improved YOLOv7. They embedded the Le-HorBlock module in the backbone network to achieve second-order spatial interaction to retain high-order features, and introduced a coordinate attention mechanism to strengthen target representation. Their mAP@0.5 reached 78.6% [9]. Gu Q et al. proposed a surface defect detection model for prefabricated steel pipes based on YOLOv7-tiny. This model integrated a squeeze-and-excitation attention mechanism and distribution-shifted convolution. The calculation amount decreased by 39.4%, and the mAP increased by 1.9% [10]. Barman J et al. modified YOLOv7-tiny for fabric defect detection. They used an enhanced residual convolutional network and a double-splicing structure to strengthen hierarchical feature extraction. Their accuracy reached 98%, and the model could identify 1 mm micro defects [11]. Huang Y C et al. designed anchor-free light-weight YOLOv7-tiny network architecture. In this work, a multi-scale attention module and adaptive spatial feature fusion bidirectional feature pyramid were developed. Its accuracy has been improved by 6.24% and the speed is above 90 FPS [12].

Similarly, Ancha V K et al. developed TRSBi-YOLO, GSS-YOLO, and CSPGhost-YOLO for PCB defect detection through module-level optimizations such as C3TR, GSConv, GhostC3, and attention mechanisms [13-15]. The researches have demonstrated the potential of YOLOv7-tiny in industrial defect detection. Nevertheless, most existing studies have been conducted on flat or nearly planar weld surfaces, where defect appearance and imaging conditions are relatively stable. In contrast, complex curved surfaces introduce geometric distortion, non-uniform reflection, and apparent scale changes of defects, which significantly increase the difficulty of feature extraction and localization. Therefore, curved-surface weld defect detection should not be treated as a simple extension of flat-surface detection.

In addition, the introduction of attention mechanisms proved effective in enhancing the feature response of the model to defect areas. Narayanasamy S A et al. combined focus super-resolution with a self-attention-based single-candidate optimization strategy to construct an adversarial network for enhancing images of laser beam welding defects. This method effectively improved the visibility of defect regions [16]. Mo H et al. introduced the concept of decoupled focused self-attention for disease detection on cotton leaf. This approach ensured the dynamic tuning of self-attention weights and improved the relationship between features [17]. Guo Y et al. suggested a decoupled and self-attention routing in the capsule network. They modeled the relationships between channel and spatial dimension separately. This approach resulted in fewer parameter calculations [18]. These studies indicated that the Decoupled Self-Attention (DSA) design could effectively reduce computational complexity and improve the relevance of feature extraction. However, its application to laser welding defect detection for large-scale shipbuilding components with complex curvatures remains unexplored.

Therefore, to break through the technical bottlenecks of difficult fine-grained feature extraction, insufficient multi-scale fusion, and limited lightweight deployment in complex curved surface laser welding defect detection, this study uses YOLOv7-tiny as the basic architecture. This study proposes the IYOLOv7-tiny-Weld (Improved YOLOv7-tiny for Weld Defect Detection) model for laser welding defect detection of large components with complex curved surfaces. By integrating DSA and multi-scale feature enhancement, the model achieves accurate recognition and lightweight deployment of weld defects. It provides efficient and reliable technical support for online inspection of laser welding quality for large ship components with complex curved surfaces. The main contributions of this study are as follows:

- A curved-surface laser welding defect dataset for large shipbuilding components is constructed, covering typical defects such as pores, cracks, undercut, and incomplete penetration under complex imaging conditions.
- Group Shuffle Convolution (GSConv) and Decoupled Self-Attention (DSA) are introduced to enhance lightweight feature extraction. GSConv reduces computational redundancy, while DSA strengthens the modeling of long-range dependencies and irregular defect morphology.
- A BiFPN-based feature fusion structure combined with Content-Aware Reassembly of Features (CARAFE) up sampling is designed to improve multi-scale representation. Their joint use helps normalize the scale variance caused by curved surfaces and preserves fine defect boundaries during feature reconstruction.
- A Decoupled Detection Head (DDH) is adopted to separate classification and localization tasks, reducing the conflict between category recognition and bounding-box regression and improving detection robustness for defects with ambiguous boundaries.

2. Defect Detection Model Based on Improved Yolov7-Tiny

2.1. Overall Architecture of Improved Yolov7-Tiny

Online industrial weld seam inspection demands strict real-time algorithm performance and tiny defect recognition capabilities. Especially in the shipbuilding field, a large number of large components with complex curved surfaces are involved. Their laser welding seams extend along three-dimensional spatial curves, which feature large curvature changes and complex spatial directions. Moreover, the significant difference in surface reflection characteristics between the weld seam and the base metal further increases the difficulty of defect feature extraction [19]. The lightweight YOLOv7-tiny has a fast inference speed and balanced feature extraction performance, which fits on-site deployment requirements. Therefore, this study adopts it as the baseline model. Meanwhile, aiming at its shortboards in feature representation and localization, this study optimizes the backbone network, neck network, and detection heads respectively to construct the IYOLOv7-tiny-Weld model. Its structure is shown in Figure 1.

As shown in Figure 1, the input layer performs uniform preprocessing on laser welding images and crops them to a size of 640×640 pixels. The backbone network embeds DSA inside the original Efficient Layer Aggregation Network (ELAN) module to reinforce the feature extraction of fine-grained weld defects. Meanwhile, GSConv replaces part of standard convolutions to compensate for the computational increment brought by DSA, ensuring the lightweight characteristics of the model. Furthermore, to enhance feature reconstruction, the original interpolation upsampling is replaced with the CARAFE operator.

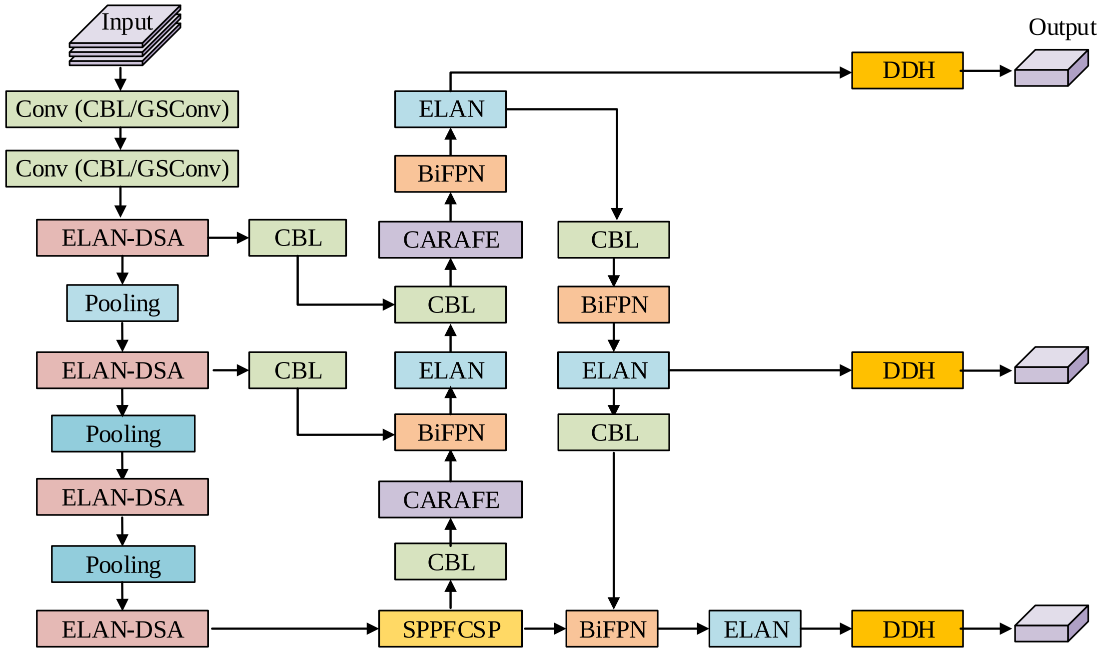


Fig. 1. Overall architecture of the IYOLOv7-tiny-Weld model

In the original coupled detection head, classification and localization share the same convolutional features. However, classification mainly relies on semantic information for category discrimination, whereas localization requires geometry-sensitive features for accurate boundary regression. Previous studies have shown that this inconsistency in feature requirements can lead to task conflict in object detection heads [20]. Therefore, DDH is introduced to separate the classification and localization branches, allowing each branch to learn task-specific representations. The structural difference between the traditional coupled detection head and the proposed DDH is illustrated in Figure 2.

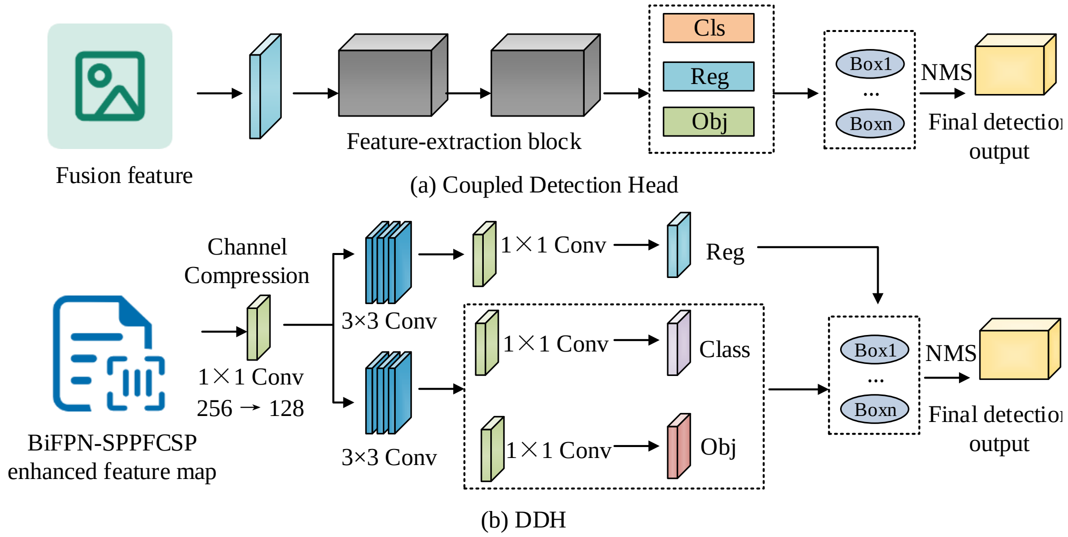


Fig. 2. Structural comparison between standard coupled detection head and DDH

As shown in Figure 2(a), the traditional coupled detection head uses shared convolutional layers to perform classification, objectness prediction, and bounding-box regression simultaneously. Since classification mainly depends on semantic features, whereas localization requires geometry-sensitive boundary features, the shared convolutional structure may cause feature conflict between the two tasks. In contrast, Figure 2(b) shows the proposed DDH structure. After channel compression through a 1×1 convolution, the enhanced feature map is fed into independent branches for classification, localization regression, and objectness prediction. The classification branch focuses on semantic defect-category information, while the regression branch learns geometry-related features for accurate bounding-box localization. Finally, Non-Maximum Suppression (NMS) is applied to remove redundant candidate boxes and retain the optimal detection results.

2.2. Improved Backbone Network Based on DSA And Gsconv

The native ELAN module of YOLOv7-tiny only relies on standard convolutions to extract features. Limited by local receptive fields, it is difficult to establish long-distance dependencies between pixels. Laser weld defects are scattered and irregular in morphology, and local textures are closely linked with global information. Therefore, this study embeds the DSA structure inside the ELAN module to strengthen feature expression in defect areas, as shown in Figure 3.

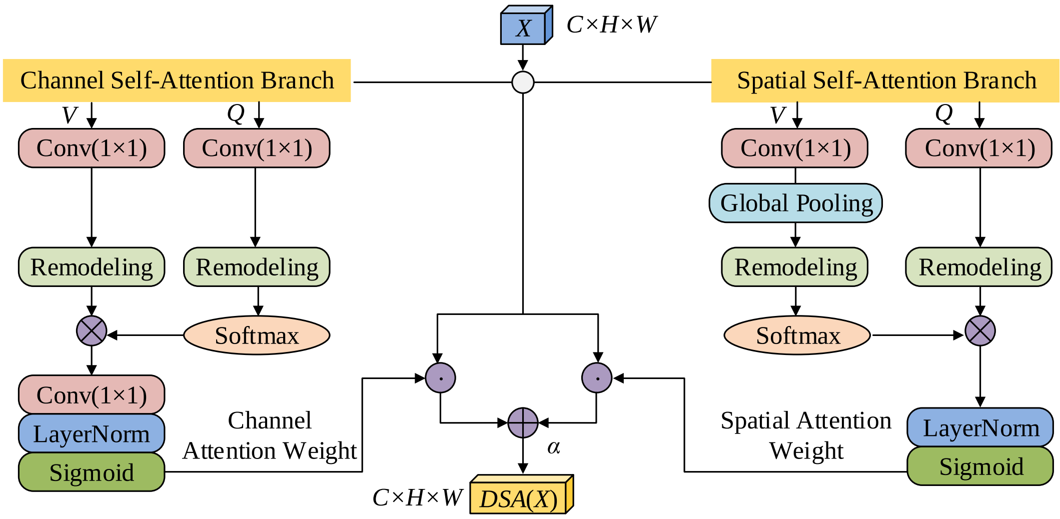


Fig. 3. Structure of the improved ELAN module with DSA

In Figure 3, DSA splits the self-attention mechanism into two independent branches: spatial attention and channel attention. They model the spatial position correlation and channel semantic correlation of feature maps respectively. Then, with the help of learnable parameters, the two outputs are adaptively fused, controlling computational overhead while strengthening feature representation capabilities. The feature tensor input to the backbone network undergoes linear mapping. The Remodeling blocks in each branch implement Reshape and Permute operations to rearrange feature tensor dimensions for attention computation. The basic scaled dot-product attention calculation method of the module is shown in Equation (1).

$$Attn(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

After the basic attention construction is completed, the spatial attention branch models contextual information within a single channel to extract structural features such as weld seam direction, contour morphology, and spatial distribution of defects. The channel attention branch fixes spatial coordinates and strengthens exclusive semantic features corresponding to different defects through cross-channel fusion. The operation method is shown in Equation (2).

$$\begin{cases} Attn_{sp}(X) = [Attn_{sp}(X_c)]_{c=1}^{C} \\ Attn_{ch}(X) = [Attn_{sp}(X_{hw})]_{h=1,w=1}^{H,W} \end{cases} \tag{2}$$

In Equation (2), $Attn_{sp}$ and $Attn_{ch}$ represent the operation results of spatial-dimension self-attention and channel-dimension self-attention, respectively. X_c is the two-dimensional space of the c -th channel, and X_{hw} is the cross-channel feature set under the same spatial position. To achieve adaptive fusion of the two types of attention features, this study introduces learnable weights and combines network layers to regulate the intensity of cross-channel interaction. Let l

be the current module layer index, and $\mathcal{M}$ ={4,6,8} be the set of network layers that enable cross-channel interaction. This selection is based on empirical experiments: shallow layers (1-3) primarily extract basic spatial features and are sensitive to channel attention, while deep layers (10-12) contain high-level semantics where cross-channel fusion brings limited gains. The middle layers (4,6,8) provide a balance between representational capacity and computational cost. When $l \in \mathcal{M}$, DSA simultaneously enables both branches and completes feature fusion, as shown in Equation (3).

$$DSA(X) = Proj(\alpha \cdot Attn_{ch}(X) + (1-\alpha) \cdot Attn_{sp}(X)) \tag{3}$$

In Equation (3), α is the learnable weight parameter, initialized to 0.5, which dynamically allocates the feature proportion of spatial attention and channel attention. Proj is the feature output projection layer, which can unify the dimension specifications of fused features to guarantee the continuity of feature transmission in subsequent network layers. The introduction of the DSA module improves detection accuracy but also brings extra computational overhead. To maintain the lightweight level of the model, this study introduces the GSConv module to replace part of standard convolutions, as shown in Figure 4.

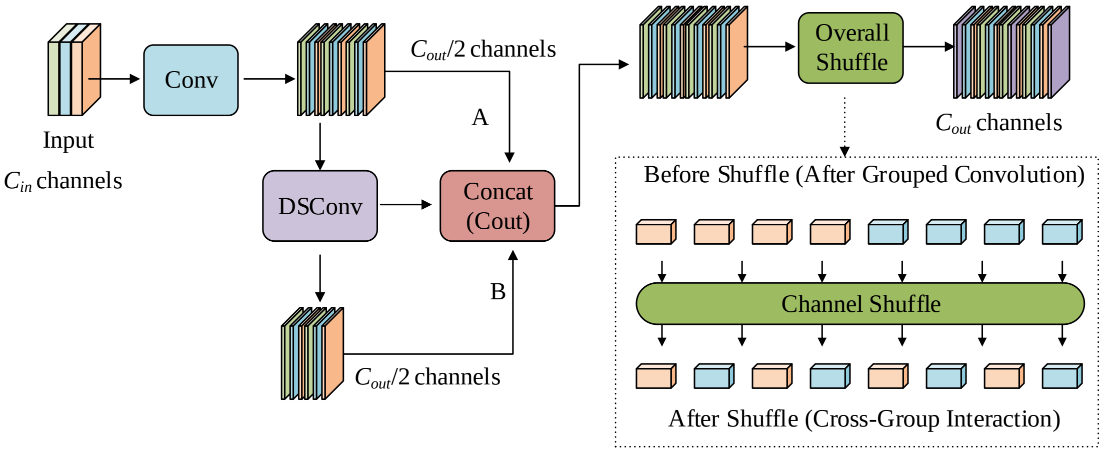


Fig. 4. Structure of the GSConv module

In Figure 4, GSConv consists of three sub-modules: standard convolution, Depthwise Separable Convolution (DSC), and channel shuffle. The Conv and DSConv branches generate Cout /2 channels, denoted as A and B, respectively. After concatenation, the feature map contains Cout channels, and the Overall Shuffle preserves the channel number while enabling cross-group information interaction. Its design is based on reducing the redundant computations of standard convolution through DSC, followed by restoring cross-group information flow via channel shuffle operations, thereby significantly lowering the parameter count and computational cost without sacrificing feature representation capability [21]. Among them, the parameter volume P_{SC} and computation volume F_{SC} of standard convolution is shown in Equation (4).

$$\begin{cases} P_{SC} = C_{in} \times C_{out} \times K \times K \\ F_{SC} = C_{in} \times C_{out} \times K \times K \times H' \times W' \end{cases} \tag{4}$$

In Equation (4), C_{in} and C_{out} represent the number of input channels and output channels, and K represents the side length of the convolutional kernel. When the convolutional stride is 1, $H' = H$, $W' = W$; when the stride is 2, $H' = H/2$, $W' = W/2$. DSC combines depthwise and pointwise convolutions, in which the former processes each input channel independently, while the latter performs cross-channel feature fusion using 1×1 convolution. Its calculation is shown in Equation (5).

$$\begin{cases} P_{DSC} = C_{in} \times K \times K + C_{in} \times C_{out} \\ F_{DSC} = C_{in} \times K \times K \times H' \times W' + C_{in} \times C_{out} \times H' \times W' \end{cases} \tag{5}$$

In Equation (5), P_{DSC} and F_{DSC} are the parameter volume and total floating-point operations of DSC. Thus, the computing power ratio of DSC relative to standard convolution is calculated as shown in Equation (6).

$$\frac{F_{DSC}}{F_{SC}} = \frac{1}{C_{out}} + \frac{1}{K^2} \tag{6}$$

In Equation (6), the value of C_{out} is usually large, so the $\frac{1}{C_{out}}$ term can be approximately ignored. The overall operation volume drops to about $\frac{1}{K^2}$ of standard convolution. Therefore, GSConv combines standard convolution with DSC and pairs them with channel shuffle to open up information interaction between grouped channels, which can effectively offset the extra power consumption brought by DSA. It is worth noting that GSConv is only applied to partial standard convolutions within the backbone rather than all convolutional layers. Replacing all backbone convolutions with GSConv would excessively weaken the model's feature extraction capacity for tiny, low-contrast weld defects, leading to obvious degradation of overall detection precision. This hybrid architecture achieves a favorable trade-off: GSConv offsets the extra computational overhead introduced by the DSA module, while retained standard convolutions preserve sufficient feature extraction capability for subtle micro-defect textures and edges.

2.3 Feature Fusion Network Based on Multi-Scale Feature Enhancement

In the original YOLOv7-tiny, the nearest neighbor interpolation is used in the process of up sampling. This technique can fill the features by fixed principles depending on their spatial positions but is unable to understand any semantics of weld defects. Thus, it will lead to missing some important information like small porosities and cracks' edges. This is why the CARAFE approach is used in up sampling, as illustrated in Figure 5.

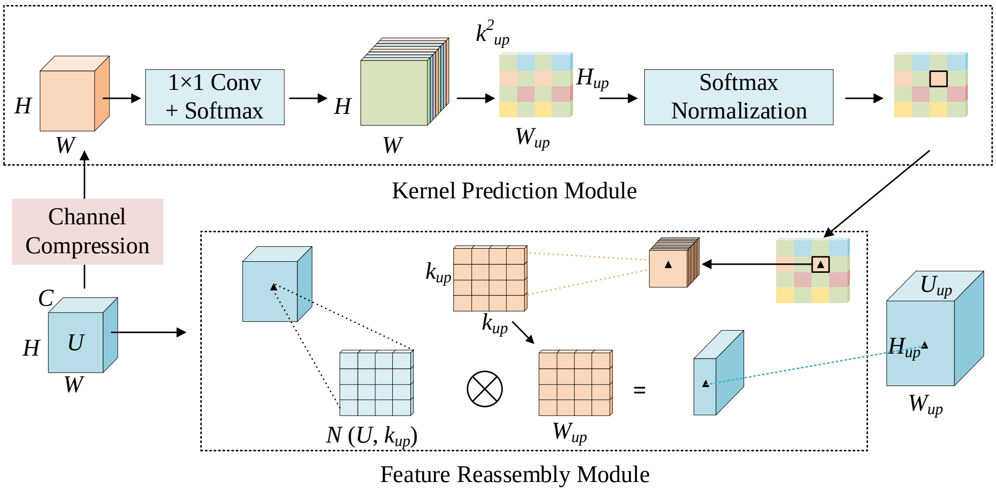


Fig. 5. Structure of the CARAFE module

Figure 5 illustrates the architecture of CARAFE, which includes an up-sampling kernel prediction module and a feature reassembly module. The former adaptively generates position-specific kernels from the input features, while the latter uses these kernels to reconstruct the feature map. Inspired by the effectiveness of feature reassembly and content-aware up sampling in steel surface defect detection and dense object detection [22-23], CARAFE is introduced in this study to better preserve boundary details in weld images. This is particularly beneficial for pores, cracks, and lack-of-fusion defects, which usually exhibit irregular shapes and weak boundaries. The upsampling kernel prediction process is shown in Equation (7).

$$\begin{cases} W_{up} = \psi(U) \\ W_{up} \in \mathbb{R}^{k_{up}^2 \times H_{up} \times W_{up}} \end{cases} \tag{7}$$

In Equation (7), $U \in \mathbb{R}^{C\times H\times W}$ represents the input feature map of the upsampling module, $\begin{cases} W_{up} = \psi(U) \\ W_{up} \in \mathbb{R}^{k_{up}^2 \times H_{up} \times W_{up}} \end{cases}$ represents the generated sampling weight kernel, k_{up}^2 represents the upsampling kernel size, and H_{up} and W_{up} are the height and width of the upsampled feature map respectively. The feature reassembly module performs weighted reassembly of input features based on W_{up}. Its operation is shown in Equation (8).

$$U_{up}(i,j) = \sum_{p=1}^{k_{up}} \sum_{q=1}^{k_{up}} W_{up}(p,q,i,j) \cdot U(i + p - \lfloor k_{up}/2 \rfloor, j + q - \lfloor k_{up}/2 \rfloor) \tag{8}$$

In Equation (8), $U_{up} \in \mathbb{R}^{C\times H\times W}$ is the upsampling output feature map. To improve multi-scale feature fusion efficiency, this study embeds BiFPN into the neck network. The high-resolution features after CARAFE up sampling are sent into the BiFPN bidirectional fusion branch to perform adaptive weighted fusion with shallow features of the same scale. The multi-layer feature fusion calculation rule is shown in Equation (9).

$$O_{OUT} = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i \tag{9}$$

In Equation (9), ε is a small constant set to 0.0001 to avoid division by zero. w_i is the feature weight updated autonomously during the network training process, and O_{OUT} is the composite feature after weighted fusion. The shallow layers contain molten pool boundary details, the middle layers correspond to defect contours, and the high layers carry global semantic information. To further enlarge the receptive field and enhance the recognition capability for defects of different specifications, this study further introduces the SPPFCSPC module. Its input feature map is $F \in$ ⍰$^{C\times H\times W}$. Based on the SPPFCSPC structure in YOLOv7, the module gradually expands the feature perception area through multi-layer max-pooling operations [24]. The three-layer serial pooling operation is shown in Equation (10).

$$\begin{cases} R_1(F) = MaxPool_{k=5}^{p=2}(F) \\ R_2(R_1) = MaxPool_{k=5}^{p=2}(R_1) \\ R_3(R_2) = MaxPool_{k=5}^{p=2}(R_2) \end{cases} \tag{10}$$

In Equation (10), R_1 represents the first pooling output feature, used to extract fine-grained local features such as molten pool edges and tiny porosity. R_2 represents the second pooling output, specifically capturing the complete spatial structure of medium-sized defects like cracks and undercuts. R_3 represents the third pooling output, fitting global context extraction for large-scale defects such as large area lack of penetration and wide cracks. Concatenating and integrating the features of three different perception ranges along the channel dimension blends local details and global structural information, yielding the enhanced feature tensor R_4, as shown in Equation (11).

$$R_4 = R_1 \otimes R_2 \otimes R_3 \tag{11}$$

In Equation (11), $\otimes$ represents the channel-dimension concatenation operation. The SPPFCSPC module adopts a multi-branch pooling paired with a cross-stage residual fusion structure, as shown in Figure 6. As illustrated in Figure 6, the pooling branches in the three groups of SPPFCSPC module are respectively dedicated to extracting features from defects of different scales, and the residual cross-connection can make the feature reuse ability become stronger. Thus, the model not only effectively regulates the total amount of computation but also enhances the effect of multi-scale detection, which meets the lightweight demands for real-time weld seam detection of industrial production lines. The BiFPN is responsible for the weighted fusion of cross-layer features, and the SPPFCSPC is responsible for the multi-scale receptive field extension of single-layer features. The multi-scale feature maps output from the SPPFCSPC module are fed as input features to the subsequent BiFPN for further cross-layer weighted feature fusion.

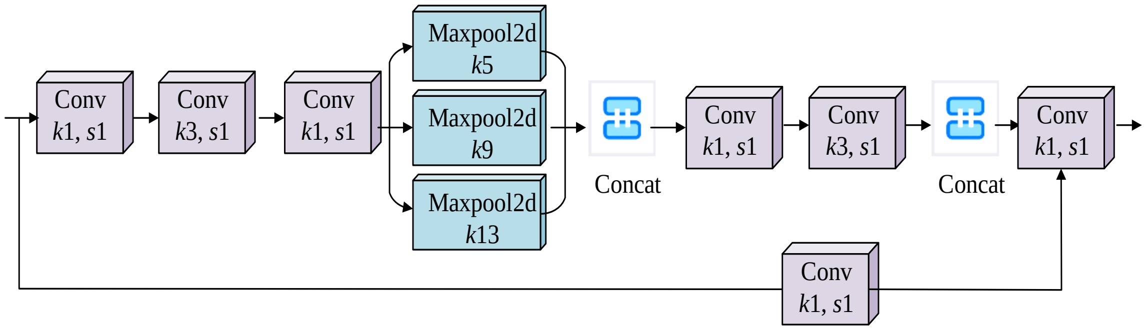


Fig. 6. Structure of the SPPFCSPC module

3. Laser Welding Defect Detection Results for Large Component with Complex Curved Surface

3.1 Performance Verification of Iyolov7-Tiny-Weld Model

Experiments were implemented using the PyTorch 1.11.0 deep learning framework with CUDA 11.3 on a platform equipped with an NVIDIA GeForce RTX 3090 GPU and an Intel Xeon Platinum 8255C CPU to evaluate the detection performance of the IYOLOv7-tiny-Weld model. During training, all input images were resized to 640×640 pixels. The model was optimized using SGD with an initial learning rate of 0.01, a momentum of 0.937, a weight decay of 0.0005, a batch size of 32, and 300 training epochs. Mosaic and MixUp data augmentation were employed to improve the model's generalization capability. Figure 7 provides schematic examples of the original weld sample, Mosaic stitching, MixUp blending, and the final network-fed training image.

The WELD dataset, publicly available on IEEE DataPort (DOI: 10.21227/k7rj-fk73), was used to evaluate the improved model. It contains high-resolution weld surface images with pixel-level annotations for defective regions, covering porosity, cracks, lack of fusion, undercuts, and pits. The dataset was randomly divided into training, validation, and test sets at a ratio of 7:1.5:1.5. The pixel-level annotations were converted into bounding-box annotations in YOLO format for object detection. To further investigate the limitations of IYOLOv7-tiny-Weld, representative failure cases from the test set were analyzed, as shown in Figure 8.

As shown in Figure 8, weld ripples easily trigger false positives, while tiny low-contrast defects often lead to missed detections. Complex weld textures also interfere with detection precision. These drawbacks reveal that micro-defect feature extraction needs further optimization. Ablation experiments were conducted to evaluate the contribution of each module. Starting from the baseline YOLOv7-tiny, the proposed modules (DSA, GSConv, CARAFE, BiFPN, SPPFCSPC, and DDH) were progressively integrated into the network. The results are shown in Table 1.

Table 1. Ablation experimental results on the public WELD dataset

No.	Modules	mAP@0.5 (%)	Params (M)	GFLOPs	Model Size (MB)	Latency (ms)	FPS
1	Baseline	76.31 ± 0.42	3.26	6.52	12.43	11.20	89.27
2	+ DSA	79.52 ± 0.38	3.41	6.98	13.01	11.53	86.74
3	+ GSConv	79.24 ± 0.35	3.28	6.37	12.51	11.31	88.51
4	+ CARAFE	81.43 ± 0.40	3.30	6.49	12.59	11.48	87.13
5	+ BiFPN	83.62 ± 0.36	3.33	6.61	12.73	11.65	85.86
6	+ SPPFCSPC	85.76 ± 0.33	3.35	6.68	12.78	11.82	84.62
7	+ DDH	88.93 ± 0.29	3.37	6.75	12.86	12.16	83.31

Note: All experiments were conducted on an NVIDIA RTX 3090 GPU with 640×640 input resolution. The mAP@0.5 values are reported as mean ± standard deviation over three independent training runs.

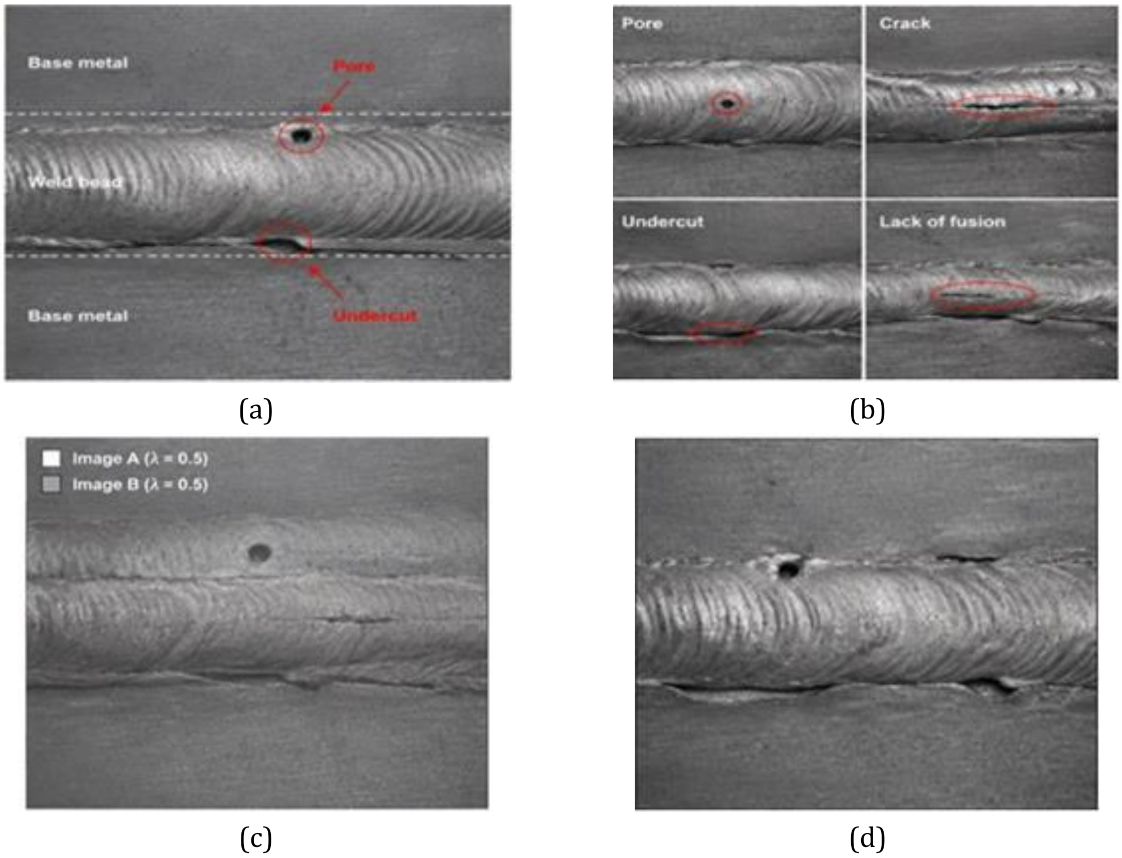


Fig. 7. Illustration of data augmentation strategies for weld defect detection (a) Original image (b) Mosaic augmentation (c) MixUp augmentation (d) Final training sample

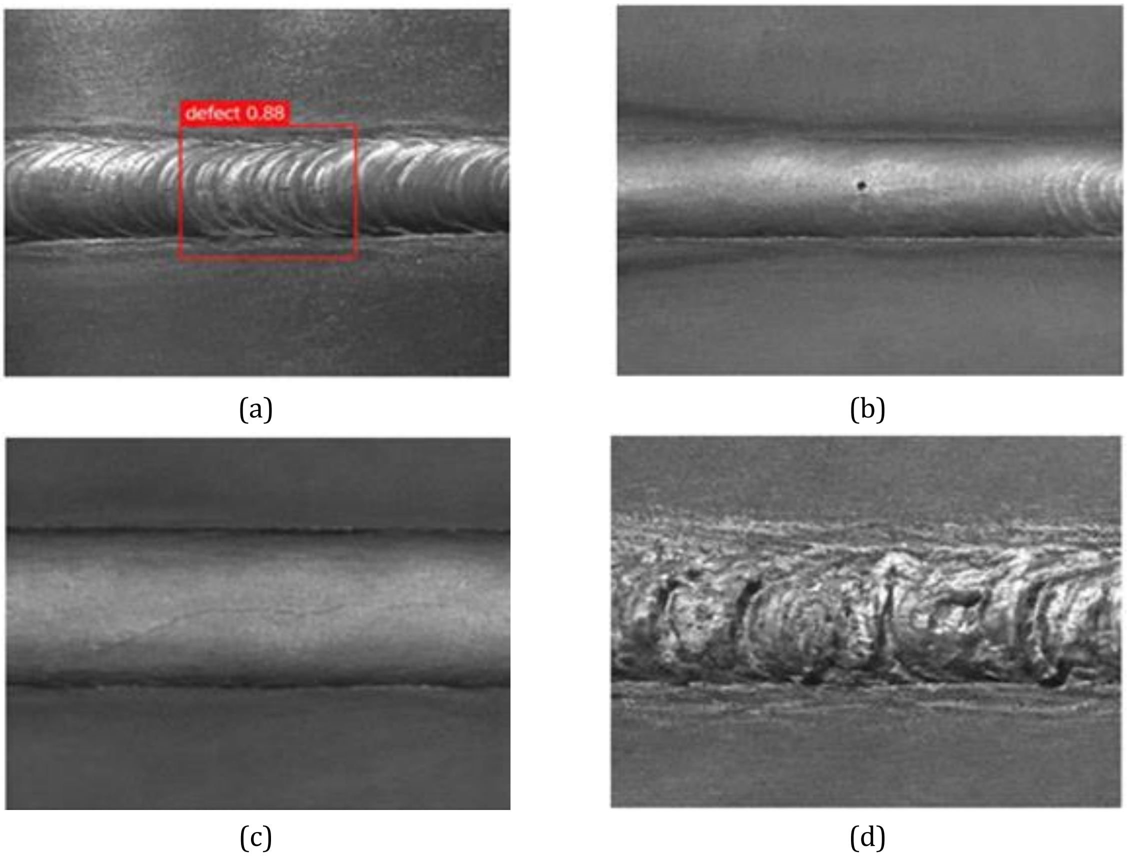


Fig. 8. Representative error and challenging detection cases (a) False alarm case (b) Missed detection case (c) Difficult sample: low-contrast micro-crack (d) Difficult sample: complex texture interference

Table 1 shows that the baseline YOLOv7-tiny achieves 76.31 ± 0.42% mAP@0.5 with 3.26 M parameters, 6.52 G FLOPs, 12.43 MB, 11.20 ms, and 89.27 FPS. Introducing DSA improves mAP by 3.21% with a slight increase in computational cost, confirming its effectiveness in enhancing fine-grained weld features. Adding GSConv reduces parameters to 3.28 M, FLOPs to 6.37 G, and model size to 12.51 MB while recovering FPS to 88.51 and latency to 11.31 ms, with only a 0.28% mAP drop, demonstrating its role in compensating for the overhead introduced by DSA. With the progressive integration of CARAFE, BiFPN, SPPFCSPC, and DDH, the model achieves 88.93% mAP@0.5 with 3.37 M parameters, 6.75 G FLOPs, 12.86 MB, 12.16 ms, and 83.31 FPS. Standalone ablation shows that DSA alone reaches 79.44% and DDH alone reaches 78.15%, confirming their independent contributions beyond the cumulative gains. To comprehensively evaluate the detection performance of IYOLOv7-tiny-Weld, this study analyzed the inference speed under different input resolutions and the changes in Average Precision (AP) under different Intersection over Union (IoU) thresholds. The proposed model was compared with several representative detectors: WDFCOS, a weld-specific detector based on FCOS; Cascade R-CNN, a classical two-stage detector; and two YOLO-series models—YOLOv7 and YOLOv8s—as strong single-stage baselines. These models were selected to cover diverse detection paradigms for a comprehensive evaluation. For fair comparison, all baseline models were retrained from scratch on the same dataset under identical experimental conditions, without using any pre-trained weights. The same input resolution, optimizer, learning rate, batch size, and training epochs were applied consistently across all models. No additional data augmentation or training strategies were employed beyond those used for the proposed model. The results are shown in Figure 9.

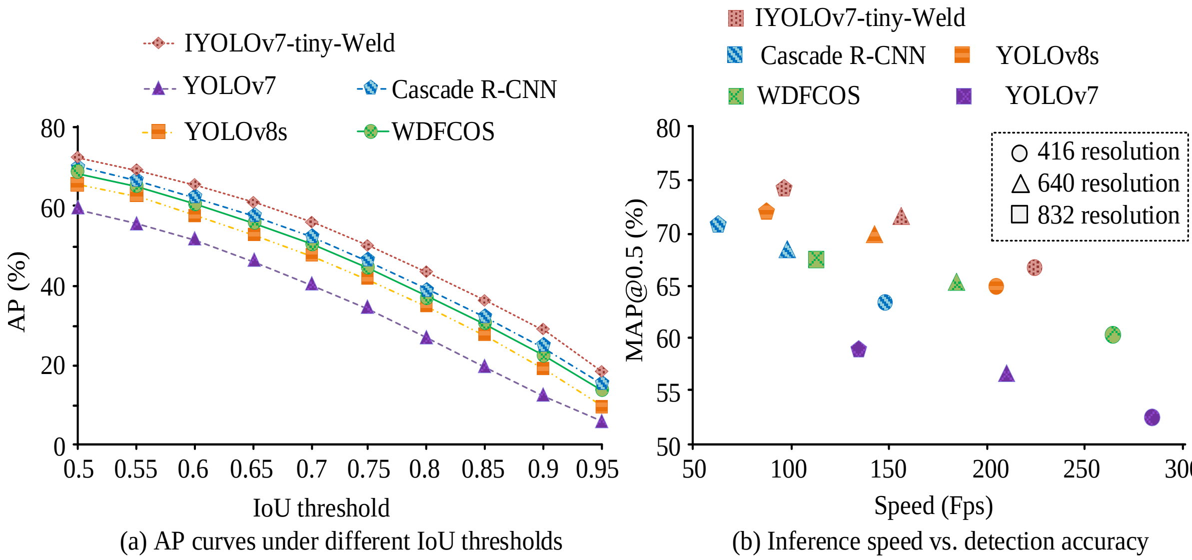
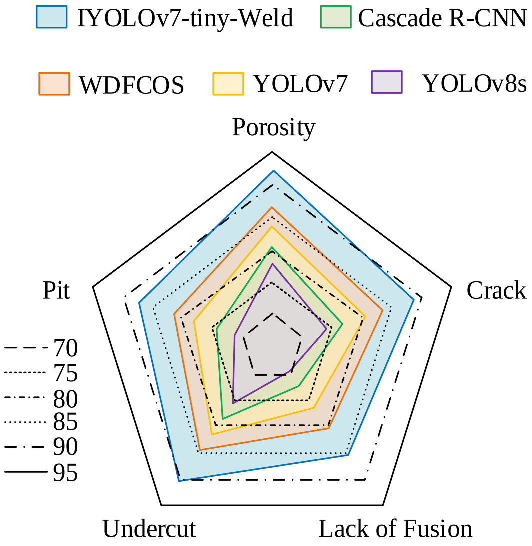
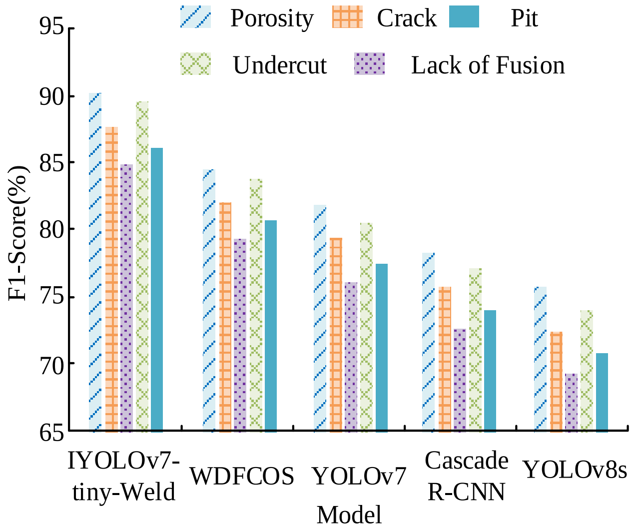


Fig. 9. Comparison of detection performance

Figure 9(a) shows that the curve of IYOLOv7-tiny-Weld had the gentlest downward trend. When the IoU thresholds were set to 0.5, 0.7, and 0.95 respectively, its AP values reached 72.86%, 61.27%, and 19.22% in sequence. This performance was clearly superior to the comparison models. In Figure 9(b), under different input scales, the inference frame rate of the IYOLOv7-tiny-Weld model reached 94.61 FPS and above, and the corresponding mAP@0.5 reached 67.58% and above, which was also better than the comparison models. This indicated that while IYOLOv7-tiny-Weld achieved high detection precision, its inference speed also increased significantly. Therefore, it is more suitable for deployment in actual online inspection scenarios of welding workpieces. To further evaluate the detection performance of IYOLOv7-tiny-Weld on different types of weld defects, this study counted the AP values and F1-Scores of each model for five categories of defects, including porosity, cracks, lack of fusion, undercut, and pits. The results are shown in Figure 10.



(a) AP comparison on various defect types



(b) F1-Score comparison on various defect types

Fig. 10. Comparison of detection performance

Figure 10(a) shows that the AP of IYOLOv7-tiny-Weld on all five categories of defects was higher than that of other comparison models. Specifically, the AP of porosity and undercut reached 91.92% and 90.13% respectively, the AP of cracks and pits was 88.73% and 87.26% respectively, and the AP of lack of fusion defects was 85.44%. This advantage originated from the fact that DSA and DDH could effectively capture the differentiated features of different morphology defects. Especially, they had a stronger extraction capability for the edge semantic information of fine-grained defects. Figure 10(b) shows that the F1-Scores of IYOLOv7-tiny-Weld for the detection of porosity, cracks, lack of fusion, undercut, and pits were 90.22%, 87.63%, 84.81%, 89.55%, and 86.18% in sequence. This further verified the effectiveness of its collaborative design of multi-scale feature enhancement and DDH in complex weld defect recognition. To visually demonstrate the effectiveness of the proposed DSA module, Grad-CAM was employed to generate attention heatmaps, as shown in Figure 11.

As can be observed in Figure 11, the heatmap's high-response region precisely covers the porosity defect and overlaps the predicted bounding box (confidence 0.92). Only scattered weak activation occurs on surrounding weld ripples. This confirms that the network concentrates feature response on real defects rather than background textures, verifying the reliability and interpretability of IYOLOv7-tiny-Weld. To further verify the generalization capability of IYOLOv7-tiny-Weld, lightweight YOLO variants from YOLOv5-tiny to YOLOv12-tiny are selected for comparison on the public NEU-DET dataset, covering successive model iterations. All competing models are trained from scratch under identical configurations on the same dataset. The results are summarized in Table 2.

Table 2. Comparison with state-of-the-art lightweight detection models on the NEU-DET dataset

Model	mAP@0.5 (%)	Params (M)	GFLOPs	Model Size (MB)	Latency (ms)	FPS
YOLOv5-tiny	72.56	3.12	6.43	11.86	11.32	88.34
YOLOv7-tiny	73.84	3.26	6.52	12.43	11.20	89.27
YOLOv8-tiny	74.91	3.44	7.19	13.07	12.15	82.31
YOLOv10-tiny	75.68	3.56	7.36	13.53	12.42	80.52
YOLOv12-tiny	76.52	3.68	7.61	13.98	12.70	78.74
WDFCOS	79.35	14.86	49.72	56.47	20.41	49.00
Cascade R-CNN	82.18	68.75	198.64	261.25	125.00	8.00
IYOLOv7-tiny-Weld	85.66	3.37	6.75	12.86	12.16	83.31

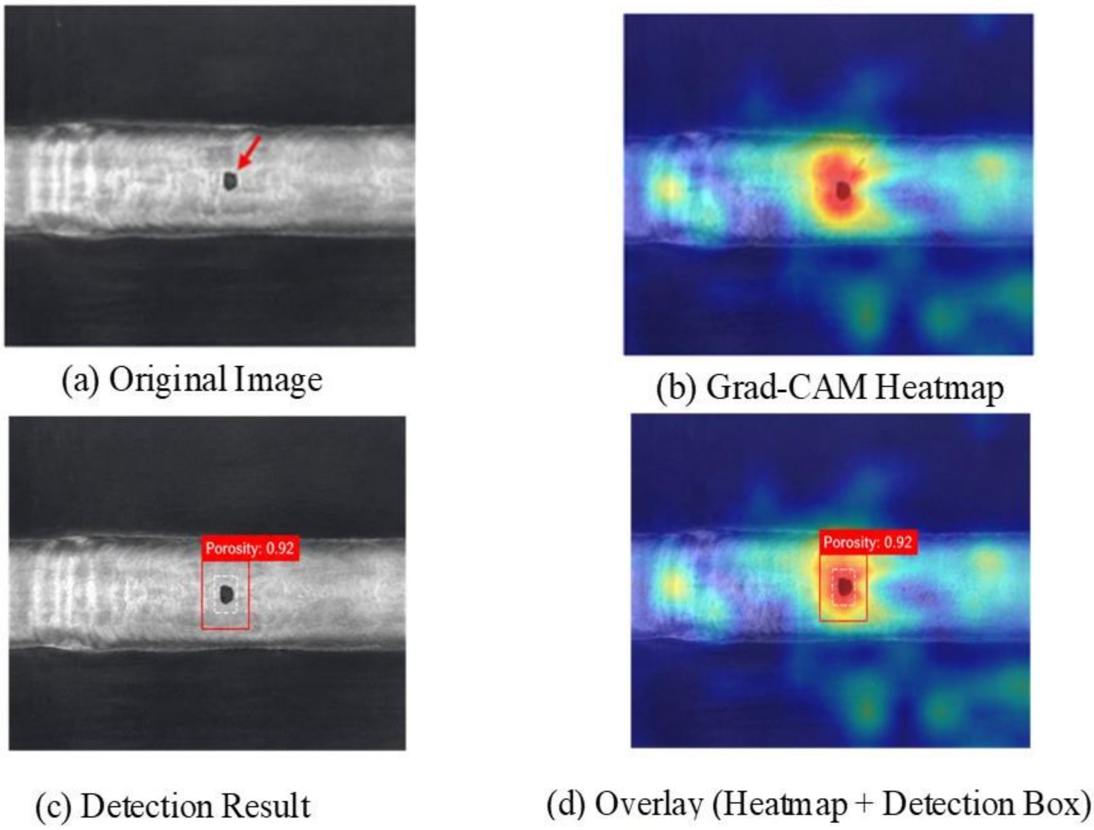


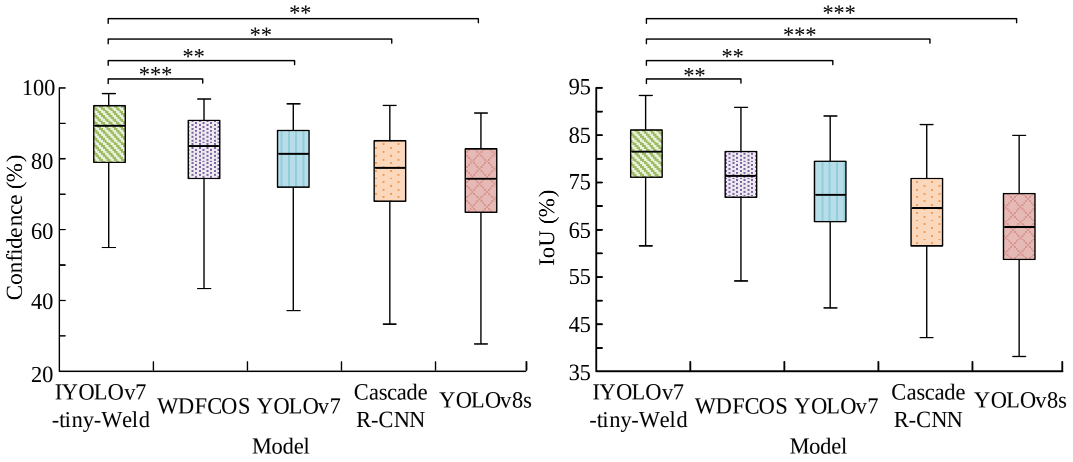
Fig. 11. Grad-CAM visualization of weld defect detection from independent test set

As shown in Table 2, IYOLOv7-tiny-Weld achieves the highest mAP@0.5 (85.66%) on the NEU-DET dataset among all competitors, while maintaining only 3.37M parameters and 6.75G FLOPs, demonstrating its superior balance between detection accuracy and computational efficiency, as well as strong generalization capability across different defect detection tasks.

3.2 Application Detection Results of Ship Laser Welding Defects

To verify the applicability of IYOLOv7-tiny-Weld in actual industrial scenarios, a self-built laser welding defect dataset was used in the experiment. The images came from a laser-arc hybrid welding production line of AH36 marine steel plates in a shipbuilding enterprise. A total of 3,200 weld surface images were collected, covering five categories: porosity (680), undercut (540), lack of fusion (490), spatter (510), and normal welds (980). To prevent overfitting on this relatively limited dataset, Mosaic and MixUp augmentations were applied during training. To evaluate the detection reliability and localization precision of IYOLOv7-tiny-Weld, the confidence scores of predicted bounding boxes and the corresponding IoU values with the ground truth were analyzed for each method, as illustrated in Figure 12.

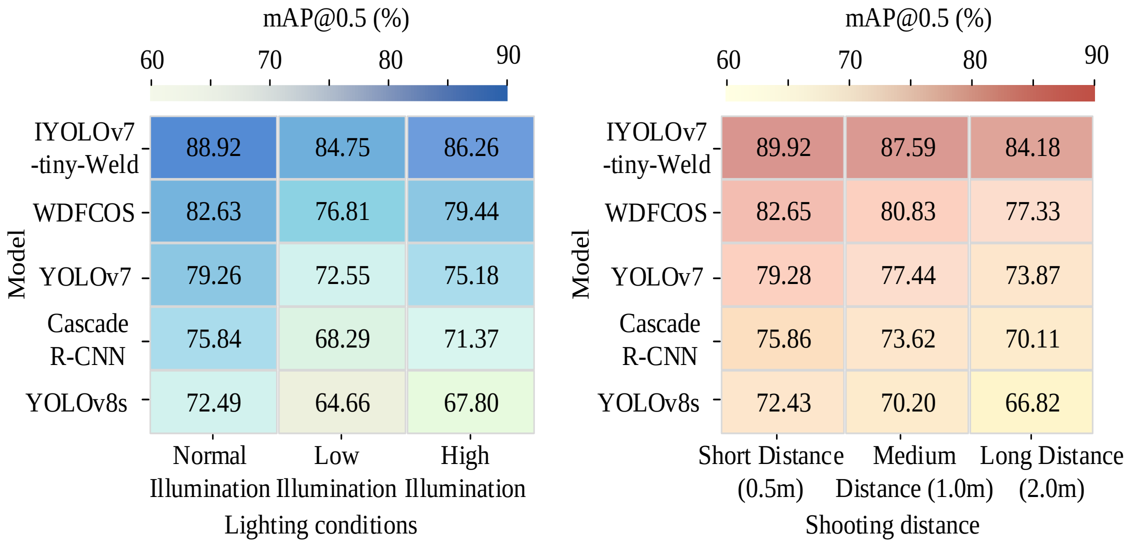
Figure 12(a) shows that the median detection confidence of IYOLOv7-tiny-Weld reached 89.75%, and the confidence distribution was concentrated in the high confidence interval of 80%–95%. Significant differences existed when compared with other models (p<0.05). This indicated that IYOLOv7-tiny-Weld had a higher prediction reliability in weld defect detection. Figure 12(b) shows that the median IoU of IYOLOv7-tiny-Weld reached 82.16%, and the IoU distribution was concentrated in the high precision interval of 75%–87%, which was significantly better than other comparison models (p<0.05). This demonstrated that the DDH decoupled detection head effectively improved the precise localization capability of the model for weld defect boundaries by independently optimizing the regression task, reducing the risk of localization offset. To verify the detection robustness of IYOLOv7-tiny-Weld, this study counted its detection precision under two working conditions: illumination changes and shooting distance changes. Illumination conditions were categorized as normal (500–800 lux), low (50–100 lux), and high (1500–2000 lux), representing typical lighting variations on shipbuilding shop floors. Shooting distances were categorized as short (0.5 m), medium (1.0 m), and long (2.0 m). The results are shown in Figure 13.



(a) Confidence Distribution (b) IoU Distribution

Fig. 12. Distribution of detection box confidence and IoU across different models

*Note: * and ** indicate statistically significant differences at p<0.05 and p<0.01, respectively, based on independent samples t-test.*



(a) Detection accuracy under different lighting conditions (b) Detection accuracy at different shooting distances

Fig. 13. Detection precision under different illumination intensities and shooting distances

Figure 13(a) shows that IYOLOv7-tiny-Weld achieved an mAP@0.5 of 88.92% under normal illumination. When the illumination was reduced, its mAP@0.5 decreased by 4.17%, representing a smaller decline than those of the other models. Under strong illumination conditions, the mAP of IYOLOv7-tiny-Weld was 86.26%, which was higher than other models. Figure 13(b) shows that the mAP of IYOLOv7-tiny-Weld at a long distance was 84.18%, which only dropped by 3.41% compared with the medium distance. This decrease was smaller than that of other comparison models. These results indicated that IYOLOv7-tiny-Weld possessed strong adaptability to actual working conditions, satisfying the weld defect detection needs in complex environments on shipbuilding sites. To compare the performance of different models across five weld defect categories, the AP and F1-score of each defect type, including porosity, cracks, lack of fusion, undercut, and pits, were analyzed. The results are shown in Figure 14.

Figure 14(a) shows that the AP of IYOLOv7-tiny-Weld on all five defect categories was higher than that of other comparison models. Among them, the AP of porosity and undercut reached 92.56% and 89.14%, respectively, and the AP of lack of fusion was 85.43%. This verifies the effectiveness of the multi-scale feature fusion module. The fine-grained features R_1 benefit small defects such as porosity, while R_2 and R_3 provide contextual information for medium and large defects like cracks, undercut, and lack of fusion. Figure 14(b) shows that the F1-Scores of IYOLOv7-tiny-Weld on the five categories of defects were also superior to the comparison models. Its F1-Scores for porosity, cracks, lack of fusion, undercut, and pits were 90.28%, 87.55%, 84.83%, 89.51%, and 86.16% respectively. This demonstrated that the model possessed stable and balanced recognition performance for all types of weld defects. Especially, it could maintain high detection precision for easily confused defects like lack of fusion and pits, providing reliable technical support for ship laser welding quality detection.

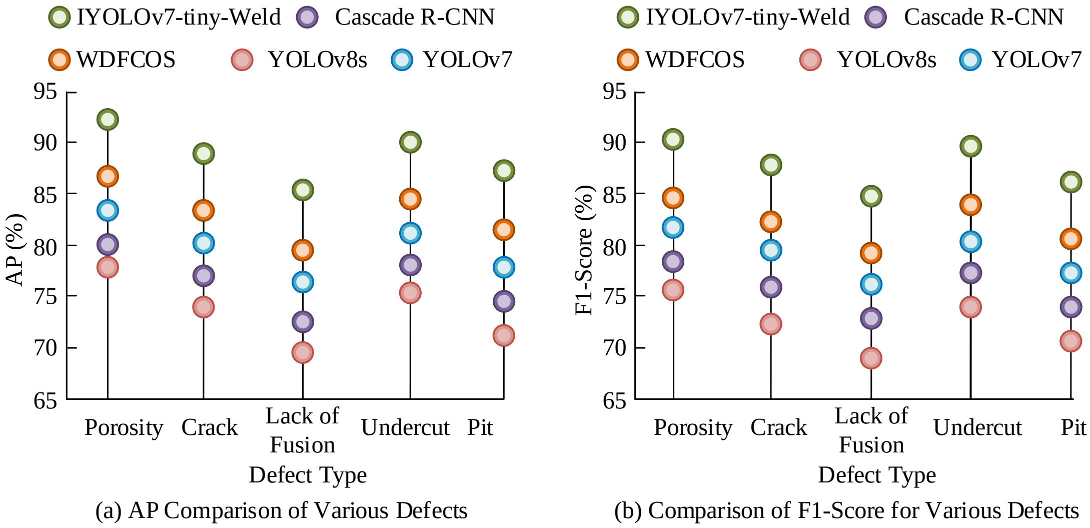


Fig. 14. Model performance comparison regarding AP values and F1-Scores on self-built ship laser welding dataset

4. Conclusions and Perspectives

Aiming at the laser welding defect detection needs for large components with complex curved surfaces, this study proposed IYOLOv7-tiny-Weld. The proposed model optimized the backbone network through DSA and GSConv, and accomplished multi-scale feature enhancement relying on BiFPN, CARAFE, and SPPFCSPC. By incorporating DSA to capture long-range dependencies and BiFPN with CARAFE to normalize multi-scale features, the model effectively handles geometric distortion and varying surface reflections caused by curved geometries. Meanwhile, it separated detection tasks through DDH, which effectively improved defect recognition and localization precision. Ablation experiments on the public WELD dataset reveal that the full proposed model achieves a 12.62% increase in mAP@0.5 compared with vanilla YOLOv7-tiny, while increasing the parameter count by only 0.11 M and maintaining an FPS of 83.31. Compared with mainstream detection models, its AP curve on the public WELD dataset was the best throughout the entire process, demonstrating a significant comprehensive advantage in inference speed and detection precision. In actual ship laser welding defect detection, the median detection confidence of the model reached 89.75%, and the median IoU was 82.16%. Its AP values and F1-Scores for multiple categories of defects, such as porosity, undercut, and lack of fusion, were all superior to the comparison models. These results demonstrate that IYOLOv7-tiny-Weld has a good adaptability to different types of defects and complex working conditions. It provides an effective technical solution for the online inspection of laser welding quality in industrial scenarios like shipbuilding. However, its detection precision under low illumination (50–100 lux) and long-distance (2.0 m) conditions remains a key challenge that will be addressed in future work, where the performance drops by 4.17% and 3.41%, respectively. Future work will enhance robustness under extreme conditions via data augmentation and domain adaptation. Additionally, edge deployment will be optimized through pruning and quantization, with evaluation on devices such as NVIDIA Jetson Nano to validate its practicality on real-world shipbuilding production lines.

Acknowledgement

The research is supported by: Zhejiang Province Visiting Engineer Project:"Research on High-Power Fiber Laser Robotic Welding Process"(No. FG2022057).

References

[1] Herzog T, Brandt M, Trinchi A, Sola A, Molotnikov A. Process monitoring and machine learning for defect detection in laser-based metal additive manufacturing. Journal of Intelligent Manufacturing,2024, 35(4): 1407-1437. https://doi.org/10.1007/s10845-023-02119-y

[2] Huang W, Gao X, Huang Y, Zhang Y. Improved convolutional neural network for laser welding defect prediction. International Journal of Precision Engineering and Manufacturing, 2023, 24(1): 33-41. https://doi.org/10.1007/s12541-022-00729-9

[3] Baek D, Moon H S. Robust laser vision sensing technology for adaptive weld seam tracking in shipbuilding automation. The International Journal of Advanced Manufacturing Technology, 2025, 139(3): 2003-2026. https://doi.org/10.1007/s00170-025-15896-y

[4] Kabir S M H, Rahman M T, Mridul A H. Software Defect Prediction Using Traditional Machine Learning and Ensemble Learning Algorithms. Smart Wearable Technology, 2025, 1: A9. https://doi.org/10.47852/bonviewSWT52025645

[5] Yusof M F M, Quazi M M, Aleem S A A, Ishak M, Ghazali M F. Identification of weld defect through the application of denoising method to the sound signal acquired during pulse mode laser welding. Welding in the World, 2023, 67(5): 1267-1281. https://doi.org/10.1007/s40194-023-01472-z

[6] Qiu Y, Ping J, Shu L, Song M, Ma D, Yan X, Li S. Defect monitoring of high-power laser-arc hybrid welding process based on an improved channel attention convolutional neural network. Journal of Intelligent Manufacturing, 2025, 36(4): 2657-2676. https://doi.org/10.1007/s10845-024-02354-x

[7] Li H, Ma Y, Duan M, Wang X, Che T. Defects detection of GMAW process based on convolutional neural network algorithm. Scientific Reports, 2023, 13(1): 21219-21235. https://doi.org/10.1038/s41598-023-48698-x

[8] Zhang M, Feng M, Chen C, Yu X, Lian G. Weld defect detection: deep learning-based image processing and the mechanisms of defect formation. Archives of Computational Methods in Engineering, 2026, 33(3): 3525-3563. https://doi.org/10.1007/s11831-025-10407-4

[9] Xu X, Li X. Research on surface defect detection algorithm of pipeline weld based on YOLOv7. Scientific reports, 2024, 14(1): 1881-1900. https://doi.org/10.1038/s41598-024-52451-3

[10] Gu Q, Yue X, Huang Y, Jian A, Huang X. D-YOLOv7-tiny: a lightweight network for defect detection of prefabricated steel pipe. Journal of Electronic Imaging, 2024, 33(4): 043017. https://doi.org/10.1117/1.JEI.33.4.043017

[11] Barman J, Kuo C F J. Fully automatic and precisely woven fabric defect detection using improved YOLOv7-tiny model utilizing enhanced residual convolutional network. Fibers and Polymers, 2025, 26(1): 353-368. https://doi.org/10.1007/s12221-024-00811-1

[12] Huang Y C, Lin J C, Wu Y Z. Design of lightweight metal surface defect detection technology for YOLOv7-tiny using Anchor-Free algorithm. Scientific Reports, 2026, 16(1): 8601-8628. https://doi.org/10.1038/s41598-026-39233-9

[13] Ancha V K, Gonuguntla V, Vaddi R. TRSBi-YOLO: Transformer based lightweight and high-performance model for PCB defects detection. The Journal of Supercomputing, 2025, 81(9): 1277. https://doi.org/10.1007/s11227-025-07771-0

[14] Ancha V K, Gonuguntla V, Vaddi R. GSS-YOLO: an improved YOLOV5 prediction head with slim-neck for defect detection in printed circuit board assembly. Signal, Image and Video Processing, 2025, 19(9): 915. https://doi.org/10.1007/s11760-025-04511-y

[15] Ancha V K, Gonuguntla V, Vaddi R. CSPGhost-YOLO: a lightweight and robust model for real-time mixed PCB defect detection system. Measurement, 2026, 236: 117230. https://doi.org/10.1016/j.measurement.2026.117230

[16] Narayanasamy S A, Kulandaivel R. Laser beam weld defect detection by image enhancement using focused super resolution with self attention single candidate optimizer and generative adversarial networks. International Journal on Interactive Design and Manufacturing (IJIDeM), 2025, 19(11): 8109-8126. https://doi.org/10.1007/s12008-025-02290-x

[17] Mo H, Wei L. Lightweight detection of cotton leaf diseases using StyleGAN2-ADA and decoupled focused self-attention. Journal of King Saud University Computer and Information Sciences, 2025, 37(4): 41-62. https://doi.org/10.1007/s44443-025-00054-x

[18] Guo Y, Zhang S, Zhang C, Gao H, Li H. A lightweight capsule network via channel-space decoupling and self-attention routing. Multimedia tools and applications, 2024, 83(35): 83513-83533. https://doi.org/10.1007/s11042-024-18861-1

[19] Nilsen M, Sikström F. Integrated vision-based seam tracking system for robotic laser welding of curved closed square butt joints. The International Journal of Advanced Manufacturing Technology, 2025, 137(7): 3387-3399. https://doi.org/10.1007/s00170-025-15357-6

[20] Song G, Liu Y, Wang X. Revisiting the sibling head in object detector. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11563-11572. https://doi.org/10.1109/CVPR42600.2020.01158

[21] Yang L, LI M, HU J, WANG D. Strip steel surface defect detection algorithm based on improved yolov7-tiny. Computer Engineering, 2025, 51(1): 208-215. https://doi.org/10.19678/j.issn.1000-3428.0068397

[22] Subburaj K, Mazroa A A, Alotaibi F A, Alnfiai M M. DBCW-YOLO: an advanced yolov5 framework for precision detection of surface defects in steel. Matéria (Rio de Janeiro), 2024, 29(4), e20240549. https://doi.org/10.1590/1517-7076-RMAT-2024-0549

[23] Bae M H, Park S W, Park J, Jung S H, Sim C B. YOLO-RACE: reassembly and convolutional block attention for enhanced dense object detection. Pattern Analysis and Applications, 2025, 28(2): 90-106. https://doi.org/10.1007/s10044-025-01471-4

[24] 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. 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Ieee, 2023: 7464-7475. https://doi.org/10.1109/CVPR52729.2023.00721