Computer Science and Information Systems 23(1):255–276 https://doi.org/10.2298/CSIS250922007Z

Weld proximity defect detection model for steel thin plates based on EP-YOLOv7

Runmei Zhang[1], Jingwei Fan[2], Zihua Chen[1,3,⋆], Zhong Chen[2], and Bin Yuan[2]

[1] School of Electronic and Information Engineering, Anhui Jianzhu University, 230000 Hefei, China
zhangrong@ahjzu.edu.cn
czh1619@ahjzu.edu.cn

[2] School of Mechanical and Electrical Engineering, Anhui Jianzhu University, 230000 Hefei, China
fjw3516@stu.ahjzu.edu.cn
cz1982@ahjzu.edu.cn
yuanbinwork@163.com

[3] Provincial and Ministerial Key Laboratory, Chang'an University, 710000 Xi'an, China
czh1619@ahjzu.edu.cn

Abstract. Quality inspection of steel plate welding is critical in industrial manufacturing. However, weld proximity defects often present diverse morphologies, overlapping regions, and dense distributions, posing challenges to accurate industrial defect inspection. Therefore, we propose an industrial detector based on the EP-YOLOv7. First, an Efficient Multiscale Channel Attention (EMCA) is introduced to strengthen multi-scale feature perception and improve the model's focus on weld proximity defects. Second, the EMCA module is integrated into the Efficient Layer Aggregation Network to enhance feature fusion and defect representation. Finally, a Partial-Bottleneck Decoupling Predictor Head (P-BD Head) is designed to significantly improve localization accuracy and reduce missed detections of small targets. Experimental evaluations on both a self-built weld proximity defect dataset and a public generalization dataset show that EP-YOLOv7 achieves mAP of 85.2%/56.2% and F1 scores of 80.3%/43.3%. Meanwhile, the model size increases by only 0.6 MB (total 37.9 MB), demonstrating that the proposed approach delivers substantial accuracy gains while maintaining lightweight computational complexity, suitable for practical industrial applications.

Keywords: Weld proximity defects, YOLOv7, EMCA, P-BD Head, machine learning

1. Introduction

Welding is a fundamental joining technology in modern industrial manufacturing, and its quality directly affects the service life and structural reliability of components. Among commonly used welded materials, steel thin plates are widely applied in industrial construction, aerospace, and intelligent manufacturing due to their lightweight properties,

⋆ Corresponding author

high strength, and recyclability [1]. However, their small thickness and narrow heat-affected zones make them highly vulnerable to thermal deformation during welding, often leading to weld proximity defects such as weld tumors, burning-through, shrinkage cavity, and black-gray oxidation[2], as shown in Figure 1. With the advancement of manufacturing technologies toward higher precision and automation, the requirements for weld quality inspection have become increasingly stringent. Therefore, achieving efficient and accurate detection and localization of weld proximity defects in steel thin plate welds is essential to ensuring welding safety and enhancing overall production efficiency.

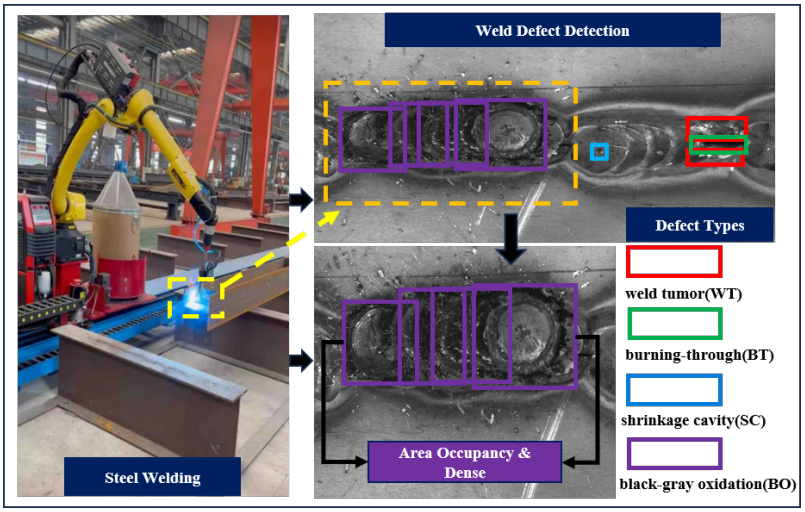

Fig. 1. Weld proximity defects

In recent years, several advanced defect detection techniques have emerged, mainly categorized into traditional classification-based methods [3] and deep learning-based object detection methods [4]. Traditional methods often combine handcrafted features with shallow classifiers, such as the Support Vector Machine (SVM) [5] and Histogram of Oriented Gradients (HOG) [6]. However, these approaches heavily rely on manually designed prior knowledge and shallow classifiers, limiting their ability to automatically learn multi-level and contextual semantic representations. Consequently, their performance deteriorates in complex visual scenes with dense and small defects, such as weld proximity defects.

Deep learning-based object detection has achieved remarkable progress in industrial defect detection, primarily categorized into two-stage algorithms(*e.g.*, Faster R-CNN [7,8,9] and Mask R-CNN [10]), and single-stage algorithms (*e.g.*, SSD [11] and YOLO [12,13,14] series). The two-stage detectors achieve target identification and localization by first generating candidate regions and then classifying and regressing the edges of each region. Chen et al [15] propose an improved Faster R-CNN model with a Res2Net backbone and a weighted fusion module to improve the accuracy of welding defect detection effectively. Although the two-stage detectors can enhance the defect recognition accu-

racy, their detection efficiency remains relatively low. It struggles to achieve a reasonable balance between model accuracy and speed, rendering it unsuitable for weld proximity defect detection scenarios in industrial manufacturing.

In contrast, single-stage detectors perform detection in a single network pass without generating candidate regions, enabling superior efficiency. Researchers have continuously optimized YOLO-based models to establish an effective trade-off between accuracy and speed in industrial defect detection.

For example, Wang et al [16] incorporate the ConvNeXt module into the backbone and introduce an attention mechanism in the pooling stage, effectively enhancing small-target feature extraction. Lu et al [17] design a mixed spatial pyramid pooling module (MIX-PCSPC) to preserve fine details that are often lost in traditional spatial pyramid pooling. Zhang et al [18] further improve small-target localization accuracy by integrating the attention mechanism and optimizing the ELAN module. Zhang et al [19] use Partial Convolution and Switchable Atrous Convolution to replace the 3×3 convolution in efficient layer aggregation networks, thereby improving adaptability to varying defect scales [20] and reducing the misdetection rate. Li et al [21] introduce a Coordinate Attention–based detection head and feedback connections in feature fusion, improving feature recognition and localization precision. Zhang et al [22] introduce an omnidimensional dynamic convolution that is highly sensitive to small targets, achieving significant improvements in model performance.

Although existing defect detection models have achieved notable progress, most approaches still focus on relatively simple weld defect scenarios and lack robustness in complex industrial environments where defects exhibit diverse morphologies, mutual occlusion, and high-density distributions. Therefore, this paper proposes a weld proximity defect detection model for steel thin plates based on EP-YOLOv7. The main contributions are summarized as follows:

1)We propose an Efficient Multiscale Channel Attention (EMCA) that combines fixed multiscale and dynamic adaptive convolutional branches in parallel for feature extraction. It adaptively captures multiscale information while maintaining computational efficiency, thereby enhancing the model's ability to locate weld proximity defects accurately.

2)To strengthen feature fusion and improve the discriminability of dense and overlapping defects, the EMCA module is integrated into the original efficient layer aggregation network module to form EM-ELAN and EM-ELAN-H modules. It optimizes intra-channel and inter-channel weight allocation, reinforces defect feature representation, and enhances the model's ability to extract and distinguish complex weld features.

3)To address localization inaccuracies caused by the coupled prediction head in baseline, we propose a Partial-Bottleneck Decoupling Predictor Head (P-BD Head) based on partial convolutions (PConv) and a dual-branch bottleneck layer. It effectively alleviates small-target miss detections and improves localization accuracy with minimal computational overhead.

The rest of this paper is organized as follows: Section 2 reviews some related works of the attention mechanism for target feature extraction and the YOLO detector for weld defect detection. Section 3 introduces the structure of the EP-YOLOv7 model. Section 4 gives the experimental results and some theoretical analysis. Finally, the conclusion and future work are given in Section 5.

2. Related Works

2.1. Attention Mechanism For Target Feature Extraction

In recent years, attention mechanisms[23,24,25] have been widely applied in object detection tasks. Their core idea is to dynamically assign weights so that the model focuses on the most relevant features, thereby improving feature extraction and object recognition accuracy. Existing attention mechanisms mainly include channel attention, spatial attention, and self-attention.

Channel attention (*e.g.*, Squeeze and Excitation (SE) Blocks[26]) reweights channel features by calculating dependencies between channels, thereby highlighting information-rich channels. For example, Xu et al [27]introduce SE Blocks into the final stage of the YOLOv5 backbone, assigning weights to image locations based on the channel domain perspective to obtain more feature information. Similarly, Li et al [28]based on the YOLOv8, introduced the SEV2 (Squeeze-and-Excitation Version 2) attention mechanism to effectively filter out general interference factors.

Spatial attention generates spatial attention maps, assigning different weights to each position in the feature map.(*e.g.*, Convolutional Block Attention Module (CBAM)[29]). Zoubir. et al[30] incorporate CBAM into the decoder to enhance feature fusion during upsampling by weighing encoder and Gabor filter features.

Self-attention mechanisms establish global dependencies between features, enabling long-range feature interactions. Recent Transformer-based detectors(*e.g.*, Detection Transformer (DETR)[31]) replace traditional components with global self-attention, achieving outstanding performance on COCO.

In summary, although existing attention mechanisms enhance feature extraction from different perspectives, they still struggle to balance multi-scale modeling, computational efficiency, and robustness in dense weld proximity defect scenarios.

2.2. YOLO Detectors For The Welding Defects

Weld defect detection methods are generally divided into two-stage and single-stage approaches. Single-stage detectors directly perform classification and regression on the full image, providing higher inference speed and better meeting industrial weld detection tasks. For example, Liang et al [32] propose an improved YOLOv4 algorithm for wheel weld detection by optimizing anchor sizes with k-means and employing the Distance-IoU loss function, which achieves significant accuracy improvement in both its homemade dataset and public dataset. Liu et al [33] propose an LF-YOLO model based on a reinforced multiscale feature (RMF) module and an efficient feature extraction (EFE) module, effectively enhancing multiscale feature extraction and addressing dimensional disparities in weld defect detection with extremely low consumption. To improve performance on more complex scenarios, Chen et al [34] developed DDT-YOLOv5 with DT-CSPNet and YOLO Di-head, significantly boosting detection accuracy. Xu et al [35] proposed enhancements to YOLOv7 by designing the Le-HorBlock module, coordinate attention, and the SIOU loss function, achieving a 15.9% increase in mAP and effectively reducing missed detections. A summary of the relevant literature is provided in the Table 1. However, although these improved YOLO-based methods perform well on general weld

defect datasets, they are still limited in detecting weld proximity defects. Therefore, to address these shortcomings and improve the detection of complex weld proximity defects, we propose the EP-YOLOv7 model.

Table 1. Summary of YOLO-based Weld Defect Detection Methods

Reference No.	Year	Model	Improvements	Shortcoming
[32]	2022	Improved YOLOv4	K-means and Distance-IoU loss	Limited multi-scale feature
[33]	2023	LF-YOLO	RMF and EFE modules	Limited dense-defect feature
[34]	2024	DDT-YOLOv5	DT-CSPNet and YOLO Di-head	Heavy computation
[35]	2024	Improved YOLOv7	Le-HorBlock module, CA, and SIOU loss	Insufficient generalization

3. Proposed EP-YOLOv7 Model

YOLOv7 is one of the single-stage target detection algorithms, and its network architecture is mainly composed of four parts: the input layer (Input), the backbone network (Backbone), the neck module (Neck), and the detection head (Head), as shown in Figure 2.

The Input module performs image preprocessing and enhancement [36], the Backbone extracts multi-level semantic features, the Neck fuses multi-scale information, and the Head completes final localization and classification.

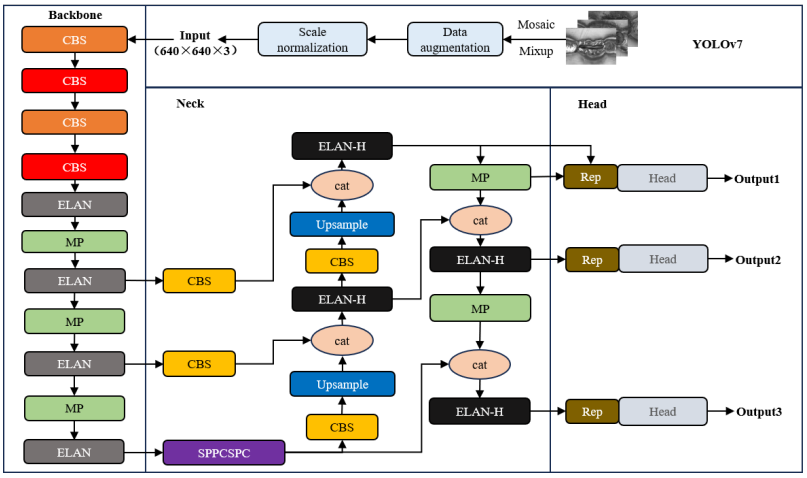

Fig. 2. Structure of YOLOv7

Although YOLOv7 performs well in many detection tasks, its performance remains limited when dealing with weld proximity defects in steel thin plates. Therefore, this paper proposes the EP-YOLOv7 model, as shown in Figure 3.

In Figure 3, we need to explain that CBS is the Convolution + Batch Normalization(BN) + Sigmoid Linear Unit(SiLU), SPPCSPC is the Spatial Pyramid Pooling + Cross-Stage Partial Connection, EM-ELAN & EM-ELAN-H are the improved Efficient Layer Aggregation Network modules, P-BD Head is the Partial-Bottleneck Decoupling Predictor Head, MP is the Maxpooling Layer, Upsample is the Upsampling Module, and Rep is the Reparameterized Convolutional Module.

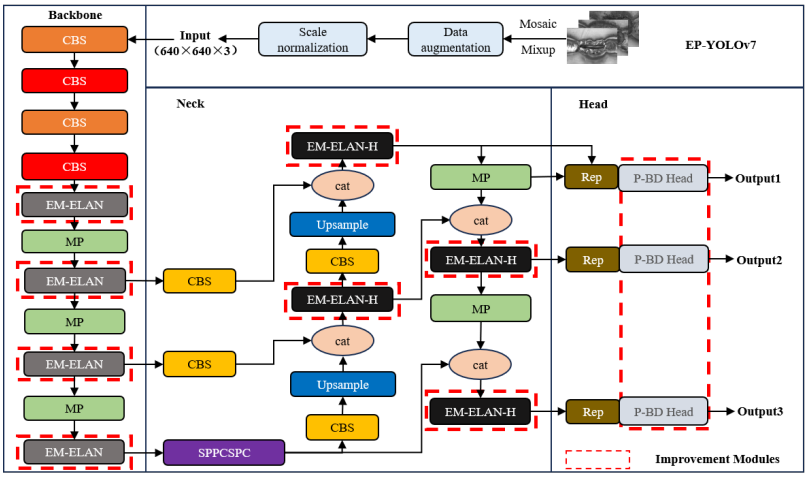

Fig. 3. Structure of EP-YOLOv7

Unlike the conventional YOLOv7, the proposed model incorporates three improvements to enhance the ability to detect weld proximity defects. First, an Efficient Multiscale Channel Attention (EMCA) module is introduced to enhance multiscale feature perception. By combining fixed multiscale and dynamic adaptive convolution branches, EMCA improves the model's sensitivity to subtle weld proximity defect features. Second, EMCA is embedded into the original ELAN and ELAN-H structures to form EM-ELAN and EM-ELAN-H modules. This integration strengthens feature fusion and improves the extraction of dense and overlapping defect regions. Finally, a Partial-Bottleneck Decoupling Predictor Head (P-BD Head) is proposed to reduce task conflict in the prediction stage. With a dual-branch bottleneck and PConv structure, it enhances localization accuracy and alleviates small-target missed detections with minimal computational overhead. Together, these improvements significantly enhance the capability of YOLOv7 in detecting complex weld proximity defects while maintaining lightweight model complexity.

3.1. EMCA Module

In convolutional neural networks, the attention mechanism filters extract key features from large amounts of information by assigning different weights to each part of the input feature map. The Efficient Channel Attention (ECA) module [37] achieves cross-channel interaction through adaptively tuned one-dimensional convolution, significantly improving detection accuracy. However, due to the use of a single convolutional kernel and the

neglect of multi-scale variations among channel features, ECA often struggles to capture key defect features.

To address this limitation, we propose an Efficient Multiscale Channel Attention mechanism (EMCA), which employs two parallel one-dimensional convolutional branches: a fixed multiscale branch and a dynamic adaptive branch, as shown in Figure 4.

The fixed branch contains 1×3, 1×5, and 1×7 convolution kernels to capture local, medium, and long-range channel dependencies, thereby covering channel correlations at different receptive-field ranges. The dynamic branch adaptively adjusts the convolutional kernels according to the number of channels of the input features, enabling flexible allocation of attention weights. Through the synergistic interaction of these two branches, EMCA enhances the representation of global contextual information. It allows the network to more accurately highlight channels associated with defect boundaries and suppress redundant background responses, leading to improved bounding-box regression precision and reduced false localization in weld proximity defects.

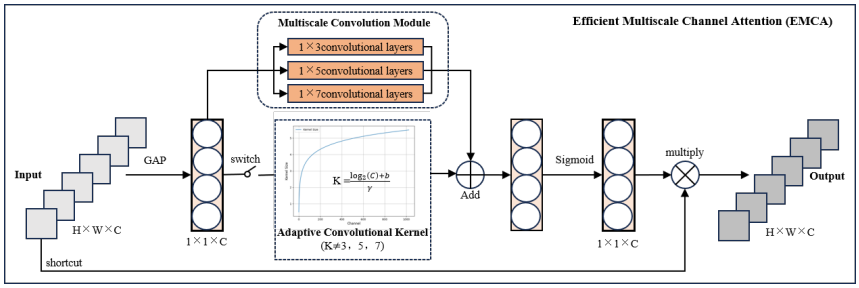

Fig. 4. Structure of the EMCA module

First, for the feature map with input dimension $H \times W \times C$, global average pooling is performed to aggregate spatial information within each channel and obtain a global channel of size $1 \times 1 \times C$, as shown in equation 1.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j} \,. \tag{1}$$

where z_c denotes the global average pooling value of the channel C, and $X_{c,i,j}$ denotes the eigenvalue of the channel C at position (i, j).

Then, two parallel branches are constructed to perform one-dimensional convolution for the new feature vector, including an adaptive convolution kernel branch and a fixed multi-scale convolution branch. A switch mechanism is introduced to dynamically control branch activation. If both branches adopt the same kernel size, the adaptive branch is deactivated to reduce redundant computation. Otherwise, their feature extraction vectors are summed. The kernel size of the adaptive convolution ensures flexible modeling of cross-channel dependencies for different feature scales, as computed as shown in equation 2.

$$K = \left| \frac{\log_2(C) + b}{\gamma} \right|_{odd} . \tag{2}$$

where b and γ denote hyperparameters, and odd denotes the convolution kernel size taken as an odd number.

Next, the output vector obtained from the 1D convolution layers is activated by a non-linear function to generate the attention weights of each channel, allowing the network to highlight informative channels while suppressing less relevant ones, as shown in equation 3.

$$s_c = \sigma(Conv1D(z)) \,. \tag{3}$$

where σ denotes the Sigmoid activation function, $Conv1D(z)$denotes the one-dimensional convolution operation, and s_c denotes the final channel weight.

Finally, the learned attention weights are multiplied with the original input feature map through element-wise channel multiplication, realizing the selective enhancement of critical weld defect features, as shown in equation 4

$$\hat{X}_c = s_c \cdot X_c \,. \tag{4}$$

where X_c denotes the channel characteristic of the input feature, and $\hat{X}_c$ denotes the channel characteristics of the weighted output feature.

3.2. EM-ELAN And EM-ELAN-H Modules

ELAN and ELAN-H serve as efficient layer aggregation networks in the YOLOv7, employing multi-branch stacking and feature fusion to extract multi-level features. However, after fusion, they still lack fine-grained intra-channel weight allocation, which leads to insufficient feature refinement and limits the representation of weld proximity defects. To address these issues, we propose the EM-ELAN and EM-ELAN-H modules, as shown in Figure 5.

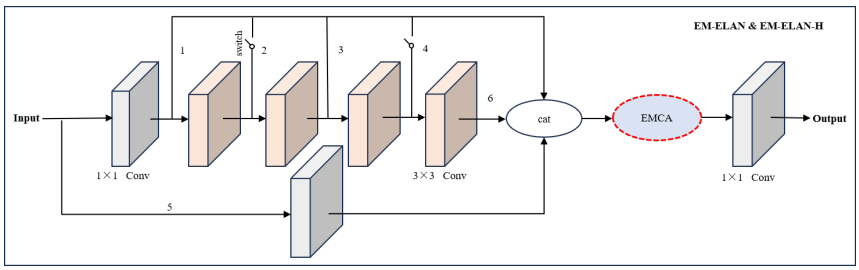

Fig. 5. Structure of EM-ELAN and EM-ELAN-H modules

In Figure 5, we need to explain that EM-ELAN and EM-ELAN-H follow the original multi-branch fusion structure but incorporate the proposed EMCA module after feature aggregation. The key difference lies in their fusion scope: EM-ELAN fuses partial branches(*i.e.,* Branches 1, 3, 5, and 6) for lightweight shallow-layer extraction in the backbone, while EM-ELAN-H fuses all branches (*i.e.,* Branches 1-6)in the neck to integrate richer semantic and multi-scale features.

Specifically, the input features are first processed through the ELAN/ELAN-H backbone to obtain deep-level weld proximity defect representations via multi-branch stacking and fusion. Then, the EMCA module adaptively assigns channel weights to the fused feature maps, enhancing both intra- and inter-channel interactions. Finally, a 1×1 convolution generates the output feature map. This enhanced design improves feature fusion quality and significantly strengthens the model's capability for representing and extracting complex weld proximity defects.

3.3. P-BD Head Module

In the traditional YOLOV7 architecture, the coupled detection head predicts classification, regression, and confidence simultaneously through a shared 1×1 convolution. This structure often causes conflicts between classification and localization objectives, leading to suboptimal performance in weld proximity defect detection. To address this limitation, we propose a Partial-Bottleneck Decoupling Predictor Head (P-BD Head), as shown in Figure 6.

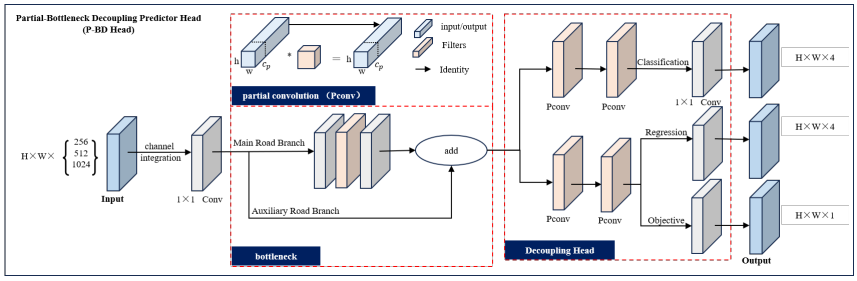

Fig. 6. P-BD Head structure

First, the decoupling prediction structure is employed to replace the original coupling detection head. By compressing channels and assigning independent convolutional branches for classification, regression, and confidence estimation, this architectural separation alleviates task interference and improves the training convergence speed[38].

Second, a dual-branch bottleneck layer is introduced after the input stage to enhance feature quality. The main branch performs channel reduction, local feature extraction, and channel restoration. Using a 1×1 convolution to reduce channels lowers computational cost while suppressing redundant information. The following convolutional layers focus on key local regions and capture deeper correlations, after which the channels are restored to maintain sufficient representation ability. Meanwhile, the auxiliary branch directly retains the original features, reducing gradient attenuation and preventing information loss during transmission.

Finally, the conventional 3×3 convolution is replaced with the PConv module in the overall structure. PConv extracts features from only part of the input channels while passing the remaining channels unchanged, and the outputs are fused through channel concatenation. This design strengthens local feature extraction for small weld defects while retaining global contextual information.

In summary, the P-BD Head introduces a decoupled architecture with a dual-branch bottleneck and Pconv. Together, these enhancements effectively reduce task conflicts, improve feature quality, and reinforce local detail representation, thereby boosting both localization accuracy and detection performance for weld proximity defects.

4. Experiments

4.1. Experimental Environment

The experimental environment is based on the Windows 10 operating system, the CPU is Intel Core i7-13700KF, the GPU is NVIDIA GeForce RTX 4090 (24 GB video memory), the Python version is 3.7, the CUDA version is 11.0, and the input image is 640×640 pixels.

The total number of training epochs is set to 100, divided into two phases: a frozen phase and an unfrozen phase, each lasting for 50 epochs. During the frozen training phase, the backbone network is frozen, and a batch size of 16 is used. In the unfrozen phase, all layers are trainable, and the batch size is reduced to 8. The Stochastic Gradient Descent (SGD) optimizer is employed with a momentum of 0.937 and a weight decay of 0.0005. The initial learning rate is set to 0.01, and a cosine annealing learning rate schedule is applied to gradually reduce the learning rate.

4.2. Stell Thin Plate Weld Defect Image Acquisition And Dataset Construction

In the welding process of steel thin plate, due to the limited number of samples of weld proximity defects, the main dataset in this paper is constructed as a real-life shooting homemade dataset, and the overall process is shown in Figure 7. The image acquisition location is the key laboratory of Anhui Province for intelligent manufacturing of construction machinery, and this experimental acquisition platform consists of a mechanical arm bracket, an industrial focal length, an industrial camera, a computer, and a steel thin plate. The welding samples mainly come from the real steel thin plate in the daily industrial welding process. The industrial focal length adopts a computar-fixed focus 45 million F2524-MPT industrial lens, and the industrial camera adopts Daheng Image ME2P-900-13GC/GM-P Mercury II Pro 9 megapixel color industrial camera. In this paper, image acquisition is carried out by mounting the industrial camera on the robotic arm, and the sampling platform is shown in Figure 7(a).

The collected defect data contains four kinds of weld proximity defects: Weld Tumor (WT), Burn-Through (BT), Shrinkage Cavity (SC), and Black-gray Oxidation (BO). The specific diagram of each defect classification is shown in Figure 7(b).

We select the 2,100 sample images in JPG format as the original dataset. Further, to enhance the generalization ability and robustness in detecting various types of weld defects, this dataset is augmented through geometric transformations, resulting in a total of 5,400 images. The final dataset is labeled using LabelImg software, and text data for training can be generated by drawing rectangular boxes in the defective regions in the images. Each image corresponds to a text file, and the labeling results are saved in XML format. Figure 7(c)shows the data labeling interface and the target box information generated after labeling.

In this experiment, 5400 images are divided into 4374 training sets, 486 validation sets, and 540 test sets. Among them, the number of weld proximity defects is 29124, including 14361 WT defects, 2400 BT defects, 9148 SC defects, and 3215 BO defects. The statistics of the number of defects in each category are shown in Figure 7(d).

In addition, the open-source weld defect detection dataset Welding-Defect-Final0 on the Robflow website is selected for generalization experiments, including five types of weld defects: Geometric defect, Non-fusion defect, crack, porosity, and spatters. Considering the balance of defect samples and data quality, a total of 1,000 images covering all five types of weld defects are selected, such as 200 samples of each category.

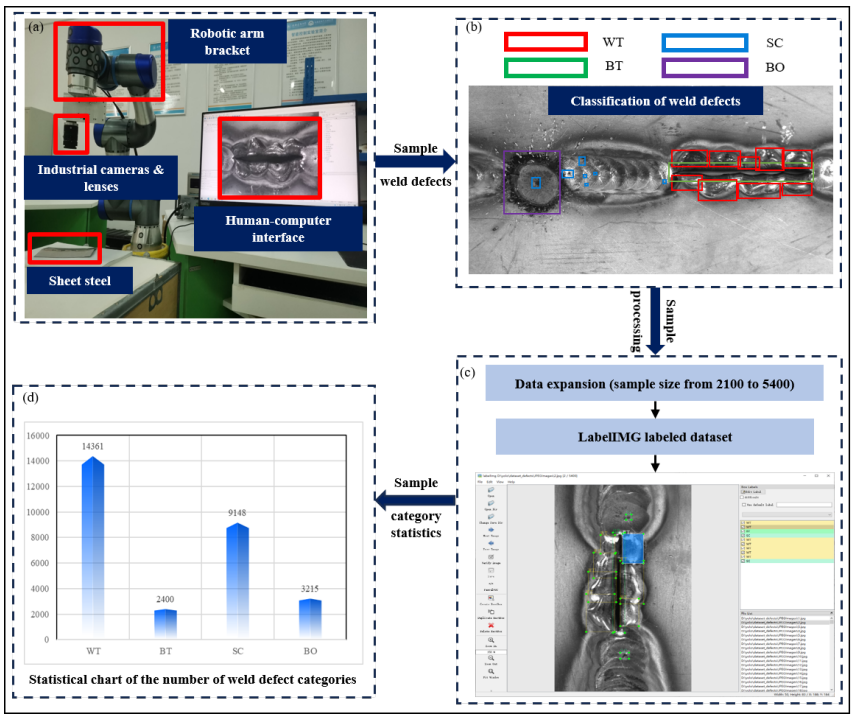

Fig. 7. The overall process of dataset construction. (a) Data collection platform for weld defects in steel sheet (b) Classification of weld defects in steel sheet (c) Expansion of weld defect samples and labeling of dataset (d) Statistical graph of the number of weld defects in each category

4.3. Evaluation Metrics

In this study, the following metrics are introduced: Precision (P), Recall (R), F1 Score (F1), mean Average Precision (mAP), model Parameters (Params), Average (Ave), and Inference Times(I-Time) for the evaluation of the model.

P characterizes the proportion of correctly identified positive samples among all predicted positives, as shown in equation 5. R characterizes the proportion of correctly detected defect samples among all actual defects, as shown in equation 6. F1 characterizes the harmonic mean of precision and recall, providing a balanced measure that considers both detection completeness and accuracy, as shown in equation 7.

$$P = \frac{TP}{TP + FP}\,. \tag{5}$$

$$R = \frac{TP}{TP + FN}\,. \tag{6}$$

$$F1 = \frac{2 \times P \times R}{P + R}\,. \tag{7}$$

For the weld dataset, four categories of proximity defects are considered, namely WT, BT, SC, and BO. For each category, the area under the Precision–Recall curve is calculated to obtain the AP, as shown in equation 8. The mAP is then obtained by averaging the AP values across all defect categories, serving as a comprehensive indicator of overall detection accuracy, as shown in equation 9.

$$AP = \int_0^1 P(R)dR\,. \tag{8}$$

$$mAP = \frac{\sum_1^4 AP}{4}\,. \tag{9}$$

To further evaluate the holistic performance of the proposed model, the Ave is introduced, which aggregates the key evaluation indicators to intuitively reflect the model's overall detection capability, as shown in equation 10.

$$Ave = \frac{P + R + F1 + mAP}{4}\,. \tag{10}$$

TP, FP, and FN denote the number of true positives, false positives, and false negatives, representing cases where a positive sample is correctly identified, incorrectly identified, or missed by the detector.

4.4. Ablation Experiments For EP-YOLOV7

This section aims to analyze the effect of each improved module on weld proximity defect detection. To this end, ablation experiments are performed using the original YOLOv7 as the baseline, and the improved modules are gradually added. The specific experimental results are shown in Table 2.

In Table 2, we need to explain that A is the improved efficient layer aggregation network module and B is the P-BD Head module.

Compared with the original YOLOv7, replacing the original module with the improved efficiency layer aggregation network leads to increases of 2.3%, 2.1%, 1.7%, and 1.4% in mAP, P, F1, and R. This improvement is achieved with nearly the same number

Table 2. Performance comparison of ablation experiments

Model	Epoch (/)	mAP (%)	P (%)	F1 (%)	R (%)	Params (M)	I-Time(s)
YOLOv7	100	81.7	89.1	77.5	68.6	37.3	0.0119
YOLOv7+A	100	84.0	91.2	79.2	70.0	37.3	0.0135
YOLOv7+B	100	84.8	89.5	80.0	72.1	37.9	0.0125
YOLOv7+A+B	100	**85.2**	90.7	**80.3**	**72.1**	37.9	0.0131

of parameters and only a slight increase in inference time, demonstrating that the EMCA module enhances multi-scale channel interaction and aggregation, thereby improving the model's sensitivity to weld surface defects and strengthening its feature extraction capability.

When replacing the original detection head with the P-BD Head, the performance is further improved, with mAP, P, F1, and R increased by 3.1%, 0.4%, 2.5%, and 3.5%. Although the P-BD Head introduces additional computation, its optimized decoupled design contributes to more accurate localization and detection of weld defects.

After integrating both modules, the overall improvement becomes more significant. Compared with the baseline, the proposed EP-YOLOv7 achieves improvements of mAP, P, F1, and R by 3.5%, 1.6%, 2.8%, and 3.5%. The parameter increases by only 0.6M, and inference time (0.0131s) shows only a minor rise, indicating that the proposed improvements achieve a better balance between accuracy and efficiency.

4.5. Efficient Layer Aggregation Network Module Comparison Experiment

This section aims to investigate the impact of different attention mechanisms on the performance of the ELAN and ELAN-H modules. To this end, we introduce improved ELAN and ELAN-H modules that integrate EMCA, ECA, SimAM, CBAM, and SE for comparative experiments. The specific experimental results are shown in Table 3.

Table 3. Comparison of efficient layer aggregation networks with different attention mechanisms

Model	Epoch (/)	mAP (%)	P (%)	F1 (%)	R (%)	Params (M)
YOLOv7	100	81.7	89.1	77.5	68.6	37.3
YOLOv7+ECA	100	82.9	89.1	79.8	72.3	37.3
YOLOv7+CBAM	100	83.3	89.1	79.6	71.9	37.3
YOLOv7+SimAM	100	83.4	89.6	78.3	69.6	39.5
YOLOv7+SE	100	82.2	88.1	78.4	70.7	38.4
YOLOv7+EMCA	100	**84.0**	**91.2**	79.2	70.0	37.3

Table 3 shows that both mAP and P are significantly improved in YOLOv7 when incorporating the EMCA into the module, compared to other attention-based variants, without increasing the overall network complexity. In Figure 8, we visualize and compare the two attention mechanisms with the highest mAP scores, CBAM and the EMCA. The results show that, compared to CBAM, EMCA effectively reduces missed detections and enables more accurate localization of weld proximity defects. The combined results of

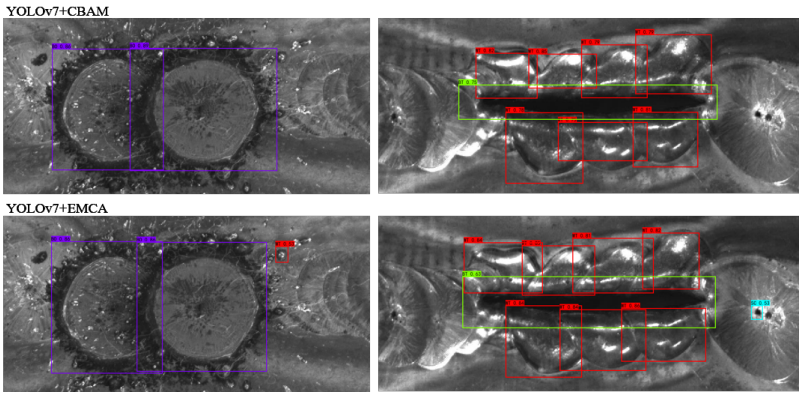

Fig. 8. Comparative visualization of different attention mechanisms

the above experiments show that EMCA strengthens the ability to focus on the information of weld proximity defects by adding fixed convolutional kernels. Furthermore, when combined with the efficient layer aggregation module, it markedly improves the model's feature extraction capability, thereby enabling accurate localization and identification of defects.

4.6. Ablation Experiment For P-BD Head

This section aims to investigate the contribution of each improved component within the P-BD Head to the overall detection performance. To this end, ablation experiments are carried out using the original YOLOv7 as the baseline model, where the decoupling head, bottleneck layer, and Pconv module are sequentially integrated for comparative analysis. The specific experimental results are shown in Table 4.

Table 4. P-BD head module ablation experiments

Model	Epoch (/)	mAP (%)	P (%)	F1 (%)	R (%)	Params (M)
YOLOv7	100	81.7	89.1	77.5	68.6	37.3
YOLOv7+C	100	83.3	89.3	76.0	66.1	44.8
YOLOv7+D	100	84.0	90.1	76.7	66.7	45.4
YOLOv7+B	100	**84.8**	89.5	**80.0**	**72.1**	37.9

In Table 4, we need to explain that C is an improved detection header with only the decoupling header added, D is an improved detection header with the bottleneck layer and decoupling header added, and B is an improved detection header with the bottleneck layer, decoupling header, and Pconv module added.

Compared to the original YOLOv7, the YOLOv7 with the decoupling head improves metrics mAP and P by 1.6% and 0.2%. The experimental results show that the decoupling head enhances detection accuracy by enabling each prediction task to be processed independently. However, due to the subdivision of the prediction task by the decoupling head,

the model's ability to detect low-confidence targets is slightly weakened, resulting in a minor decrease in R. Additionally, the introduction of a 3×3 convolutional layer increases the model size by 7.5 MB.

Furthermore, when both the bottleneck layer and the decoupling head are integrated into YOLOv7, mAP and P show further improvements. The bottleneck layer refines the input features to provide more feature information for subsequent decoupling processing.

Finally, after incorporating the P-BD Head into YOLOv7, mAP, F1, and R are improved by 0.8%, 3.3%, and 5.4%. While the model parameters are reduced by 7.5 MB compared with the version containing both the bottleneck layer and the decoupled head, with only a slight decrease in P. The ablation experiments show that with only a minimal parameter increase, the P-BD Head substantially enhances the localization and detection accuracy for the weld proximity defects, while significantly improving the model's recall capability.

4.7. Performance Comparison Of Different Target Detection Models

This section aims to comprehensively evaluate the overall performance of the proposed EP-YOLOv7 model. To this end, a series of comparative experiments is conducted on object detection algorithms, including two-stage detectors (*e.g.,*Faster R-CNN) and single-stage detectors (*e.g.,*SSD and the YOLO series), as well as recent References[12], [18], and transformer-based models (*e.g.,* RT-DETR). The specific results are shown in Table 5 and Figure 9, where Figure 9(a) shows loss function curves for the training and validation sets of the EP-YOLOv7 model, and Figure 9(b) shows the change curves of each evaluation index.

Table 5. Performance comparison of different target detection models

Model	Epoch (/)	mAP (%)	P (%)	F1 (%)	R (%)	Ave (%)	Params (M)
SSD	100	66.8	84.0	61.0	47.9	64.9	24.1
Faster-RCNN	100	73.1	62.9	69.5	77.7	70.8	28.4
YOLOv5	100	71.5	85.8	61.5	47.9	66.7	46.7
YOLOv7	100	81.7	89.1	77.5	68.6	79.2	37.3
YOLOv7-Tiny	100	73.5	86.9	66.5	53.8	70.2	6.0
[18]	100	74.5	87.5	67.2	54.6	71.0	8.6
YOLOv8n	100	80.2	90.4	77.7	68.2	79.1	3.0
YOLOv8l	100	80.7	88.1	77.5	69.1	79.0	43.6
YOLOv10n	100	77.7	90.7	73.7	62.1	76.1	2.3
YOLOv10l	100	81.3	90.4	78.8	69.9	80.1	24.3
YOLOv12	100	79.6	90.1	76.7	66.8	78.3	2.5
[12]	100	81.1	88.1	77.6	69.3	79.0	5.6
RT-DETR	100	81.9	76.7	75.8	74.9	77.3	28.5
EP-YOLOv7	100	**85.2**	**90.7**	**80.3**	72.1	**82.1**	37.9

Table 5 shows that the proposed EP-YOLOv7 model outperforms several mainstream detection models across multiple evaluation metrics(*e.g.,* mAP:85.2%, P:90.7%, F1:80.3%, Ave:82.1%). Compared to the original YOLOv7 model, the number of parameters in EP-YOLOv7 increases by only 0.6 MB, while the detection accuracy improves by 3.5%.

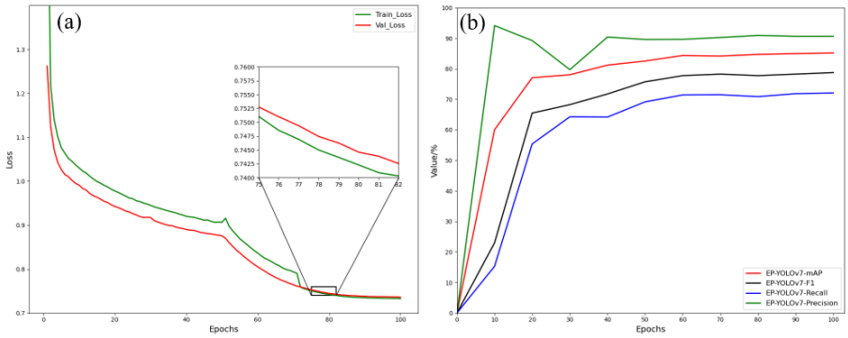

Fig. 9. EP-YOLOv7 effect plots.(a) training set and validation set loss function curves (b) P, R, mAP, and F1 variation curves

Although the R and parameter values are not the highest among all models, the Ave value surpasses those of the compared models, demonstrating its stronger overall performance in accurate weld proximity defect detection.

Meanwhile, according to Figure 9, both the training and validation loss functions exhibit a clear convergence trend, and all evaluation metrics (mAP, precision, recall, and F1) stabilize in the later training stages, indicating that the model achieves reliable convergence. In summary, the experimental results confirm that the proposed EP-YOLOv7 model achieves stable convergence and superior performance in the weld proximity defect detection task.

4.8. Comparison Chart Of Visualization Model Effect Detection

This section aims to provide an intuitive comparison of the detection performance among different models. We select three models with the highest Ave value(YOLOv7, YOLOv10l, and the proposed EP-YOLOv7)to visualize and compare their weld proximity defect detection results.

Due to the weld defects in this dataset of a variety of defects densely occupy the position, so this paper selects several representative defect combinations for comparison. The corresponding detection results are shown in Figure 10, where Figure 10(a)–(d) represent the combinations WT&BT&SC, WT&BO&SC classes, WT&BT&BO class, WT&BT&BO&SC classes. As shown in the figure, the proposed EP-YOLOv7 model outperforms both YOLOv7 and YOLOv10l in detecting weld proximity defects on thin steel plates.

Specifically, YOLOv7 and YOLOv10l exhibit noticeable missed detections across multiple scenarios. For instance, both models fail to WT defects in Figure 10(a). In Figure 10(b), the YOLOv7 model misses the WT and SC defects, while YOLOv10l misses the BO defects. Similarly, in Figure 10(c), YOLOv10l fails to detect BO defects. In Figure 10(d), YOLOv7 and YOLOv10l miss SC defects, and YOLOv10l additionally fails to detect BO defects. These results indicate that existing models struggle with dense multi-defect environments.

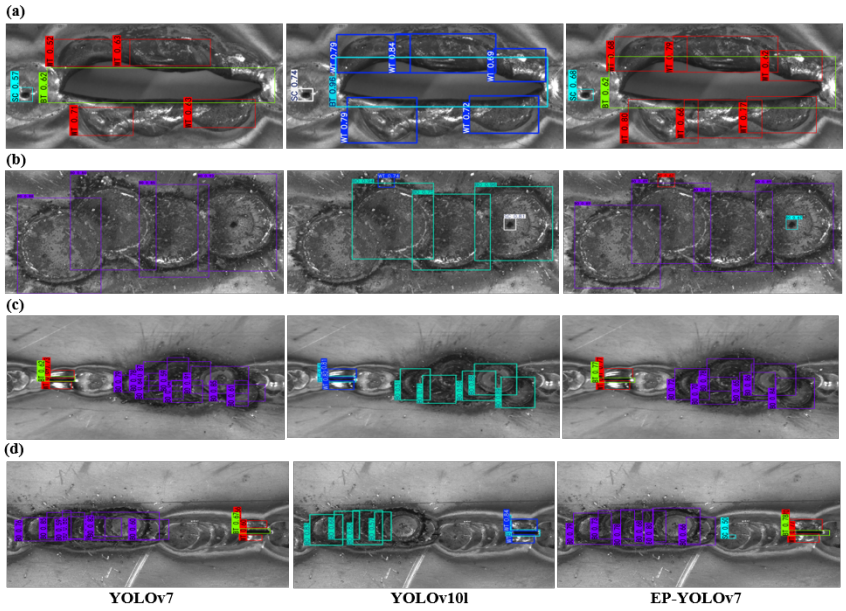

Fig. 10. Comparison of the detection effect of weld proximity defects (a) WT&BT&SC classes (b) WT&BO&SC classes (c) WT&BT&BO classes (d) WT&BT&BO&SC classes

In addition to missed detections, YOLOv7 exhibits inaccurate localization. In Figure 10(a), the predicted bounding boxes for WT and BT defects fail to fully cover the target regions. In Figure 10(b), the bounding box for the BO defect significantly exceeds the actual defect area. In contrast, the proposed EP-YOLOv7 model effectively addresses both missed detection and localization inaccuracies, achieving more accurate and reliable detection of weld proximity defects in steel thin plates.

4.9. Generalizability Experiments

This section aims to evaluate the generalization performance of the proposed EP-YOLOv7 model. To this end, we select the three models with the highest Ave value in the comparative experiments (YOLOv7, YOLOv10l, and the proposed EP-YOLOv7) and conduct experiments on the public dataset Welding-Defect-Final0. As shown in Table 6, EP-YOLOv7 achieves an mAP of 56.2%, representing improvements of 4.6% and 1.8% over the baseline YOLOv7 and YOLOv10l models.

Meanwhile, Figure 11 shows the detection results for the five weld defect categories in the dataset. Overall, the proposed model exhibits fewer missed detections and higher confidence scores, confirming the superior generalization ability of EP-YOLOv7 in weld defect detection.

Table 6. Results of generalization experiment data

Model	Epoch (/)	mAP (%)	P (%)	F1 (%)	R (%)	Params (M)
YOLOv7	100	51.6	88.0	40.0	25.7	37.3
YOLOv10l	100	54.4	89.8	41.4	26.9	24.3
EP-YOLOv7	100	**56.2**	**91.3**	**43.3**	**28.4**	37.9

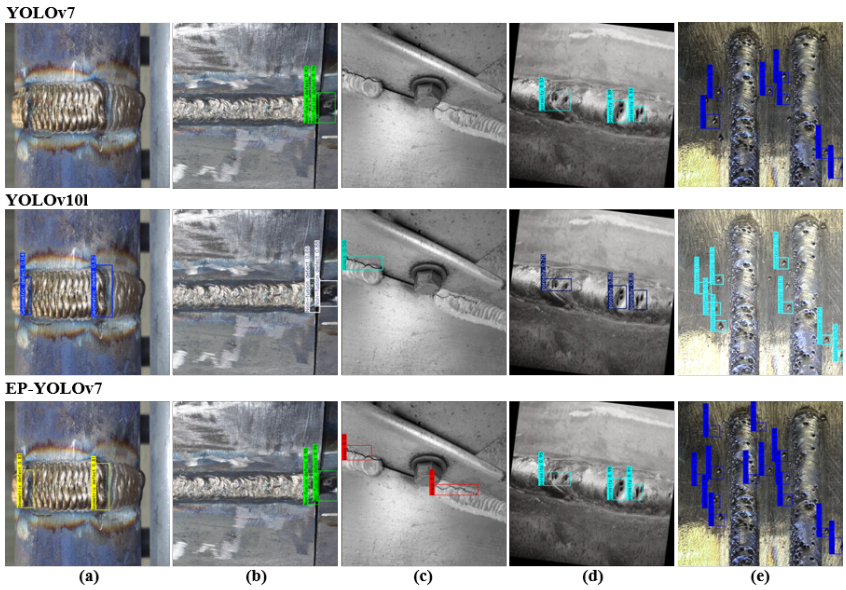

Fig. 11. Generalized experimental visualization results (a) Geometric defect (b) Non-fusion defect (c) crack (d) porosity (e) spatters

5. Conclusion

In this paper, we propose the EP-YOLOv7 model to achieve accurate localization and identification of weld proximity defects in steel thin plates. In the EM-ELAN and EM-ELAN-H modules, we incorporate the proposed EMCA module based on multi-scale convolution into the original module to effectively optimize the allocation of the feature weights in the channels and to improve the sensitivity of the model to the weld proximity defects and the feature extraction capability. In the P-BD Head module, the optimized decoupled head structure is used to effectively improve the detector's positioning and detection accuracy for weld proximity defects. Experimental results show that EP-YOLOv7 achieves superior detection performance compared with mainstream models. With only a 0.6MB increase in parameters and a minor rise in inference time, it achieves a 3.5% improvement in mAP over the baseline YOLOv7, along with notable gains in precision, F1 score, and recall. The model also exhibits good generalizability in publicly available datasets.

In the future, we will further expand the weld defect dataset to include additional materials (*e.g.,*aluminum, stainless steel, and galvanized steel) and more diverse industrial application scenarios (*e.g.,* shipbuilding, pressure vessel manufacturing, and rail transit equipment production) to enhance the model's adaptability and generalization. Moreover, we will continue to optimize EP-YOLOv7 through lightweight structural design, network pruning, and quantization strategies, aiming to improve detection efficiency and facilitate deployment in resource-constrained industrial environments.

Acknowledgments. This work was supported in part by the Provincial Natural Science Research Programs for Universities in Anhui Province (2024AH050234), in part by the open subject of Anhui Jianzhu University research platform (IMCM2023KF05), in part by the Chang'an University Special Funds for Basic Research Operating Expenses of the Central Universities (300102254501-202412), in part by the Scientific Research Program of Higher Education Institutions in Anhui Province (2024AH050245), in part by the Anhui Jianzhu University Research Reserve Project (2022XMK03), and in part by the Corporate Commissioned Project (JR202242).

References

1. Matthew Rumsa, Michele John, and Wahidul Biswas. Global steel decarbonisation roadmaps: Near-zero by 2050. *Environmental Impact Assessment Review*, 112:107807, 2025.
2. Xiaopeng Wang, Uwe Zscherpel, Paolo Tripicchio, Salvatore D'Avella, Baoxin Zhang, Juntao Wu, Zhimin Liang, Shaoxin Zhou, and Xinghua Yu. A comprehensive review of welding defect recognition from x-ray images. *Journal of Manufacturing Processes*, 140:161–180, 2025.
3. Changlu Xu, Linsheng Li, Jiwei Li, and Chuanbo Wen. Surface defects detection and identification of lithium battery pole piece based on multi-feature fusion and pso-svm. *Ieee Access*, 9:85232–85239, 2021.
4. Ruichen Ma, Jinglong Chen, Yong Feng, Zitong Zhou, and Jingsong Xie. Ela-yolo: An efficient method with linear attention for steel surface defect detection during manufacturing. *Advanced Engineering Informatics*, 65:103377, 2025.
5. Dorra Zaibi, Maroua Salhi, Khaoula Tbarki, and Riadh Ksantini. An incremental software defect detection model based on support vector machine. *Engineering Computations*, 42(1):76–95, 2025.

6. Yirga Kene Molla and Emebet Abeje Mitiku. Cnn-hog based hybrid feature mining for classification of coffee bean varieties using image processing. *Multimedia Tools and Applications*, 84(2):749–764, 2025.
7. Qianwen Ye, Yiwei Dong, Xiaoxin Zhang, Duo Zhang, and Siyi Wang. Robustness defect detection: improving the performance of surface defect detection in interference environment. *Optics and Lasers in Engineering*, 175:108035, 2024.
8. Xianghe Zou, Chongyang Wu, Hongen Liu, Zhangwei Yu, and Xianyan Kuang. An accurate object detection of wood defects using an improved faster r-cnn model. *Wood Material Science & Engineering*, 20(2):413–419, 2025.
9. Zhige He and Yuanqing He. As-faster-rcnn: An improved object detection algorithm for airport scene based on faster r-cnn. *IEEE Access*, 2025.
10. Xin Weng, Qihua Ma, Qilin Li, and Wenchao Wang. Improved mask r-cnn algorithm: Multi-ore detection and positioning based multi-sensor fusion in complex field environment. *Measurement*, 246:116602, 2025.
11. Ling Tan, Hui Wu, Zifeng Xu, and Jingming Xia. Multi-object garbage image detection algorithm based on sp-ssd. *Expert Systems with Applications*, 263:125773, 2025.
12. Beilong Chen, Mingjun Wei, Jianuo Liu, Hui Li, Chenxu Dai, Jinyun Liu, and Zhanlin Ji. Efs-yolo: a lightweight network based on steel strip surface defect detection. *Measurement Science and Technology*, 35(11):116003, 2024.
13. Wenqi Cui, Zhenye Li, Anning Duanmu, Sheng Xue, Yiren Guo, Chao Ni, Tingting Zhu, and Yajun Zhang. Ccg-yolov7: A wood defect detection model for small targets using improved yolov7. *IEEE Access*, 12:10575–10585, 2024.
14. Keji Mao, Runhui Jin, Lingkang Ying, Xingda Yao, Guanglin Dai, and Kai Fang. Sc-yolo: Provide application-level recognition and perception capabilities for smart city industrial cyber-physical system. *IEEE Systems Journal*, 17(4):5118–5129, 2023.
15. Yongbin Chen, Jingran Wang, and Guitang Wang. Intelligent welding defect detection model on improved r-cnn. *IETE Journal of Research*, 69(12):9235–9244, 2023.
16. Rijun Wang, Fulong Liang, Xiangwei Mou, Lintao Chen, Xinye Yu, Zhujing Peng, and Hongyang Chen. Development of an improved yolov7-based model for detecting defects on strip steel surfaces. *Coatings*, 13(3):536, 2023.
17. Quan Lu, Kehong Lin, and Linfei Yin. 3d attention-focused pure convolutional target detection algorithm for insulator defect detection. *Expert Systems with Applications*, 249:123720, 2024.
18. Qingqi Zhang, Xinbo Wang, Hao Wang, and Jie Li. Vehicle small target detection algorithm for remote sensing images based on improved yolov7-tiny. In *2024 5th International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pages 183–187. IEEE, 2024.
19. Jinlai Zhang, Wenjie Yang, Yumei Chen, Mingkang Ding, Huiling Huang, Bingkun Wang, Kai Gao, Shuhan Chen, and Ronghua Du. Fast object detection of anomaly photovoltaic (pv) cells using deep neural networks. *Applied Energy*, 372:123759, 2024.
20. Jierun Chen, Shiu-hong Kao, Hao He, Weipeng Zhuo, Song Wen, Chul-Ho Lee, and S-H Gary Chan. Run, don't walk: chasing higher flops for faster neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12021–12031, 2023.
21. Zhijin Li, Jinfeng Yan, Jie Zhou, Xiaozhen Fan, and Jiahui Tang. An efficient smd-pcba detection based on yolov7 network model. *Engineering Applications of Artificial Intelligence*, 124:106492, 2023.
22. Yi Zhang and Qingjian Ni. A novel weld-seam defect detection algorithm based on the s-yolo model. *Axioms*, 12(7):697, 2023.
23. Shunxiang Zhang, Meng Chen, Kuan-Ching Li, Hua Wen, and Liang Sun. Fire detection models based on attention mechanisms and multiscale features. *Computer Science and Information Systems*, (00):59–59, 2025.
24. Zihua Chen, Runmei Zhang, Meng-Yen Hsieh, Alireza Souri, and Kuan-Ching Li. Average sigmoid-tanh attention and multi-filter partially decoupled mechanism via yolov7 for detecting weld proximity defects. *Metallurgical and Materials Transactions B*, pages 1–15, 2025.

25. Yushan Zhao, Kuan-Ching Li, Shunxiang Zhang, and Tongzhou Ye. Hyperbolic graph attention network fusing long-context for technical keyphrase extraction. *Information Fusion*, 120:103061, 2025.
26. Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
27. Junpeng Xu, Xiangbo Zhu, Lei Shi, Jin Li, and Ziman Guo. Squeeze-and-excitation attention and bi-directional feature pyramid network for filter screens surface detection. *Journal of Electronic Imaging*, 33(4):043044–043044, 2024.
28. Zhaochao Li, Linxuan Xiao, Meiling Shen, and Xiya Tang. A lightweight yolov8-based model with squeeze-and-excitation version 2 for crack detection of pipelines. *Applied Soft Computing*, page 113260, 2025.
29. Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
30. Hajar Zoubir, Mustapha Rguig, Mohamed El Aroussi, Rachid Saadane, and Abdellah Chehri. Pixel-level concrete bridge crack detection using convolutional neural networks, gabor filters, and attention mechanisms. *Engineering Structures*, 314:118343, 2024.
31. Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.
32. Liang Tian Jiao, Pan Wei Guo, Bao Hong, and Pan Feng. Vehicle wheel weld detection based on improved yolo v4 algorithm. , 46(2):271–279, 2022.
33. Moyun Liu, Youping Chen, Jingming Xie, Lei He, and Yang Zhang. Lf-yolo: A lighter and faster yolo for weld defect detection of x-ray image. *IEEE sensors journal*, 23(7):7430–7439, 2023.
34. Zihua Chen, Yu Zheng, Ling-Huey Li, Tien-Hsiung Weng, Kuan-Ching Li, and Aneta Poniszewska-Maranda. An improved dilated-transposed convolution detector of weld proximity defects. *IEEE Access*, 2024.
35. Xiangqian Xu and Xing Li. Research on surface defect detection algorithm of pipeline weld based on yolov7. *Scientific reports*, 14(1):1881, 2024.
36. Yong Ren, Dong Liu, and Sanhong Gu. A lightweight defect classification method for latex gloves based on image enhancement. *Computer Science and Information Systems*, (00):7–7, 2025.
37. Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.
38. Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.

Runmei Zhang is a professor in the School of Electronic and Information Engineering at Anhui Jianzhu University. She received her PhD degree from the Hefei University of Technology in 2015. Her current research interests are robotics, artificial intelligence, and big data.

Jingwei Fan is a graduate student in the School of Mechanical and Electrical Engineering at Anhui Jianzhu University. His research interests include deep learning and image processing.

Zihua Chen (corresponding author) is a lecturer in the School of Electronic and Information Engineering at Anhui Jianzhu University. He received his PhD degree in the College

of Mechanical Engineering, Anhui University of Science and Technology (AUST), in 2022. His current research interests are the Internet of Things, multi-agent cooperative control, and computer vision.

Zhong Chen is a senior engineer in the School of Mechanical and Electrical Engineering at Anhui Jianzhu University. He received his PhD degree from the Hefei University of Technology in 2015. His current research interests are artificial intelligence, electrical control.

Bin Yuan is a lecturer in the School of Mechanical and Electrical Engineering at Anhui Jianzhu University. He received his PhD degree from the Hefei University of Technology in 2018. His current research interests are intelligent manufacturing and construction technology.

Received: September 22, 2025; Accepted: December 14, 2025.