High-Precision Welding Defect Detection Practice: An Innovative YOLOv12 Model with Pinwheel Convolution and Adaptive Attention

Md Helal Miah[1], *Shashi Kant Gupta*[1]

[1] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

Abstract: This research introduces an innovative approach to enhance the accuracy and speed of weld defect detection by integrating attention mechanisms and Pinwheel-Shaped Convolution (PConv) into the YOLOv12 framework. Addressing the limitations of traditional CNNs and transformer-based models, which often struggle with either missing subtle defects or incurring high computational costs, this study achieves precise, real-time detection, outperforming existing solutions in complex industrial environments. A systematic methodology was adopted to train and validate the proposed model on four benchmark datasets, employing diverse data augmentation techniques such as Mosaic, Mixup, and Copy-Paste to improve generalization. Quantitative performance evaluation was conducted using established metrics, including recall, F1-score, and mean Average Precision (mAP), enabling rigorous comparison with baseline YOLO models and attention-based architectures under realistic industrial inspection conditions. The innovative PConv based YOLOv12 model demonstrated outstanding performance in weld defect detection. It achieved an F1-score of 0.941 and mAP@0.5 of 0.989 in single-class detection, and an F1-score of 0.848 with 0.887 recall in multi-class detection. Mosaic augmentation boosted mAP@0.5:0.95 to 0.854, enhancing generalization. The model converged rapidly, reaching mAP@0.5 of 0.905 in just 10 epochs and stabilizing near 0.996, proving its robustness and suitability for industrial real-time applications.

Keywords: Weld Defect Detection; YOLOv12; Lightweight Attention Module; Image Processing; Deep Learning Method.

Introduction

While traditional convolutional neural networks are quite effective at detecting weld defects in images, they sometimes struggle with capturing complex patterns and long-range dependencies in the data. To address this, some approaches use techniques to preprocess the images to improve detection accuracy. For example, a Gaussian kernel can be used for blurring to help with image extraction [1,2]. In contrast, transformer-based models leverage an attention mechanism that allows them to grasp the broader context within an image, enabling them to identify subtle or scattered defects that CNNs may overlook. For instance, transformers are able to model global relations in images, unlike CNNs. By using an attention mechanism, transformers can focus on the most important information about defects. One approach is to inject CNN features into different stages of the transformer network to capture detailed features and reduce background noise, facilitating accurate defect detection [3]. However, a significant limitation of attention mechanisms is their computational cost. The demands on computing power and memory, particularly when processing high-resolution images, can create an efficiency bottleneck [4,5]. The aforementioned issue results in a bottleneck concerning efficiency, thereby reducing the model's speed and practicality for real-time weld defect detection within industrial applications.

While attention mechanisms, like those in transformers, are excellent at discerning complex patterns useful for identifying subtle weld defects, they demand significant computational power [6,7]. This can make it difficult to use attention-based models for real-time weld defect detection, where speed is crucial. YOLO models, on the other hand, are known for their speed and efficiency in real-time object detection [8–10]. However, there is room to improve YOLO models for detecting very small or hard-to-see weld defects. Combining the strengths of YOLO frameworks with attention mechanisms could lead to better performance without sacrificing speed [11–16]. Manual weld detection can be time-consuming, so automated methods like YOLO are gaining traction.

Introduced in April 2024 by Joseph Redmon's team, YOLOv12 marks a significant progression from predecessors like YOLOv5, YOLOv7, and YOLOv8. Key enhancements include an improved backbone and neck design for better feature extraction, especially for small to medium objects. It uses multi-scale feature fusion techniques like Dynamic Head and BiFPN++, along with dynamic attention mechanisms and global context modelling for precise focus on critical image regions, unlike simpler attention structures in earlier models. Enhanced training strategies incorporate data augmentation methods like Mosaic++ and CopyPaste, plus optimization techniques such as Exponential Moving Average updates and refined learning rate scheduling. A major innovation is its hybrid detection approach, combining anchor-based and anchor-free mechanisms to handle various object sizes and shapes. It also introduces improved loss functions, including a modified VariFocal loss for objectness prediction and the SIoU loss for more accurate localization. Despite these increased capabilities, YOLOv12 maintains real-time speed with a higher mean Average Precision (mAP) than YOLOv8, without significant loss in frames per second. Its modular framework allows easier adaptation to specialized domains like medical imaging and industrial inspection, and it incorporates lightweight transformer modules in deeper layers for enhanced global reasoning while keeping computational demands low [17]. Overall, YOLOv12 offers greater accuracy and faster performance relative to its complexity, with improved feature focus, while maintaining the efficiency needed for real-time applications.

While YOLOv12 models are popular for object detection due to their speed and accuracy balance, which is important for real-time industrial uses, current research hasn't focused on using YOLOv12 specifically for weld detection. A potential limitation of these models is their underutilization of attention mechanisms, which help focus on key image regions [18,19]. Consequently, YOLOv12 models might miss subtle or complex weld defects that attention-based methods could detect more effectively [20,21].

The research aims to improve the accuracy of weld defect detection by incorporating attention mechanisms into the YOLOv12 framework. Traditional YOLOv12 models struggle with detecting subtle defects in complex welding environments. This research introduces a novel approach, using Pinwheel-Shaped Convolution and attention mechanisms, to enable the model to focus on critical areas and fine-grained anomalies, enhancing detection accuracy while maintaining real-time processing speeds suitable for industrial applications.

Method and Models

The YOLOv12 architecture is meticulously designed to optimize visual data processing efficiency while maintaining high accuracy in defect detection. Its key innovations significantly enhance weld flaw detection performance by refining the capture and processing of critical image details, enabling precise identification of even minor defects. Consequently, YOLOv12 excels in delivering rapid and accurate detection, making it a valuable solution for real-time industrial inspection applications [22]. The main innovations of the YOLOv12 architecture are as follows:

- Area Attention (A2) Module: The Area Attention (A2) module enhances weld defect detection by segmenting feature processing through spatial reshaping, enabling a more focused analysis of critical regions. By integrating Flash Attention, the module achieves a significant reduction in computational demands (reducing complexity by approximately 50%) while maintaining a broad contextual understanding via a large receptive field. This efficient design facilitates real-time operation at a fixed resolution of 640, accomplished through optimized memory access strategies that improve processing speed without sacrificing detection accuracy. Many researchers are focusing on automated X-ray welding image defect detection methods [23].

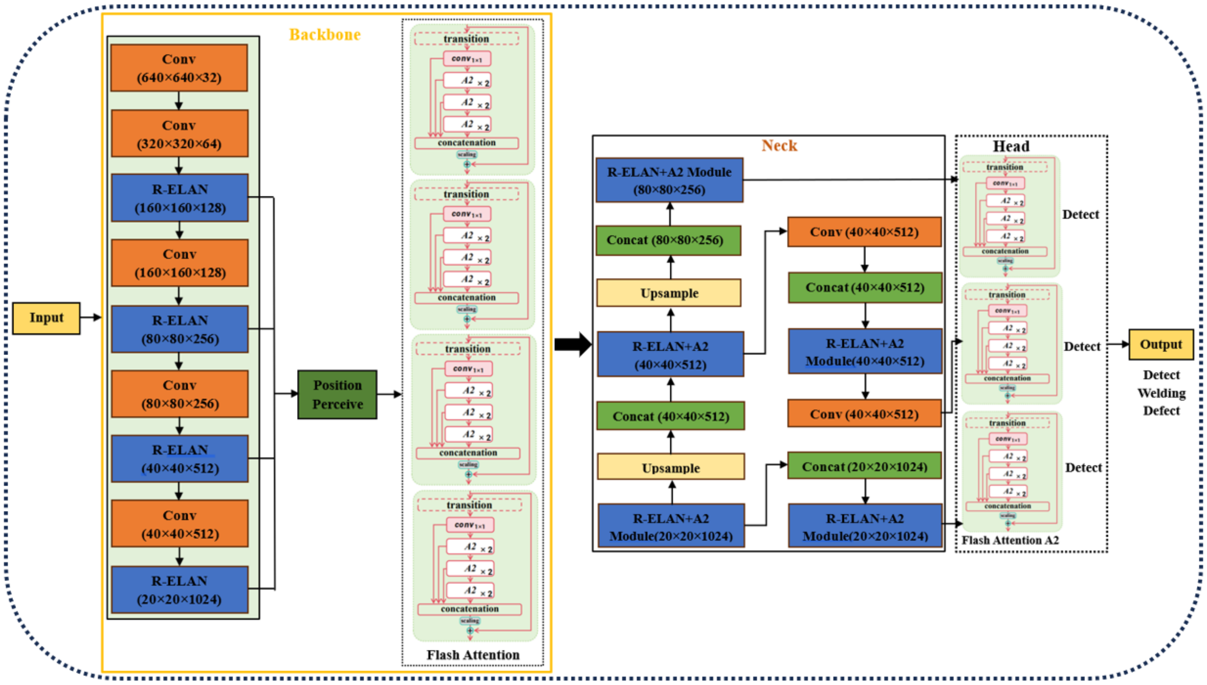


Figure 1. Architectural mechanism of YOLO v12 for weld defect detection.

- Residual ELAN (R-ELAN) Hierarchy: Within the weld defect detection system, the Area Attention (A2) module incorporates the R-ELAN structure, utilizing residual shortcuts with a scaling factor of 0.01 and a dual-branch design to effectively address the vanishing gradient problem during training. This architectural improvement enhances feature propagation and bolsters stability in deeper networks. Furthermore, the model employs an efficient final aggregation phase, achieving an 18% reduction in parameters and a 24% decrease in floating-point operations compared to the standard baseline, thereby enhancing overall computational efficiency without compromising performance. Some approaches use attention mechanisms to improve feature fusion [24].

- Efficient architecture modification: In this weld defect detection system, YOLOv12 enhances spatial awareness through the replacement of traditional position encoding with a 7x7 depth-wise convolution, facilitating implicit spatial feature extraction. To optimize computational efficiency, it integrates an adaptive MLP scaling factor of 1.2x alongside a shallow stacking of blocks, effectively managing processing demands. Collectively, these design choices enable a rapid inference time of only 4.1 milliseconds when deployed on V100 hardware, rendering the model well-suited for real-time inspection tasks. Indeed, deep learning methods based on the YOLO network are being optimized for weld defect detection [25]. Some approaches focus on improving detection with lightweight and faster YOLO implementations [26].
- Optimized training framework: Trained over 600 epochs using stochastic gradient descent with a cosine learning rate schedule starting at 0.01, the weld defect detection model incorporated advanced data augmentation techniques such as Mosaic-9 and Mixup to enhance generalization and detection accuracy. These techniques contributed to a 12.8% improvement in mean Average Precision (mAP) on the COCO benchmark. Despite these enhancements, the model maintained its real-time processing capability, owing to the efficient use of selective kernel convolution, which balances accuracy with speed. Data augmentation is commonly used to improve welding defect detection [27].

The YOLOv12 architecture centers around the A2C2f module, which is crucial for enhancing the model's overall performance. This key component is specifically designed to improve feature representation and boost detection accuracy, making it highly effective for identifying weld defects in real-time applications. The overall structure of the model is composed of several key components, as illustrated in Figure 2. Optimizations to YOLO are being explored to account for the particular nature of weld defects, and research is being done to create lighter and faster implementations.

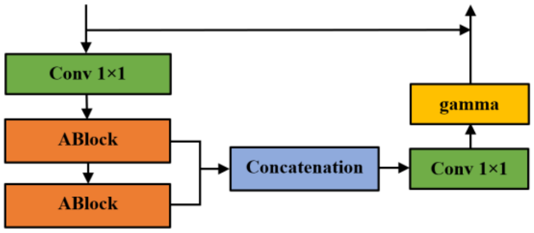


Figure 2. A2C2f module.

The A2C2f module is structurally composed of key elements. It includes two 1×1 convolutional layers, named cv1 and cv2, that reduce the dimensionality of the input features while expanding the dimensions of the output features. The core of the module features the ABlock, incorporating area attention and a multi-layer perceptron. This design enables rapid and refined feature enhancement by focusing on spatially significant regions. Furthermore, the module includes an optional residual connection, which improves training stability and strengthens feature representation, ensuring sustained high performance across complex visual tasks. The arrangement of modules allows for reasonable scaling up in certain applications.

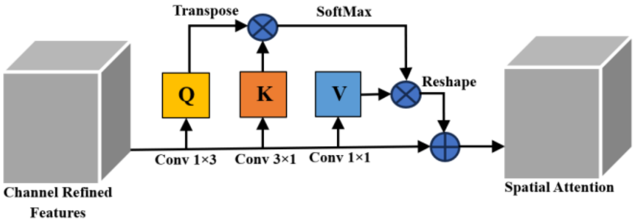


Figure 3. Architectural model of A2C2f mechanism.

The A2C2f mechanism employs a channel refinement strategy to enhance feature extraction by emphasizing relevant information in both spatial and channel dimensions. Initially, channel-refined features are generated to accentuate significant feature maps while suppressing less informative ones. These refined features then undergo processing through a series of convolutional layers, specifically Conc 1×3, Conc 3×1, and Conc 1×1, designed to efficiently capture directional and cross-channel dependencies. Following the convolutional layers, a transpose operation rearranges the feature dimensions in preparation for attention computation. The rearranged features are then normalized using a SoftMax function to produce attention weights, ensuring the model prioritizes the most relevant spatial locations. Subsequently, a reshape operation aligns the tensor dimensions back to their original form. Finally, spatial attention is applied to focus on critical regions within the feature map, refining the overall representation and improving downstream detection task performance. Consider that attention mechanisms can be roughly divided into channel attention (enhancing important channels and suppressing unnecessary ones) and spatial attention (highlighting areas of interest and suppressing background information). When applying both, it's important to consider that calculating the two attentions separately not only increases the computational complexity, but may also result in conflicting importance of feature representations [28].

Dataset Explanation

This study utilizes a weld inspection image dataset comprising 1,658 images, each with a resolution of 640×640 pixels. The dataset was built using a combination of original data collection and publicly available images from online sources. The dataset is designed to detect four key features on weld surfaces: defects, welding lines, workpieces, and porosity and is annotated in YOLO format for object detection tasks. Example images from the validated dataset, illustrating a range of welding conditions with variations in defect types, sizes, and surface characteristics, are shown in Figure 4. This validated dataset is a benchmark for assessing the defect detection model's ability to generalize. By including real-world complexities like inconsistent lighting, material variations, and background noise, the validated images provide a rigorous evaluation environment, ensuring the model's performance is comprehensive, reliable, and reflective of practical industrial scenarios. This strengthens the credibility of the model's reported performance.

To improve the model's robustness and generalization, several data augmentation techniques were applied during training, as illustrated in Figure 4 (b). The original dataset, consisting of unaltered images, served as the baseline. A simple copy-paste augmentation was used, where image regions were randomly duplicated and inserted elsewhere to introduce variability. More advanced techniques, like Mosaic augmentation, were also employed. Mosaic augmentation involves randomly cropping and combining four different images into a single composite, significantly increasing training sample diversity and reducing background dominance. This enhances the model's ability to detect targets in complex environments and reduces overfitting. Additionally, the Mixup technique was used, combining two images and their labels through linear interpolation. This encourages the model to learn smoother, more generalized decision boundaries, improving performance on varied and unseen data.

The dataset used in this research provides a broad representation of common welding defects, with careful annotation of each defect category to ensure accuracy and diversity. The images effectively capture a wide range of quality issues found during welding, making the dataset representative and relevant for real-world applications. For effective model development and evaluation, the dataset was divided into training and validation sets in a 3:1 ratio. Specifically, 904 images were used for training, and 224 images were used for validation. This division provides ample data for learning while maintaining a sufficient sample for reliably assessing the model's generalization.

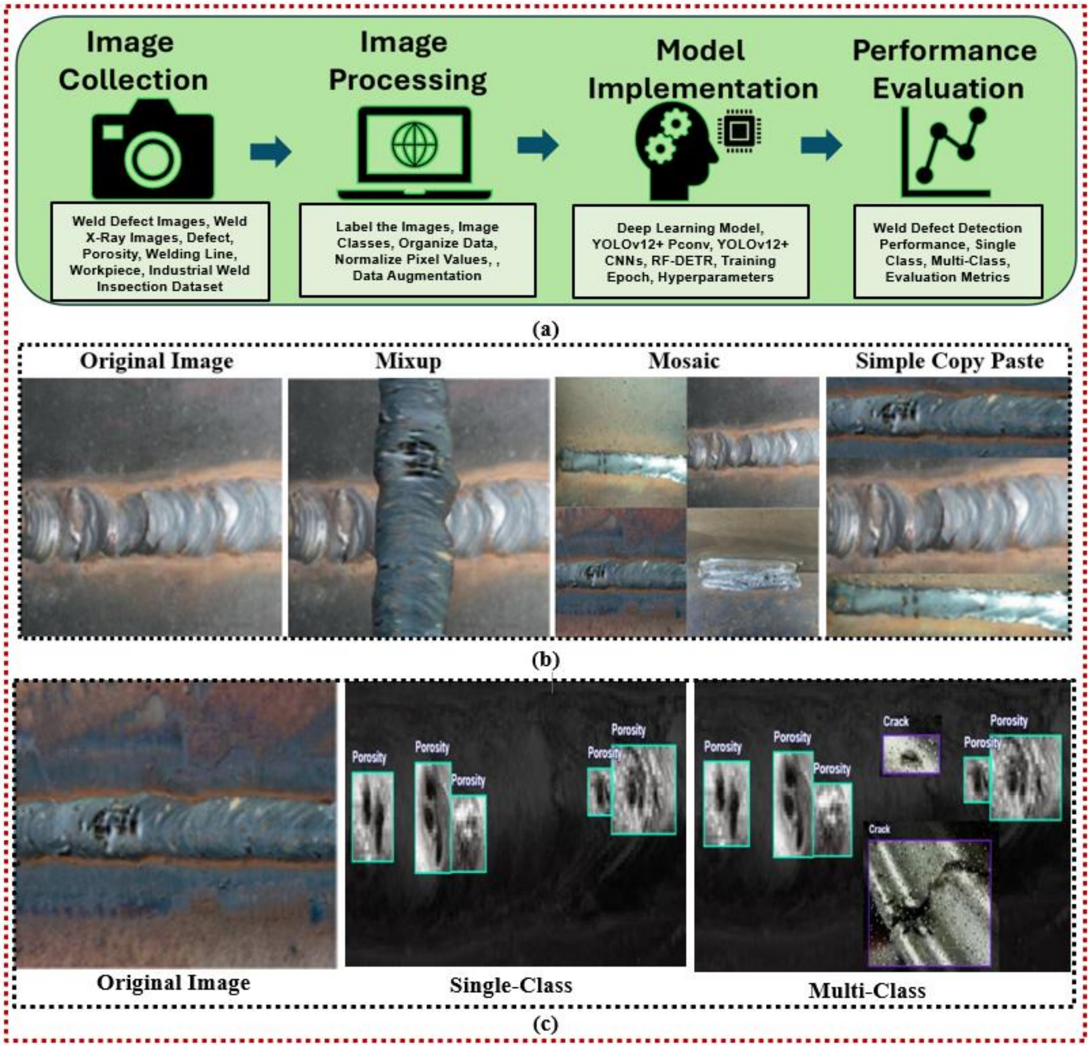


Image Collection
Image Processing
Model Implementation
Performance Evaluation
Weld Defect Images, Weld X-Ray Images, Defect, Porosity, Welding Line, Workpiece, Industrial Weld Inspection Dataset
Label the Images, Image Classes, Organize Data, Normalize Pixel Values, , Data Augmentation
Deep Learning Model, YOLOv12+ Pconv, YOLOv12+ CNNs, RF-DETR, Training Epoch, Hyperparameters
Weld Defect Detection Performance, Single Class, Multi-Class, Evaluation Metrics
(a)
Original Image
Mixup
Mosaic
Simple Copy Paste
(b)
Porosity
Crack
Original Image
Single-Class
Multi-Class
(c)

Figure 4. A details representation of data collection setup and procedures: (a) Flow diagram of the image processing for weld defect detection, (b) Data augmentation for weld defect detection, (c) Single and multi-class image for weld defect detection.

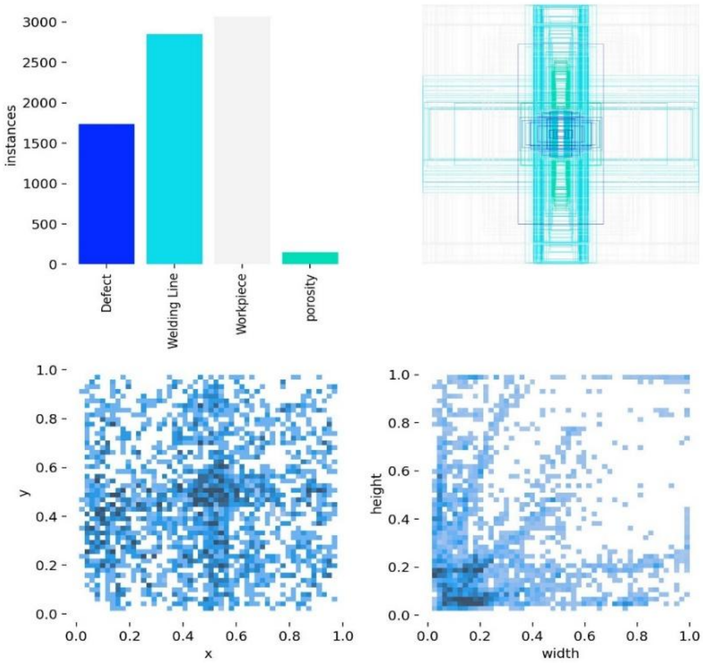


Figure 5. The distribution of welding images across each class in the dataset for welding defect detection.

Figure 5 provides a visual representation of the welding image dataset's composition, showcasing the distribution of images across different categories. This illustration clearly indicates the proportion of each class within the dataset, offering insights into the balance and representation of various welding characteristics and potential defects. Such a visual aid is crucial for understanding the dataset's structure and for assessing its suitability for training and evaluating machine learning models designed for weld inspection.

Experiments and Results

In this research, deep learning models for weld defect detection were developed using Python and implemented within the PyTorch framework. To maintain consistency, the same software environment was used for both training and evaluation. Experiments were performed on a Windows 10 system, which included 16 GB of RAM, an Intel Core i7-11370H CPU, and an NVIDIA GeForce RTX 3050 Ti GPU, providing adequate computational resources for the deep learning tasks. The training process utilized the following hyperparameters: 100 epochs, a batch size of 8, a momentum of 0.937, a weight decay of 0.0005, and a learning rate of 0.01.

The confusion matrix results (As shown in Figure 6) indicate that the proposed model exhibits high accuracy in detecting weld defects, as evidenced by elevated true positive rates across all categories. Specifically, the model demonstrates strong recall values for various classes: 0.89 for Defect, 0.93 for Welding Line, 0.99 for Workpiece, and 0.95 for Porosity. These values, concentrated along the matrix diagonal, confirm the model's ability to effectively discriminate between classes. While minor misclassifications were observed, primarily between the Defect and Welding Line categories (likely due to visual similarities) the model shows near-perfect accuracy in identifying Workpiece and Porosity, with minimal confusion involving other categories. Overall, the classifier performs excellently, displaying only limited cross-class confusion.

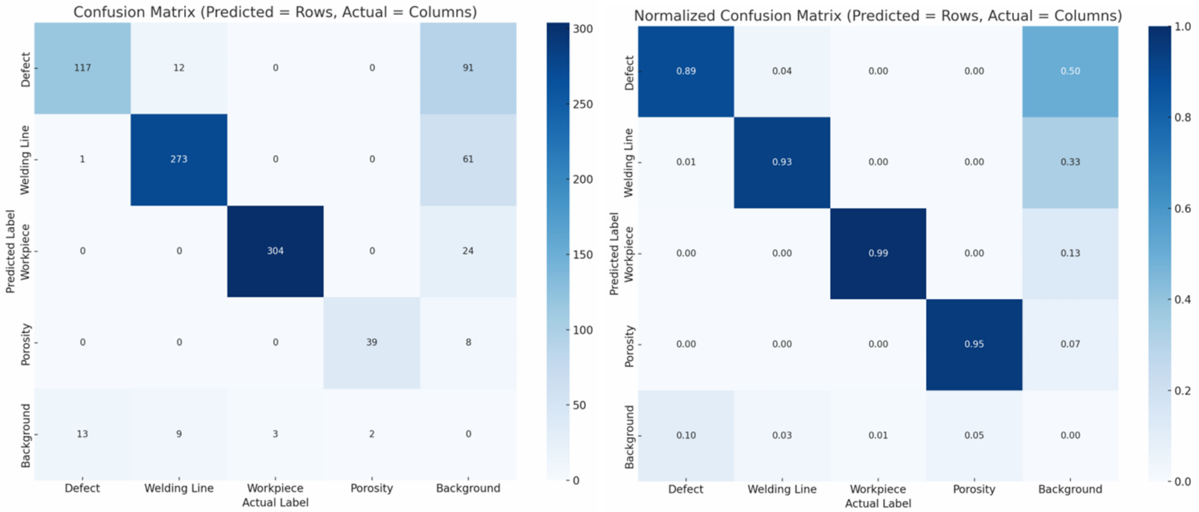


Figure 6. Confusion matrix and normalized confusion matrix for welding defect detection.

Figure 7 provides a detailed analysis of the proposed welding defect detection model’s performance, illustrating the progression of key performance indicators throughout the training, validation, and prediction phases. The graph depicts the evolution of box loss, object loss, and class loss across successive epochs, demonstrating a consistent downward trend indicative of effective learning and model convergence. Concurrently, evaluation metrics such as accuracy and recall exhibit steady improvement as training advances. The increasing values of accuracy and recall reflect the model's enhanced ability to correctly detect and classify welding defects over time. The simultaneous decline in various loss functions, coupled with the rise in performance metrics, confirms the stability and effectiveness of the model's optimization process. Overall, Figure 8 visually validates the robust learning behavior of the proposed model and its capability to achieve high prediction reliability during deployment.

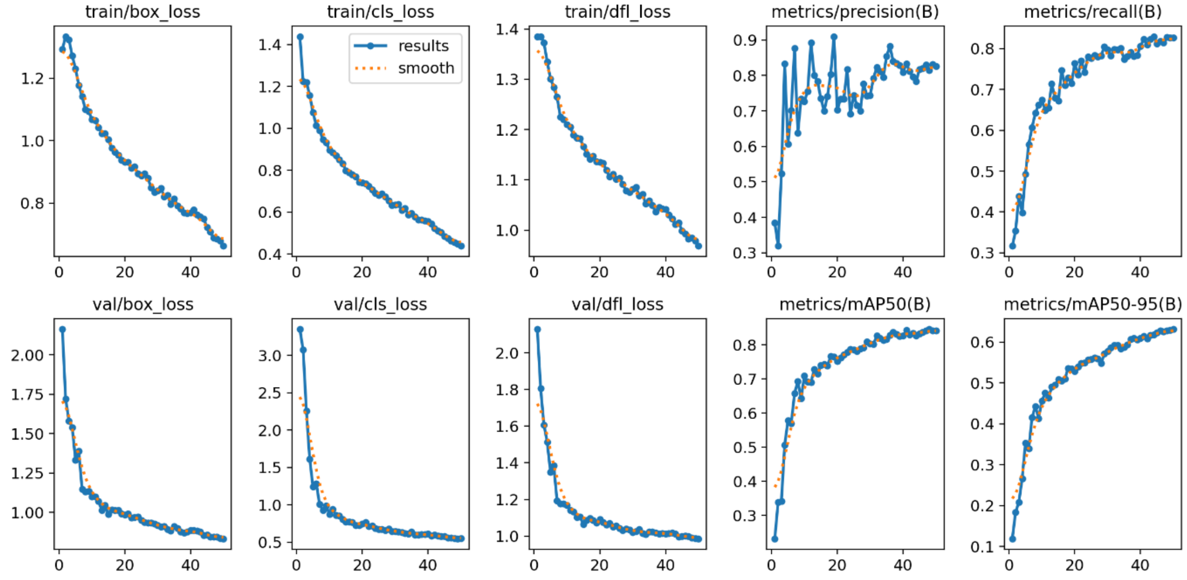


Figure 7. Performance graph analysis between training and prediction model for welding defect detection.

As shown in Figure 8, the training performance of the innovative YOLOv12 model with PConv and Adaptive Attention for high-precision welding defect detection demonstrates a consistent improvement in mAP@0.5 over 100 epochs. The model starts with a low mAP of 0.037 at epoch 1 but shows rapid learning in the first 10 epochs, reaching 0.905. This fast early improvement reflects the effective feature extraction and learning capabilities introduced by the novel architecture. From epochs 11 to 30, the mAP steadily climbs into the high 0.99 range, indicating the model is nearing optimal detection performance. Beyond epoch 30, the mAP stabilizes around 0.996, with minor fluctuations, and even hits a perfect score of 1.0 at epoch 57, demonstrating exceptional detection accuracy. However, slight oscillations between 0.99 and 1.0 in later epochs suggest the model maintains robustness but faces diminishing returns from further training. Overall, the data reflects that the YOLOv12 variant efficiently converges to a near-perfect defect detection solution, confirming the effectiveness of pinwheel convolution and adaptive attention mechanisms in enhancing welding defect identification.

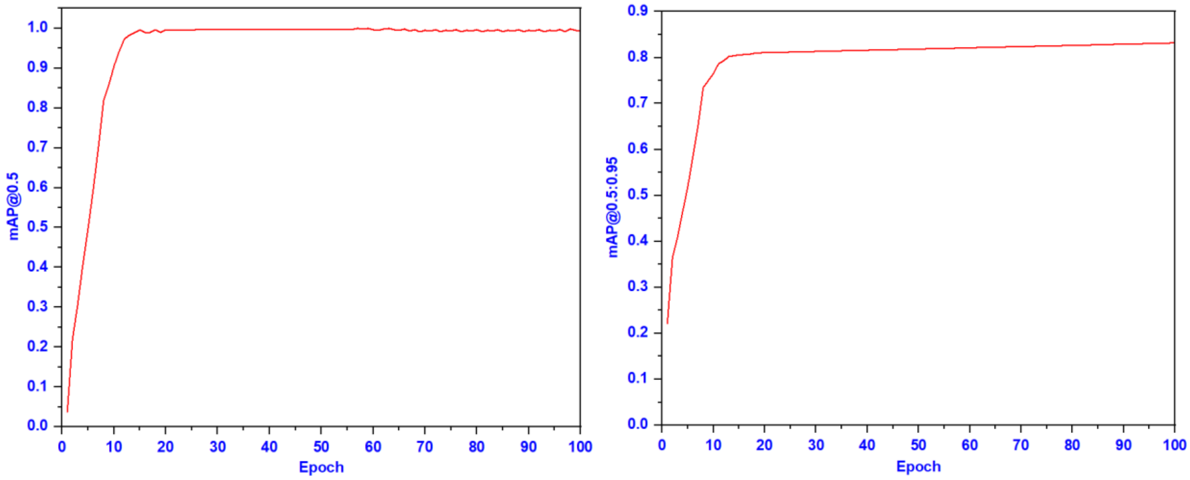


Figure 8. The mAP evolution during the training process of YOLOv12 (PConv Based) for welding defect detection.

The training results for the YOLOv12 model with PConv and Adaptive Attention, measured by mAP@0.5:0.95 over 100 epochs, demonstrate a consistent improvement in detection precision for high-precision welding defect detection. Beginning with a mAP of 0.22 in epoch 1, the model exhibits rapid learning during the initial 10 epochs, achieving a value of approximately 0.76, indicative of effective early-stage feature extraction and detection capability. Subsequently, the increase in mAP decelerates yet remains consistent, gradually ascending from approximately 0.80 at epoch 15 to around 0.83 by epoch 100. This gradual ascent suggests that the model continues to refine its performance at more challenging detection thresholds (higher IoU levels), demonstrating improved localization and classification accuracy over time. The overall trend indicates that the model effectively learns to detect welding defects with high precision, although gains become incremental as it approaches its performance ceiling. This sustained progression underscores the efficacy of the model's innovative components in enhancing defect detection accuracy under stricter evaluation metrics.

Visual inspection of Figure 9 indicates that high-precision welding defect detection has been significantly enhanced through the introduction of an innovative YOLOv12 model integrated with PConv and Adaptive Attention mechanisms. This advanced model architecture demonstrates superior detection performance, highlighting its capability to identify welding defects with high accuracy while minimizing both false positives and false negatives. The integration of PConv allows the model to better capture detailed spatial features, while Adaptive Attention ensures a focus on the most relevant defect regions, collectively making YOLOv12 a powerful tool for real-time, reliable welding defect detection in industrial applications.

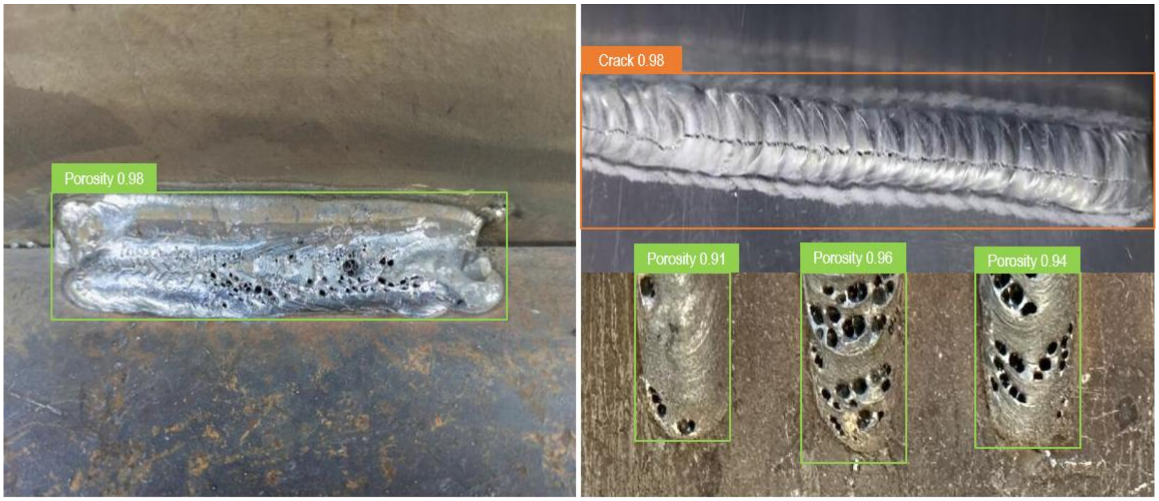


Figure 9. Detective result of YOLOv12 (PConv Based) for welding defect detection.

Figure 10 illustrates the impact of various datasets on the mean Average Precision (mAP) during the training of the proposed PConv based YOLOv12 model. The highest performance, as measured by the mAP@0.5 metric, was observed when training with both the original image dataset and the simple copy-paste dataset. This result exceeded the performance achieved with Mosaic and Mixup data augmentation techniques. In contrast, the Mixup augmentation method yielded the lowest performance across all datasets. However, when evaluating the mAP@0.5:0.95 metric, Mosaic augmentation demonstrated the best performance, indicating its effectiveness in improving the model's ability to generalize across various Intersection over Union thresholds. The Mixup augmentation technique again resulted in the lowest mAP@0.5:0.95 performance, suggesting its limited effectiveness in this context.

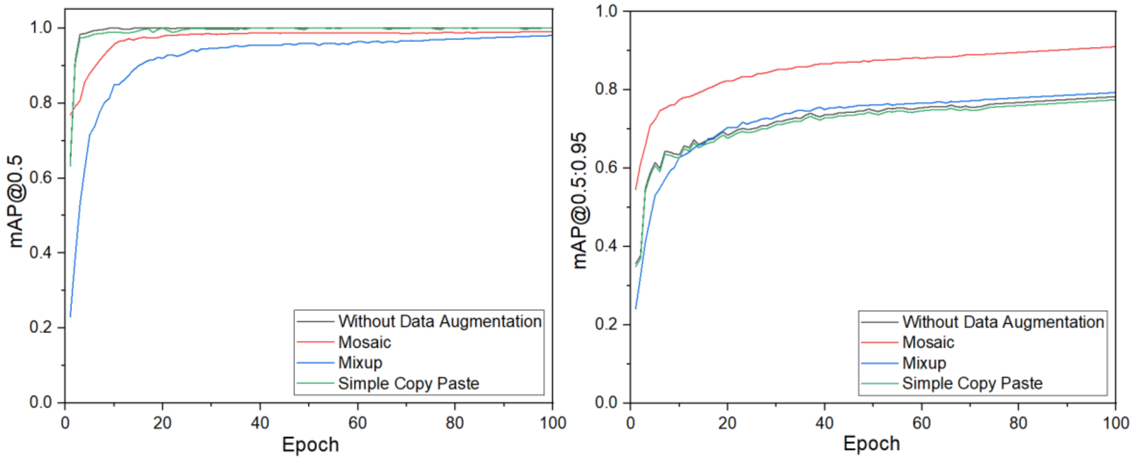


Figure 10. Average precision evolution of various data augmentation techniques with original dataset during YOLOv12 (PConv based) training process for welding defect detection.

Table 1 presents a comparative analysis of the proposed PConv based YOLOv12 model's performance with different data augmentation techniques, focusing on precision, recall, F1-score, mAP@0.5, and mAP@0.5:0.95 metrics. Training with the original images resulted in a high precision of 0.921, a high recall of 0.962, and the highest F1-score of 0.941, indicating a strong balance between minimizing false positives and false negatives. Additionally, the model achieved the highest mAP@0.5 of 0.989, demonstrating high detection accuracy at a single IoU threshold. However, its mAP@0.5:0.95 score of 0.724 was moderate compared to that of the Mosaic augmentation, suggesting that while the model was highly accurate at standard thresholds, its adaptability to stricter IoU conditions was somewhat lower. The Mixup augmentation showed the lowest overall performance, with a precision of 0.873, a recall of 0.863, and an F1-score of 0.869. Although its mAP@0.5 score of 0.944 was significantly lower than that of the original images and simple copy-paste methods, its mAP@0.5:0.95 score of 0.727 was slightly better than the original, implying improved generalization despite a decline in basic detection performance. Mosaic augmentation stood out with the highest mAP@0.5:0.95 score of 0.854, indicating excellent generalization across a range of IoU thresholds. It also achieved a high precision of 0.934 and a strong recall of 0.924, resulting in an F1-score of 0.929, which is close to that of the original images. This suggests that Mosaic significantly improved the model's robustness and adaptability to varying object scales and occlusions. The simple copy-paste method yielded moderate improvements, with a precision of 0.891, a recall of 0.872, and an F1-score of 0.881. Its mAP@0.5 score of 0.971 was relatively high, but the mAP@0.5:0.95 score of 0.716 was the lowest among the evaluated techniques, indicating limited improvement in generalization. In summary, while the original images provided the highest precision at standard thresholds, Mosaic augmentation demonstrated the best overall generalization and adaptability, making it the most effective augmentation strategy for the proposed YOLOv12 model in complex detection scenarios.

Table 1. The performance evaluation of various data augmentation techniques with original dataset for welding defect detection.

Method	Precision	Recall	F1-Score	mAP@0.5	mAP@0.5:0.95
Original Image	0.921	0.962	0.941	0.989	0.724
Mixup	0.873	0.863	0.869	0.944	0.727
Mosaic	0.934	0.924	0.929	0.952	0.854
Simple Copy-Paste	0.891	0.872	0.881	0.971	0.716

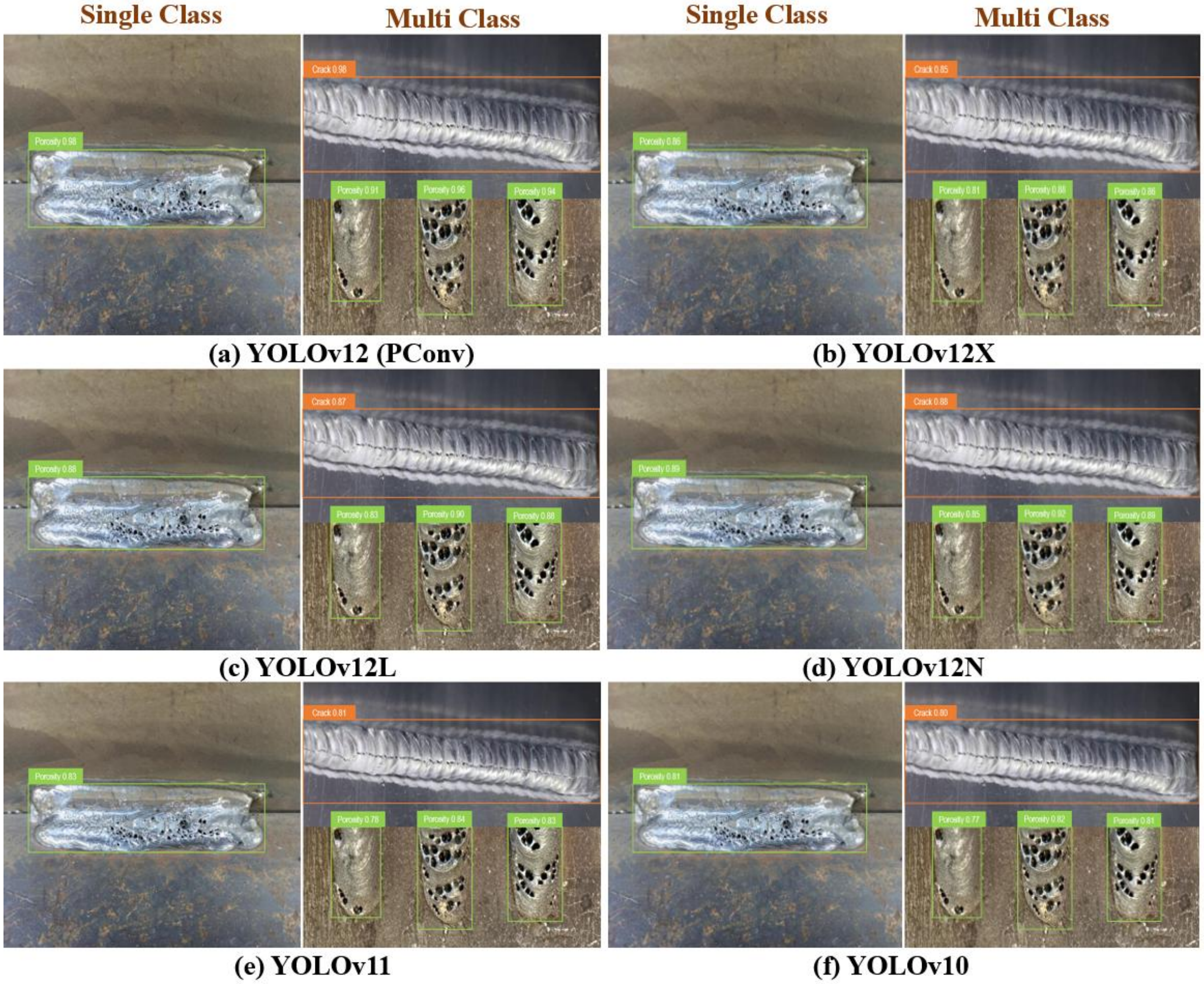
The enhanced performance of the PConv based YOLOv12 model in welding defect detection is substantiated by a comparative analysis against previous YOLO variants, as detailed in Table 2 and Figure 11. In single-class detection scenarios, the model attained a precision of 0.921, a recall of 0.962, and an F1 score of 0.941, surpassing the performance of other models, including YOLOv12N, which achieved an F1 score of 0.883. The notable improvement in recall indicates the model's enhanced sensitivity in identifying subtle or complex defect characteristics. This is attributed to the Pinwheel Convolution's improved spatial feature extraction and the Adaptive Attention mechanism's targeted feature weighting, rendering the model suitable for applications where defect detection is critical to prevent structural failures.

Table 2. Performance comparison of the YOLOv12 (PConv Based) detection performance across various existing YOLO models for welding defect detection.

	Single Class			Multi Class		
Model	Precision	Recall	F1 Score	Precision	Recall	F1 Score
YOLOv10*	0.782	0.851	0.815	0.712	.806	0.756
YOLOv11*	0.802	0.856	0.828	0.721	0.820	0.767
YOLOv12X	0.857	0.859	0.858	0.738	0.832	0.782
YOLOv12L	0.879	0.862	0.870	0.772	0.791	0.781
YOLOv12N	0.898	0.868	0.883	0.788	0.807	0.797

YOLOv12 (PConv Based)	0.921	0.962	0.941	0.812	0.887	0.848

In multi-class detection scenarios, which emulate more realistic conditions, the PConv based YOLOv12 model maintained a leading position with a precision of 0.812, a recall of 0.887, and an F1 score of 0.848. These metrics suggest robust performance and effective generalization across various defect types. Compared to the YOLOv12N model, which demonstrated an F1 score of 0.797, the PConv-based architecture exhibited advancements in multi-class discrimination and detection reliability. The progressive performance enhancement from earlier YOLO versions to the PConv-enhanced YOLOv12, as illustrated in Table 2 and Figure 11, supports the integration of advanced convolutional operations and attention mechanisms to improve model accuracy and adaptability for industrial welding inspection systems.



(a) YOLOv12 (PConv) (b) YOLOv12X

(c) YOLOv12L (d) YOLOv12N

(e) YOLOv11 (f) YOLOv10

Figure 11. Detective results comparison with different YOLO model for welding defect detection.

Comparison between PConv Based YOLOv12 Model and RF-DETR

RF-DETR and YOLOv12 represent distinct approaches to object detection, with RF-DETR leveraging transformer-based architectures and YOLOv12 utilizing CNN-based architectures. A study focusing on green fruit detection in complex orchard environments compared the performance of RF-DETR and YOLOv12. The RF-DETR model, with a DINOv2 backbone and deformable attention, demonstrated strength in global context modelling, which allowed it to effectively identify partially occluded or ambiguous green fruits. YOLOv12, on the other hand, used CNN-based attention to enhance local feature extraction, making it suitable for computational efficiency and edge deployment [29].

Figure 12 illustrates a comparative analysis of the PConv based YOLOv12 model and the RF-DETR model for high-precision weld defect detection, highlighting their respective advantages. YOLOv12 exhibits a consistent increase in detection accuracy throughout training, reaching a mAP@0.5:0.95 of approximately 0.83, demonstrating effective feature extraction and adaptive attention mechanisms that improve local defect identification. Conversely, RF-DETR achieves a higher initial accuracy with rapid convergence, exceeding a mAP@0.5 of 0.97 within 100 epochs. This is attributed to its transformer-based architecture, which excels at capturing global spatial relationships and complex contextual information relevant for weld defect localization. While PConv based YOLOv12 provides a computationally efficient and stable learning process suitable for real-time applications, RF-DETR offers superior detection precision by utilizing region-based transformers, albeit potentially at a higher computational cost. The optimal model selection depends on the desired balance between speed and accuracy, with PConv based YOLOv12 prioritizing efficiency and RF-DETR emphasizing precision in weld defect detection.

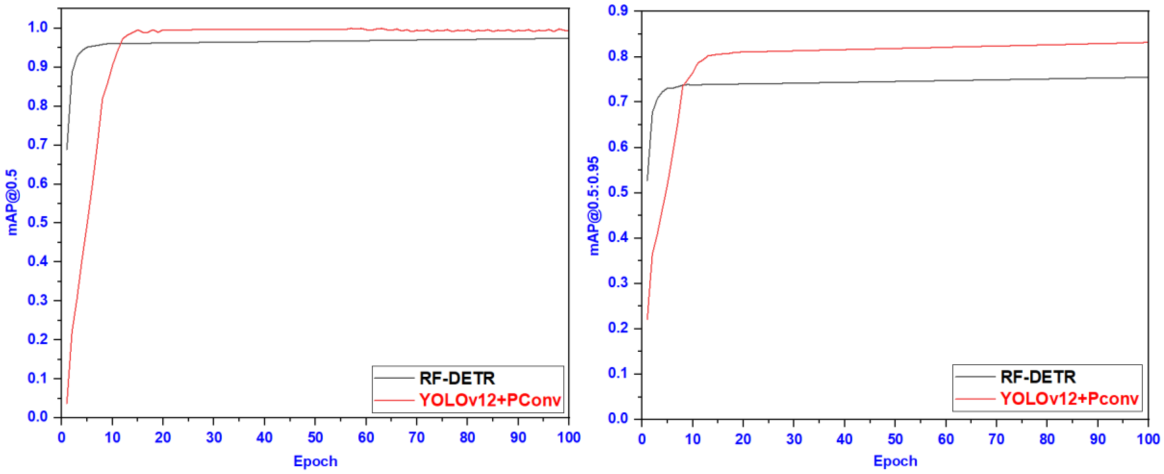
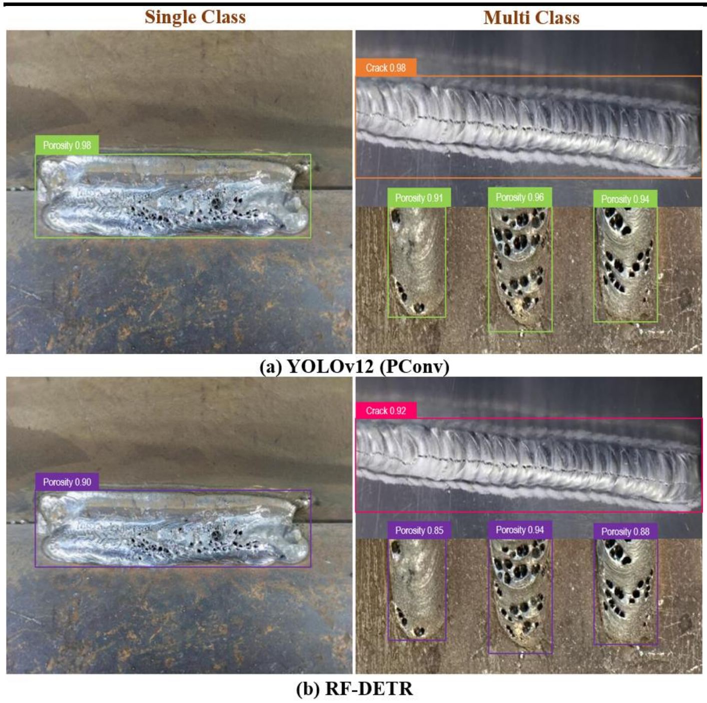


Figure 12. Average precision (AP) comparison between YOLOv12 (PConv Based) and RF-DETR for welding defect detection.

Table 3 provides a detailed performance comparison between the PConv based YOLOv12 model and the RF-DETR model for welding defect detection in single-class and multi-class scenarios. In the single-class setting, YOLOv12 outperforms RF-DETR, achieving higher precision (0.921 vs. 0.902), recall (0.962 vs. 0.916), and F1 score (0.941 vs. 0.909), indicating its superior ability to accurately identify welding defects when focusing on a single defect type. In the more complex multi-class detection task, PConv based YOLOv12 maintains its lead with better precision (0.812 vs. 0.792), recall (0.887 vs. 0.842), and F1 score (0.848 vs. 0.816), demonstrating enhanced robustness and generalization in detecting multiple defect categories simultaneously. These results demonstrate the effectiveness of the PConv-based YOLOv12 model in delivering high-precision and reliable weld defect detection compared to the transformer-based RF-DETR approach.

Table 3. Performance comparison between YOLOv12 (PConv Based) and RF-DETR for welding defect detection.

	Single Class			Multi Class		
Model	Precision	Recall	F1 Score	Precision	Recall	F1 Score
RF-DETR	0.902	0.916	0.909	0.792	0.842	0.816
YOLOv12 (PConv Based)	0.921	0.962	0.941	0.812	0.887	0.848



(a) YOLOv12 (PConv)

(b) RF-DETR

Figure 13. Detective results comparison between YOLOv12 (PConv Based) and RF-DETR for welding defect detection.

The visual detection results from the Figure 13 validation tests reveal performance differences between PConv based YOLOv12 and the RF-DETR model in welding defect detection. PConv based YOLOv12 demonstrates superior localization accuracy and sharper boundary delineation, especially in complex or overlapping defect regions. The model consistently identifies small and irregular defects with higher clarity and fewer false positives, because of the enhanced feature extraction capabilities of the PConv module. RF-DETR, while effective in broader defect identification, occasionally misclassifies or overlooks subtle defects, particularly in multi-class scenarios with low-contrast images. The qualitative comparison emphasizes PConv based YOLOv12's robustness and reliability in practical settings, making it more suitable for high-precision weld defect inspection tasks where visual clarity and detection granularity are critical.

Discussions and Future Research Implication

The improved YOLOv12 model, enhanced with Pinwheel Convolution and Adaptive Attention mechanisms, represents a significant leap forward in welding defect detection technology. This model demonstrates superior detection accuracy and enhanced class discrimination, evidenced by its high recall rates across various defect classes, achieving 0.99 for Workpiece and 0.95 for Porosity. High F1 scores of 0.941 for single-class and 0.848 for multi-class detection scenarios confirm its effectiveness in distinguishing between weld defect categories with minimal confusion, notably outperforming previous

YOLO iterations such as YOLOv12N. Such advancements are crucial in ensuring the reliability and safety of welded structures, reducing the potential for catastrophic failures due to undetected flaws. The enhancements not only improve detection rates but also contribute to a more robust and reliable automated inspection process.

The model's efficient and stable learning process is characterized by rapid convergence, achieving a mAP@0.5 of 0.905 within the first 10 epochs and stabilizing around 0.996, showcasing its ability to quickly adapt and learn from the provided data. Furthermore, the strategic use of data augmentation techniques, particularly Mosaic augmentation, significantly boosts the model's generalization capabilities, achieving a high mAP@0.5:0.95 of 0.854. This PConv-based YOLOv12 achieves a balanced trade-off between speed and precision, rendering it suitable for real-time applications while maintaining competitive precision. The integration of PConv enhances spatial feature extraction, while the Adaptive Attention mechanism improves focus on defect-prone regions, leading to superior detection reliability and model generalization. Such improvements align with the broader goals of enhancing the precision and efficiency of automated defect detection system.

Although the proposed PConv-based YOLOv12 attention-centric model for welding defect detection achieved high accuracy and robustness, several limitations warrant acknowledgment. First, despite the dataset's diversity, its limited sample size and variation, particularly concerning the subtle distinctions between "Defect" and "Welding Line" classes, suggest the need for more complex and varied real-world samples to enhance the model's differentiation capabilities. The experiments' reliance on a specific hardware configuration also implies that performance may vary on systems with fewer computational resources, potentially affecting real-time detection. Furthermore, the inconsistent effectiveness of data augmentation techniques, such as the superior performance of Mosaic compared to Mixup, underscores the necessity for targeted approaches tailored to welding defect image characteristics. Another limitation lies in the model's focus on static image analysis, neglecting the dynamic aspects of welding processes suitable for real-time video analysis. Future research should address these limitations by expanding the dataset to encompass a broader range of welding types, materials, defect patterns, and operational conditions. Integrating real-time video stream analysis could significantly enhance the model's practical applicability in industrial settings. Exploring advanced attention mechanisms, semi-supervised learning, and domain adaptation techniques could further improve performance in diverse and unseen environments. Investigating lightweight model versions optimized for deployment on edge devices could also pave the way for advancements in smart manufacturing and real-time quality control applications.

Conclusions

This research presents an innovative approach to automated weld defect detection, focusing on an enhanced YOLOv12 model that incorporates Pinwheel Convolution and Adaptive Attention mechanisms. By integrating PConv and Adaptive Attention, the model achieves superior spatial feature extraction and focuses more precisely on defect-prone regions. This leads to higher detection accuracy and improved robustness, which are essential for reliable performance in industrial settings. The enhanced model demonstrates excellent performance in both single-class and multi-class detection scenarios, achieving high F1-scores and recall rates across all defect categories. Furthermore, the efficient training dynamics, characterized by rapid convergence and stable learning, highlight the practical applicability of this approach. The model's adaptability is further enhanced through effective data augmentation strategies, with Mosaic augmentation playing a key role in improving its generalization capability. Overall, this research offers a powerful, real-time, and computationally efficient solution, which advances the state-of-the-art in intelligent weld defect detection systems and contributes to automated defect recognition in industrial applications. The use of X-ray imaging combined with the enhanced YOLOv12 model offers a promising direction for improving the reliability of welding processes.

- Data Augmentation Process: Data augmentation significantly influences the model's performance and ability to generalize to new, unseen data. Specifically, the study found that Mosaic augmentation led to the highest mAP@0.5:0.95 (0.854), demonstrating its effectiveness in enhancing the model's adaptability to varied defect scenarios. Conversely, training the model with original, unaugmented images resulted in the highest mAP@0.5 (0.989) and F1-score (0.941), highlighting the importance of high-quality, original data for achieving precision.
- Single-Class Detection: In single-class defect detection, the advanced YOLOv12 model achieved an F1-score of 0.941 and a mAP@0.5 of 0.989. These scores indicate excellent accuracy and reliability in identifying single defect types. The high recall rates across different defect classes confirm that the model minimizes misclassification, effectively detecting nearly all instances of the targeted defect. Furthermore, the model's rapid convergence during training underscores its robustness and suitability for precise, single-defect identification in welding applications.
- Multi-Class Detection: In multi-class detection, the advanced YOLOv12 model achieved an F1-score of 0.848 and a recall of 0.887, outperforming baseline YOLO variants. This indicates that the model is highly effective at distinguishing between multiple types of weld defects, which is crucial for comprehensive quality control. The model also demonstrated high localization accuracy, with a mAP@0.5:0.95 of approximately 0.83, ensuring reliable classification and precise location of defects. The enhancements from adaptive attention and PConv contribute to better defect separation and improved generalization, enabling the model to perform well on unseen data.
- Model Training Dynamics and Convergence: The advanced YOLOv12 model exhibits efficient training dynamics and rapid convergence. It achieves a high mAP@0.5 of 0.905 within just 10 epochs and stabilizes around 0.996, indicating stable learning behavior. The progressive increase in mAP@0.5:0.95 from 0.22 to approximately 0.83 reflects enhanced localization accuracy under stricter Intersection over Union thresholds, demonstrating effective model optimization. This rapid convergence not only saves computational resources but also makes the model more practical for real-world applications where timely results are essential. YOLOv12 offers scalable solutions and achieves consistent gains in mean average precision (mAP) and inference speed.

References

1. Vasan V, Sridharan NV, Balasundaram RJ, Vaithiyanathan S. Ensemble-based deep learning model for welding defect detection and classification. Eng Appl Artif Intell 2024;136:108961. https://doi.org/10.1016/J.ENGAPPAI.2024.108961.

2. Mobaraki M. Vision-based seam tracking and multi-modal defect detection in GMAW fillet welding using artificial intelligence 2025. https://doi.org/10.14288/1.0448316.
3. Chen X, Wu Y, He X, Ming W. A Comprehensive Review of Deep Learning-Based PCB Defect Detection. IEEE Access 2023;11:139017–38. https://doi.org/10.1109/ACCESS.2023.3339561.
4. Zhu H, Xie C, Fei Y, Tao H. Attention Mechanisms in CNN-Based Single Image Super-Resolution: A Brief Review and a New Perspective. Electronics 2021, Vol 10, Page 1187 2021;10:1187. https://doi.org/10.3390/ELECTRONICS10101187.
5. Xu R, Tao Y, Lu Z, Zhong Y. Attention-Mechanism-Containing Neural Networks for High-Resolution Remote Sensing Image Classification. Remote Sensing 2018, Vol 10, Page 1602 2018;10:1602. https://doi.org/10.3390/RS10101602.
6. Pozzi A, Barbierato E, Aldein A, Ibrahim MS, Tapamo J-R. A Survey of Vision-Based Methods for Surface Defects' Detection and Classification in Steel Products. Informatics 2024, Vol 11, Page 25 2024;11:25. https://doi.org/10.3390/INFORMATICS11020025.
7. Sohag Mia M, Li C. STD2: Swin Transformer-Based Defect Detector for Surface Anomaly Detection. IEEE Trans Instrum Meas 2025;74:1–15. https://doi.org/10.1109/TIM.2024.3492728.
8. Diwan T, Anirudh G, Tembhurne J V. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimed Tools Appl 2023;82:9243–75. https://doi.org/10.1007/S11042-022-13644-Y/METRICS.
9. Jiang C, Ren H, Ye X, Zhu J, Zeng H, Nan Y, et al. Object detection from UAV thermal infrared images and videos using YOLO models. International Journal of Applied Earth Observation and Geoinformation 2022;112:102912. https://doi.org/10.1016/J.JAG.2022.102912.
10. Alahdal NM, Abukhodair F, Meftah LH, Cherif A. Real-time Object Detection in Autonomous Vehicles with YOLO. Procedia Comput Sci 2024;246:2792–801. https://doi.org/10.1016/J.PROCS.2024.09.392.
11. Pan K, Hu H, Gu P. WD-YOLO: A More Accurate YOLO for Defect Detection in Weld X-ray Images. Sensors 2023, Vol 23, Page 8677 2023;23:8677. https://doi.org/10.3390/S23218677.
12. Luo Y, Ling J, Wang J, Zhang H, Chen F, Xiao X, et al. SFW-YOLO: A lightweight multi-scale dynamic attention network for weld defect detection in steel bridge inspection. Measurement 2025;253:117608. https://doi.org/10.1016/J.MEASUREMENT.2025.117608.
13. Wang GQ, Zhang CZ, Chen MS, Lin YC, Tan XH, Liang P, et al. Yolo-MSAPF: Multiscale Alignment Fusion with Parallel Feature Filtering Model for High Accuracy Weld Defect Detection. IEEE Trans Instrum Meas 2023;72. https://doi.org/10.1109/TIM.2023.3302372.
14. Kwon JE, Park JH, Kim JH, Lee YH, Cho SI. Context and scale-aware YOLO for welding defect detection. NDT & E International 2023;139:102919. https://doi.org/10.1016/J.NDTEINT.2023.102919.
15. Liu M, Chen Y, Xie J, He L, Zhang Y. LF-YOLO: A Lighter and Faster YOLO for Weld Defect Detection of X-Ray Image. IEEE Sens J 2023;23:7430–9. https://doi.org/10.1109/JSEN.2023.3247006.
16. Chen Y, Yuan X, Wang J, Wu R, Li X, Hou Q, et al. YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-time Object Detection. IEEE Trans Pattern Anal Mach Intell 2025. https://doi.org/10.1109/TPAMI.2025.3538473.

17. Khanam R, Hussain M. A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions 2025.
18. Li C, Li L, Jiang H, Weng K, Geng Y, Li L, et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications 2022.
19. Alif MAR, Hussain M. YOLOv12: A Breakdown of the Key Architectural Features 2025.
20. Tian Y, Ye Q, Doermann D. YOLOv12: Attention-Centric Real-Time Object Detectors 2025. https://doi.org/10.0.
21. Yang J, Lee C-H. Real-Time Data-Driven Method for Bolt Defect Detection and Size Measurement in Industrial Production. Actuators 2025, Vol 14, Page 185 2025;14:185. https://doi.org/10.3390/ACT14040185.
22. Khanam R, Hussain M. A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions 2025.
23. Wang H, Zhang C, Wang Y, Ni P, Wang Y. Segmentation of Inner Surface Defects of Stainless Steel Pipes Based on Semi-bilateral Efficient Self-Attention Network. J Nondestr Eval 2025;44:1–17. https://doi.org/10.1007/S10921-025-01176-Y/METRICS.
24. Sapkota R, Cheppally RH, Sharda A, Karkee M. RF-DETR Object Detection vs YOLOv12 : A Study of Transformer-based and CNN-based Architectures for Single-Class and Multi-Class Greenfruit Detection in Complex Orchard Environments Under Label Ambiguity 2025.
25. He Y, Yu Z, Li J, Yu L, Ma G. Discerning Weld Seam Profiles from Strong Arc Background for the Robotic Automated Welding Process via Visual Attention Features. Chinese Journal of Mechanical Engineering (English Edition) 2020;33:1–12. https://doi.org/10.1186/S10033-020-00438-2/FIGURES/17.
26. Fan J, Ling X, Liang J. Detection of Surface Defects of Steel Plate Based on ViT. J Phys Conf Ser 2021;2002:012039. https://doi.org/10.1088/1742-6596/2002/1/012039.
27. Hao Z, Wang Z, Bai D, Tao B, Tong X, Chen B. Intelligent Detection of Steel Defects Based on Improved Split Attention Networks. Front Bioeng Biotechnol 2022;9:810876. https://doi.org/10.3389/FBIOE.2021.810876/BIBTEX.
28. Miao R, Shan Z, Zhou Q, Wu Y, Ge L, Zhang J, et al. Real-time defect identification of narrow overlap welds and application based on convolutional neural networks. J Manuf Syst 2022;62:800–10. https://doi.org/10.1016/J.JMSY.2021.01.012.
29. Liu T, Wang J, Huang X, Lu Y, Bao J. 3DSMDA-Net: An improved 3DCNN with separable structure and multi-dimensional attention for welding status recognition. J Manuf Syst 2022;62:811–22. https://doi.org/10.1016/J.JMSY.2021.01.017.