applied sciences

MDPI

Article

ADFE-DET: An Adaptive Dynamic Feature Enhancement Algorithm for Weld Defect Detection

Xiaocui Wu [1], Changjun Liu [1,*], Hao Zhang [1] and Pengyu Xu [2]

[1] School of Chemical Equipment, Shenyang University of Technology, Liaoyang 111003, China
[2] School of Materials Science and Engineering, Wuhan University of Technology, Wuhan 430070, China
* Correspondence: 18941998881@126.com

Abstract

Welding is a critical joining process in modern manufacturing, with defects contributing to 50–80% of structural failures. Traditional inspection methods are often inefficient, subjective, and inconsistent. To address challenges in weld defect detection—including scale variation, morphological complexity, low contrast, and sample imbalance—this paper proposes ADFE-DET, an adaptive dynamic feature enhancement algorithm. The approach introduces three core innovations: the Dynamic Selection Cross-stage Cascade Feature Block (DSCFBlock) captures fine texture features via edge-preserving dynamic selection attention; the Adaptive Hierarchical Spatial Feature Pyramid Network (AHSFPN) achieves adaptive multi-scale feature integration through directional channel attention and hierarchical fusion; and the Multi-Directional Differential Lightweight Head (MDDLH) enables precise defect localization via multi-directional differential convolution while maintaining a lightweight architecture. Experiments on three public datasets (Weld-DET, NEU-DET, PKU-Market-PCB) show that ADFE-DET improves mAP50 by 2.16%, 2.73%, and 1.81%, respectively, over baseline YOLOv11n, while reducing parameters by 34.1%, computational complexity by 4.6%, and achieving 105 FPS inference speed. The results demonstrate that ADFE-DET provides an effective and practical solution for intelligent industrial weld quality inspection.

Keywords: weld defect detection; target detection; attention mechanism; feature pyramid network

check for updates

Academic Editor: Manoj Gupta

Received: 10 October 2025
Revised: 23 October 2025
Accepted: 27 October 2025
Published: 30 October 2025

Citation: Wu, X.; Liu, C.; Zhang, H.; Xu, P. ADFE-DET: An Adaptive Dynamic Feature Enhancement Algorithm for Weld Defect Detection. *Appl. Sci.* **2025**, *15*, 11595. https://doi.org/10.3390/app152111595

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Welding, a fundamental joining process in modern manufacturing, is extensively applied in the automotive, aerospace, construction, and petrochemical industries, where weld quality directly determines structural integrity and safety [1–5]. Statistics indicate that welding defects contribute to approximately 50–80% of all structural failures, underscoring the necessity of effective quality control measures [6]. However, conventional manual visual inspection is inefficient, subjective, and unable to maintain consistency and accuracy in high-speed production environments. Therefore, the development of automated, intelligent weld defect-detection technologies is essential to enhance product quality, reduce production costs, and ensure structural reliability [7].

Traditional weld defect detection primarily relies on non-destructive testing (NDT) techniques, including radiographic, ultrasonic, magnetic particle, and liquid penetrant inspections. Among these methods, X-ray inspection is the most widely adopted due to its intuitive visualization and high sensitivity to internal defects, enabling effective detection

of porosity, cracks, slag inclusions, and lack of fusion within welds [8]. However, these traditional approaches suffer from inherent limitations, such as low detection efficiency, high operational costs, and dependence on expert interpretation. Moreover, inspection results are easily influenced by the inspectors' expertise and subjective judgment, compromising the overall reliability of defect evaluation [9].

In recent years, deep learning has achieved remarkable advancements in weld defect detection. Madhav et al. [10] employed a deep convolutional neural network (DCNN) for automated defect identification in automotive welds, attaining 99.01% detection accuracy using 10,000 OK and Not-OK images. Wang et al. [11] introduced a depth-based learning approach for binary classification of welding defects, enhancing model performance through data augmentation and deep optimization of convolutional layers. Ngo Thi Hoa et al. [6] developed Weld-CNN, a hybrid convolutional neural network integrating sequential convolutional layers with a parallel block structure, achieving 99.83% testing accuracy on a dataset of 24,407 X-ray images. Kim et al. [12] applied the MobileNet model to ultrasonic B-scan images, achieving 98%, 92%, and 84% accuracy in defect detection, crack recognition, and crack classification, respectively. Zhang et al. [13] proposed a lightweight model combining DCGAN and MobileNet to address the issue of sample imbalance effectively. Patel et al. [14] developed a hybrid Convolutional Stacked Sparse Self-Encoder with a Dual Attention Mechanism (HAT-CS2E), enabling accurate classification of multiple welding defect types.

With the rapid advancement of the YOLO series of object detection algorithms, weld defect detection methods based on these models have become increasingly mature and widely adopted. Pan et al. [15] proposed the WD-YOLO model, which integrates the NeXt backbone network with a dual-attention mechanism, achieving a detection accuracy of 92.6% mAP@0.5 and a processing speed of 98 FPS on weld X-ray images. Feng et al. [16] developed the YOLO-Weld model for laser weld spot detection by introducing a diverse category normalized loss function and an adaptive hierarchical intersection-parallel ratio loss function, resulting in 15.6% and 15.8% improvements in mAP50 and mAP50:95 metrics, respectively. Li et al. [17] proposed the lightweight multiscale dynamic attention network SFW-YOLO, which incorporates the P2 layer and DyHead module to enhance the detection of small-scale defects. Chen et al. [18] designed the DSF-YOLO framework, which significantly improves the recognition of fuzzy-boundary defects through a dynamic staged fusion feature extraction module and a dual multiscale feature fusion strategy. Xu et al. [19] introduced an improved YOLOv7 model incorporating a coordinate attention mechanism and SIoU loss function, achieving outstanding performance in pipeline weld surface defect detection. With the release of YOLOv11 in 2024, the introduction of the C3k2 block and C2PSA attention mechanism has further enhanced its feature extraction capability and computational efficiency [20].

Although target detection-based weld defect detection methods have achieved notable improvements in accuracy and efficiency, several challenges persist in practical applications. First, weld defects exhibit significant scale variation and morphological complexity—small defects are often overlooked, while blurred boundaries in irregularly shaped defects lead to inaccurate localization. Second, the complex backgrounds of weld images and the low contrast between defects and their surroundings hinder effective feature extraction. Third, the imbalance in sample distribution across defect categories reduces the model's ability to identify minority defect types accurately. Finally, most existing methods involve high computational complexity to maintain accuracy, limiting their suitability for real-time industrial inspection.

To address the aforementioned challenges, we propose ADFE-DET, an adaptive dynamic feature enhancement algorithm specifically designed for weld defect detection. The

algorithm incorporates several welding-specific design considerations: the edge-preserving attention mechanism targets the sharp boundaries characteristic of weld discontinuities, the multi-directional differential convolution captures the linear and directional nature of common weld defects (cracks, lack of fusion), and the direction-aware attention modeling addresses the anisotropic characteristics of welding-induced defects. The algorithm's key contributions include:

(1) Dynamic Selection Cross-stage Cascade Feature Block (DSCFBlock): This module effectively captures subtle texture variations and geometric morphology features on weld surfaces by incorporating an edge-preserving dynamic selective attention mechanism and a cross-cascade residual connection strategy. It significantly enhances the precision and spatial localization of multi-scale defect features.

(2) Adaptive Hierarchical Spatial Feature Pyramid Network (AHSFPN): By integrating the Direction-Aware Channel Attention (DACA) mechanism and the Hierarchical Spatial Feature Fusion (HSFF) strategy, this network enables adaptive weighting and efficient fusion of multi-scale features, thereby improving the perception of defects across scales—particularly enhancing the detection of small-size defects.

Multi-Directional Differential Lightweight Head (MDDLH): This head employs multi-directional differential convolution and a parameter-sharing mechanism to accurately locate and classify various weld defect types while maintaining a lightweight architecture, achieving an optimal balance between detection accuracy and computational efficiency.

2. Related Work

YOLOv11, the latest version of the YOLO series released by the Ultralytics team in 2024, marks a new milestone in the accuracy and efficiency of single-stage object detection algorithms. Building on the real-time detection advantages of its predecessors, YOLOv11 introduces several architectural innovations, including the C3k2 module, Spatial Pyramid Pooling Fast (SPPF), and Cross Stage Partial with Spatial Attention (C2PSA), which collectively enhance feature extraction capability and multi-scale target recognition accuracy.

The network adopts the classic Backbone–Neck–Head three-stage architecture. The backbone performs hierarchical feature extraction through multilayer convolutional modules and C3k2 blocks, generating feature maps of varying resolutions (P1–P5). As a core innovation, the C3k2 block replaces the C2f block used in YOLOv8 with a smaller convolutional kernel (kernel size = 2), significantly reducing computational complexity while maintaining detection performance through the Cross Stage Partial (CSP) design. The SPPF module, positioned at the end of the backbone, effectively extracts multi-scale contextual information via multi-scale max pooling, providing rich semantic representations for subsequent feature fusion and target detection.

The key innovation of YOLOv11 lies in the introduction of the C2PSA module, which performs cross-scale pixel spatial attention operations following the SPPF module. This partial spatial attention mechanism enables the model to adaptively focus on critical regions within an image, substantially improving the detection of small and occluded targets. The Neck employs a hybrid architecture combining the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN), achieving efficient multi-scale feature aggregation through up-sampling, concatenation, and C3k2 blocks. This design generates three feature maps at different scales (feature 1, feature 2, and feature 3) for detecting large, medium, and small targets, respectively. The detection head adopts a decoupled design consisting of a distribution focal loss (DFL) branch, a bounding box regression branch, and a classification branch, enabling direct end-to-end output of object category and location information. Experimental results demonstrate that YOLOv11 achieves higher mAP and faster inference speed on the COCO dataset than YOLOv8 and YOLOv10, particularly for

small objects. YOLOv11 is selected as the baseline model in this study for three reasons: (1) its C2PSA attention mechanism aligns closely with the adaptive feature enhancement concept proposed herein, providing a solid foundation for further optimization; (2) the lightweight C3k2 block offers architectural advantages for developing an efficient defect detection model; and (3) its proven high performance and deployment efficiency across multiple public datasets make YOLOv11 an ideal starting point for industrial weld defect detection applications.

3. Methods

The ADFE-DET weld defect detection algorithm proposed in this study comprises three core innovative modules, Dynamic Selection Cross-stage Cascade Feature Block (DSCFBlock), Adaptive Hierarchical Spatial Feature Pyramid Network (AHSFPN), and Multi-Directional Differential Lightweight Head (MDDLH), as illustrated in Figure 1.

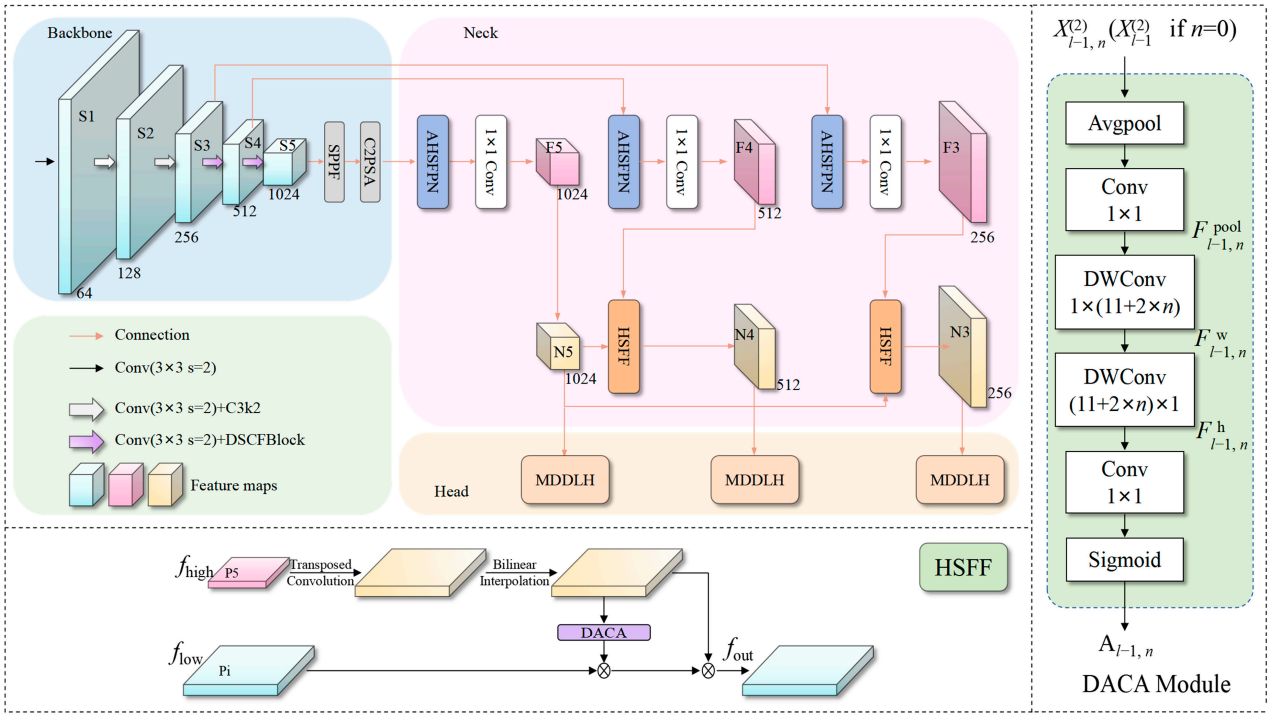


Figure 1. Overall framework of the ADFE-DET algorithm and architecture of the AHSFPN module.

The algorithm first extracts deep features from the input weld image using the DSCF-Block, which leverages an edge-preserving dynamic selective attention mechanism and a cross-cascade residual connection strategy to effectively capture subtle texture variations and geometric features on the weld surface. Next, the AHSFPN module performs adaptive weighting and multi-scale feature fusion by integrating a direction-aware channel attention mechanism with a hierarchical feature fusion strategy, substantially enhancing the perception of defect targets across scales. Finally, the MDDLH module employs multi-directional differential convolution and a parameter-sharing mechanism to accurately localize and classify various weld defect types while maintaining a lightweight network architecture, enabling high-precision and efficient weld seam quality inspection.

3.1. DSCFBlock Module

The traditional C3k2 module primarily relies on local convolution operations for feature extraction, which limits its ability to capture long-range spatial dependencies on weld surfaces with complex geometries and varying lighting conditions, resulting in

insufficient perception of the global context of defect regions. Additionally, C3k2 lacks an adaptive feature weighting mechanism, making it prone to feature redundancy when processing weld defects of different scales and types, and reducing sensitivity to minor defects. To address these limitations, this study proposes the DSCFBlock feature-extraction module, which substantially enhances the network's ability to perceive and discriminate weld defect features accurately. By incorporating a single-head self-attention mechanism and an edge-preserving global optimization strategy, DSCFBlock effectively overcomes the shortcomings of traditional methods, including insufficient feature extraction and limited characterization capability in complex industrial environments. The structure of DSCFBlock is illustrated in Figure 2.

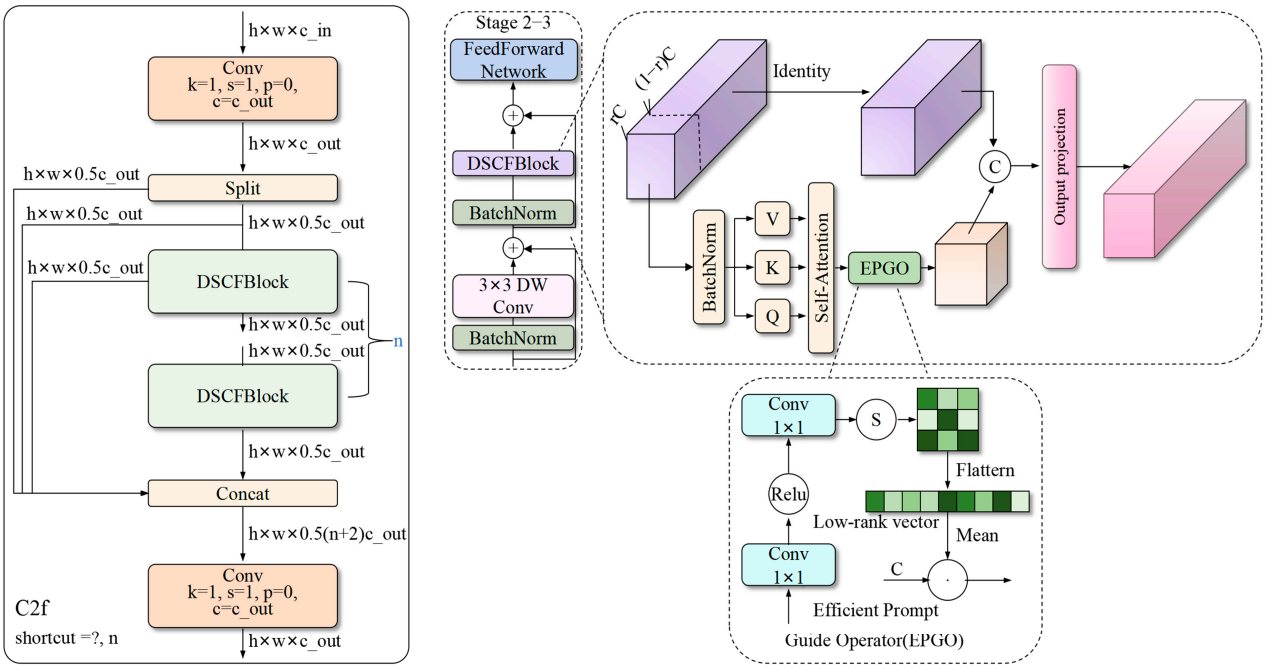


Figure 2. Structure of the DSCFBlock module: (**a**) overall architecture of DSCFBlock; (**b**) detailed design of the EPDS module.

The DSCFBlock module employs a multi-branch parallel processing architecture that integrates the Cross Stage Partial (CSP) design with an advanced self-attention mechanism to achieve efficient feature extraction and representation learning. The workflow begins with channel-wise decomposition of the input feature map, followed by mapping the original features into a higher-dimensional space using a 1×1 convolution. The feature map is then partitioned into multiple parallel branches for independent processing. Within each branch, cascaded Edge-Preserving Dynamic Selection (EPDS) units perform deep transformations on the grouped features, capturing hierarchical semantic information and spatial relationships at different levels. Finally, the outputs from all branches are fused via a feature-cascade operation, and the resulting feature representation is refined via a dimensionality-reduction convolution. The mathematical formulation of this module is expressed as:

$$Y = F_{conv}^{(2)}\left(\text{Concat}\left(\left[y_0, y_1\right] \cup \left\{F_{EPDS}^{(i)}\left(y_{i-1}\right) | i = 1, 2, \ldots, n\right\}\right)\right)$$

Here, y_0, y_1 denote the base feature groups obtained from the initial convolutional $F_{conv}^{(1)}$ segmentation, $F_{EPDS}^{(i)}$ represents the i-th EPDS Module processing unit, n is the total

number of processing units, and $F_{conv}^{(2)}$ corresponds to the final feature fusion convolution operation.

The EPDS Module employs a three-layer cascaded architecture to achieve multi-dimensional enhancement and optimization of input features. First, local spatial features are extracted using a depthwise separable convolutional layer, effectively capturing fine texture and edge details on the weld surface. Next, the Dynamic K-Selection Attention (DKSA) mechanism models long-range spatial dependencies, improving the network's perception and localization of critical defect regions. Finally, feature transformation and nonlinear mapping are performed via a feed-forward network structured as a multilayer perceptron (MLP), further enriching the feature representation. The complete forward propagation process is mathematically expressed as:

$$Z = X + F_{FFN}(X + F_{DKSA}(X + F_{DWConv}(X)))$$

In this expression, F_{DWConv} represents the depthwise separable convolution operation, which efficiently extracts local features through a combination of grouped and pointwise convolutions; F_{DKSA} denotes the single-head self-attention mechanism proposed in this study, responsible for modeling global spatial relationships; and F_{FFN} corresponds to the feed-forward network, which performs nonlinear transformations and feature dimensionality adjustments.

The DKSA (Single-Head Self-Attention with Edge-Preserving Global Optimization) mechanism represents the core technological innovation of this study. By integrating single-head attention computation with a dynamic sparsification strategy, it significantly enhances feature representation while maintaining computational efficiency and discriminative capability. The workflow begins with intelligent channel-wise segmentation of the input feature map, decomposing it into a central part for attention computation and an auxiliary part that directly participates in feature fusion. The central feature portion is group-normalized and then transformed linearly to generate Query (Q), Key (K), and Value (V) vectors, ensuring numerical stability during attention computation. During attention weight calculation, a novel dynamic K-selection strategy is employed to adaptively determine the relevant neighborhood for each spatial location via a gating network, achieving intelligent sparsification of attention weights. The generation of the query, key, and value vectors can be expressed mathematically as:

$$Q, K, V = \text{Split}\left(W_{qkv} \cdot F_{GroupNorm}(X_1), \left[d_q, d_k, d_p\right]\right)$$

Here, $W_{qkv} \in R^{(2d_q+d_p)\times C_p}$ denotes the learnable linear transformation matrix, $F_{GroupNorm}$ represents the group normalization operation, and d_q and d_p correspond to the query/key dimension and projection dimension, respectively. The attention weight matrix is computed under the constraints of the dynamic sparse mask as follows:

$$A = \text{Softmax}\left(M_{dynamic} \odot \frac{Q^T K}{\sqrt{d_q}} + \left(1 - M_{dynamic}\right) \odot (-\infty)\right)$$

In this formulation, $M_{dynamic}$ denotes the dynamically generated binary mask matrix, $\odot$ represents element-wise multiplication, and the temperature parameter $\sqrt{d_q}$ controls the sharpness of the attention distribution. The generation of the dynamic sparse mask, which embodies the core innovation of this method, is calculated as follows:

$$M_{dynamic} = I_{TopK}\left(Q^T K, k^*\right), \quad k^* = H \cdot W \cdot \sigma(F_{Gate}(X))$$

Here, I_{TopK} denotes the Top-K selection operation, F_{Gate} represents the gating network, which consists of two convolutional layers and a nonlinear activation function, σ is the Sigmoid activation function, and $H \cdot W$ corresponds to the spatial dimension of the feature map. The final feature fusion output is obtained by integrating the attention-weighted features with the original auxiliary features as follows:

$$Y = F_{Proj}\left(\text{Concat}\left[\text{Reshape}\left(VA^{T}, (B, d_p, H, W)\right), X_2\right]\right)$$

Here, F_{Proj} denotes the feature projection function, which consists of a SiLU activation and a convolutional layer, X_2 represents the auxiliary feature branch, and corresponds to the batch size.

By systematically incorporating the DSCFBlock feature extraction module, the improved YOLOv11 algorithm proposed in this study achieves a significant performance leap in weld defect detection. By integrating a single-head self-attention mechanism, a dynamic sparsification strategy, and edge-preserving global optimization, this innovative module not only substantially enhances the network's ability to capture multi-scale defect features and accurately localize spatial locations but also markedly improves model robustness and generalization in complex industrial environments.

3.2. AHSFPN Module

The traditional Path Aggregation Feature Pyramid Network (PAFPN) employs a simple feature-cascading approach for multi-scale information fusion, lacking the ability to evaluate feature importance across different levels adaptively. This limitation reduces its effectiveness in suppressing redundant information and highlighting critical features during fusion. Additionally, when handling multi-scale targets, PAFPN's fixed-weight fusion strategy cannot dynamically adapt to the scale characteristics of different defect types, leading to a gradual loss of small-defect information during deep feature propagation. To overcome these limitations, this study proposes the Adaptive Hierarchical Spatial Feature Pyramid Network (AHSFPN), which integrates a channel–spatial attention mechanism. By employing adaptive attention-weight allocation and hierarchical spatial-feature fusion, AHSFPN significantly enhances the accuracy and robustness of weld defect detection, particularly improving the detection of small targets and suppressing complex background interference.

The AHSFPN network employs a hierarchical bi-directional feature fusion strategy, constructing a multi-scale feature fusion framework with adaptive weight allocation by integrating the Directional Adaptive Channel Attention (DACA) module and the Hierarchical Spatial Feature Fusion (HSFF) module at different backbone layers. The core innovation of AHSFPN lies in its attention-guided hierarchical feature propagation mechanism, in which high-level semantic features are weighted and fused with attention-enhanced low-level features, followed by spatial resolution recovery via transposed convolution. The mathematical formulation of the complete feature fusion process is:

$$F_{fused}^{(i)} = \text{HSFF}\left(\text{DACA}\left(U\left(F_{high}^{(i+1)}\right)\right) \odot \text{DACA}\left(F_{low}^{(i)}\right) + \alpha \cdot U\left(F_{high}^{(i+1)}\right)\right)$$

Here, $F_{high}^{(i+1)}$ and $F_{low}^{(i)}$ denote the feature maps of layers $i + 1$ and i, respectively; $U(\cdot)$ represents the transposed convolutional upsampling operation; $\odot$ denotes the Hadamard product; and is the residual weight coefficient. To further optimize feature consistency and

semantic coherence across layers, the network employs a multi-scale feature alignment loss function:

$$L_{HSFPN} = \sum_{i=1}^{N} \omega_i L_{det}\left(F_{fused}^{(i)}, Y^{(i)}\right) + \lambda \sum_{i=1}^{N-1} \| \text{Align}\left(F_{fused}^{(i)}\right) - \text{Align}\left(F_{fused}^{(i+1)}\right) \|_2^2$$

Here, L_{det} denotes the detection loss function, $Y^{(i)}$ represents the true label of layer i, Align$(\cdot)$ corresponds to the feature alignment function, and ω_i and λ are the layer weights and consistency regularization coefficients, respectively.

The DACA module enhances the input feature map in multiple dimensions by integrating global pooling, direction-separation convolution, and adaptive activation mechanisms. Its design leverages the spatial distribution and orientation characteristics of weld defects, capturing linear defect features by modeling spatial dependencies along horizontal and vertical directions. Specifically, DACA first extracts global contextual information via adaptive average pooling, then encodes horizontal and vertical spatial features using depthwise separable convolution kernels of appropriate dimensions, and finally generates an adaptive attention weight map through cascaded convolutional layers followed by a sigmoid activation. The complete mathematical formulation of the DACA module is expressed as:

$$\text{DACA}(X) = \sigma\big(\text{Conv}_{1\times1}\big(\text{DWConv}_{k_v\times1}\big(\text{DWConv}_{1\times k_h}(\text{Conv}_{1\times1}(\text{GAP}(X)))\big)\big)\big) \odot X$$

Here, $X \in R^{C\times H\times W}$ denotes the input feature map, GAP$(\cdot)$ represents the global adaptive average pooling operation, $\text{DWConv}_{k_v\times1}$ and $\text{DWConv}_{1\times k_h}$ correspond to the depthwise separable convolutions in the vertical and horizontal directions, respectively, $\text{Conv}_{1\times1}(\cdot)$ denotes the 1×1 convolution for channel dimension transformation, $\sigma(\cdot)$ is the sigmoid activation function, and $\odot$ represents element-wise multiplication.

The HSFF module is responsible for effective integration and information transfer across multi-scale feature maps. It employs a two-branch parallel processing architecture: the main branch evaluates the importance of fused features through an attention weighting mechanism, while the auxiliary branch preserves the integrity of the original high-level semantic information via residual concatenation. The module first performs element-wise multiplication of the attention-weighted high- and low-level features to emphasize regions of common interest, then fuses the resampled high-level features via transposed convolution using an additive operation, and finally refines features and adjusts dimensions through the C3k2 module. The mathematical formulation of the HSFF module is given by:

$$\text{HSFF}(F_h, F_l) = \text{C3k2}(\text{DACA}(U(F_h)) \odot \text{DACA}(F_l) + \beta \cdot U(F_h))$$

Here, $F_h \in R^{C_h\times H_h\times W_h}$ and $F_l \in R^{C_l\times H_l\times W_l}$ denote the high-level and low-level input feature maps, respectively; $U(\cdot)$ represents the bilinear interpolation upsampling operation; and $U(F_h) \in R^{C_l\times H_l\times W_l}$ and β are the learnable residual weight parameters.

The AHSFPN network significantly enhances overall weld defect detection performance while preserving computational efficiency by integrating a channel–spatial dual-attention mechanism with a hierarchical spatial feature fusion strategy. It addresses the limitations of traditional PAFPN, including suboptimal weight allocation and loss of spatial information during multi-scale feature fusion. The DACA module enables adaptive focus on defects across different directions and scales. In contrast, the hierarchical fusion strategy of the HSFF module effectively combines high-level semantic information with low-level detail features, maintaining the network's ability to detect minor defects without compromising recognition of significant defects.

3.3. MDDLH Module

The conventional YOLOv11 detection head relies on standard convolutional operations for feature extraction, which capture spatial correlations only within a fixed receptive field and fail to effectively perceive multi-directional edge information, subtle texture variations, and irregular geometries commonly present in weld defects. Additionally, traditional approaches lack adaptive assessment of feature importance during multi-scale fusion, preventing dynamic adjustment of each feature level's contribution based on defect type and scale. This often leads to dilution of practical features and amplification of redundant information, adversely affecting detection accuracy and generalization. To address these challenges, we propose the MDDLH module. As shown in Figure 3, MDDLH substantially improves the network's perception of multi-dimensional weld defect features by integrating a multi-directional differential convolution operator group with an adaptive feature weight adjustment mechanism, while controlling computational complexity through parameter-sharing and reparameterization techniques.

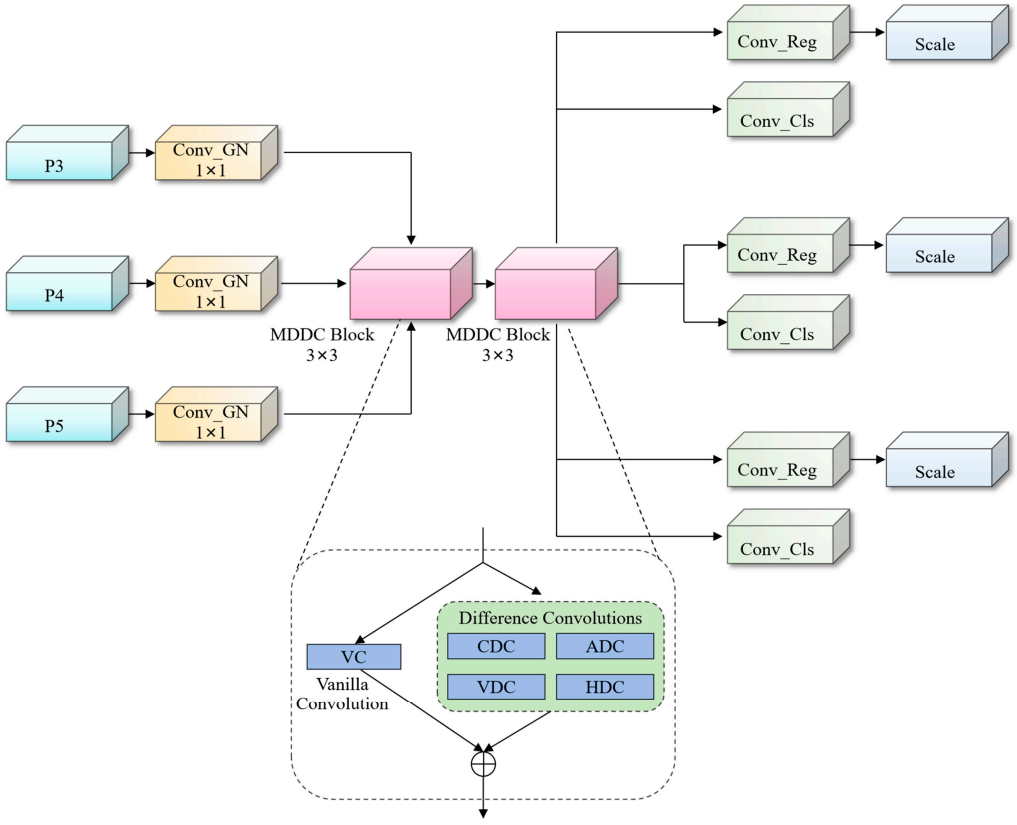


Figure 3. Structure of the MDDLH detection head and multi-directional differential convolution.

The MDDLH detection head enables efficient multi-scale feature extraction, enhancement, and fusion through a shared convolutional architecture. Its core workflow consists of three key phases: feature dimension unification, shared detail enhancement, and adaptive output generation, each tailored to the specific requirements of weld defect detection. In the feature dimension unification stage, multi-scale features from different backbone layers are first projected into a unified hidden feature space via independent 1×1 convolution operations. This not only standardizes channel dimensions but also establishes a consistent representation for subsequent shared processing. The mathematical formulation of this stage is expressed as:

$$H_i = F_{conv}(X_i; \Theta_i) = SiLU(GN(W_i * X_i + b_i))$$

Here, $X_i \in R^{C_i \times H_i \times W_i}$ denotes the input feature map from the i-th feature layer, $\Theta_i = \{W_i, b_i\}$ represents the corresponding learnable parameters, $H_i \in R^{D \times H_i \times W_i}$ is the feature representation after dimensionality unification, and D corresponds to the preset hidden layer dimension. In the shared detail enhancement stage, feature maps across all scales are processed synchronously through a two-layer Multi-Directional Differential Convolution (MDDC) Block in series. In the final adaptive output generation stage, the regression branches are differentially adjusted using learnable scale factors, with the process formalized mathematically as:

$$\begin{cases} Y_{reg}^{(i)} = \sigma_i \odot W_{reg} * F_i^{(2)} + b_{reg} \\ Y_{cls}^{(i)} = W_{cls} * F_i^{(2)} + b_{cls} \end{cases}$$

Here, $F_i^{(2)}$ denotes the feature map after processing through the two MDDC Blocks, $\sigma_i \in R^{4 \times reg_max}$ is the learnable scaling parameter tensor corresponding to the i-th scale layer, and $\odot$ represents the Hadamard accumulation operation.

The MDDC Block achieves multi-directional enhancement of weld defect edge features by integrating multiple differential convolution operators. Specifically, it combines five feature extraction operators—Center Difference, Horizontal Difference, Vertical Difference, Anti-Diagonal Difference, and Standard Convolution—each with distinct orientation sensitivity and feature response characteristics. The Center Difference operator emphasizes local contrast changes by computing the difference between the center pixel and the mean of its eight neighbors; the Horizontal and Vertical Difference operators enhance edge information along their respective directions; the Anti-Diagonal Difference operator targets diagonal edge features; and the Standard Convolution preserves the original spatial correlations. During training, these operators process input features in parallel, and their outputs are fused via element-wise addition. During inference, the weights of all operators are merged into a single equivalent convolutional kernel via a reparameterization, substantially reducing computational complexity while retaining feature-extraction capability. The complete operation of the MDDC Block can be expressed as:

$$F_{out} = SiLU\left(GN\left(\sum_{k \in K} W_k^{(d)} * F_{in} + \sum_{k \in K} b_k^{(d)}\right)\right)$$

Here, $K = \{cd, hd, vd, ad, std\}$ denotes the set of five differential operators, $W_k^{(d)}$ and $b_k^{(d)}$ are the weight tensor and bias vector of the k-th operator at layer d, respectively, $GN(\cdot)$ represents the group normalization operation to enhance training stability, and $SiLU(\cdot)$ is the Swish activation function.

The MDDLH module substantially enhances weld defect detection performance while maintaining a lightweight network architecture by combining multi-directional differential convolution with an adaptive feature scaling mechanism. Its core advantage lies in precise, detailed feature capture and computational efficiency. The multi-directional differential convolution design enables accurate recognition of edge contours and directional features across various defect types. Moreover, the shared convolutional architecture not only reduces model parameters and computational complexity but also strengthens semantic consistency across feature scales via a unified feature transformation space, effectively mitigating feature conflicts and information redundancy commonly encountered in multi-scale detection tasks.

4. Experimentation

4.1. Introduction to Public Datasets

To comprehensively evaluate the effectiveness and robustness of the proposed ADFE-DET algorithm across different defect detection tasks, this study selected three representative public datasets for experimental validation.

The Weld-DET weld defect dataset [21] is constructed explicitly for weld defect detection and comprises images of welded seams captured in real industrial environments. It contains 1000 high-resolution images covering three common defect types: porosity, spatter, and cracks. The dataset is split into training, validation, and test sets at a 7:2:1 ratio, with 700 images for training, 200 for validation, and 100 for testing, ensuring a scientifically rigorous framework for model training and evaluation.

The NEU-DET Steel Surface Defect Dataset [22] is an industrial benchmark released by Northeastern University, designed for the identification and localization of surface defects on hot-rolled steel strips. It comprises 1800 authentic steel surface images covering six typical defect types: crazing (Cr), inclusion (In), patches (Pa), pitted surface (Ps), rolled-in scale (Rs), and scratches (Sc). The dataset is randomly split into training, validation, and test sets containing 1260, 270, and 270 images, respectively, with each defect type maintaining a relatively balanced distribution across the subsets.

Peking University created the PKU-Market-PCB defect-detection dataset [23] to evaluate defect-detection algorithms in PCB manufacturing. It comprises 2151 high-resolution PCB images covering six common defect types: missing hole (MH), mouse bite (MB), open circuit (OC), short circuit (SC), spur (SPR), and spurious copper (SPC). The dataset is split into 1944 images for training, 139 for validation, and 68 for testing, providing sufficient samples for algorithm training, parameter tuning, and performance evaluation.

4.2. Experimental Environment and Parameter Settings

All experiments were performed on a high-performance workstation equipped with an Intel® Xeon® Platinum 8352V CPU (Intel Corporation, Santa Clara, CA, USA), 32 GB DDR4 memory, and an NVIDIA GeForce RTX 4090 GPU with 24 GB of video memory (NVIDIA Corporation, Santa Clara, CA, USA). The software environment consisted of Ubuntu 22.04, Python 3.10, PyTorch 2.1.2, and CUDA 11.8, ensuring experimental reproducibility and reliability of results. The experimental code was developed based on the official YOLOv11 framework, incorporating the improved modules proposed in this study.

During model training, the number of epochs was set to 1000, the batch size to 32, and the initial learning rate (lr_0) to 0.01. Parameter updates were performed using stochastic gradient descent (SGD), with a momentum coefficient of 0.937 and a weight decay (weight_decay) of 0.0005 to prevent overfitting. A cosine annealing learning rate schedule was employed to adjust the learning rate for improved convergence dynamically. Data augmentation strategies included random horizontal flipping, random rotation, color jittering, and mosaic enhancement to boost model generalization. All other hyperparameters were set to the default YOLOv11 settings to ensure experimental fairness and a reliable comparative evaluation.

4.3. Assessment of Indicators

In this study, widely recognized target-detection metrics were employed to evaluate the performance of the proposed algorithms comprehensively. Precision (P), representing the proportion of predicted positive samples that are genuinely positive, reflects the accuracy of detection results, while Recall (R), indicating the proportion of true positives correctly identified, measures the model's ability to detect defective targets. mAP@0.5, calculated at an Intersection over Union (IoU) threshold of 0.5, is the most commonly used

performance metric for target detection tasks, whereas mAP@0.5:0.95, computed as the mean average precision across IoU thresholds from 0.5 to 0.95 in increments of 0.05, provides a more rigorous assessment of detection accuracy. Additionally, computational complexity is quantified in GFLOPS (giga floating-point operations), storage requirements in Params (parameter count), and inference speed in FPS (frames per second). Evaluating these metrics together offers a balanced assessment of detection accuracy, computational efficiency, and real-time performance, providing a critical reference for industrial applications.

4.4. Ablation Experiments

4.4.1. Ablation Study on Convolutional Kernel Sizes in the DACA Attention Mechanism of AHSFPN

To evaluate the impact of the DACA mechanism in AHSFPN networks on weld defect detection performance under different convolution kernel configurations, systematic ablation experiments were conducted on the Weld-DET dataset. The horizontal and vertical convolution kernel sizes in the DACA attention mechanism were controlled by adjusting the n value (0–3), corresponding to kernel sizes of 11 × 11, 13 × 13, 15 × 15, and 17 × 17, respectively. The experimental results are summarized in Table 1.

Table 1. Ablation results of the DACA attention mechanism in AHSFPN with varying convolutional kernel sizes (Weld-DET dataset).

n Value	h_Kernel_Size	v_Kernel_Size	*p*	R	mAP50	mAP50-95
0	11	11	84.2	78.5	85.1	54.8
1 (ours)	13	13	86	81.3	87.6	56.9
2	15	15	85.4	80.7	86.8	55.7
3	17	17	84.8	79.9	86.2	55.2

The experimental results indicate that our proposed optimal configuration (n = 1, 13 × 13 convolutional kernel) achieves the best performance: compared with the baseline (n = 0, 11 × 11 kernel), mAP@0.5 and mAP@0.5:0.95 improve by 2.5% and 2.1%, respectively. Notably, further increasing the kernel size to 15 × 15 (n = 2) and 17 × 17 (n = 3) results in a performance decline, suggesting that a vast receptive field may introduce noise and degrade feature extraction accuracy. These experiments demonstrate that the proposed DACA attention mechanism maximizes weld defect feature extraction while maintaining computational efficiency under the 13 × 13 kernel configuration, confirming its effectiveness in direction-sensitive feature modeling and multi-scale spatial attention allocation.

4.4.2. Integral Module Ablation Experiments

To comprehensively evaluate the effectiveness and synergistic impact of the three core innovative modules in the ADFE-DET algorithm, a systematic ablation study was conducted. Using a control-variable approach, each module—DSCFBlock, AHSFPN, MDDLH—and the transfer learning strategy were sequentially integrated into the baseline YOLOv11 model. Detailed performance evaluations were carried out on the Weld-DET and NEU-DET datasets, with the results summarized in Table 2.

Table 2. Results of comprehensive model ablation experiments on the Weld-DET and NEU-DET datasets.

Datasets	YOLOv11	DSCFBlock	AHSFPN	MDDLH	Transfer Learning	GFLOPS	Params/M	P	R	mAP50	mAP50-95
Weld	√	-	-	-	-	6.5	2.63	84.4	79.6	86.5	55.7
	√	√	-	-	-	6.3	2.51	85.7	80.9	87.4	56.2
	√	√	√	-	-	6.4	1.91	86.0	81.8	87.6	56.9
	√	√	√	√	-	6.2	1.72	86.5	81.3	87.6	57.2
	√	√	√	√	√	6.2	1.72	86.1	81.9	88.7	57.4
NEU-DET	√	-	-	-	-	6.5	2.63	69.5	80.6	80.6	48.1
	√	√	-	-	-	6.3	2.51	73.8	76.3	81.5	48.9
	√	√	√	-	-	6.4	1.91	76.1	73.5	82.2	49.4
	√	√	√	√	-	6.2	1.72	74.2	74.8	82.7	49.7
	√	√	√	√	√	6.2	1.72	78.4	75.8	83.4	50.2

Specifically, integrating the DSCFBlock module improves mAP@0.5 on the Weld-DET dataset by 0.9% and mAP@0.5:0.95 by 0.5%, demonstrating the effectiveness of the single-head self-attention mechanism and the edge-preserving dynamic selection strategy in weld feature extraction. The addition of the AHSFPN module further boosts mAP@0.5 to 87.6% and mAP@0.5:0.95 to 56.9%, while reducing the number of parameters from 2.51 M to 1.91 M, highlighting the advantages of the direction-adaptive channel attention mechanism and hierarchical spatial feature fusion for multi-scale defect detection. The MDDLH module enhances computational efficiency and parameter utilization through multi-directional differential convolution and parameter-sharing, while maintaining detection accuracy, demonstrating the effectiveness of its lightweight design. Finally, incorporating the transfer learning strategy allows the complete ADFE-DET algorithm to achieve optimal performance: 88.7% mAP@0.5 and 57.4% mAP@0.5:0.95 on the Weld-DET dataset, representing improvements of 2.2% and 1.7% over the baseline YOLOv11. On the NEU-DET dataset, the model attains 83.4% mAP@0.5 and 50.2% mAP@0.5:0.95, improving 2.8% and 2.1% relative to the baseline, fully demonstrating the superior performance and generalization capability of the proposed multi-module co-optimization strategy across different industrial defect detection tasks.

4.5. Comparative Experiments

4.5.1. Performance Comparison of Different C3k2 Variant Modules

To evaluate the superiority of the proposed DSCFBlock module over existing state-of-the-art C3k2 variants, a systematic comparative study was conducted, as summarized in Table 3. The experiments included several mainstream C3k2 improvement methods as baselines, such as C3k2-SWC (Sliding Window Convolution), C3k2-iRMB (Inverse Residual Moving Block), C3k2-Star (Star Convolution), C3k2-LFEM (Lightweight Feature Enhancement Module), and C3k2-EfficientVIM (Efficient Visual Attention Module). All comparisons were performed under identical experimental conditions and parameter settings, employing a consistent training strategy and evaluation metrics. The results of these experiments are presented in Table 3.

Table 3. Performance comparison of various C3k2 variant modules on the Weld-DET dataset.

Model	GFLOPs	Params(M)	P	R	mAP50	mAP50-95
C3k2 (base)	6.5	2.63	84.4	79.6	86.5	55.7
C3k2-SWC [24]	8.9	2.82	84.9	80.2	87.1	55.9
C3k2-iRMB [25]	6.2	2.46	84.6	79.7	86.4	55.5
C3k2-Star [26]	6.5	2.50	84.5	79.5	86.3	55.4
C3k2-LFEM [27]	9.2	2.67	85.1	80.4	87.2	56.0
C3k2-EfficientVIM [28]	6.8	2.50	84.7	79.8	86.6	55.6
DSCFBlock (ours)	6.3	2.51	85.7	80.9	87.4	56.2

The comparison results clearly demonstrate the substantial advantages and technological advancements of the DSCFBlock module. Quantitative analysis indicates that although C3k2-LFEM achieves a high mAP50 of 87.2%, its computational overhead is significantly increased; similarly, C3k2-SWC reaches 87.1% in mAP50 but incurs 8.9 GFLOPs, reflecting poor computational efficiency. In contrast, the proposed DSCFBlock achieves an mAP50 that is only 0.2% lower than the highest-performing baseline, while maintaining relatively low computational complexity and a moderate parameter count, and it improves mAP50-95 by 0.5% over the standard C3k2. These results confirm that DSCFBlock effectively balances detection accuracy and computational efficiency by integrating a single-head self-attention mechanism and an edge-preserving dynamic selection strategy, highlighting its superior performance and practical value in weld defect detection tasks.

4.5.2. Comprehensive Comparison with Mainstream SOTA Methods

To comprehensively assess the detection performance and computational efficiency of the proposed ADFE-DET model, comparison experiments were conducted against current mainstream target detection architectures on the Weld-DET dataset, with results summarized in Table 4. Quantitative analysis shows that ADFE-DET achieves substantial improvements across multiple evaluation metrics while maintaining optimal model complexity. Specifically, compared with the baseline YOLOv11n, ADFE-DET improves mAP50 by 2.16%, reduces the parameter count by 34.1%, and lowers computational complexity by 4.6%, demonstrating the effectiveness of the adaptive feature enhancement module and the decoupled detection head in capturing fine-grained defect features. Comparisons with other lightweight architectures indicate that ADFE-DET achieves the best accuracy-efficiency trade-off: mAP50 is increased by 2.86% over YOLOv12n with similar computational complexity, and by 2.66% over Mamba-YOLO while reducing the number of parameters by 70.3%. Although its inference speed is slightly lower than that of the ultra-lightweight YOLOv12n (142 FPS), it still exceeds the real-time industrial requirement (>30 FPS). Overall, these results confirm that ADFE-DET achieves an optimal balance among detection accuracy, model compactness, and inference efficiency, highlighting its technological advancement and practical applicability in industrial weld defect detection.

Table 4. Performance comparison with state-of-the-art (SOTA) methods on the Weld-DET dataset.

Model	GFLOPs	Params (M)	P	R	mAP50	mAP50-95	FPS
YOLOv5n [29]	7.1	2.5	82.61	77.98	84.63	53.85	132
YOLOv8n [30]	8.1	2.9	84.21	79.78	86.93	56.05	98
YOLOv9t [31]	7.7	2.1	83.11	77.58	84.53	53.45	102
YOLOv10n [32]	6.5	2.3	82.21	76.88	84.03	52.95	113
YOLOv12n [33]	5.8	2.5	83.41	79.18	85.83	55.05	125
YOLOv13n [34]	6.1	2.4	82.71	78.28	85.13	53.75	118
Hyper-yolo-t [35]	9.5	3.6	83.01	78.68	85.63	54.95	88
Mamba-yolo [36]	13.2	5.8	83.51	79.08	86.03	54.85	72
YOLOv11n (base)	6.5	2.61	84.41	79.58	86.53	55.65	112
ADFE-DET (ours)	6.2	1.72	85.13	81.87	88.69	57.43	105

To further validate the generalization ability and robustness of the proposed ADFE-DET model across different industrial defect detection scenarios, cross-dataset evaluation experiments were conducted on the NEU-DET steel surface defects dataset, with results presented in Table 5. The NEU-DET dataset contains six defect types—Cr (cracks), In (inclusions), Pa (plaques), Ps (pitting), Rs (indented oxides), and Sc (scratches)—providing a challenging benchmark for assessing model transferability. Experimental results indicate that ADFE-DET achieves an impressive mAP50 of 83.35%, representing a 2.73% improvement over the baseline YOLOv11n, while maintaining a compact architecture with a 34.1% reduction in parameters and a 4.6% reduction in computational complexity. The model exhibits balanced detection performance across all defect classes, particularly on challenging courses such as Cr, Pa, and Ps, demonstrating the effectiveness of the adaptive feature enhancement mechanism in capturing diverse defect patterns. Compared with other state-of-the-art methods, ADFE-DET achieves competitive accuracy with optimal parameter efficiency: a 1.05% mAP50 improvement over DDN, and substantial gains of 9.75%, 9.55%, and 5.55% compared to lightweight models YOLOv5n (73.6%), YOLOv10n, and YOLOv8n, respectively. Furthermore, it consistently outperforms dedicated architectures such as REDef-DETR and DCA_RFCN, demonstrating strong generalization across heterogeneous defect detection tasks. These cross-dataset results fully demonstrate that ADFE-DET maintains robust performance and excellent transferability, validating its practical applicability in diverse industrial defect inspection scenarios.

Table 5. Performance comparison with mainstream SOTA methods on NEU-DET dataset.

Method	mAP50	Cr	In	Pa	Ps	Rs	Sc	GFLOPS	Params/M
SCN-CNN [37]	79.89	59.39	83.44	91.14	82.45	67.95	95	-	-
YOLOv5n [29]	73.6	42.4	82	90	77.8	59.4	89.9	7.1	2.5
RSTD-YOLOv7 [38]	79.3	51.2	78.1	92.7	92.4	67.2	94	-	16.5
REDef-DETR [39]	80.6	40.5	87.7	92.6	99.5	75.2	88.2	57	20.4
DDN [40]	82.3	62.4	84.7	90.7	89.7	76.3	90.1	-	-
EFD-YOLOv4 [41]	79.8	45.6	85.3	97.0	84.9	72.6	93.5	-	-
MSD-YOLO [42]	80.9	58.2	82.3	95.7	85.6	75.9	87.9	-	-
YOLOv10n [32]	73.8	41.9	84.7	94.6	82.5	54	85.1	6.5	2.3
YOLOv8n [30]	77.8	44.4	81.7	94.8	82.1	70.5	93.6	8.1	2.9
YOLOv12n [33]	79.9	42.9	96.6	93.1	53	97.1	96.6	5.8	2.5
YOLOv11n (base)	80.62	71.34	71.67	99.22	98.17	58.3	84.9	6.5	2.61
ADFE-DET (ours)	83.35	71.84	73.14	99.72	98.17	71.8	85.4	6.2	1.72

To comprehensively evaluate the effectiveness of ADFE-DET in printed circuit board (PCB) defect detection, extensive experiments were conducted on the PKU-Market-PCB dataset, with results summarized in Table 6. ADFE-DET achieves an outstanding mAP50 of 96.19%, representing a 1.81% improvement over the baseline YOLOv11n, while reducing the number of parameters by 24.6% and computational complexity by 6.2%. Detailed category analysis shows significant improvements in challenging defect types: Mb detection increased by 3.94%, SPR detection by 3.62%, and SC detection by 1.27%, demonstrating the effectiveness of the adaptive feature enhancement mechanism in capturing fine-grained PCB defect features. The model maintains consistently high accuracy across all categories, reaching 99.5% for MH detection and 99.48% for SPC detection. Compared to other state-of-the-art methods, ADFE-DET outperforms MFYOLO by 2.09% with 60% fewer parameters, and significantly surpasses SSD by 6.69% while maintaining efficient computation. These results confirm that ADFE-DET achieves an optimal balance between detection accuracy and model compactness, demonstrating strong applicability for precision-critical PCB defect inspection in industrial manufacturing scenarios.

Table 6. Performance comparison with mainstream SOTA methods on PKU-Market-PCB dataset.

Method	MH	MB	OC	SC	SPR	SPC	mAP50	Params/M	GFLOPS
MFYOLO [43]	98.6	99	99.2	96.5	98.8	99	94.1	5.68	15.1
SSD [44]	92.2	89.6	87.9	91.9	85.7	89.8	89.5	24.7	30.6
YOLOv5s [29]	98.9	98.2	98.2	99.5	96.7	99.2	94.4	7.06	16.3
YOLOv11n (base)	99.5	89.13	91.32	97.66	90.13	98.56	94.38	3.01	8.1
ADFE-DET (ours)	99.5	93.07	92.38	98.93	93.75	99.48	96.19	2.27	7.6

4.6. Visualization Analysis

4.6.1. Detection Results and Heatmap Visualization on the Weld-DET Dataset

To visualize the detection performance and feature characterization capabilities of ADFE-DET, qualitative experiments were conducted on three typical weld defect types—Crack, Spatter, and Porosity—with results shown in Figure 4. The left side of the figure displays the detection results with bounding boxes, while the right side shows the corresponding Grad-CAM heatmaps to illustrate the spatial attention distribution. Comparative analysis indicates that ADFE-DET achieves superior localization accuracy and feature discrimination across all defect types. For crack defects with long and narrow morphology, the model produces bounding boxes that tightly fit the defect, and the heatmaps show concentrated activation in the crack region with minimal background interference, demonstrating enhanced fine-grained feature extraction. For spatter defects with irregular distribution, the model effectively captures multi-scale spatial features, with attention weights uniformly covering the defect region. For minor, densely distributed porosity defects, ADFE-DET achieves high detection completeness, and the heatmaps reveal stronger, more focused feature responses than those of the YOLOv13 and YOLOv11 baselines. Overall, the adaptive feature enhancement module produces more discriminative feature representations, with clearer, more concentrated activation regions in Grad-CAM visualizations, validating the model's effectiveness in capturing critical defect features for accurate weld defect detection.

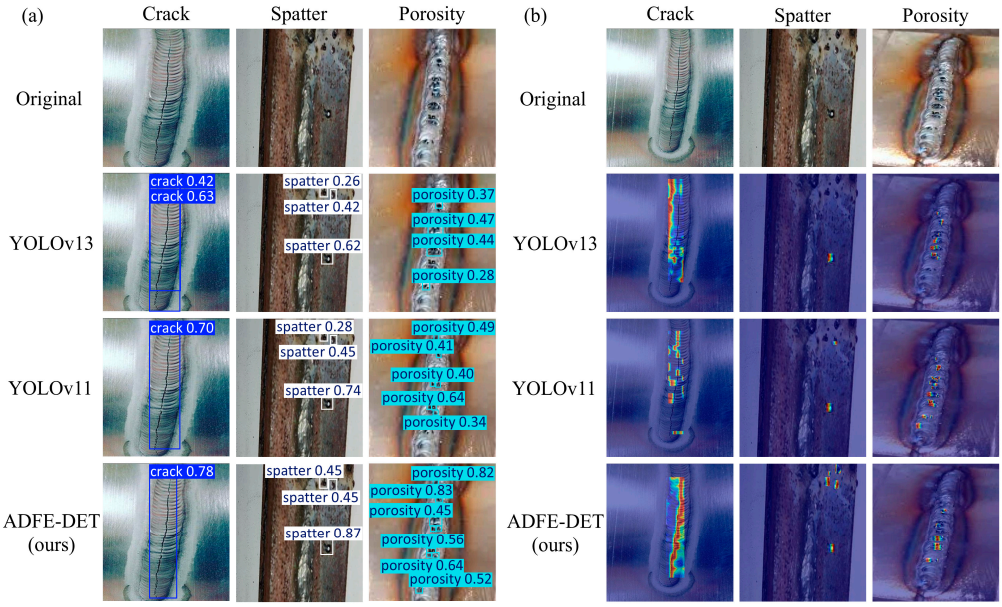


Figure 4. Visualization of detection results on the Weld-DET dataset: (**a**) Comparison of detection accuracy across models; (**b**) Grad-CAM heatmap visualization, where warmer colors (yellow to red) indicate regions of higher model attention and activation intensity, while cooler colors (blue to purple) represent areas of lower activation, effectively illustrating the spatial focus of each model during defect detection, Figures 6 and 8 also follow this principle.

4.6.2. Detection Results and Heatmap Visualization on the PKU-Market-PCB Dataset

To demonstrate the generalizability and robustness of the ADFE-DET algorithm beyond weld defect detection, we evaluate its performance on printed circuit board (PCB) defect detection. This cross-domain validation is crucial for weld defect detection research as it: (1) validates that the adaptive feature enhancement mechanisms are not overfitted to weld-specific patterns, (2) demonstrates the transferability of edge-preserving attention and multi-directional differential convolution to other fine-grained industrial defects, and (3) provides evidence that the algorithm can potentially be adapted for detecting defects in welded electronic components or hybrid manufacturing scenarios where both welding and PCB assembly occur.

Figure 5 presents the qualitative detection results of different models on representative PCB defect samples, further demonstrating the superior performance of ADFE-DET. The visualizations indicate that ADFE-DET achieves more accurate bounding box localization and higher confidence scores across all defect categories. For missing hole defects, the model produces tightly fitting bounding boxes with confidence scores of 0.79 and 0.73, accurately reflecting defect presence, whereas baseline methods exhibit missed detections or inaccurate localization. In mouse bite defects, ADFE-DET successfully identifies all instances with high confidence, highlighting its robust edge anomaly detection. For open circuit defects, the model demonstrates excellent accuracy with confidence scores of 0.83 and 0.78, precisely outlining discontinuous circuit alignments. Notably, in challenging cases such as spurious copper and short circuits, where multiple defects of varying scales coexist, ADFE-DET maintains consistent detection integrity with balanced confidence distributions, whereas SSD shows false positives and YOLOv11 produces lower confidence scores. Overall, these qualitative evaluations confirm that ADFE-DET achieves superior detection accuracy, precise localization, and robust confidence estimation, demonstrating its practical applicability for precision-critical PCB defect detection in industrial manufacturing scenarios.

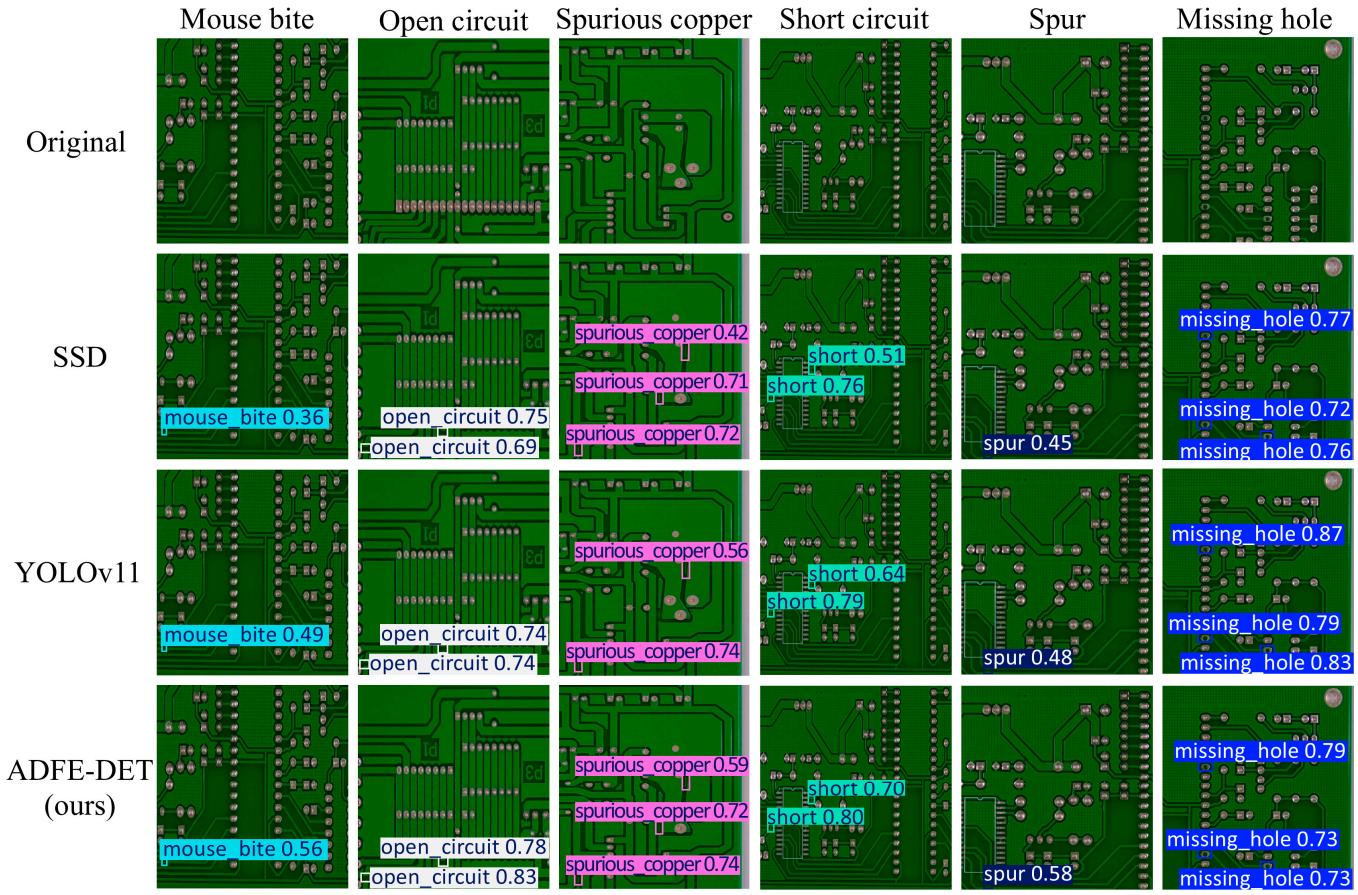


Figure 5. Comparison of detection results of different models on PKU-Market-PCB dataset.

Figure 6 illustrates the Grad-CAM visualizations for six types of PCB defects, highlighting the ADFE-DET's feature characterization and attention mechanism. The visualizations demonstrate that ADFE-DET exhibits superior feature discrimination and precise localization compared to SSD and YOLOv11 baseline models. For missing hole defects, characterized by circular voids, ADFE-DET shows a highly concentrated activation response with minimal background interference, indicating accurate defect localization. In mouse bite defects with irregular notched PCB edges, the thermograms reveal that attention is tightly focused on the true defect areas, whereas competing methods produce scattered activations in non-defective regions. For spur defects, which present subtle raised features, ADFE-DET captures fine-grained spatial details with more precise and more discriminative activation regions, validating the effectiveness of the adaptive feature enhancement module. Similarly, for short circuit, spurious copper, and open circuit defects, the model consistently produces focused and accurate thermogram responses, with activation strengths precisely aligned with defect boundaries. Overall, the qualitative analysis confirms that ADFE-DET achieves highly discriminative feature representation and an enhanced spatial attention mechanism, enabling accurate PCB defect detection even under complex backgrounds and subtle defect patterns.

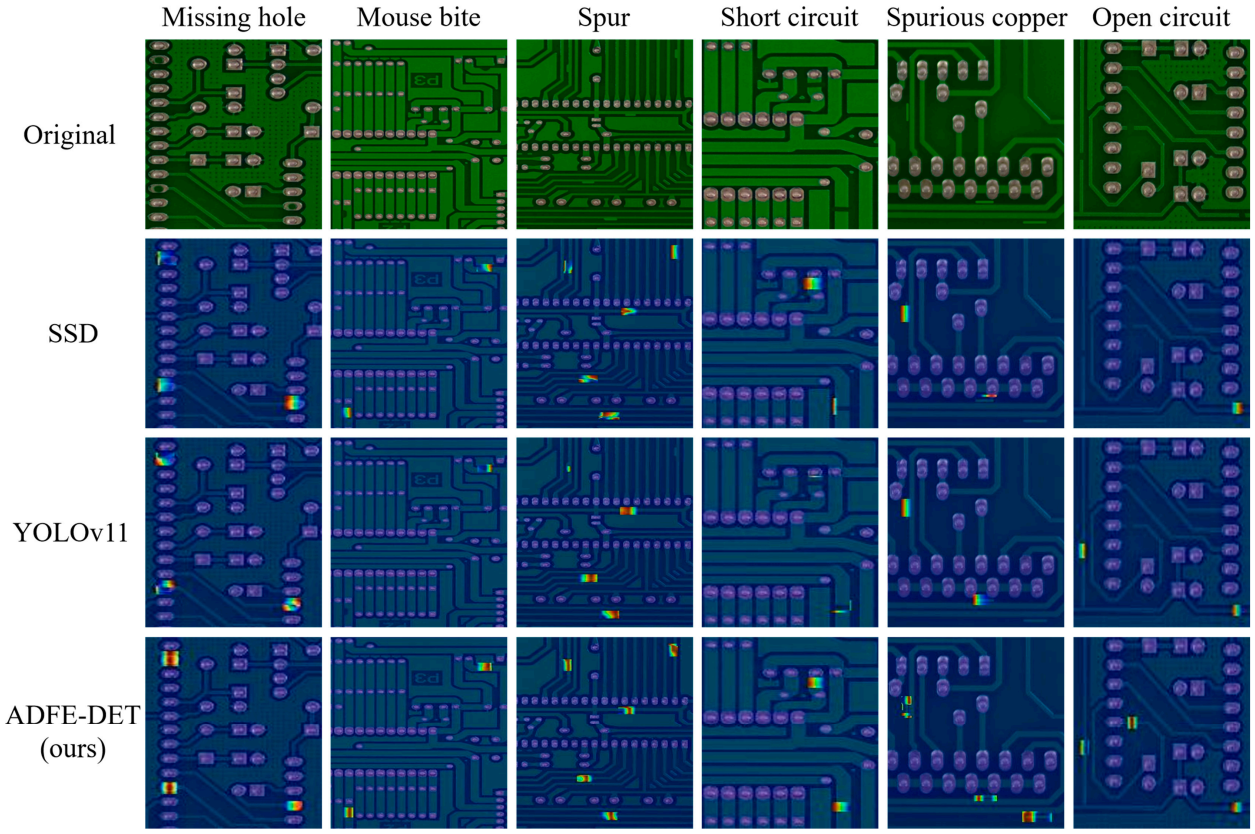


Figure 6. Grad-CAM heat map visualization for different models on PKU-Market-PCB dataset.

4.6.3. Detection Results and Heatmap Visualization on the NEU-DET Dataset

The evaluation on steel surface defects provides additional validation relevant to weld defect detection, as steel is the primary material in most welding applications. Steel surface defects (crazing, inclusions, patches, pitted surfaces, rolled-in scale, and scratches) share similar characteristics with weld defects in terms of texture variations, irregular morphologies, and challenging backgrounds. This cross-material validation demonstrates that ADFE-DET's adaptive feature enhancement mechanisms are effective for detecting defects in the base materials commonly used in welding processes, providing confidence in the algorithm's performance across the complete welding quality assessment pipeline.

Figure 7 presents the qualitative detection results of ADFE-DET on six representative steel surface defect types from the NEU-DET dataset, providing a visual validation of its detection performance. The results demonstrate that ADFE-DET consistently outperforms DDN and the YOLOv11 baseline models in localization accuracy and confidence estimation. For Crazing defects with reticulated crack patterns, ADFE-DET generates accurately fitting bounding boxes with a confidence score of 0.44, significantly higher than DDN (0.29) and YOLOv11 (0.27), highlighting its superior feature discrimination for complex textures. In inclusions defect detection, all defect instances are successfully identified with balanced confidence scores, while baseline methods show inconsistent detections and lower confidence. For pitted surface defects with distributed corrosion patterns, ADFE-DET achieves a confidence of 0.93 and precise boundary outlining, outperforming DDN (0.86) and YOLOv11 (0.83). In the case of patch defects with irregular morphology and varying intensities, the model captures all three instances with high confidence scores (0.79, 0.71, 0.66), whereas competing methods exhibit missed detections or lower confidence. Similarly, for rolled-in scale and scratch defects, ADFE-DET maintains robust performance with confidence scores of 0.53/0.64 and 0.79, respectively. Overall, the qualitative analysis confirms that ADFE-DET achieves an optimal balance of detection accuracy, precise localization, and

reliable confidence estimation, demonstrating strong robustness and practical applicability in industrial steel surface inspection.

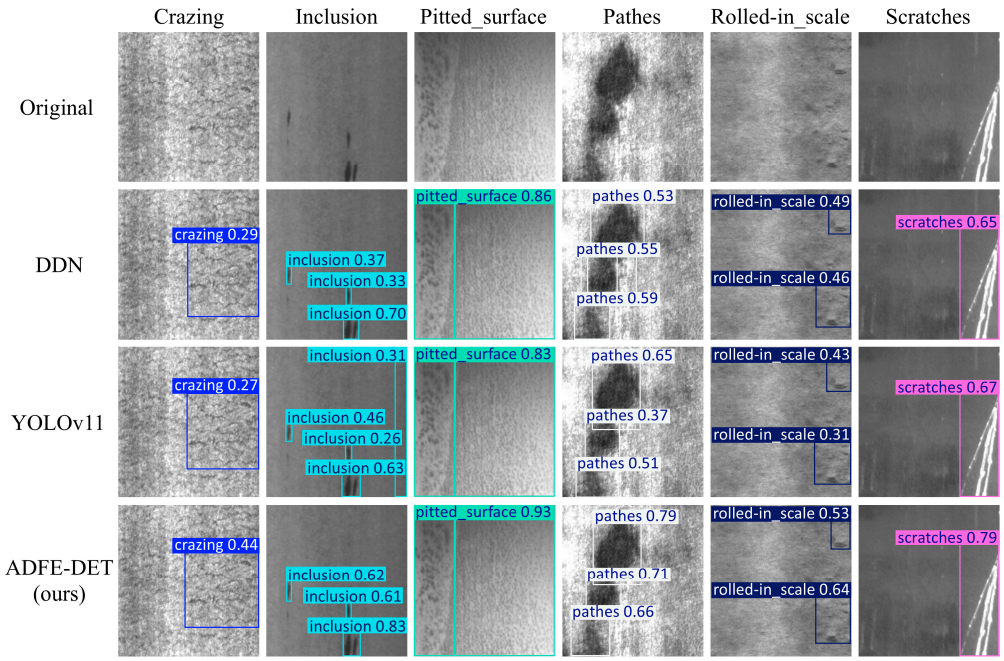


Figure 7. Comparison of detection results of different models on NEU-DET dataset.

Figure 8 presents the Grad-CAM visualization of representative steel surface defects to illustrate the internal feature representation and attention mechanism of ADFE-DET. The thermogram comparison clearly shows that ADFE-DET exhibits markedly enhanced spatial attention focus and discriminative feature activation relative to baseline models. For Crazing defects with fine-grained crack networks, ADFE-DET produces highly concentrated activation precisely aligned with the defect region, with a transparent red-to-yellow gradient indicating intense feature discrimination, whereas DDN and YOLOv11 show dispersed activation in background regions. In inclusion defects, ADFE-DET generates a focused vertical activation pattern that accurately captures linear defect features, while the baseline shows decentralized responses with reduced spatial specificity. For pitted surface defects exhibiting distributed texture anomalies, ADFE-DET achieves full coverage with uniform high-intensity activation, demonstrating the adaptive feature enhancement mechanism's ability to capture global defect patterns. For rolled-in scale defects with subtle intensity variations, the model precisely localizes defect boundaries with minimal background interference, as evidenced by sharp transitions in the thermogram. Similarly, for patches and scratches, ADFE-DET consistently produces more focused, semantically meaningful activation patterns, with thermogram intensities closely matching the actual defect locations. Overall, the Grad-CAM analysis confirms that ADFE-DET's adaptive feature augmentation architecture enables superior spatial attention and discriminative feature representation, facilitating accurate defect detection and precise localization even under complex texture backgrounds and subtle visual variations typical of steel surface inspection tasks.

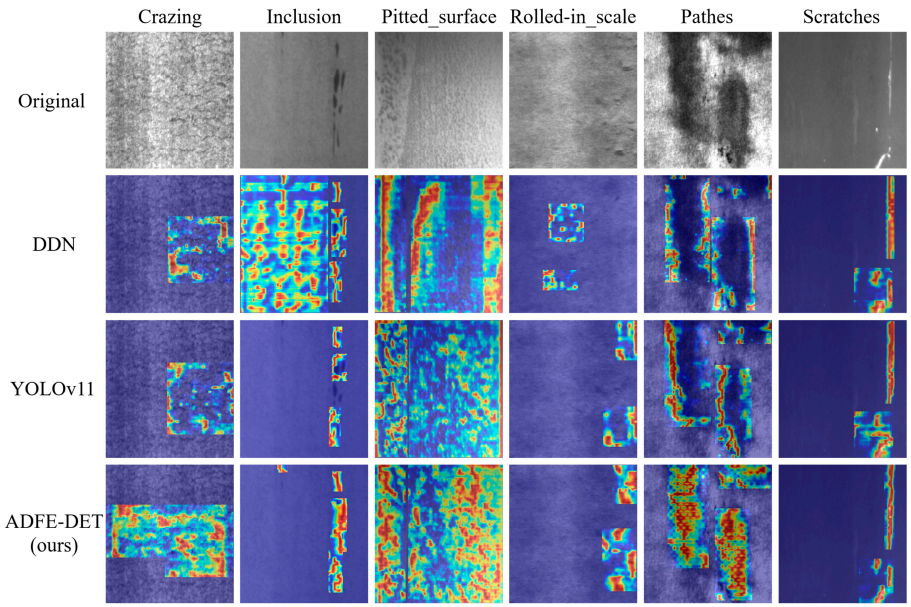


Figure 8. Grad-CAM heat map visualization for different models on NEU-DET dataset.

5. Discussion

The ADFE-DET weld defect detection algorithm proposed in this study effectively addresses key technical challenges faced by conventional methods by leveraging a synergistic design of three core modules. Experimental results demonstrate that the DSCFBlock module overcomes the limitations of the traditional C3k2 module in capturing long-range spatial dependencies by introducing the Edge-Protected Dynamic Selective (EPDS) attention mechanism, enabling the model to accurately perceive subtle texture variations and geometric morphology features of weld seams. Ablation studies show that incorporating DSCFBlock alone improves mAP50 on the Weld-DET dataset by 0.9%, validating the effectiveness of the single-head self-attention mechanism and cross-cascade residual linkage strategy in weld feature extraction. The introduction of the AHSFPN module further improves detection performance by integrating the DACA attention mechanism and the HSFF fusion strategy. The DACA module captures orientation-sensitive features by modeling horizontal and vertical spatial dependencies, while the hierarchical fusion in HSFF effectively combines high-level semantic information with low-level details, significantly enhancing small-defect detection. Notably, AHSFPN reduces the number of parameters by 23.9% while improving accuracy, thanks to its efficient feature reuse and adaptive weight allocation. The MDDLH module, a lightweight detection head, achieves fine-grained detection of minor defects by integrating five differential convolution operators—center difference, horizontal difference, vertical difference, anti-diagonal difference, and standard convolution—thereby enhancing multi-directional edge feature extraction. By applying a reparameterization technique, the module merges the weights of multiple operators during inference, significantly reducing computational complexity while maintaining strong feature extraction.

Cross-dataset validation experiments further confirm the strong generalization and robustness of ADFE-DET. On the NEU-DET steel surface defects dataset, the model achieves balanced detection performance across all six defect categories, with notable improvements in challenging categories such as cracking, patches, and pitted surfaces, demonstrating the effectiveness of the adaptive feature enhancement mechanism in capturing diverse defect patterns. On the PKU-Market-PCB PCB defect dataset, ADFE-DET significantly improves detection accuracy for fine-grained defects, including mouse bites, spurs, and short circuits, demonstrating its strong applicability to accuracy-critical industrial defect-detection tasks. The Grad-CAM visualization further shows that ADFE-DET exhibits more

focused and precise spatial attention than baseline models, with activation regions closely aligned with actual defect locations and minimal background interference. This highlights the superiority of the adaptive attention mechanism and multi-directional differential convolution in feature discrimination and localization. However, certain limitations remain. First, while ADFE-DET performs well on the three datasets, detection under extreme lighting conditions, severe occlusions, or ultra-small defects (area $< 5 \times 5$ pixels) still leaves room for improvement. Second, the model's generalization to novel or rare defect types requires further verification. Finally, although the inference speed reaches 105 FPS, it remains slower than ultra-lightweight models (e.g., YOLOv12n at 142 FPS), underscoring the need for deployment-optimized optimization.

Regarding the generalizability of our approach, while ADFE-DET incorporates welding-specific design elements, the three core modules demonstrate adaptability to other industrial defect detection scenarios with appropriate modifications. The DSCFBlock's edge-preserving dynamic attention mechanism is particularly effective for capturing fine-grained texture variations common in various material defects. The AHSFPN's adaptive hierarchical feature fusion can handle multi-scale defects across different industrial contexts. The MDDLH's multi-directional differential convolution is suited for detecting linear and directional features present in many types of structural defects. Our validation on NEU-DET (steel surface defects) and PKU-Market-PCB (printed circuit board defects) datasets demonstrates this cross-domain effectiveness, though domain-specific fine-tuning and potential architectural modifications may be beneficial for optimal performance in new applications such as gear tooth flank damage or bearing component wear detection.

The reliability of ADFE-DET is demonstrated through several key aspects: (1) Consistent performance across multiple datasets: The model achieves stable performance improvements of 2.16%, 2.73%, and 1.81% mAP50 on Weld-DET, NEU-DET, and PKU-Market-PCB datasets, respectively, demonstrating cross-domain reliability. (2) Robust feature extraction: The Grad-CAM visualizations (Figures 4–8) show consistent and focused attention on defect regions with minimal background interference, indicating reliable feature discrimination. (3) Computational stability: The model maintains a consistent inference speed of 105 FPS across different defect types and scales. (4) Parameter efficiency: The 34.1% reduction in parameters while improving accuracy indicates architectural reliability without overfitting. However, we acknowledge limitations under extreme conditions (severe lighting variations, ultra-small defects $< 5 \times 5$ pixels) that require further investigation. Future work will include comprehensive reliability testing under various environmental conditions and statistical significance analysis of detection results.

6. Conclusions

The ADFE-DET algorithm proposed in this study effectively addresses key technical challenges in weld defect detection by synergistically optimizing three core innovative modules. The DSCFBlock captures fine-grained texture features via an edge-protected dynamic attention mechanism; the AHSFPN enhances multi-scale defect perception through direction-aware channel attention and hierarchical feature fusion; and the MDDLH achieves a balance between lightweight design and accurate localization using multi-directional differential convolution. Experimental results demonstrate the superior performance of ADFE-DET: compared with YOLOv11n, mAP50 is improved by 2.16%, 2.73%, and 1.81% on the Weld-DET, NEU-DET, and PKU-Market-PCB datasets, respectively, while the number of parameters is reduced by 34.1%, computational complexity decreases by 4.6%, and inference speed reaches 105 FPS. This study provides an efficient, accurate, and lightweight solution for intelligent industrial weld quality inspection.

Future work will focus on (1) incorporating domain adaptation and adversarial training to improve model robustness under extreme conditions; (2) exploring super-resolution reconstruction techniques to enhance detection of ultra-small defects; (3) applying model compression strategies, including knowledge distillation and quantization, for efficient deployment on edge devices; (4) building large-scale, diversified datasets and investigating few-shot learning methods; and (5) extending the adaptive feature enhancement mechanism to other industrial defect detection scenarios, supporting the advancement of intelligent manufacturing.

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

Funding: This research was funded by the Key Projects of Basic Scientific Research in Higher Education Institutions of Liaoning Province (Grant Number LZGD2021037).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available from the corresponding author upon reasonable request.

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

References

1. Georgievska, M.; Malengier, B.; Roelofs, L.; Tiku, S.D.; Van Langenhove, L. Towards Woven Fabrics with Integrated Stainless Steel-Nickel-Carbon Thermopile for Sensing and Cooling Applications. *Appl. Sci.* **2025**, *15*, 9002. [CrossRef]
2. Luo, Y.; Yamane, S.; Wang, W.; Tsumori, R.; Ochiai, K.; Lu, J.; Xia, Y. Vision-Based Closed-Loop Control of Pulsed MAG Welding Using Otsu-Segmented Arc Features. *Appl. Sci.* **2025**, *15*, 8950. [CrossRef]
3. Almazán, M.Á.; Marín, M.; Almazán, J.A.; García-Domínguez, A.; Rubio, E.M. Friction Stir Welding Process Using a Manual Tool on Polylactic Acid Structures Manufactured by Additive Techniques. *Appl. Sci.* **2025**, *15*, 8155. [CrossRef]
4. Lv, K.; Liu, Z.; Zhang, H.; Jia, H.; Mao, Y.; Zhang, Y.; Bi, G. Wheeled Permanent Magnet Climbing Robot for Weld Defect Detection on Hydraulic Steel Gates. *Appl. Sci.* **2025**, *15*, 7948. [CrossRef]
5. Tomić, T.; Mihalic, T.; Groš, J.; Vugrinec, L. Experimental Evaluation of Arc Stud Welding Techniques on Structural and Stainless Steel: Effects on Penetration Depth and Weld Quality. *Appl. Sci.* **2025**, *15*, 7269. [CrossRef]
6. Hoa, N.T.; Quan, T.H.M.; Diep, Q.B. Weld-CNN: Advancing non-destructive testing with a hybrid deep learning model for weld defect detection. *Adv. Mech. Eng.* **2025**, *17*, 16878132251341615.
7. Wang, X.; Zscherpel, U.; Tripicchio, P.; D'Avella, S.; Zhang, B.; Wu, J.; Liang, Z.; Zhou, S.; Yu, X. A comprehensive review of welding defect recognition from X-ray images. *J. Manuf. Process.* **2025**, *140*, 161–180. [CrossRef]
8. Wang, G.; Liao, T.W. Automatic identification of different types of welding defects in radiographic images. *Ndt E Int.* **2002**, *35*, 519–528. [CrossRef]
9. Hou, W.; Zhang, D.; Wei, Y.; Guo, J.; Zhang, X. Review on computer aided weld defect detection from radiography images. *Appl. Sci.* **2020**, *10*, 1878. [CrossRef]
10. Madhav, M.; Ambekar, S.S.; Hudnurkar, M. Weld defect detection with convolutional neural network: An application of deep learning. *Ann. Oper. Res.* **2023**, *350*, 579–602. [CrossRef]
11. Wang, X.; Wang, X.; Zhang, B.; Cui, J.; Lu, X.; Ren, C.; Cai, W.; Yu, X. Binary classification of welding defect based on deep learning. *Sci. Technol. Weld. Join.* **2022**, *27*, 407–417. [CrossRef]
12. Kim, Y.; Lee, H.; Jang, J.; Choi, W.; Kim, K.-B. Data-Driven Approach to Weld Defect Evaluation Using Convolutional Neural Network and Ultrasonic B-Scan Images. Available online: https://ssrn.com/abstract=5085738 (accessed on 7 January 2025).
13. Zhang, L.; Pan, H.; Jia, B.; Li, L.; Pan, M.; Chen, L. Lightweight DCGAN and MobileNet based model for detecting X-ray welding defects under unbalanced samples. *Sci. Rep.* **2025**, *15*, 6221. [CrossRef]
14. Vasan, V.; Sridharan, N.V.; Balasundaram, R.J.; Vaithiyanathan, S. Ensemble-based deep learning model for welding defect detection and classification. *Eng. Appl. Artif. Intell.* **2024**, *136*, 108961. [CrossRef]

15. Pan, K.; Hu, H.; Gu, P. Wd-yolo: A more accurate yolo for defect detection in weld X-ray images. *Sensors* **2023**, *23*, 8677. [CrossRef]
16. Feng, J.; Wang, J.; Zhao, X.; Liu, Z.; Ding, Y. Laser weld spot detection based on YOLO-weld. *Sci. Rep.* **2024**, *14*, 29403. [CrossRef]
17. Luo, Y.; Ling, J.; Wang, J.; Zhang, H.; Chen, F.; Xiao, X.; Lu, N. SFW-YOLO: A lightweight multi-scale dynamic attention network for weld defect detection in steel bridge inspection. *Measurement* **2025**, *253*, 117608. [CrossRef]
18. Zhang, M.; Hu, Y.; Xu, B.; Luo, L.; Wang, S. DSF-YOLO for weld defect detection in X-ray images with dynamic staged fusion. *Sci. Rep.* **2025**, *15*, 23305. [CrossRef]
19. Xu, X.; Li, X. Research on surface defect detection algorithm of pipeline weld based on YOLOv7. *Sci. Rep.* **2024**, *14*, 1881. [CrossRef] [PubMed]
20. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
21. DELVIN. Welddefect Dataset [DS/OL]. Roboflow. 2025. Available online: https://universe.roboflow.com/delvin-9qaxw/welddefect-xocdz (accessed on 9 October 2025).
22. He, Y.; Song, K.; Meng, Q.; Yan, Y. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1493–1504. [CrossRef]
23. Liu, J.; Li, H.; Zuo, F.; Zhao, Z.; Lu, S. Kd-lightnet: A lightweight network based on knowledge distillation for industrial defect detection. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–13. [CrossRef]
24. Li, D.; Li, L.; Chen, Z.; Li, J. ShiftwiseConv: Small Convolutional Kernel with Large Kernel Effect. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 10–17 June 2025; pp. 25281–25291.
25. Zhang, J.; Li, X.; Li, J.; Liu, L.; Xue, Z.; Zhang, B.; Jiang, Z.; Huang, T.; Wang, Y.; Wang, C. Rethinking mobile block for efficient attention-based models. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023; IEEE Computer Society: Washington, DC, USA, 2023; pp. 1389–1400.
26. Ma, X.; Dai, X.; Bai, Y.; Wang, Y.; Fu, Y. Rewrite the stars. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 5694–5703.
27. Lu, W.; Chen, S.-B.; Li, H.-D.; Shu, Q.-L.; Ding, C.H.; Tang, J.; Luo, B. Legnet: Lightweight edge-Gaussian driven network for low-quality remote sensing image object detection. *arXiv* **2025**, arXiv:2503.14012.
28. Lee, S.; Choi, J.; Kim, H.J. Efficientvim: Efficient vision mamba with hidden state mixer based state space duality. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 14923–14933.
29. Zhang, Y.; Guo, Z.; Wu, J.; Tian, Y.; Tang, H.; Guo, X. Real-time vehicle detection based on improved yolo v5. *Sustainability* **2022**, *14*, 12274. [CrossRef]
30. Wang, G.; Chen, Y.; An, P.; Hong, H.; Hu, J.; Huang, T. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **2023**, *23*, 7190. [CrossRef]
31. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. Yolov9: Learning what you want to learn using programmable gradient information. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 1–21.
32. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J. Yolov10: Real-time end-to-end object detection. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 107984–108011.
33. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524.
34. Lei, M.; Li, S.; Wu, Y.; Hu, H.; Zhou, Y.; Zheng, X.; Ding, G.; Du, S.; Wu, Z.; Gao, Y. YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception. *arXiv* **2025**, arXiv:2506.17733.
35. Feng, Y.; Huang, J.; Du, S.; Ying, S.; Yong, J.-H.; Li, Y.; Ding, G.; Ji, R.; Gao, Y. Hyper-yolo: When visual object detection meets hypergraph computation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**. [CrossRef]
36. Wang, Z.; Li, C.; Xu, H.; Zhu, X. Mamba YOLO: SSMs-based YOLO for object detection. *arXiv* **2024**, arXiv:2406.05835. [CrossRef]
37. Xing, J.; Jia, M. A convolutional neural network-based method for workpiece surface defect detection. *Measurement* **2021**, *176*, 109185. [CrossRef]
38. Song, H. RSTD-YOLOv7: A steel surface defect detection based on improved YOLOv7. *Sci. Rep.* **2025**, *15*, 19649. [CrossRef]
39. Li, D.; Jiang, C.; Liang, T. REDef-DETR: Real-time and efficient DETR for industrial surface defect detection. *Meas. Sci. Technol.* **2024**, *35*, 105411. [CrossRef]
40. Benzaoui, N.; Gonzalez, M.S.; Estarán, J.M.; Mardoyan, H.; Lautenschlaeger, W.; Gebhard, U.; Dembeck, L.; Bigo, S.; Pointurier, Y. Deterministic dynamic networks (DDN). *J. Light. Technol.* **2019**, *37*, 3465–3474. [CrossRef]
41. Li, S.; Kong, F.; Wang, R.; Luo, T.; Shi, Z. Efd-yolov4: A steel surface defect detection network with encoder-decoder residual block and feature alignment module. *Measurement* **2023**, *220*, 113359. [CrossRef]
42. Fu, M.; Wu, J.; Wang, Q.; Sun, L.; Ma, Z.; Zhang, C.; Guan, W.; Li, W.; Chen, N.; Wang, D. Region-based fully convolutional networks with deformable convolution and attention fusion for steel surface defect detection in industrial Internet of Things. *IET Signal Process.* **2023**, *17*, e12208. [CrossRef]

43. Wang, C.; Lin, M.; Shao, L.; Xiang, J. MF-YOLO: A lightweight method for real-time dangerous driving behavior detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5035213. [CrossRef]
44. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

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