

Article

FALW-YOLOv8: A Lightweight Model for Detecting Pipeline Defects

Huazhong Wang , Xuetao Wang  and Lihua Sun * 

School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China; hzwang@ecust.edu.cn (H.W.); y80240224@mail.ecust.edu.cn (X.W.)

* Correspondence: sunlihua@ecust.edu.cn; Tel.: +86-187-2120-8677

Abstract

Pipelines are critical infrastructures in both industrial production and daily life. However, defects frequently arise due to environmental and manufacturing factors, which may lead to severe safety risks. To overcome the limitations of traditional object detection methods, such as inefficient feature extraction and the loss of critical information, this paper proposes an improved algorithm, termed FALW-YOLOv8, built upon the YOLOv8 architecture. Specifically, the FasterBlock is incorporated into the C2f module to replace standard convolutional layers, effectively reducing computational redundancy while improving feature extraction efficiency. In addition, the ADown module is employed to enhance multi-scale feature preservation, while the LSKA attention mechanism is introduced to improve detection accuracy, particularly for small defects. The Wise-IoU v2 loss function is further adopted to refine bounding box regression for complex samples. Experimental results demonstrate that the proposed FALW-YOLOv8 achieves a 5.8% improvement in mAP50, along with a 34.8% reduction in model parameters and a 30.86% decrease in computational cost. These results indicate that the proposed method achieves a favorable balance between accuracy and efficiency, making it well-suited for real-time industrial pipeline inspection applications.

Keywords: YOLOv8; object detection; FasterBlock; ADown; LSKA; Wise-IoU v2

1. Introduction

Pipelines serve as vital infrastructure for both industrial sectors and daily life, facilitating critical functions such as energy transmission, water resource distribution, and industrial fluid conveyance. The safe operation of these systems is intrinsically linked to national security, public safety, and property protection. However, due to harsh environmental conditions and prolonged operational lifespans, pipelines are susceptible to defects such as corrosion, deformation, and fractures. These issues not only compromise structural integrity but also pose significant safety hazards. Statistics indicate that in the United States alone, pipeline defects result in annual economic losses exceeding \$130 billion. Similarly, as a major industrial power, China relies on an extensive pipeline network, where direct annual economic damages caused by corrosion and defects are estimated to reach over \$200 billion [1]. Consequently, the development of efficient and high-precision defect detection methods is crucial for preventing catastrophic failures and mitigating associated risks. Traditional detection approaches, particularly manual inspection, rely heavily on subjective human judgment. This reliance introduces significant variability and increases the likelihood of missed detections or false alarms. Furthermore, these methods



Academic Editor: Taeshik Shon

Received: 8 December 2025

Revised: 25 December 2025

Accepted: 28 December 2025

Published: 1 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

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

are often constrained by environmental factors, resulting in low inspection efficiency and poor adaptability to extreme operating conditions. As a result, manual inspection techniques are becoming increasingly inadequate for meeting the rigorous demands of modern industrial production.

In addition, mainstream nondestructive testing methods, such as ultrasonic testing, magnetic flux leakage testing, and eddy current testing, have been widely applied for defect detection since the 20th century [2]. Although these techniques significantly improve detection accuracy and efficiency compared to manual inspection, they are heavily constrained by the material properties of the tested objects. This limitation results in restricted versatility, rendering them unsuitable for pipelines composed of non-metallic materials. Furthermore, certain methods suffer from operational complexity; for instance, magnetic flux leakage testing requires the magnetization of the object prior to inspection, a process that is both time-consuming and labor-intensive. In recent years, pipeline materials have diversified to meet industrial production demands, encompassing materials such as concrete, polyvinyl chloride (PVC), and fiberglass-reinforced plastic (FRP). Unfortunately, these emerging materials are largely incompatible with the aforementioned conventional detection techniques.

Parallel to physical inspection methods, data-driven fault diagnosis techniques have also attracted significant attention in the field of pipeline monitoring. Researchers have established systematic frameworks that integrate Artificial Neural Networks (ANN) with Neuro-Fuzzy systems and Extended Kalman Filters (EKF) to process time-series sensor data for pipeline leak diagnosis [3,4]. By leveraging system state estimation and fuzzy logic, these approaches effectively manage non-linear dynamics and operational uncertainties, providing robust solutions for internal fault identification. However, such methods rely predominantly on continuous sensor data streams (e.g., pressure or flow rates) and face limitations in visually characterizing the specific morphology of exterior surface defects—a task that constitutes the primary focus of computer vision-based approaches.

In recent years, with the continuous development of deep learning and machine vision, it has gradually become the mainstream method in pipeline defect detection due to its advantages of non-contact detection, fast response speed and high detection accuracy. The current mainstream object detection algorithm framework has evolved into two dominant paradigms: the first comprises well-established convolutional neural network (CNN)-based methods [5], while the second features Transformer-based approaches that have demonstrated remarkable potential in recent years [6].

The Transformer-based object detection algorithm treats object detection as a Set Prediction Problem, where the attention mechanism directly models global features interactively to effectively capture global contextual information in images. In recent years, the research interest in this algorithm for object detection has remained consistently high. Carion et al. [7] pioneered the DETR (Detection Transformer) algorithm, which transforms object detection into a Set Prediction Problem. By combining Bipartite Matching Loss with Transformer architecture, the method eliminates the need for complex anchor design and Non-Maximum Suppression (NMS) post-processing in traditional detectors, achieving a fully end-to-end detection process. This approach achieves detection accuracy comparable to Faster R-CNN on the COCO dataset. To address the slow convergence during training and poor small object detection performance of DETR, Zhu et al. [8] subsequently proposed Deformable DETR. By introducing a multi-scale deformable attention module, the attention mechanism now focuses solely on a limited number of key sampling points around reference points rather than scanning all global pixels. This innovation achieved a tenfold acceleration in training convergence speed while significantly enhancing the model's ability to capture small objects across multi-scale feature maps. To further en-

hance Transformer's versatility and computational efficiency in dense prediction tasks, Liu et al. [9] introduced Swin Transformer. By constructing a hierarchical feature pyramid and employing a self-attention mechanism with shifted windows, the model effectively balances local feature extraction with long-range dependency modeling. This approach significantly reduces computational complexity while outperforming mainstream CNN backbone networks in downstream tasks such as object detection and instance segmentation. While Transformer-based object detection algorithms overcome the local limitations of convolutional operations by incorporating the self-attention mechanism from natural language processing, their architecture lacks the inductive bias inherent in Convolutional Neural Networks (CNNs). This necessitates massive training datasets and extended training periods to fine-tune model parameters, while the high computational costs also pose challenges for real-time deployment on edge devices [10].

The core of object detection algorithms based on Convolutional Neural Networks (CNNs) lies in the extraction of spatial features from images through convolutional operations. These algorithms are typically categorized into two distinct paradigms: two-stage methods (e.g., the R-CNN series) and one-stage methods (e.g., the YOLO series and SSD). Two-stage algorithms prioritize the generation of Region Proposals (RPs) prior to classification and regression, a strategy that generally yields higher detection accuracy. In contrast, one-stage methods bypass the RP generation phase to perform dense predictions directly on input images. Consequently, their streamlined network architectures offer significant advantages in inference speed, rendering them particularly suitable for industrial deployment [11]. The YOLO series has gained significant prominence in the field of object detection due to its numerous advantages, including low computational overhead, rapid processing speeds, real-time performance, and ease of training and deployment. These attributes make it highly effective in meeting the rigorous demands of modern industrial production. Consequently, there has been sustained research interest in optimizing and refining YOLO models. Zhang et al. [12] enhanced the YOLOv5 algorithm by integrating the Enhanced Convolutional Block Attention Module (ECBAM) and Switchable Atrous Convolution (SAC). These additions effectively strengthened the model's focus on key features while suppressing irrelevant background noise. Furthermore, the adoption of the SIOU loss function provided a more comprehensive assessment of the alignment between predicted and ground truth bounding boxes. Collectively, these modifications led to significant performance enhancements across various metrics in pipeline defect detection tasks. Similarly, Wang et al. [13] proposed an improved model based on YOLOv5s that incorporates the Squeeze-and-Excitation (SE) module and GSConv structures within the backbone and feature fusion networks. This design not only enhanced detection accuracy but also streamlined the model architecture. By integrating the CBAM attention mechanism, the model's ability to recognize objects against complex backgrounds was bolstered. Moreover, the application of knowledge distillation further elevated performance, effectively addressing challenges related to subjectivity, inefficiency, and deployment in CCTV pipeline defect detection. In another study, Zhao et al. [14] introduced CEM-YOLO, an algorithm based on YOLOv7. This model integrates the CARAFE sampling strategy, which maintains strong feature extraction capabilities while effectively reducing computational costs and accelerating detection speed. The authors also introduced an Enhanced Variance-Center Feature Pyramid (EVC) module, which significantly improved the detection and recognition of small-scale targets. Additionally, the MPDIoU loss function was implemented to expedite model convergence and enhance localization accuracy. More recently, Wu et al. [15] developed an improved drainage pipe defect detection model by integrating EfficientViT with YOLOv8. By replacing the YOLOv8 backbone with the EfficientViT feature extraction network, the number of parameters was effectively reduced. Subsequently, the SE attention mechanism

was introduced to capture key features more effectively, thereby enhancing robustness. Finally, Focal Loss was employed to mitigate the impact of easy negative samples, resulting in more stable convergence for the optimized model.

While the aforementioned model optimization methods offer distinct advantages in terms of enhancing detection accuracy, simplifying model architectures, and strengthening feature extraction capabilities, they often entail trade-offs such as increased parameter counts, higher computational overhead, and potential compromises in robustness. These limitations render them suboptimal for meeting the rigorous, resource-constrained, and real-time requirements of modern industrial production environments. To address these challenges, this paper proposes a lightweight pipeline defect detection algorithm named FALW-YOLOv8. The major contributions of our work are as follows:

- (1) FasterBlock is integrated into the C2f module of YOLOv8's backbone and neck, enabling accelerated feature propagation while conserving computational resources.
- (2) ADown downsampling is used instead of traditional downsampling convolution to reduce the feature loss of small targets.
- (3) The LSKA attention mechanism is incorporated into the neck model to suppress complex background interference, thereby enhancing the model's feature response capability.
- (4) The Wise-IoU v2 loss function is employed to optimize the regression accuracy of challenging samples, thereby accelerating model convergence and enhancing its robustness.

2. Materials and Methods

2.1. Dataset

The dataset for this experiment was collected from municipal drainage networks, comprising approximately 700 images. To enhance the model's generalization capability, data augmentation techniques [16], including rotation, occlusion, and cropping, were employed to expand the dataset to 2000 images. The dataset encompasses common pipeline defect types such as break, hole, collapse-kink, surface corrosion, deformation, spalling, crack, joint, and degradation. Due to the scarcity of samples for certain defect types and to avoid confusion in anchor frame color during detection, break, hole and collapse-kink were grouped under a single label; this grouping represents a "Severe Structural Defect" category. Consequently, the final dataset contains seven label categories. To ensure the objectivity and reliability of model evaluation, the dataset was divided into training and validation sets at an 8:2 ratio. Figure 1 illustrates representative examples of each defect category in the dataset.



Figure 1. Examples of defect types.

2.2. FALW-YOLOv8

While the baseline YOLOv8 is powerful, its standard convolution operations treat all image regions equally. In the specific scenario of industrial pipeline inspection, images are characterized by highly repetitive background textures and sparse, minute defects [17]. Standard models suffer from substantial “computational redundancy” in processing these backgrounds, and lightweight models often fail to retain the high-frequency features of small cracks during downsampling.

To address these specific industrial challenges and overcome the bottleneck between precision and efficiency, we propose FALW-YOLOv8, with its network architecture illustrated in Figure 2.

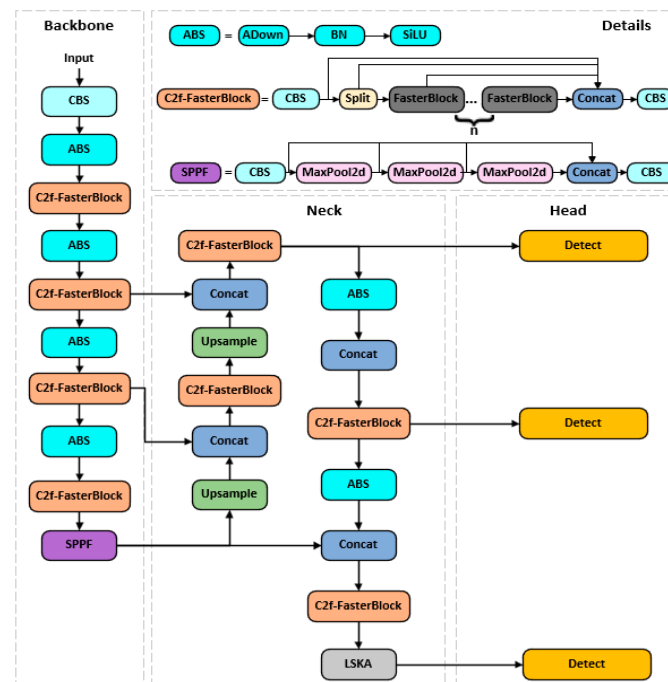


Figure 2. Network structure of FALW-YOLOv8.

The design philosophy centers on three key dimensions:

- (1) **Eliminating Redundancy:** We introduce the FasterBlock into the C2f module. By leveraging Partial Convolutions (PConv) [18], the model minimizes redundant computations on the repetitive pipe wall backgrounds, significantly reducing GFLOPs and parameter counts to suit embedded industrial devices.
- (2) **Enhanced Perception at Low Cost:** We replace standard downsampling with ADown to prevent feature loss of small defects and incorporate the LSKA mechanism. LSKA [19] decomposes large kernels to simulate human-like global scanning for distinguishing defects from water stains, enhancing feature extraction in complex environments without the heavy computational penalty of traditional transformers.
- (3) **Wise-IoU v2 [20]** replaces CIoU with a dynamic focusing mechanism that evaluates the quality of anchor boxes. Unlike standard loss functions, Wise-IoU v2 introduces a focusing coefficient based on the outlier degree, which effectively reduces the weight of simple examples and amplifies the gradient penalty for difficult samples. This aligns with the goal of prioritizing precision in detecting small, low-contrast defects in complex environments.

This synergistic design allows the model to achieve higher precision for fine defects while maintaining the lightweight characteristics required for real-time edge computing.

2.3. Experimental Setup

2.3.1. Implementation Details

To clearly describe the basic conditions relied on in this experiment, the specific hardware configuration, software platform, and core training parameters are shown in Table 1.

Table 1. Experimental environment.

Categories	Configuration Item	Specification
Hardware	GPU	NVIDIA GeForce RTX 4090
	Video Memory	24 GB
	Device Type	High-performance Server
Software	Operating System	Ubuntu 18.04
	Programming Language	Python 3.9.21
	Deep Learning Framework	PyTorch 2.0.1
	CUDA Toolkit	Version 11.7
Training Parameters	Training Epochs	400
	Input Resolution	640 × 640
	Batch Size	64
	Optimizer	SGD
	Momentum	0.937
	Initial Learning Rate	0.01

2.3.2. Evaluation Metrics

After model training is completed, to evaluate the quality of training results and assess its performance, the experiment employs Mean Average Precision (mAP), Parameter quantities and computational cost (GFLOPs) as performance metrics. The calculation formulas are as follows:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(R)d(R) \quad (3)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (4)$$

In these formulas, P represents Precision, R represents Recall, TP denotes the number of correctly detected instances, FP represents the number of false positives, FN indicates the number of missed detections and AP is the average precision, whose value is obtained by integrating the area under the P-R curve. mAP stands for mean average precision, a performance metric calculated as the arithmetic mean of AP values across all classes when the intersection-over-union ratio (IoU) between predicted and ground-truth boxes exceeds a specified threshold. n denotes the total number of target classes.

3. Experimental Results and Analysis

3.1. Ablation Experiment

To verify that each individual improvement module positively contributes to the overall model, this experiment designed five ablation studies. Each study employed

identical training parameters and environmental conditions, varying only the number of active improvement modules. The results are presented in Table 2.

Table 2. Ablation experiment, “✓” denotes that the corresponding module is included in the configuration.

Baseline	C2f-FasterBlock	ADown	LSKA	Wise-IoU v2	mAP50 (%)	Parameters/M	GFLOPs/G
✓					72.1	3.00	8.1
✓	✓				73.4	2.31	6.4
✓	✓	✓			75.1	1.89	5.5
✓	✓	✓	✓		77.7	1.96	5.6
✓	✓	✓	✓	✓	77.9	1.96	5.6

Analysis of the ablation experiment data reveals that the baseline YOLOv8 model achieves a mAP50 of 72.1%, with 3.00 M parameters and 8.1 G GFLOPs. After introducing the C2f-FasterBlock module, mAP50 improved to 73.4%, parameters decreased to 2.31 M, and GFLOPs dropped to 6.4 G. This demonstrates that the module effectively reduces computational complexity and parameter count while minimizing memory usage and enhancing the model’s ability to extract key features. Further integration of the ADown module elevated mAP50 to 75.1%, reduced parameters to 1.89 M, and lowered GFLOPs to 5.5 G. This demonstrates that the module, through optimized combinations of asymmetric convolution kernels and adaptive channel mechanisms, effectively reduces model parameters and computational overhead while enhancing feature expression capabilities to capture richer semantic information. Subsequently, integrating the LSKA attention mechanism further elevated the model’s mAP50 to 77.7%, though parameters and GFLOPs saw a slight increase. This demonstrates that the LSKA module significantly enhances the model’s spatial perception of defects across different scales and its feature discrimination capabilities with minimal computational overhead, substantially improving detection accuracy. Finally, integrating the C2f-FasterBlock, ADown, LSKA, and Wise-IoU v2 modules slightly improved the model’s mAP50 to 77.9%. This demonstrates that the Wise-IoU v2 loss function enhances the model’s adaptability to complex scenes, improves localization accuracy, strengthens robustness, and elevates overall detection precision.

3.2. Comparison Experiments

To validate the effectiveness of the model improvement in this study, we first conducted a comparative experiment with the baseline model. The experimental results indicate that the proposed FALW-YOLOv8 lightweight model demonstrates superior performance across all metrics compared to the YOLOv8 baseline. Specifically, the mean Average Precision (mAP50) increased by 5.8 percentage points, from 72.1% to 77.9%. The number of parameters was reduced by approximately 34.7%, from 3 M to 1.96 M, and the computational complexity (GFLOPs) decreased by about 30.9%, from 8.1 G to 5.6 G. Figure 3 presents a comparison of the Precision-Recall (P-R) curves for defect detection before and after the improvement.

Comparing the P-R curves before and after the improvements reveals that the enhanced model achieves higher detection accuracy across all defect categories. Overall, the detection performance is more balanced, demonstrating an improved equilibrium between precision and recall.

To evaluate whether the improved algorithm outperforms other methods in pipeline defect detection, this study conducted a model comparison experiment. The experiment compared multiple mainstream object detection algorithms, including RT-DETR [21] and common YOLO models such as YOLOv3-tiny [22], YOLOv5, YOLOv6 [23], YOLOv8n, YOLOv10n [24], and YOLO11n [25]. The data dependency inherent in Transformer architectures makes zero convergence challenging on small datasets. To address this, we

implemented COCO pre-trained [26] weights in RT-DETR training to accelerate convergence. Meanwhile, given YOLOv8's outstanding performance with its CNN architecture in small-sample scenarios, we adopted a zero-training strategy in our experiments. The experimental results are detailed in Table 3.

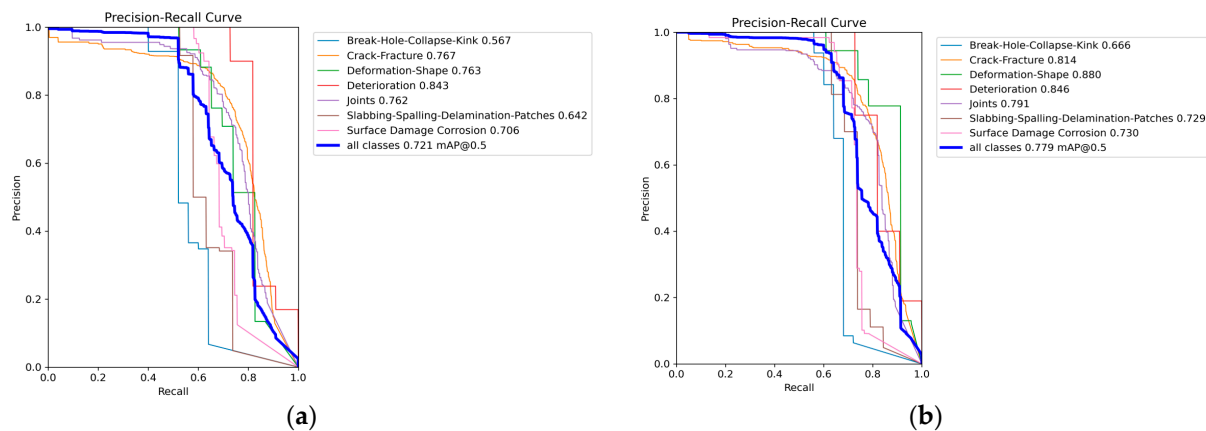


Figure 3. Comparison of P-R curves before and after model improvement. (a) YOLOv8n Precision-Recall curve; (b) FALW-YOLOv8 Precision-Recall curve.

Table 3. Model comparison experiment.

Model	Precision	Recall	mAP50 (%)	mAP50-95 (%)	Parameters/M	GFLOPs/G	Inference Time/ms	FPS
RT-DETR-resnet50	0.887	0.679	74.9	48.3	32.00	103.5	40.25	24.8
YOLOv3-t	0.835	0.610	68.7	41.9	12.13	18.9	2.24	445.9
YOLOv5	0.904	0.596	71.1	42.9	2.50	7.1	1.61	621.5
YOLOv6	0.779	0.557	60.9	36.4	4.23	11.8	1.64	608.9
YOLOv8n	0.884	0.638	72.1	45.4	3.01	8.1	1.57	637.6
YOLOv10n	0.798	0.626	70.4	47.2	2.69	8.2	2.37	421.2
YOLOv11n	0.895	0.639	74.2	46.8	2.58	6.3	2.35	424.9
FALW-YOLOv8	0.889	0.671	77.9	48.9	1.96	5.6	2.30	433.9

The results reveal that while the Transformer-based RT-DETR-resnet50 model demonstrates competitive detection performance, its complex structure incurs substantial computational overhead and requires a large number of parameters. This heavy resource consumption limits its deployment in lightweight, resource-constrained scenarios. In contrast, the widely adopted YOLO series models demonstrate exceptional advantages in lightweight design and real-time capabilities. The proposed FALW-YOLOv8 model outperforms other reference models in pipeline defect detection tasks, achieving an optimal balance between performance, lightweight design, and computational efficiency. Regarding core detection accuracy, FALW-YOLOv8 achieves mAP50 and mAP50-95 scores of 77.9% and 48.9%, respectively, ranking first among all reference models for both metrics. Compared to the YOLOv8 baseline model, these represent improvements of 5.8 percentage points and 3.5 percentage points, respectively. This demonstrates the model's enhanced stability in identifying targets across varying overlap levels (from low to high IoU), particularly excelling in complex scenarios involving small or occluded objects.

In balancing target capture and classification reliability, this model achieves a recall rate of 67.1%. Although slightly lower than the significantly heavier RT-DETR-resnet50, this represents a 3.3 percentage point improvement over the YOLOv8 baseline model, effectively reducing the risk of missed detections with minimal computational cost. Simultaneously, the model maintains a high precision of 88.9%, effectively avoiding the sharp increase

in false positive rates that often accompanies the pursuit of high recall. This achieves an optimal dual balance of low missed detections and low false positives.

Regarding deployment adaptability, FALW-YOLOv8 features only 1.96 million parameters and a computational load (GFLOPs) as low as 5.6 G. Both metrics rank lowest among all reference models, enabling efficient adaptation to resource-constrained scenarios such as embedded devices and mobile platforms. A comparative analysis of inference time and FPS data shows that FALW-YOLOv8 outperforms RT-DETR-resnet50 by a significant margin. While its specialized operators exhibit lower parallel efficiency on GPUs compared to traditional 3×3 standard convolutions and incur higher memory access costs, these factors do not translate into measurable advantages in inference time or frame rates. The difference remains within milliseconds, a negligible margin in industrial applications. Thus, FALW-YOLOv8 achieves substantial accuracy improvements without compromising real-time performance, maintaining its capability for real-time inference on low-power hardware.

The confusion matrices [27] of different models are shown in Figure 4. The results demonstrate that the proposed FALW-YOLOv8 model shows significant improvement in detecting critical structural defects that pose the greatest threat to pipeline safety. The optimized feature extraction module has notably enhanced sensitivity to local edge features and small-scale geometric anomalies, ensuring robust performance in identifying severe structural failures. Additionally, FALW-YOLOv8 achieves a high recall rate of 0.91 in the deformation category, highlighting its exceptional capability in capturing contour deformations.

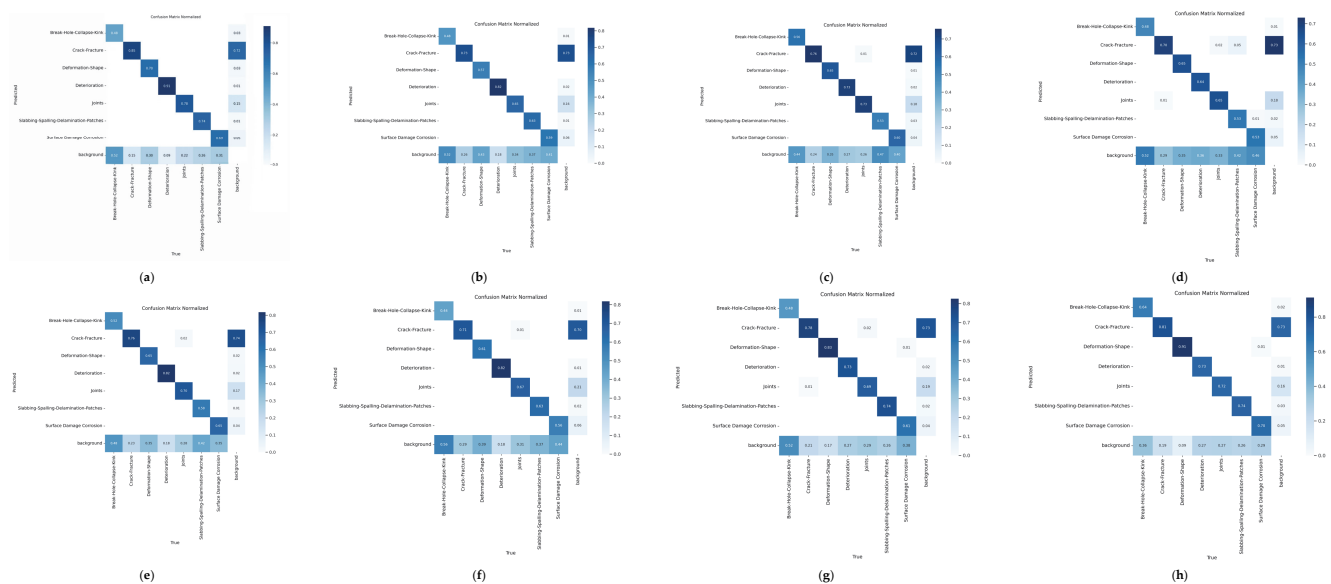


Figure 4. Comparison of different model confusion matrices. (a) RT-DETR-resnet50. (b) YOLOv3-t. (c) YOLOv5. (d) YOLOv6. (e) YOLOv8n. (f) YOLOv10n. (g) YOLOv11n. (h) FALW-YOLOv8.

However, in the field of texture-based defect detection—particularly for degradation-related issues—FALW-YOLOv8 achieves a mere 0.73 recall rate, significantly lower than RT-DETR’s 0.91. This disparity stems from their architectural differences: RT-DETR leverages self-attention mechanisms to capture global semantic information, giving it a natural advantage in identifying large-scale texture features like degradation. In contrast, FALW-YOLOv8, built on a CNN architecture, focuses on local feature extraction. While this approach sacrifices some texture recognition capabilities, we strike a balance between reducing false negatives and lightening real-time inference load, considering practical engineering constraints like limited deployment resources.

3.3. Visualization

To visually evaluate the real-world detection performance of the proposed FALW-YOLOv8 algorithm in industrial environments, Figure 5 presents a comparison of detection results across different models on typical samples from the validation set.

Model	Break-Hole-Cloapse-Kink	Crack-Fracture	Deformation-Shape	Deteriora-tion	Joints	Slabbing-Spalling-Delamina-tion-Patches	Surface-Damage-Corrosion
Ground Truth							
RT-DETR							
YOLOv3-t							
YOLOv5							
YOLOv6							
YOLOv8n							
YOLOv10n							
YOLO11							
FALW-YOLOv8							

Figure 5. Defect detection results of different models.

The comparative analysis demonstrates that the proposed FALW-YOLOv8 model achieves significant improvements in detection accuracy over other YOLO models, with notable reductions in both missed and false detections. Furthermore, its bounding boxes

exhibit superior defect edge alignment and higher overlap rates compared to models like RT-DETR and YOLOv6.

To further validate the feature extraction advantages of the improved model in complex pipeline environments and evaluate its detection reliability in the presence of background noise, this section employs the EigenCAM algorithm to generate class activation heatmaps, conducting a qualitative analysis of the visual attention mechanisms across all reference models. The experimental results are presented in Figure 6.

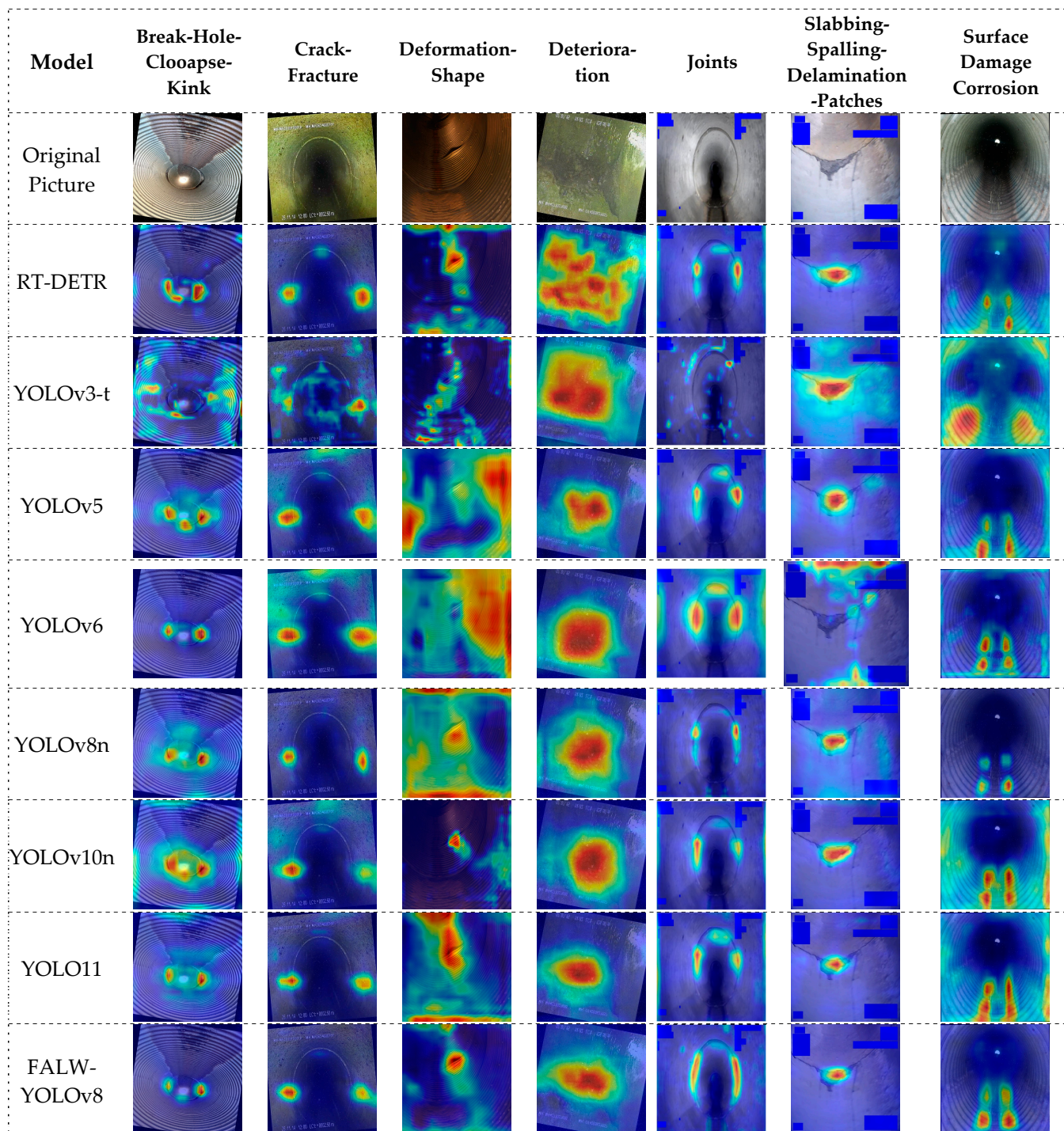


Figure 6. Thermal maps of different models.

Comparative analysis demonstrates that FALW-YOLOv8 effectively suppresses background interference while precisely focusing attention on the core region of defect targets.

This enhanced feature localization capability reduces feature misalignment, explaining why the model exhibits lower false positive and false negative rates in the confusion matrix. The results validate the success of the model improvement strategy in enhancing feature extraction robustness under complex conditions.

3.4. Downsampling Strategy Selection

In industrial pipeline defect detection applications, this model faces dual challenges: achieving high-precision capture of subtle defect features while meeting the stringent computational resource constraints of edge deployment. To address these requirements, the study selected four down-sampling strategies with distinct emphases—RFAConv [28], SPDConv [29], SCDwn [24], and ADown—for comparative experiments. The experimental results are shown in Table 4.

Table 4. Comparison of Downsampling Strategies.

Strategies	mAP50 (%)	Parameters/M	GFLOPs/G
Base	72.1	3.01	8.1
RFAConv	71.8	3.03	8.3
SPDConv	72.0	2.69	7.4
SCDwn	73.0	2.50	7.5
ADown	74.2	2.59	7.2

Experimental results demonstrate that ADown emerges as the optimal downsample strategy for this application scenario, effectively resolving the traditional trade-off between precision and computational efficiency. Unlike SPDConv (which prioritizes small target information preservation), RFAConv (focusing on spatial receptive field attention), and SCDwn (designated for ultra-lightweight architectures), ADown achieves dual objectives through asymmetric convolution design: preserving multi-scale features while maintaining computational efficiency. This approach delivers peak accuracy with minimal GFLOPs, establishing it as the most theoretically and practically compatible solution for real-time defect detection in resource-constrained edge devices.

3.5. K-Fold Cross-Validation Analysis

To validate the robustness and generalization capability of the FALW-YOLOv8 model and exclude performance improvements due to the randomness of specific dataset partitioning, this study employed a 5-fold cross-validation experiment [30]. The complete dataset containing 2000 images was randomly shuffled and evenly divided into five subsets. During each iteration, one subset was alternately selected as the validation set, while the remaining four served as the training set. This process was repeated five times to ensure each sample was validated once, with the final result being the average of the five validation sets. The training parameters and experimental environment remained consistent with the specifications described in Section 3.3. Experimental results are detailed in Table 5.

Experimental results demonstrate that the proposed model maintains robust performance across various data partitioning methods. While performance fluctuations occur between different folds, this variation primarily stems from the inherent heterogeneity of industrial production line datasets—certain subsets may contain complex samples with severe occlusion or insufficient illumination. The model achieves 79.3% mAP50 and 50.2% mAP50-95 in 5-fold cross-validation, showing only marginal differences from the baseline fixed partitioning results. This indicates that the performance improvement of the FALW-YOLOv8 model over the baseline model is not dependent on the randomness of dataset partitioning.

Table 5. Five-fold cross-validation results.

Fold Count	mAP50 (%)	mAP50-95 (%)	Precision	Recall
Fold1	81.3	50.2	0.853	0.719
Fold2	80.0	51.6	0.887	0.699
Fold3	82.1	52.2	0.854	0.765
Fold4	72.5	45.7	0.886	0.649
Fold5	80.4	51.4	0.844	0.794
Average	79.3	50.2	0.865	0.725
FALW-YOLOv8	77.9	48.9	0.889	0.671

4. Summary and Conclusions

4.1. Summary

To address critical challenges in industrial pipeline defect detection—specifically inefficient feature extraction, the loss of key information, and the difficulty of balancing lightweight design with accuracy—this paper proposes the FALW-YOLOv8 algorithm, built upon the YOLOv8 architecture. The effectiveness of the proposed method is rigorously verified through ablation studies, comparative experiments, and visualization analysis. Firstly, by integrating the FasterBlock into the C2f module of both the backbone and neck networks, the model leverages partial convolutions and lightweight Multi-Layer Perceptrons (MLPs). This combination significantly reduces computational redundancy and memory access costs, thereby enabling efficient spatial feature extraction and the deep fusion of channel information. Secondly, replacing traditional downsampling convolutions with the ADown module enhances multi-scale feature retention through an asymmetric kernel design, which effectively mitigates the loss of features associated with small defects. Thirdly, the incorporation of the LSKA attention mechanism in the neck network utilizes lightweight large-kernel attention to bolster the model’s responsiveness to minute defect features and enhance spatial perception capabilities, ultimately optimizing multi-scale feature fusion. Additionally, the original CIoU loss function is replaced with Wise-IoU v2. Through dynamic weight adjustment and a focus on hard examples, this function significantly improves bounding box regression accuracy for complex samples—particularly enhancing localization precision for small targets. This modification effectively addresses the issue of localization inaccuracy inherent in traditional models when detecting minute pipeline defects, thereby ensuring greater detection reliability.

4.2. Future Perspectives

In our future work, we will focus on migrating the FALW-YOLOv8 model from laboratory settings to real-world applications. Specifically, we will deploy this model on pipeline inspection robotic systems equipped with NVIDIA Jetson Orin Nano embedded platforms, conducting field tests and performance evaluations under real-world complex conditions to validate the algorithm’s long-term stability at the edge. Additionally, detecting extremely small defects remains one of the primary challenges we currently face. Future research will prioritize this area, exploring the introduction of super-resolution reconstruction or adaptive small object enhancement paradigms to further enhance the model’s perception limits and robustness for detecting minute targets.

4.3. Conclusions

Experimental results demonstrate that, compared to the YOLOv8 baseline, FALW-YOLOv8 achieves a 5.8% improvement in mAP50 while simultaneously reducing the

parameter count by 34.8% and computational cost by 30.86%. These results reflect a synergistic optimization of detection accuracy, computational efficiency, and deployment flexibility. Consequently, the FALW-YOLOv8 model not only satisfies rigorous industrial inspection demands for accuracy and robustness but also, thanks to its lightweight architecture, proves highly adaptable to resource-constrained scenarios such as embedded devices and industrial edge computing terminals. Ultimately, this approach facilitates real-time pipeline defect detection, providing a robust technical foundation for the safe operation and maintenance of industrial infrastructure.

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

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this experiment was collected and created by Professor Jiang Qingchao's team at the School of Information Science and Engineering, East China University of Science and Technology. This dataset is intended for hardware deployment testing and is not yet publicly available for upload, though it may be made available upon request. The proposed model code is an enhancement of the open-source YOLOv8 baseline model.

Acknowledgments: The authors have reviewed and approved the final manuscript and take full responsibility for its content. The authors also gratefully acknowledge Jiang Qingchao for his valuable insights and constructive discussions.

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

Abbreviations

The following abbreviations are used in this manuscript:

CARAFE	Content-Aware ReAssembly of FEatures
MPDIoU	Minimum Point Distance Intersection over Union
LSKA	Large Selective Kernel Attention
WiseIoU-v2	Wise Intersection over Union v2
SGD	Stochastic Gradient Descent
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives
mAP	Mean Average Precision
EigenCAM	Eigen-Class Activation Map using Principal Components

References

1. Hou, B.; Li, X.; Ma, X.; Du, C.; Zhang, D.; Zheng, M.; Xu, W.; Lu, D.; Ma, F. The cost of corrosion in China. *Npj Mater. Degrad.* **2017**, *1*, 4. [CrossRef]
2. Qingjun, G.; Teng, Z.; Wei, L.; Zhu, Z.; Bo, X. Research status of oil and gas pipeline inspection technology and data analysis methods. *Pipeline Prot.* **2025**, *2*, 35–42+73.
3. Torres, L.; Jiménez-Cabas, J.; González, O.; Molina, L.; López-Estrada, F.-R. Kalman Filters for Leak Diagnosis in Pipelines: Brief History and Future Research. *J. Mar. Sci. Eng.* **2020**, *8*, 173. [CrossRef]
4. Pérez-Pérez, E.J.; González-Baldizón, Y.J.; Santos-Ruiz, I.; López-Estrada, F.R.; Guerrero-Sánchez, M.E. Data-Driven Fault Diagnosis in Water Pipelines Based on Neuro-Fuzzy Zonotopic Kalman Filters. *Math. Comput. Appl.* **2021**, *26*, 2. [CrossRef]

5. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object Detection in 20 Years: A Survey. *Proc. IEEE* **2023**, *111*, 257–276. [[CrossRef](#)]
6. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*; IEEE Computer Society: Los Alamitos, CA, USA, 2023; Volume 45, pp. 87–110.
7. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, UK, 23–28 August 2020; pp. 213–229.
8. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable Transformers for End-to-End Object Detection. In *Proceedings of the International Conference on Learning Representations (ICLR)*, Vienna, Austria, 3–7 May 2021.
9. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada, 11–17 October 2021; pp. 10012–10022.
10. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in Vision: A Survey. *ACM Comput. Surv.* **2022**, *54*, 200. [[CrossRef](#)]
11. Terven, J.; Cordova-Esparza, D.; Romero-Gonzalez, J.A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [[CrossRef](#)]
12. Zhang, J.; Zhang, Y.; Jiang, X.; Li, W.; Zhou, J. Research on Improved YOLOv5 Pipeline Defect Detection Algorithm. *J. Pipeline Syst. Eng. Pract.* **2025**, *16*, 04025008. [[CrossRef](#)]
13. Wang, T.; Li, Y.; Zhai, Y.; Wang, H. A Sewer Pipeline Defect Detection Method Based on Improved YOLOv5. *Processes* **2023**, *11*, 2508. [[CrossRef](#)]
14. Zhao, X.; Guo, F.; Wang, Y.; Shen, R. CEM-YOLO: Defect Detection in Large Hydropower Station Water Conveyance Pipelines Based on YOLOv7. *J. Phys. Conf. Ser.* **2025**, *2988*, 012012. [[CrossRef](#)]
15. Wu, Z.; Guo, Y.; Huang, S.; Ma, B. Improve the Defect Detection Model of YOLOv8 Drainage Pipe Based on EfficientViT. *Water Wastewater Eng.* **2025**, *61*, 125–130.
16. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
17. Wang, Q.; Lu, L.; Liu, S.; Hu, Q.; Zhong, G.; Su, Z.; Xu, S. A Method for Improving the Efficiency and Effectiveness of Automatic Image Analysis of Water Pipes. *Water* **2025**, *17*, 2781. [[CrossRef](#)]
18. Chen, J.; Kao, S.-H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 17–24 June 2023; pp. 12021–12031.
19. Lau, K.W.; Po, L.-M.; Rehman, Y.A. Large Separable Kernel Attention: Rethinking the Large Kernel Attention Design in CNN. *arXiv* **2023**, arXiv:2309.01439. [[CrossRef](#)]
20. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
21. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs Beat YOLOs on Real-time Object Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
22. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [[CrossRef](#)]
23. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976. [[CrossRef](#)]
24. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458.
25. Shi, W.; Dai, J.; Li, C.; Niu, N. YOLOv11-EMD: An Enhanced Object Detection Algorithm Assisted by Multi-Stage Transfer Learning for Industrial Steel Surface Defect Detection. *Mathematics* **2025**, *13*, 2769. [[CrossRef](#)]
26. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755.
27. Powers, D.M.W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2011**, arXiv:2010.16061.
28. Zhang, X.; Liu, C.; Yang, D.; Song, T.; Ye, Y.; Li, K.; Song, Y. RFACnv: Innovating Spatial Attention and Standard Convolutional Operation. *arXiv* **2023**, arXiv:2304.03198.

29. Sunkara, R.; Luo, T. No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. In *Machine Learning and Knowledge Discovery in Databases*; Amini, M.R., Canu, S., Fischer, A., Guns, T., Kralj Novak, P., Tsoumakas, G., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2023; Volume 13716, pp. 443–459.
30. Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv* **2018**, arXiv:1811.12808.

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.