

Article

Surface Defect Detection Algorithm for Workpieces Based on Improved YOLOv8

Da An ^{1,2}, Ng Kok Why ^{1,*} and Fangfang Chua ¹

¹ Faculty of Computing and Informatics, Multimedia University, Cyberjaya 63100, Selangor, Malaysia; 15122453209@163.com (D.A.); ffchua@mmu.edu.my (F.C.)

² School of Data Science and Big Data, Jining Normal University, Ulanqab 012000, China

* Correspondence: kwng@mmu.edu.my

Abstract

Industrial surface defect detection is crucial for quality control in manufacturing, yet remains challenging due to the small scale, low contrast, and texture variability of defects. While YOLOv8n offers high inference speed and efficiency, its accuracy is limited by insufficient feature representation and inadequate data diversity. This paper proposes a detection framework integrating Channel–Spatial Modulation Attention (CASM) and Small-Scale Grid Texture Shuffling Augmentation (SG-TSA) into YOLOv8n to improve detection performance without sacrificing efficiency. CASM introduces a parallel channel–spatial attention structure with adaptive fusion to better capture fine-grained defect features, while SG-TSA increases sample diversity by introducing realistic texture perturbations within defect regions. Experiments on the NEU-DET dataset show that our method improves mAP@0.5:0.95 by 3.01% and mAP@0.5 by 2.84% over baseline YOLOv8n. These results highlight the importance of architecture-specific optimization for lightweight detectors in industrial scenarios.

Keywords: defect detection; deep-learning; computer vision

1. Introduction

Currently, advanced industrial systems require increasingly improved product performance along with an increased need for quality control during production [1]. However, defects such as cracks, inclusions, patches, pitted surfaces, rolled-in scales, and scratches can significantly compromise product performance, aesthetics, and safety [2]. As a result, automated and accurate defect detection has become an indispensable step in manufacturing quality control.

Traditional methods, such as manual inspection and traditional image processing techniques, have low reliability, low accuracy, long detection time, and different conclusions drawn by operators under different production conditions [3], which cannot provide the accuracy and efficiency required by modern production environments.

In recent years, as deep learning has developed rapidly and object detection tasks have been successfully carried out in many fields, object detection technology based on deep learning has also made remarkable progress. According to the steps of obtaining detection results, it can be divided into two categories: two-stage object detectors [4,5] and one-stage object detectors [6,7]. In 2014, Ross et al. [8] proposed using CNN to obtain candidate regions, and since then, object detection has developed at an unprecedented speed.

Academic Editor: En Li

Received: 30 October 2025

Revised: 26 December 2025

Accepted: 22 January 2026

Published: 12 February 2026

Copyright: © 2026 by the author. 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.

Li J. et al. [9] proposed a new method that integrates an improved faster region convolutional neural network (R-CNN) and an adaptive Canny algorithm to enhance the defect detection performance of ECT. Wu, Y. et al. [10] chose the improved faster R-CNN deep learning model to classify steel plate surface defects. It notably increased the classification accuracy of small defects. These improved methods achieve high detection accuracy, but they have difficulty meeting the real-time detection requirements for surface defects on steel strips. Compared to two-stage algorithms, one-stage algorithms treat object detection tasks as regression tasks and retrieve all objects through one-step inference, but at a faster speed. There are obvious advantages of real-time detection, but the accuracy is slightly inferior. However, with the development of one-stage object detection technology, its performance has approached or even surpassed that of two-stage algorithms. In the field of defect detection, YOLO algorithms are popular for detection. Wang Hongyao [11] proposed an improved YOLO5 algorithm that uses a fast adaptive weighted median filter for image preprocessing, improving the recognition accuracy of wire rope damage images. Zhao et al. [12] proposed a model, named RDD-YOLO, based on YOLOv5 for steel surface defect detection. The extensive experimental results illustrate that the accuracies of RDD-YOLO reach 81.1 mAP on NEU-DET and 75.2 mAP on GC10-DET, improvements of 4.3% and 5.8% over YOLOv5. Although the method has improved the accuracy of surface defect detection on strip steel to a certain extent, it still faces problems such as interference from complex backgrounds of defect targets, small targets, and variable scales. This results in missed and false detections, especially for industrial workpieces such as steel plates with highly textured or reflective surfaces [13]. However, traditional deep learning-based methods, including CNN, faster R-CNN, and standard YOLO models, often fail to distinguish between actual defects and naturally occurring texture patterns [14].

Therefore, in order to solve the above problems, this article improves the YOLOv8n model as the benchmark. We propose a dual improvement strategy. First, we introduce Channel-Spatial Modulation Attention (CASM), a parallel attention mechanism combining channel and spatial information with an adaptive fusion gate, enhancing the network's sensitivity to critical defect features while suppressing background noise. Second, we design Small-Scale Grid Texture Shuffling Augmentation (SG-TSA), a targeted data augmentation method that perturbs defect region textures at the micro-structure level, increasing feature diversity and improving robustness to intra-class variation.

2. Related Work

In recent years, with the swift development of deep learning and the successful implementation of object detection tasks in numerous fields, the object detection technology founded on deep learning has also achieved significant advancements. Depending on the diverse kinds of data labels, defect detection grounded in deep learning can be classified into supervised learning models, unsupervised learning models, and semi-supervised learning models [15]. Among these, the supervision method is currently the most commonly use.

Recently, unsupervised and self-supervised anomaly detection (UAD) has advanced rapidly for industrial inspection. Unsupervised learning approaches employ a small amount of labeled data and a large quantity of unlabeled data for training to boost the predictive capability of the model, thus helping to lessen human involvement in training the model. Jun Long et al. [16] proposed a finer-grained self-supervised proxy task to extract differences between normal and abnormal patterns via augmented patch segmentation; SSAPS achieves excellent performance in surface defect detection and localization for high-resolution industrial images without matching simulated anomalies to real scenarios. Daitao Wang et al. [17] put forward a pixel-level real-time detection scheme based on autoencoder morphology to tackle the issue of insufficient sample data and high-cost

datasets due to the need for high-precision marking of a large number of abnormal samples by current deep learning algorithms in textile defect detection. Experimental results show effective detection when using the dataset. Compared with methods such as RCNN and YOLO, the training process greatly reduces errors. Moreover, the accuracy and F1-score both exceed 0.95 on four different datasets, and its FPS is 36.2. Sa Xiao et al. [18] proposed a lightweight CNN based on MobileNetV3, which can be trained with only a few normal extract feature gradients input images, achieving 99% accuracy. This algorithm is trained by using normal samples to obtain an overfitting model. Then, all kinds of images are fed into the overfitting model.

Unsupervised learning falls between fully supervised learning and semi-supervised learning. Compared to supervised learning models, it uses less labeled data for training and uses data with noisy labels. These labels are typically generated by computers, which combine unlabeled data with signals using heuristic methods to create their labels. The disadvantage is that this requires a large amount of unlabeled data, and the model may overfit the unlabeled data, resulting in inaccurate prediction results. It also cannot process data of unknown categories. Therefore, this article uses the YOLO series as the baseline model.

2.1. Deep Learning-Based Object Detection

YOLOv8 features a novel backbone using C2f modules, a fully anchor-free decoupled detection head, and a new label assignment strategy (TAL). These improvements make YOLOv8 highly competitive in both speed and precision. YOLOv8n is the lightest version in the YOLOv8 series, and due to its minimal parameters and computational footprint, it is naturally suitable for edge deployment. In actual industrial strip surface defect detection, only lightweight models are often required, so this article uses YOLOv8n as the benchmark model.

2.2. Attention Mechanisms in Object Detection

In the field of computer vision, the emergence of attention mechanism technology has brought new opportunities to improve the performance of models. Attention mechanism can help the model allocate more “attention” to the parts that may have defects, improving the accuracy and efficiency of detection [15]. Typical attention mechanism algorithms include Squeeze-and-Excitation (SE) [17], Efficient Channel Attention (ECA) [18], and CBAM [19]. In addition, there are some other improved versions available [20–23]. A novel attention mechanism combining Shuffle Attention (SA) and SENetV2 (SaE) was proposed by Xiang Li et al. [24], which integrates channel–spatial feature shuffling with channel–global feature optimization into YOLOv8n, thereby enhancing feature extraction capability and improving the classification accuracy of steel. Song X. et al. [25] integrated the polarized self-attention (PSA) mechanism to improve the model’s feature representation and enhance its ability to focus on relevant information. In addition, we incorporated the BAM attention mechanism into the C2F module to enhance the feature selection function of the model. Ching-Chi Yeung et al. [26] proposed a fused-attention module (FAM) to address the shape variations of defects. This module can enhance the channel and spatial feature information to perform precise localization and classification of defects with shape variations. In order to improve the accuracy of steel defect detection, Shi, J. et al. [27] introduced SENet, CBAM, and ECANet attention mechanism algorithms on the basis of YOLOv5. Experimental results showed that the introduction of SENet, CBAM, and ECANet improved the average accuracy by 4.05%, 4.57%, and 2.83%, respectively.

However, conventional attention mechanisms often treat channel and spatial attention sequentially, which may lead to suboptimal feature fusion. Our proposed CASM addresses this limitation by processing channel and spatial attention in parallel and combining them through an adaptive fusion gate. This design allows the network to dynamically balance the contributions of channel and spatial cues based on the input feature characteristics, which is particularly beneficial for detecting low-contrast defects against complex backgrounds.

2.3. Data Augmentation for Defect Detection

Data augmentation is essential in industrial defect detection due to the limited availability of defect samples and the significant intra-class variation in real-world conditions. Common augmentation strategies include geometric transformations (rotation, scaling, flipping) [28], photometric adjustments (brightness, contrast, hue changes), and advanced synthetic approaches like MixUp, CutMix, and Mosaic, which combine multiple images to improve robustness.

By using defect free steel images and masks containing real defects, the defects in the steel images were simulated, and a synthetic defect (SDG) technique was proposed by Guclu, E. and Akin, E., [29] to enhance the dataset and improve model robustness. Yang, B. et al. [30] proposed a new data augmentation algorithm called Mask2Defect. Through knowledge-based data injection, this method can generate defects with multiple functionalities. It can generate a large number of defects with different shapes, severity, scales, rotation angles, spatial positions, and part numbers in a controllable manner, improving the performance of defect classification and localization inspection models. Ran, G. et al [31] proposed an image synthesis framework based on Generative Adversarial Networks (GANs). Defect edges and background texture information are selectively extracted from the original image and used as network inputs through a spatial adaptive (DE) normalization (Spade) module. This enriches the input information, thus greatly reducing the amount of training data for GANs in image generation and enhancing the background details and defect boundaries in the generated images. Other scholars have also proposed some data augmentation algorithms [32–34].

For small-scale defects, localized augmentations can improve feature diversity more effectively than global transformations. Grid-based methods, such as GridMask, randomly mask out square regions in images to encourage the model to learn more discriminative features. Building upon this idea, our SG-TSA focuses on perturbations within defect regions, modifying local textures to increase intra-class diversity while preserving global structure. This targeted augmentation strategy is particularly effective for small and texture-sensitive defect detection.

3. Materials and Methods

3.1. Overall Framework

The proposed method is built upon the YOLOv8n architecture, which consists of three main components: the backbone for feature extraction, the neck for feature aggregation, and the head for multi-scale prediction. To enhance the model's ability to capture small and low-contrast defect features, we integrated the Channel–Spatial Modulation Attention (CASM) module into multiple feature extraction stages and introduced a targeted data augmentation strategy, Small-Scale Grid Texture Shuffling Augmentation (SG-TSA), during training.

CASM modules are inserted after the C2f blocks in the P3, P4, and P5 feature layers of the backbone. This placement ensures that both shallow and deep features benefit from improved attention to defect-relevant patterns. SG-TSA, on the other hand, operates at the

data preprocessing stage, selectively perturbing local textures in defect regions to increase feature diversity without altering the global structure.

The combination of CASM and SG-TSA enables YOLOv8n to achieve higher detection accuracy while maintaining its lightweight and fast inference characteristics, making it suitable for real-time industrial defect detection.

3.2. Channel–Spatial Modulation Attention (CASM)

The CASM module is designed to simultaneously leverage channel and spatial cues while enabling adaptive fusion between them. Unlike CBAM, which applies channel attention followed by spatial attention sequentially, CASM processes them in parallel, thus preserving independent information streams before combination.

The CASM module is composed of two parallel submodules: Lite Channel Attention (LCA) and Guided Spatial Attention (GSA). A fusion modulation unit is integrated at the end to dynamically balance and combine the contributions of channel and spatial features through a soft attention weighting strategy. Figure 1 presents the structural diagram of CASM.

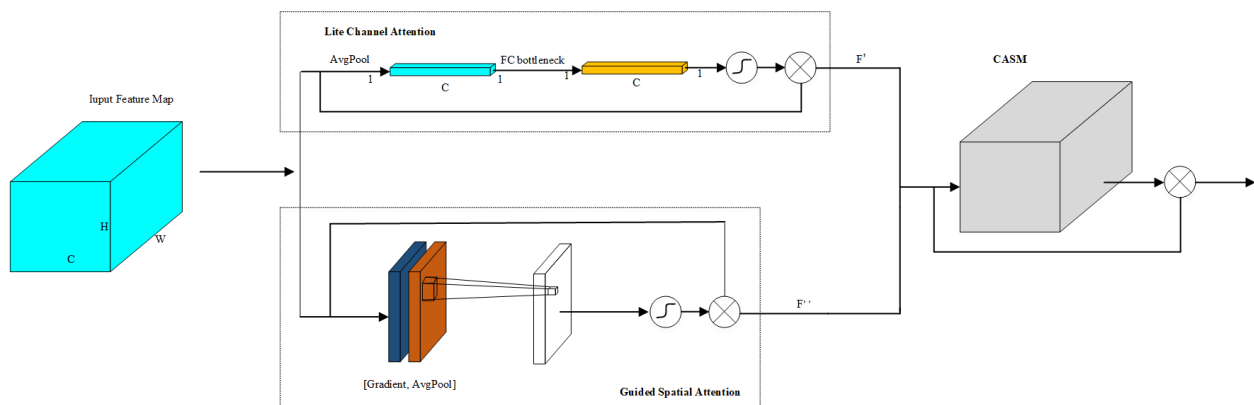


Figure 1. Structural diagram of CASM.

3.2.1. Lightweight Channel Attention (LCA)

Given the input feature map, LCA first applies global average pooling to summarize the feature distribution along the spatial dimension:

$$F_{avg}^c = Avgpool(F), \quad (1)$$

Then, a lightweight fully convolutional bottleneck is applied to introduce non-linearity and inter-channel dependencies:

$$M_{LCA} = \sigma(W_2 \cdot ReLU(W_1 \cdot F_{avg}^c)), \quad (2)$$

where

- $W_1 \in R^{C \times C/r}$, $W_2 \in R^{C/r \times C}$;
- r is the reduction ratio;
- $\sigma(x) = \frac{1}{1+e^{-x}}$ is the sigmoid activation function.

The LCA attention-refined feature is obtained via element-wise multiplication:

$$F' = F \otimes M_{LCA}, \quad (3)$$

This module consists of two fully connected layers and two excitation functions: the first fully connected layer and ReLU function achieve dimensionality reduction, reduce parameter and computational complexity, and introduce non-linear mapping to enhance

the model's ability to express complex features. The second fully connected layer and Sigmoid function restore the dimensionality-reduced features to the original number of channels, generating the final channel attention weights.

3.2.2. Guided Spatial Attention (GSA)

GSA is designed to highlight important spatial locations by combining average activation and texture gradient information. Given the same input, the following steps are taken:

Compute average feature activation across channels:

$$F_{avg}^S = \frac{1}{C} \sum_{i=1}^C F_i, \quad (4)$$

Compute texture gradients along the vertical axis:

$$F_{grad} = |F[:, :, 1, :, :] - F[:, :, :, -1, :]|, \quad (5)$$

$$F_{grad}^S = \text{Mean}(F_{grad}, \text{dim} = 1), \quad (6)$$

Concatenate and convolve:

$$M_{GSA} = \sigma(\text{Conv}_k(\text{Concat}(F_{avg}^S, F_{grad}^S))), \quad (7)$$

where

- Conv_k denotes a convolution with kernel size k ;
- The concatenated result has 2 channels and is reduced to 1 channel via convolution.

The spatially refined feature map is

$$F'' = F \otimes M_{GSA}, \quad (8)$$

3.2.3. Attention Fusion Modulation

To effectively integrate the refined channel and spatial features F' and F'' , CASM introduces a learnable fusion strategy:

Stack both outputs:

$$F_{stack} = \text{Stack}(F', F'') \in R^{2 \times C \times H \times W}, \quad (9)$$

Compute soft weights via global averaging and softmax:

$$\alpha = \text{Softmax}(\text{GAP}(F_{stack})) \in R^{2 \times 1 \times 1 \times 1}, \quad (10)$$

Perform weighted fusion:

$$F_{CASM} = \alpha_1 \cdot F' + \alpha_2 \cdot F'', \quad (11)$$

where

- $\alpha_1 + \alpha_2 = 1$, automatically balancing channel vs. spatial emphasis;
- This fusion strategy enhances the adaptive feature selection capability of the attention mechanism.

3.3. Small-Scale Grid Texture Shuffling Augmentation

SG-TSA is a targeted augmentation strategy designed to increase intra-class variation for defect samples, particularly small or texture-sensitive defects. This method enhances the diversity and saliency of small defects during training by applying texture-level perturbations within the defect region, while preserving the overall location and boundary structure of the object.

The key idea behind SG-TSA is to divide a defect region into multiple small grids, shuffle the positions of these grids randomly, and reconstruct the region with altered internal textures. Unlike traditional geometric transformations or pixel-level perturbations, SG-TSA focuses on fine-grained texture-level disturbances that simulate real-world industrial variations, such as surface inconsistencies, minor occlusions, or material shifts. The pseudocode is as follows:

Pseudocode

Input:

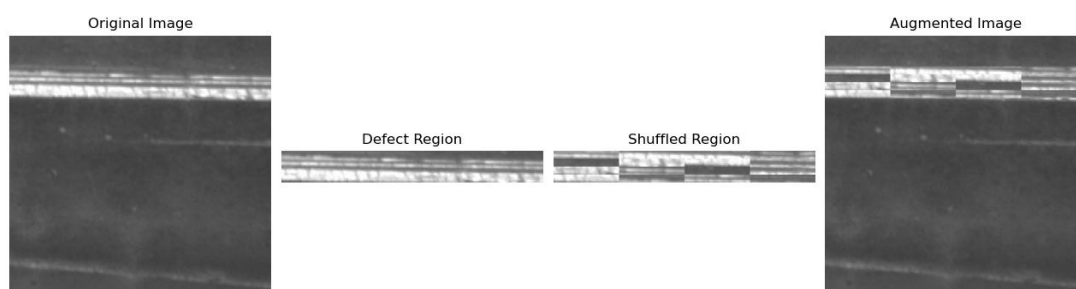
- Original image I of size $H \times W$
- Bounding box $B = (x_{\min}, y_{\min}, x_{\max}, y_{\max})$
- Grid division size $G = (4, 4)$

Output:

- Augmented image I_{aug}
-

1. Extract the defect region $R = I[y_{\min}:y_{\max}, x_{\min}:x_{\max}]$
 2. Get height h and width w of R
 3. Compute patch height: $\text{step}_h = \text{floor}(h/4)$, patch width: $\text{step}_w = \text{floor}(w/4)$
 4. Crop R to $(\text{step}_h \times 4, \text{step}_w \times 4)$
 5. Initialize empty patch list $P = []$
 6. For $i = 0$ to 3:
 - For $j = 0$ to 3:
 - $P_{ij} = R[i*\text{step}_h: (i+1)*\text{step}_h, j*\text{step}_w: (j+1)*\text{step}_w]$
 - Append P_{ij} to P
 7. Shuffle patch list P randomly
 8. Reconstruct R_{shuffled} by placing shuffled patches back in order
 9. Replace original region in I with $R_{\text{shuffled}} \rightarrow I_{\text{aug}}$
 10. Return I_{aug}
-

To facilitate a more intuitive understanding of the operating mechanism and augmentation strategy of SG-TSA, this study adopts a graphical representation to illustrate the complete process of the proposed Small-Scale Grid Texture Shuffling Augmentation (SG-TSA) algorithm. As shown in Figure 2, the flow proceeds from left to right through the following four stages: (1) the original defect image; (2) the defect region cropped based on YOLO-format annotations; (3) the texture-shuffled patch generated by partitioning the cropped region into a 4-by-4 grid and randomly permuting the patches; (4) the final augmented image with the disturbed region pasted back into its original position. This method enhances the local texture complexity of defect regions without altering the global structure or semantic consistency. This approach increases the visual variability of small target defects and mitigates the risk of overfitting to fixed texture patterns.



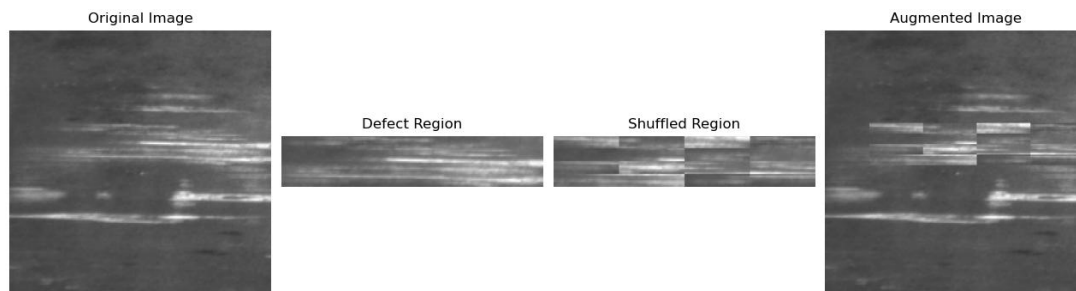


Figure 2. Visualization of the SG-TSA augmentation pipeline on defect images.

4. Results and Discussion

4.1. Dataset and Experimental Setup

We evaluated our proposed method on the NEU-DET dataset, a widely used benchmark for steel surface defect detection. The dataset contains 1800 grayscale images (300×300 pixels), evenly distributed among six defect categories: crazing (Cr), inclusion (In), patches (Pa), pitted surface (PS), rolled-in scale (RS), and scratches (Sc), with 300 samples per category.

4.2. Evaluation Metrics

To quantitatively evaluate detection performance, we adopted the COCO-style mean Average Precision (mAP) at multiple Intersection over Union (IoU) thresholds:

mAP@0.5: Average precision at IoU = 0.5, reflecting the detection capability with moderate localization accuracy.

mAP@0.5:0.95: Average precision over IoU thresholds from 0.5 to 0.95 in steps of 0.05, providing a comprehensive measure of both classification and localization quality.

We also report the Precision (P), Recall (R), and F1-score to assess the trade-off between false positives and false negatives.

4.3. Implementation Details

We evaluated the detection performance using the mean Average Precision (mAP) at two IoU thresholds: mAP@0.5 and mAP@0.5:0.95. Precision (P), Recall (R), and F1-score are also reported. We did not perform evaluation based on small, medium, and large object sizes, as the NEU-DET dataset consists of defect regions with relatively consistent scales and minimal size variation, making such categorization less informative.

The metrics are defined as

$$Precision = \frac{TP}{TP + FP'} \quad (12)$$

$$Recall = \frac{TP}{TP + FN'} \quad (13)$$

where TP , FP , and FN denote true positives, false positives, and false negatives, respectively.

4.4. Combined Effect of CASM and SG-TSA on YOLOv8n

To verify the synergistic gain effect of CASM module and SG-TSA data augmentation method in the YOLOv8n framework, this paper designed a stepwise introduction ablation experiment. Firstly, based on the original YOLOv8n, the performance differences of the three-stage models were compared by sequentially adding the CASM module and SG-TSA module. The contribution of each module to the improvement in detection accuracy was analyzed. The specific experimental results are shown in Table 1 and Figures 3–5.

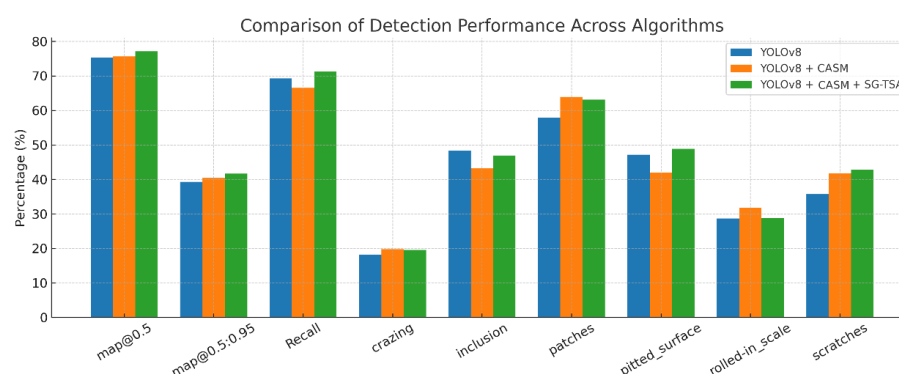
Table 1. Comparison of ablation experiment results.

Improvement	CASM	√		√
	SG-TSA			√
Performance	map50	75.27%	75.47%	76.28%
	map50-95	39.31%	40.9%	42.32%
	Recall	69.28%	69.72%	71.82%
	Crazing	18.16%	20.93%	20.92%
	Inclusion	48.35%	49.56%	44.05%
	Patches	57.91%	62.56%	64.48%
	Pitted surface	47.12%	46.58%	49.04%
	Rolled-in scale	28.64%	29.37%	28.4%
	Scratches	35.69%	36.41%	47.04%

As shown in Table 1, the original YOLOv8n achieved a mAP@0.5:0.95 of 39.31%. After incorporating the CASM module, the accuracy improved to 40.90%, and with the additional application of SG-TSA, the performance further increased to 42.32%. This highlights the effectiveness of the proposed improvements.

Figure 3 and Table 1 illustrate the ablation experimental results of integrating the proposed CASM module and SG-TSA augmentation method into the YOLOv8n detection framework. The experiments were conducted in three configurations: baseline YOLOv8n, YOLOv8n+CASM, and YOLOv8n+CASM+SG-TSA.

From the bar chart in Figure 3, it is clear that performance improvements are particularly evident in metrics such as Recall, and in categories like “patches” and “scratches”, which are characterized by small size or low contrast. This indicates that CASM enhances the attention to crucial features, while SG-TSA improves the model’s ability to perceive fine-grained textures and small-scale defects.

**Figure 3.** Bar chart comparison of detection performance across different algorithms.

To further analyze the effectiveness of the proposed CASM module, attention feature maps at different detection layers (P3, P4, and P5) were visualized for both the baseline YOLOv8 and the improved YOLOv8+CASM model, as shown in Figure 4. The input image used for visualization contains multiple linear scratch defects, which pose a challenge due to their subtle texture and elongated structure.

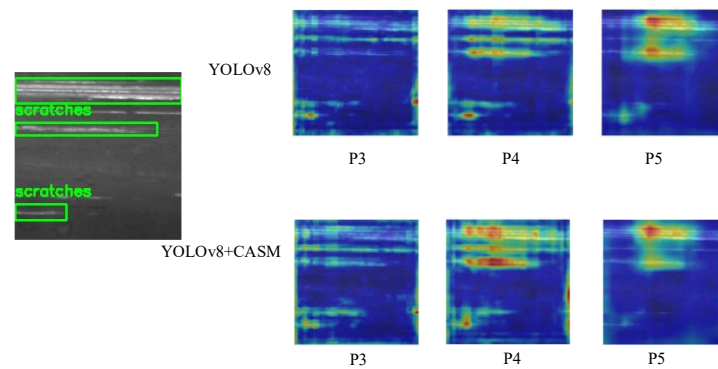


Figure 4. Comparison of feature heatmaps before and after using CASM.

In contrast, the YOLOv8+CASM model demonstrated significantly enhanced and more focused activation across all three layers. Especially in P4 and P5, the attention maps show clearer and more continuous responses along the defect regions. This suggests that the CASM module effectively enhances the model's capacity to capture structural and semantic cues of the defects, thereby improving detection.

These visual results corroborate the quantitative improvements observed in the evaluation metrics and demonstrate the practical benefit of integrating contextual attention mechanisms in defect detection tasks. These visual results corroborate the quantitative improvements observed in the evaluation metrics and demonstrate the practical benefit of integrating contextual attention mechanisms in defect detection tasks.

To evaluate the effectiveness of the proposed SG-TSA (Spatial–Global Transformer Self-Attention) module, we compared the attention feature maps of the YOLOv8+CASM baseline and the improved YOLOv8+CASM+SG-TSA model at the P3, P4, and P5 detection layers, as presented in Figure 5. The input image remained the same to ensure a fair comparison.

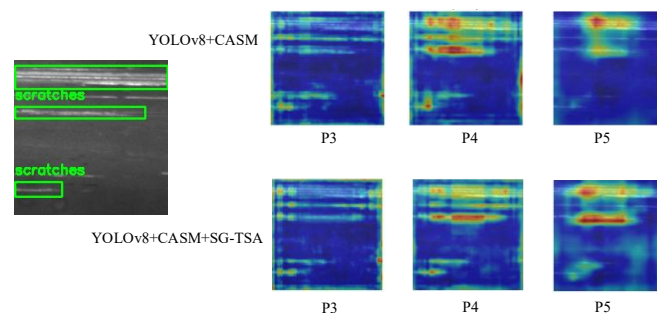


Figure 5. Comparison of feature heatmaps before and after using SG-TSA summary.

Compared to YOLOv8+CASM, the YOLOv8+CASM+SG-TSA model exhibits stronger, more concentrated, and better-aligned attention responses across all layers. Notably, the P4 and P5 feature maps display higher activation specifically around the defect areas, indicating that SG-TSA further enhances the model's ability to focus on subtle defect features by integrating spatial and global contextual information.

Additionally, the attention maps from YOLOv8+CASM+SG-TSA show reduced noise activation in non-defect regions, suggesting improved discrimination between defects and background textures. This refined attention helps minimize false positives and contributes to better detection performance, as also supported by quantitative results.

The experimental findings demonstrate that the proposed SG-TSA module effectively complements the CASM structure, leading to more robust and accurate defect detection in challenging scenarios.

To further verify the generality and adaptability of the SG-TSA data augmentation algorithm, this paper integrates the SG-TSA module with the traditional attention mechanism CBAM, forming a new enhanced detection model. By comparing the detection performance with YOLOv8n and YOLOv8n+CBAM, the wide applicability and effectiveness of SG-TSA are further confirmed, and the performance advantage of CASM module in feature selective modulation is indirectly reflected. The relevant experimental results are shown in Table 2 and Figures 6–8.

Table 2. The results of the ablation experiment.

Improvement	CBAM	√		√
	SG-TSA	√		
Performance	map50	75.27%	75.6%	77.12%
	map50-95	39.31%	40.4%	41.66%
	Recall	69.28%	66.53%	71.24%
	Crazing	18.16%	19.80%	19.55%
	Inclusion	48.35%	43.22%	46.96%
	Patches	57.91%	63.89%	63.10%
	Pitted surface	47.12%	42.02%	48.83%
	Rolled-in scale	28.64%	31.74%	28.79%
	Scratches	35.69%	41.71%	42.74%



Figure 6. Comparison of detection performance.

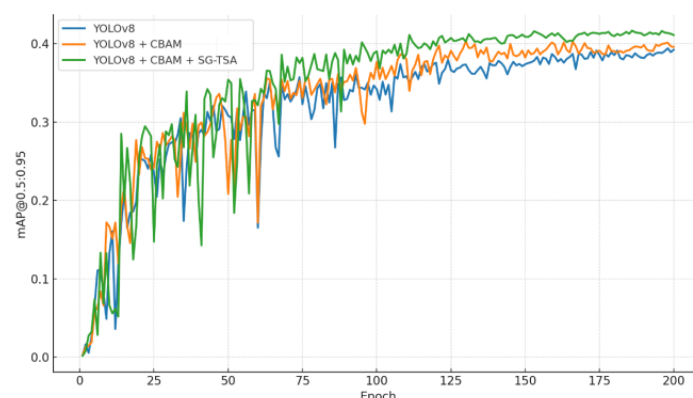


Figure 7. The mAP50-95 curve of the ablation experiment.

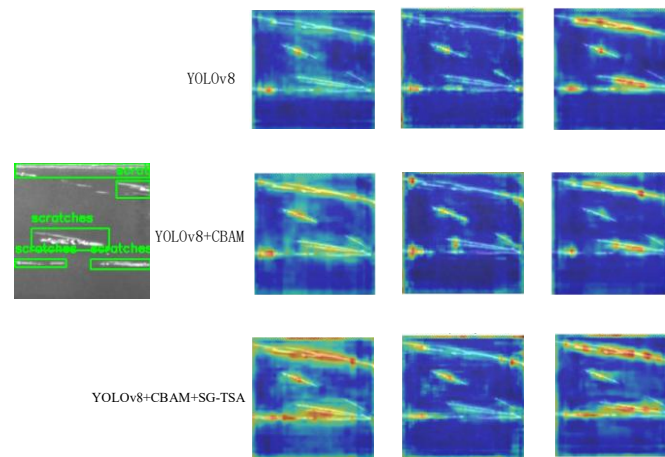


Figure 8. Comparison of feature heatmaps of using CBAM summary.

Overall, the incorporation of SG-TSA leads to consistent improvements across nearly all evaluation metrics. Specifically, mAP@0.5 increased from 75.6% to 77.12%, and mAP@0.5:0.95 rose from 40.4% to 41.66%, indicating better localization accuracy and robustness across IoU thresholds. Recall also improved from 66.53% to 71.24%, showing enhanced sensitivity in detecting true positives.

The quantitative results of the second ablation experiment are summarized in Table 2 and visualized in Figure 6. These results illustrate the impact of the SG-TSA augmentation strategy when applied to a YOLOv8n detection framework enhanced with the CBAM attention mechanism.

In terms of class-wise performance, the gains were observed in the detection of scratches (from 41.71% to 42.74%), inclusion defects (from 43.22% to 46.96%), and pitted surfaces (from 42.02% to 48.83%). These categories typically suffer from background interference or small-scale features, which validates the effectiveness of SG-TSA in enhancing texture diversity and local discriminability during training.

Although the accuracy of certain categories (such as rolling scale) has actually decreased (e.g., 31.74% to 28.79%), the overall trend confirms the universal applicability of SG-TSA in CBAM-based frameworks, further enhancing its value as a lightweight and transferable enhancement method.

These findings not only demonstrate the compatibility of SG-TSA with existing attention mechanisms like CBAM, but also indirectly support the claim that CASM offers more substantial performance gain when compared under identical augmentation strategies.

It is evident that the YOLOv8+CBAM+SG-TSA model not only achieves the highest final detection performance but also exhibits a faster and more stable convergence process. After approximately 100 epochs, the mAP of the SG-TSA-enhanced model stabilized above 41%, surpassing both the YOLOv8 and YOLOv8+CBAM baselines.

Moreover, the SG-TSA model demonstrated reduced fluctuation in intermediate epochs (e.g., between 50–150), indicating better robustness and generalization during training. In contrast, the base YOLOv8 and YOLOv8+CBAM showed greater performance jitter, especially in early stages, which may reflect sensitivity to local minima or noise in training data.

These findings provide further empirical evidence supporting the conclusion that SG-TSA not only enhances detection accuracy but also contributes to faster convergence and improved training stability, making it highly suitable for integration into lightweight industrial defect detection pipelines.

To further examine the synergistic effect of SG-TSA augmentation when applied to an attention-based detection framework, Figure 8 presents the attention heatmaps of three model variants—YOLOv8, YOLOv8+CBAM, and YOLOv8+CBAM+SG-TSA—across the P3, P4, and P5 stages. The test image contains multiple elongated, low-contrast scratch defects that challenge both spatial localization and feature discrimination.

In the baseline YOLOv8 model, the attention response is sparse and tends to miss critical regions of the scratches, especially in deeper layers (P5). While some activations align with defect locations, the responses are weak and scattered, indicating insufficient feature representation.

After integrating CBAM, the attention maps show notable improvements. The YOLOv8+CBAM model focuses more effectively on the horizontal scratch regions, especially in the P3 and P4 layers. However, background interference still remains evident, and some defect edges are not well captured.

The proposed YOLOv8+CBAM+SG-TSA configuration delivers the most focused and discriminative attention distribution. Across all three stages (P3–P5), the attention maps demonstrate dense, continuous activations that align well with the full extent of the scratch defects. Compared to the other two models, this configuration suppresses irrelevant background responses more effectively, while highlighting subtle textures with greater precision. These visual results confirm that SG-TSA improves data diversity during training, improving the model’s perception of fine-grained textures. The results confirm that SG-TSA is a transferable and universally applicable augmentation strategy.

In order to comprehensively and intuitively verify the advantages and disadvantages of the proposed algorithm in defect detection tasks, this section will design and conduct multiple sets of comparative experiments. These experiments will cover performance comparisons with classical benchmark models and mainstream attention mechanisms in order to systematically evaluate the advantages and disadvantages of our research method in terms of accuracy improvement and parameter quantity. The experimental results are shown in Table 3.

Table 3. Comparison of multiple experimental groups.

Algorithm	Map50-95	GFLOPs	Params
YOLOv8n	39.31%	8.1	3,012,018
YOLOv8n+CBAM	40.29%	8.3	3,098,776
YOLOv8n+ECA	39.86%	8.1	3,006,827
YOLOv8n+SE	41.05%	8.1	3,017,570
YOLOv8n+CBAM+SG-TSA	41.66%	8.3	3,098,776
YOLOv8n+CASM+SG-TSA	42.32%	8.2	3,023,540
YOLOv5n	39.50%	7.2	2,509,618
YOLOv5n+CASM+SG-TSA	41.12%	7.2	2,521,156

Table 3 presents the comparison results of ablation experiments conducted on the NEU-DET dataset with different improved algorithms. As shown, the baseline YOLOv8n model achieved only 39.31% mAP50-95. By incorporating CBAM and SG-TSA jointly, the performance improved to 41.66%, with only a slight increase in parameter size and computational cost. Furthermore, integrating CASM with SG-TSA into YOLOv8n boosted the mAP50-95 to 42.32%, demonstrating a significant improvement over the baseline, while maintaining a relatively low computational complexity (8.2 GFLOPs) and parameter count (approximately 3.02M). In addition, other lightweight attention mechanisms, such as ECA and SE, also showed better performance compared to the baseline YOLOv8n, but their improvement was not as substantial as CASM combined with SG-TSA. Moreover, applying CASM and SG-TSA to YOLOv5n achieved 41.12% mAP50-95, indicating the

generalization and transferability of the proposed method across different model architectures. In summary, the results in Table 3 clearly demonstrate that the proposed combination of CASM and SG-TSA significantly enhances defect detection accuracy while maintaining lightweight model characteristics, thus achieving a balance between accuracy and efficiency and highlighting both the advancement and practical applicability of the method.

To visually verify the advantages of the proposed CASM over the classical serial attention mechanism CBAM, this study conducted multi-scale visualization analysis on the feature heatmaps of YOLOv8+CASM, YOLOv8+CBAM, and the benchmark YOLOv8. The results are shown in Figures 4–9.

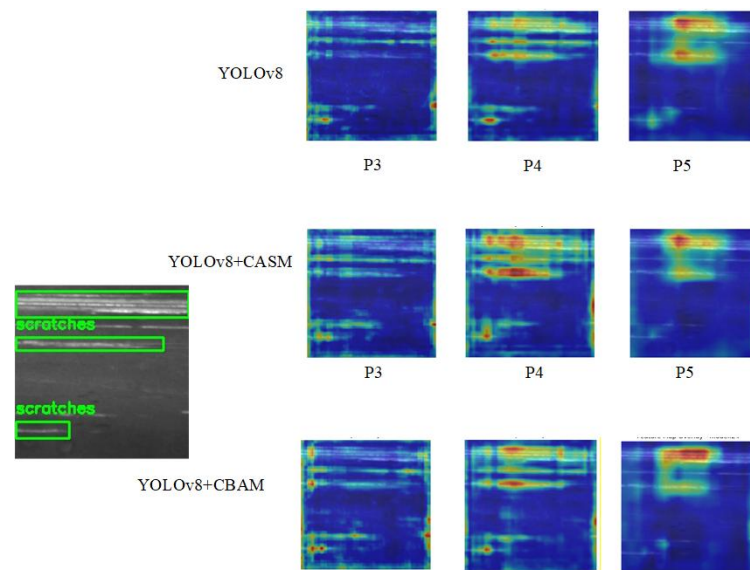


Figure 9. Comparison of feature heatmaps between CASM and CBAM.

In the feature map of benchmark YOLOv8, there is obvious background interference in the high-response area (such as redundant response in the blue background area of P3/P4). Although the feature response of YOLOv8+CBAM has been enhanced, the high-response areas exhibit a “scattered” feature (such as the deviation of the response point from the scratch area in P3, and the yellow area in P5 heatmap exceeding the scratch boundary of the original image), resulting in insufficient positioning accuracy for slender scratches. In the feature maps of YOLOv8+CASM, the red/yellow high-response areas completely overlap with the scratch defects in the original image, while the non-defect background areas always maintain low response, proving that CASM can accurately focus on the fine-grained features of real defects and avoid interference from background textures.

Compared to YOLOv8+CBM, the color contrast between “high response (red/yellow)” and “low response (blue)” in the feature map of YOLOv8+CASM is significantly improved: the response intensity in the defect area is higher, and the response suppression in the background area is more thorough. The essence of this difference is that CBAM adopts a serial process of “channel attention→spatial attention”, and there is information loss in the transmission of attention weights; The parallel attention mechanism of CASM can simultaneously weight and enhance the defect semantics in the channel dimension and the defect positions in the spatial dimension, thereby maximizing the feature differences between defects and the background and providing more discriminative feature representations for subsequent defect detection.

In summary, through the visualization comparison of multi-scale feature heatmaps, it can be intuitively proved that the proposed CASM has significant advantages over the classical serial attention mechanism CBAM in the three core dimensions of attention focusing accuracy, multi-scale feature consistency, and defect background discrimination for fine-grained defects. This also provides a feature level explanation for the improvement of YOLOv8+CASM in detection accuracy in subsequent experiments.

The YOLO series has a new version, and to demonstrate the generality and portability of the proposed method across different model architectures, a comparison was made specifically between the more mature Yolov11n and Yolov12n. The results are shown in Table 4. It can be clearly seen that the proposed combination of CASM and SG-TSA significantly improves the accuracy of defect detection, with high universality and portability. It should be noted that the following experimental results were conducted on the NVIDIA GeForce RTX 4070 Ti SUPER platform.

Table 4. Generalization of CASM+SG-TSA across different detectors.

Algorithm	Map50	Map50-95	Recall	FPS
YOLOv11n	76.69%	41.23%	75.14%	103.26
YOLOv11n+CASM+SG-TSA	78.32%	42.03%	76.03%	93.05
YOLOv12n	74.97%	40.26%	71.44%	89.15
YOLOv12n+CASM+SG-TSA	78.32%	42.88%	74.47%	81.45
YOLOv12n+CASM	75.92%	40.60%	74.59%	—
YOLOv12n+BAM	75.11%	40.16%	71.52%	70.08
RT-DETR	68.43%	35.99%	65.02%	—

To support the real-time claim, we report the actual inference throughput (FPS) measured. The results are summarized in Table 4. The baseline YOLOv11n and YOLOv12n achieved 103.26 FPS and 89.15 FPS, respectively. After integrating CASM and SG-TSA, YOLOv11n+CASM+SG-TSA maintained 93.05 FPS, while YOLOv12n+CASM+SG-TSA ran at 81.45 FPS. Although the proposed enhancements introduce a moderate speed reduction due to additional attention computation and post-processing, the resulting throughput remains within a practical real-time range for online defect inspection, while delivering consistent accuracy improvements across detectors.

To verify the effectiveness of the proposed SG-TSA, we used YOLOv12n and CASM as baselines and only replaced the data augmentation strategy while maintaining consistent training settings. We conducted comparative experiments between SG-TSA and typical general augmentation methods Mosaic and MixUp, and the results are shown in Table 5. It can be seen that compared to Mosaic and MixUp, SG-TSA achieved better (or more competitive) detection performance on the NEU-DET dataset, indicating that in defect detection scenarios, local texture perturbations targeting defect ROIs can more effectively enhance intra-class diversity while maintaining the overall structural information of defects, thereby bringing more stable performance gains.

Table 5. Augmentation comparison on NEU-DET (YOLOv12n baseline).

Algorithm	Map50	Map50-95	Recall
YOLOv12n+CASM+SG-TSA	78.32%	42.88%	74.47%
YOLOv12n+CASM+MixUP	75.97%	41.12%	73.13%
YOLOv12n+CASM+Mosaic	75.97%	40.60%	74.59%

To further validate the generalization performance of the proposed algorithm and improve the reliability and universality of the research conclusions, this study conducted

supplementary experiments on a second independent dataset of metal surface defects (DAGM 2007) [35]. DAGM 2007 [35] is a synthetic benchmark dataset for industrial surface defect detection, consisting of 10 artificially generated defects (Class 1–Class 10), each corresponding to a different texture background and defect morphology, used to simulate surface anomalies in real industrial scenes. There are differences between this dataset and the first experimental dataset in terms of data distribution, defect type coverage, and image acquisition environment. Through cross-dataset validation, the adaptability and performance stability of the algorithm can be effectively tested in different data scenarios. This experiment was also based on the NVIDIA GeForce RTX 4070 Ti SUPER platform.

As shown in Table 6, the proposed algorithm has significant advantages in comprehensive indicators, with mAP50 reaching 98.2% and mAP50-95 reaching 69.1%. Compared with the original YOLOv8n algorithm (mAP50 of 95.3%, mAP50-95 of 64.3%) and only adding CASM algorithm (mAP50 of 96.4%, mAP50-95 of 67.2%), both achieved improvements.

Table 6. Augmentation comparison for defect location on metal surfaces (YOLOv8n baseline).

Improvement	CASM		√	√
	SG-TSA			√
Performance	map50	95.3%	96.4%	98.2%
	map50-95	64.3%	67.2%	69.1%
	Recall	94%	92.6%	96.9%
	Class1	58.3%	67.1%	70.2%
	Class2	72%	77.8%	78.5%
	Class3	66.3%	76.7%	62%
	Class4	48.3%	45.3%	58.4%
	Class5	71.2%	70.9%	71.1%
	Class6	52.1%	59.5%	68.9%
	Class7	65.2%	64%	69.1%
	Class8	57.8%	58.4%	61.5%
	Class9	75.6%	76.3%	74.9%
	Class10	76%	76.3%	76.9%

In summary, both on the first dataset and the second dataset used in this article, the proposed algorithm demonstrated superior performance compared to the other algorithms in terms of comprehensive detection metrics and recognition accuracy for most single-class defects, fully verifying the effectiveness and generalization ability of the algorithm.

5. Conclusions

This paper presents a defect detection framework that enhances the YOLOv8n model with two key innovations: the Channel–Spatial Modulation Attention (CASM) module and the Small-Scale Grid Texture Shuffling Augmentation (SG-TSA) strategy. (1) CASM improves feature representation by parallelizing channel and spatial attention computations and adaptively fusing their outputs, enabling better discrimination between defects and background noise. (2) SG-TSA augments defect samples by locally perturbing textures within defect regions, increasing intra-class diversity without altering the global structure. (3) Evaluated on the NEU-DET dataset, the proposed YOLOv8n+CASM+SG-TSA outperforms the baseline YOLOv8n and YOLOv8n+CBAM variants, achieving a 2.84% increase in mAP@0.5 and a 3.01% increase in mAP@0.5:0.95, while maintaining real-time inference speeds.

In order to further characterize the portability and robustness of CASM and SG-TSA, this paper also conducted validation experiments on the DAGM 2007 [35] dataset. The experimental results show that the proposed algorithm exhibited better performance in defect detection and recognition accuracy, fully verifying the effectiveness and generalization ability of the algorithm. Future work will explore adaptive enhancement strategies that adjust disturbance intensity based on defect type, as well as lightweight attention variants tailored for ultra-low power edge devices.

Author Contributions: Methodology, D.A., N.K.W. and F.C.; Software, D.A.; Validation, D.A.; Formal analysis, D.A.; Investigation, D.A. and N.K.W.; Writing—original draft, D.A.; Writing—review & editing, N.K.W. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data is contained within the article.

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

References

1. Wang, T.; Chen, Y.; Qiao, M.; Snoussi, H. A fast and robust convolutional neural network-based defect detection model in product quality control. *Int. J. Adv. Manuf. Technol.* **2018**, *94*, 3465–3471.
2. Rao, X.; Zhang, F.; Lu, Y.; Luo, X.; Chen, F. Surface and subsurface damage of reaction-bonded silicon carbide induced by electrical discharge diamond grinding. *Int. J. Mach. Tools Manuf. Des. Res. Appl.* **2020**, *154*, 103564.
3. Hong, B.B. Research on Intelligent Detection Technology of Flat Workpiece Based on Deep Learning. Master's Thesis, Xi'an Polytechnic University, Xi'an, China, 2019. <https://doi.org/10.27390/d.cnki.gxbfc.2019.000133>.
4. Xia, B.; Luo, H.; Shi, S. Improved Faster R-CNN Based Surface Defect Detection Algorithm for Plates. *Comput. Intell. Neurosci.* **2022**, *2022*, 3248722. <https://doi.org/10.1155/2022/3248722>.
5. Xu, S.; Shao, Z.J. An improved faster RCNN based on Swin transformer for surface defect detection of metal workpieces. In Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering, Xiamen, China, 21–23 October 2022.
6. He, L.; Zheng, L.; Xiong, J. FMV-YOLO: A steel surface defect detection algorithm for real-world scenarios. *Electronics* **2025**, *14*, 1143.
7. Chen, Y.; Zhang, H. Research on PCB defect detection based on SSD. In Proceedings of the 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Dali, China, 12–14 October 2022; IEEE: Washington, DC, USA, 2022; pp. 1–5.
8. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
9. Li, J.; Zhang, L.; Zheng, W. Improved faster R-CNN and adaptive Canny algorithm for defect detection using eddy current thermography. *AIP Adv.* **2024**, *14*, 025310.
10. Shi, X.; Zhou, S.; Tai, Y.; Wang, J.; Wu, S.; Liu, J.; Xu, K.; Peng, T.; Zhang, Z. An improved faster R-CNN for steel surface defect detection. In Proceedings of the 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), Shanghai, China, 26–28 September 2022.
11. Wang, H.; Han, S.; Li, Q. Research on Experimental Method for Image Recognition of Wire Rope Damage Using Improved YOLOv5. *J. Comput. Eng. Appl.* **2023**, *59*, 99.
12. Zhao, C.; Shu, X.; Yan, X.; Zuo, X.; Zhu, F. RDD-YOLO: A modified YOLO for detection of steel surface defects. *Measurement* **2023**, *214*, 112776.
13. Khan, I.U.; Aslam, N.; Aboulmour, M.; Bashamakh, A.; Alghool, F.; Alsawayan, N.; Alturaif, R.; Gull, H.; Iqbal, S.Z.; Hussain, T. Deep Learning-Based Surface Defect Detection in Steel Products Using Convolutional Neural Networks. *Math. Model. Eng. Probl.* **2024**, *11*, 3006–3014.
14. Yang, J.; Li, S.; Wang, Z.; Dong, H.; Wang, J.; Tang, S. Using deep learning to detect defects in manufacturing: A comprehensive survey and current challenges. *Materials* **2020**, *13*, 5755.

15. Lin, Y.; Xiang, Y.; Lin, Y.; Yu, J. Defect detection system for optical element surface based on machine vision. In Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 28–30 September 2019; IEEE: Washington, DC, USA, 2019; pp. 415–418.
16. Long, J.; Yang, Y.; Hua, L.; Ou, Y. Self-supervised augmented patches segmentation for anomaly detection. In Proceedings of the Asian Conference on Computer Vision 2023, Macao, China, 4–8 December 2022.
17. Wang, D.; Yu, W.; Lian, P.; Zhang, M. Textile Defect Detection Algorithm Based on Unsupervised Learning. In Proceedings of the 2022 7th International Conference on Image, Vision and Computing (ICIVC) 2022, Xi'an, China, 26–28 July 2022. <https://doi.org/10.1109/icivc55077.2022.9887216>.
18. Xiao, S.; Liu, Z.; Yan, Z.; Wang, M. Grad-MobileNet: A Gradient-Based Unsupervised Learning Method for Laser Welding Surface Defect Classification. *Sensors* **2023**, *23*, 4563. <https://doi.org/10.3390/s23094563>.
19. Roy, A.G.; Navab, N.; Wachinger, C. Concurrent spatial and channel ‘squeeze & excitation’ in fully convolutional networks. In Proceedings of the Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, 16–20 September 2018.
20. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
21. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.
22. Lu, X.-Y.; Qu, M.-X. Steel surface defect detection based on improved YOLOv8. *Eng. Res. Express* **2025**, *7*, 015262.
23. Chen, Y.; Tang, Y.; Hao, H.; Zhou, J.; Yuan, H.; Zhang, Y.; Zhao, Y. AMFF-YOLOx: Towards an attention mechanism and multiple feature fusion based on YOLOx for industrial defect detection. *Electronics* **2023**, *12*, 20.
24. Ma, S.; Zhao, X.; Wan, L.; Zhang, Y.; Gao, H. A lightweight algorithm for steel surface defect detection using improved yolov8. *Sci. Rep.* **2025**, *15*, 8966.
25. Zhou, Z.; Yuan, M. Steelguard-yolo: Steel surface defect detection network based on improved yolov5s. *Rev. Int. De Metodos Numer. Para. Calc. Y. Diseno En. Ing.* **2024**, *40*, 29.
26. Ding, Z.; Zhao, J.; Misra, R.D.K.; Guo, F.; Xie, Z.; Wang, X.; Li, X.; Wang, J.; Shang, C. Deep Learning-Based Understanding of Defects in Continuous Casting Product. *Metals* **2023**, *13*, 1809. <https://doi.org/10.3390/met13111809>.
27. You, C.; Kong, H. Improved steel surface defect detection algorithm based on YOLOv8. *IEEE Access* **2024**, *12*, 99570–99577.
28. Song, X.; Cao, S.; Zhang, J.; Hou, Z. Steel surface defect detection algorithm based on YOLOv8. *Electronics* **2024**, *13*, 988.
29. Yeung, C.C.; Lam, K.M. Efficient fused-attention model for steel surface defect detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2510011.
30. Shi, J.; Yang, J.; Zhang, Y. Research on steel surface defect detection based on YOLOv5 with attention mechanism. *Electronics* **2022**, *11*, 3735.
31. Tang, B.; Chen, L.; Sun, W.; Lin, Z.K. Review of surface defect detection of steel products based on machine vision. *IET Image Process.* **2023**, *17*, 303–322.
32. Guclu, E.; Akin, E. Enhanced defect detection on steel surfaces using integrated residual refinement module with synthetic data augmentation. *Measurement* **2025**, *250*, 117136.
33. Yang, B.; Liu, Z.; Duan, G.; Tan, J. Mask2Defect: A prior knowledge-based data augmentation method for metal surface defect inspection. *IEEE Trans. Ind. Inform.* **2021**, *18*, 6743–6755.
34. Ran, G.; Yao, X.; Wang, K.; Ye, J.; Ou, S. Sketch-guided spatial adaptive normalization and high-level feature constraints based GAN image synthesis for steel strip defect detection data augmentation. *Meas. Sci. Technol.* **2024**, *35*, 045408.
35. Wieler, M.; Hahn, T.; Hamprecht, F.A. Weakly Supervised Learning for Industrial Optical Inspection. [Dataset]. 2007. Available online: <https://hci.iwr.uni-heidelberg.de/content/weakly-supervised-learning-industrial-optical-inspection> (accessed on 30 October 2025).

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.