SCIENTIFIC Publishing Limited

Digital Technologies Research and Applications
http://ojs.ukscip.com/index.php/dtra

Article

Research on Industrial Defect Machine Vision Detection Algorithm Based on Deep Learning and Spectral Fusion

Kaisheng Deng

School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China; 13879422779m@sina.cn

Received: 5 February 2026; **Revised:** 27 March 2026; **Accepted:** 6 April 2026; **Published:** 31 August 2026

Abstract: Industrial surface defect detection is core to intelligent manufacturing quality assurance, guaranteeing stable product quality and reliable production. Conventional defect detection deep learning algorithms relying on Red-Green-Blue (RGB) images suffer low accuracy and weak robustness amid complicated industrial scenes. This paper puts forward Multi-Spectral Fusion Defect Detection Network (MS-FDNet), a novel industrial surface defect detection framework combining deep learning and spectral imaging. It adopts a spectral-spatial attention coupling module to fully fuse multi-band spectral information and exploit distinctive feature advantages of each band. Self-calibrated convolution coupled with spatial pyramid structure constructs a multi-scale feature extraction branch to strengthen the model's perception of defects of varying sizes. A multi-task composite loss integrating Focal Loss and Dice Loss is designed to balance model memory occupation and detection precision and boost overall inference efficiency. Tests on the public NEU-det dataset reveal the proposed algorithm achieves a mean Average Precision at Intersection over Union threshold 0.5 (mAP@0.5) of 81.1%, a 4.6% improvement over the baseline while retaining fast inference. It runs at 67.9 frames per second with a tiny model volume of merely 4.22 MB, showing outstanding deployment practicability. The method greatly enhances defect identification performance and system stability under uneven illumination and cluttered backgrounds. This work offers an innovative industrial defect detection scheme with valuable theoretical research value and promising engineering application potential.

Keywords: Industrial Surface Defect Detection; Deep Learning; Spectral Fusion; Attention Mechanism; Lightweight Network

1. Introduction

1.1. Research Background and Significance

The inspection nodes for apparent defects in production systems are crucial for ensuring product compliance and accelerating manufacturing processes. Traditional manual inspection methods suffer from inherent limitations such as slow processing, ambiguous standards, and operator fatigue, making them incompatible with the high-speed, high-precision requirements of modern intelligent manufacturing [1]. With the rapid advancement of deep learning technologies, surface defect detection methods utilizing convolutional neural networks have demonstrated excellent feature extraction and pattern classification capabilities, achieving significant results across various manufacturing stages [2]. However, current RGB-based deep learning detection methods still face major challenges in complex industrial environments: Firstly, many surface defects lack distinctive visual features in the visible spectrum, particularly hidden defects beneath surfaces that traditional imaging methods struggle to detect effectively [3, 4]. Under industrial production conditions, factors such as light changes, complex background and noise interference significantly affect the accuracy and reliability of detection. These challenges have prompted researchers to combine multispectral imaging technology with deep learning. Due to the acquisition of spectral information of different wavelengths, multispectral imaging technology allows clear detection of differences between surface and internal features. For example, the reflection of the metal oxide layer in the visible range (400–700 nm) and the near-infrared range (900–1,700 nm) showed significant differences, while the cracks on the plastic surface showed a significant increase in the short-wave infrared range (Svir, 1,000–2,500 nm). These spectral changes provide a new technical method for detecting surface defects. By integrating multi-band spectral data, we can obtain a wider range of functions from the use of a single RGB image, which greatly improves the accuracy and reliability of damage detection.

This research has developed a scheme that uses deep learning and spectral fusion to detect industrial surface defects. The system adopts a separate strategy of multispectral element integration, improves the structure of the deep model, and provides an accurate and reliable diagnosis of industrial surface defects. Research and experimentation are essential theories and practices to accelerate the innovation of intelligent manufacturing quality control technology, as well as to optimize China's industrial automation and intelligence hierarchy.

1.2. Research Objectives

The main goal of this research is to develop an effective multispectral deep learning fusion system that expands the ability to extract spectral data from industrial surface control. Three major technological breakthroughs have been made: first, through the application of focus-based comprehensive data methods, the limitations of traditional deep learning in spectral data processing have been overcome. This process automatically prioritizes the spectral band to detect damage by optimizing the integration of spectral data. Secondly, we created a compact multi-step feature recognition network model to reduce system load while maintaining accuracy. Third, due to the self-calibration of convolution and the method of combining spatial pyramids, the network has significantly improved the efficiency of identifying defects of different sizes. In addition, we have developed a multitasking function to eliminate the error of uneven sample distribution, effectively eliminating the strong distortion ratio of positive and negative samples in industrial tracking. This optimization function combines the advantages of focus optimization and dice metric to provide more accurate defect detection. Finally, we created a comprehensive verification system through simulation, and proved the efficiency of the technology by analyzing manual data sets and actual production scenarios.

2. Current Research Status

Thanks to the combination of deep learning models and spectral data, breakthroughs have been made in the identification of surface defects in the industry. These advancements are mainly reflected in three core areas: multispectral data integration and performance optimization, spectral dynamic evolution deep learning architecture, and small sample detection technology. By integrating hyperspectral and multispectral terahertz imaging tools with advanced neural network solutions, coherent research has significantly improved accuracy, stability and reliability by providing reliable solutions for complex operating scenarios.

2.1. Research Progress on Lightweight Network Design and Attention Mechanism in Real-Time Defect Detection

The deep learning damage detection mechanism exhibits excellent adaptability in specialized production scenarios, allowing the network topology to be fine-tuned according to the different technical verification requirements of different industries. Raj et al. focused on the industrial weld defect detection task, systematically explored the performance balance strategy of multiple YOLO lightweight architectures in weld visual inspection, and analyzed the adaptability of different lightweight models in balancing detection speed and accuracy [5]. The research results provide practical technical guidance for the lightweight deployment of weld defect detection systems in industrial manufacturing. Vijayakumar et al. designed LightDefectNet-18, a universal multi-domain lightweight defect detection framework. Aiming at the common problems of poor generalization and excessive parameter volume of traditional detection models in cross-scene defect detection, the network adopts a streamlined structural design to realize efficient feature extraction for different types of defects, which can be widely applied to multi-industry defect detection scenarios and provides a universal lightweight solution for industrial visual inspection tasks [6].

2.2. Innovative Application of Spectral Detection Technology in Agricultural Product Quality Evaluation

Using fast and non-destructive spectral analysis, the crop quality control process meets advanced standards and is a reliable alternative to traditional manual inspections. Ding Ziyu's team used data mining and comprehensive learning methods to link the spectral parameters of corn seeds with germination ability, allowing real-time classification of grain varieties [7]. Gao Hongsheng's team used hyperspectral imaging band combination technology to conduct preliminary group detection of kiwi soft rot disease, and applied robust spectral feature separation and aggregation technology to achieve accurate coding of early lesions [8]. These examples show that the combination of spectral analysis technology with traditional classifiers and deep neural networks has a unique ability to verify the quality of on-site production, which is necessary for evaluating internal compliance and early detection of damage.

2.3. Research on Defect Detection of Multispectral and Hyperspectral Image Fusion Technology

The integration of multispectral and hyperspectral image synthesis technologies that combine spectral information of different wavelengths has significantly improved the reliability and continuity of defect detection, and has shown extraordinary value in production quality control. Deng and his team focus on the technology of controlling surface defects of chip components, and effectively improve anomaly identification through the synergistic benefits of multi-scale data [9]. Chen et al.'s research group has developed a multispectral method of visual signal integration to solve the surface damage of microelectronic components and improve the reliability and extensibility of equipment through optimized signal combination strategies [10]. The Malan Technology Alliance conducted in-depth research on multispectral and hyperspectral imaging coupling technologies, and compared the effectiveness of various hybrid methods for further development [11]. Malan's team systematically analyzed the research status of these image merging technologies, summarized their respective advantages and limitations, and put forward suggestions for future research [12]. These studies have shown that multispectral and hyperspectral synthesis technologies can make full use of spectral information of different wavelengths, have unique advantages in micro-defect detection and early damage identification, and are the key development direction of industrial damage detection.

2.4. Innovation in Feature Extraction and Fusion Strategies for Complex Industrial Scenarios

Through innovative technologies for identifying and combining data, the system has shown higher efficiency in detecting errors in complex production scenarios. By integrating cross-distributed data, more accurate fault identification is achieved. Chen et al.'s research developed a system for detecting errors in the composition of computer motherboards, which combines a parallel collection of elements with a gradual coupling of elements. This technology effectively improves the ability to detect complex defects through the multi-layer integration of components. Maintaining high-resolution reliability, the process achieves a scientific balance between calculating load and detection efficiency, and provides a practical method for quality verification of high-quality electrical equipment [10]. Well-designed feature extraction and link path optimization have significantly improved the performance of machine learning models in complex industrial environments, and have shown significant efficiency in detecting defects of different sizes.

2.5. Systematic Research and Future Prospects of Industrial Defect Detection Technology

The comprehensive research on the technology of detecting manufacturing defects has laid an important scientific foundation for the modernization of industry. Experts from different countries conducted a systematic review and conducted a comprehensive analysis of the history and future direction of technological development. Zimmerman's research team conducted a detailed review of the process of detecting visual deviations and differentiation in industrial applications, and systematically compared the advantages and limitations of traditional methods with deep learning models [12]. Saberironaghi's team studied in detail the defect recognition technology of industrial products based on machine learning, focusing on the adaptability of different deep learning architectures to specific manufacturing processes [13]. Using category activation mapping technology, Shin's team identified defects in deep learning-based defect detection, and used imaging technology to explain these problems, making the model easier to understand [14]. Overall, industrial defect detection technology is becoming more and more intelligent, combining many technologies and high practicality. Future research should give priority to improving the accuracy and effectiveness of testing, while providing people with comprehensibility.

Through a systematic review of the literature, we conclude that deep learning technology and spectral synthesis have created a comprehensive technical foundation for detecting industrial damage. From deep learning-based methods [5,6] to spectral detection methods [7,8], as well as multispectral synthesis strategies and multi-functional combinations [9–11,15]. These different methods all show unique advantages and complement each other. International academic research provides key theoretical support for the development of this field [12–14]. Future research should focus on more effective integration of these technologies, especially in the development of convenient models, the collection of information from multiple sources, and the adaptation of them to different application fields. This integration will improve the accuracy, efficiency and practicality of detecting industrial defects.

3. Algorithm Design and Improvement

3.1. Multispectral Data Preprocessing and Feature Extraction

3.1.1. Spectral Calibration and Band Selection

The preparation of multispectral data is the main component of the detection system. This process initiates the spectral alignment of the original multispectral data to eliminate hardware and spatial interference, including background current adjustment, brightness alignment, and wavelength alignment to ensure that the spectral feedback matches. Range selection is a central task that requires analysis of material properties and defect structure to ensure the accuracy of identification, which makes it possible to identify the range with the most characteristic characteristics. In this study, a mechanism based on mutual information is used to measure the mutual information between the frequency band and the damage index, thereby showing that the frequency band with higher mutual information value forms a characteristic subset. The range of 450–950 nm shows excellent efficiency in the industrial control of metal surface damage, because different defects have different spectral characteristics in this range. Metal oxide surfaces have different reflection characteristics in the visible and near-infrared spectra, while surface cracks have high reflection at certain wavelengths [15].

3.1.2. Multi-Scale Feature Extraction Network Architecture

In order to reliably identify multi-layer anomalies, a compact multi-dimensional network model with feature sets has been developed in this research. The Iolov11 architecture using C3X2 components uses self-balancing convolution to optimize feature extraction. The self-calibrating convention method dynamically establishes spatial and channel correlation over long distances, expanding the coverage of information in spatial nodes. Unlike traditional conventional methods, this technique uses four unrelated modules to process filters, each of which specializes in capturing different types of signals. X input samples are divided into two independent blocks, each processing raw information from different spatial dimensions. In the network topology design, the intelligent balanced convolution is integrated into the original C3X2 architecture by applying the optimized C3X2 module. The module dynamically adapts to environmental parameters, while using hierarchical visual data, which improves the accuracy of parameter definition and the possibility of semantic representation. The spatial coupling of pyramid features and the ReLU fast process allows the use of multi-level perceptual data to improve the model's ability to identify defect patterns at different levels. The architecture of the network is shown in **Figure 1**. The proposed multispectral image is first feature isolated through a multi-layer convolution structure, and then the several components of SC-C3k2 are used for in-depth feature analysis. In the process of gradient formation, the system connecting the spatial pyramid recognizes both small details and macroscopic nutrients. Subsequent processing methods, including increasing the number of pixels and mixing objects, allow the creation of multi-directional blocks of objects for detection.

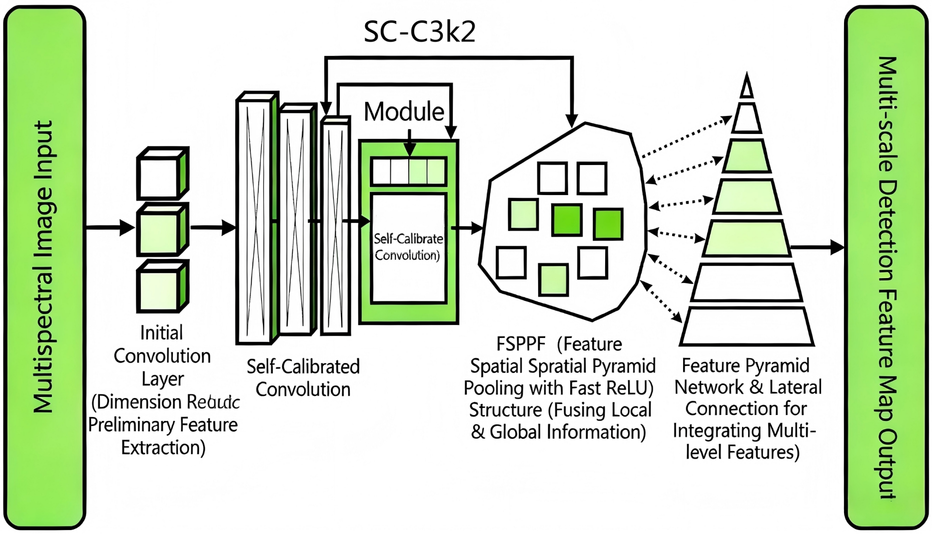

Figure 1. Architecture of Multi-Scale Feature Extraction Network.

3.2. Spectral-Spatial Attention Fusion Mechanism

3.2.1. Design of the Spectral Attention Mechanism

Spectral focusing mechanisms serve as the core approach for efficient multi-band information integration. Traditional stitching methods relying on fixed allocations or basic aggregation approaches struggle to fully analyze the unique functions of individual spectral bands [16]. This study introduces a dynamic spectral attention module that autonomously adjusts band component weights based on defect detection criteria. The system incorporates line coordination mechanisms to enhance multispectral data elements [17]. The detailed workflow involves: First, applying global mean pooling and global extreme value pooling to spectral band feature maps to obtain channel symbols. Then, a deep perception architecture built by Multilayer Perceptron (MLP) is adopted to convert the above feature sets and calculate core weight coefficients for each spectral band. Finally, correlation coefficients are applied to the original data matrix to reinforce the recognition of key wavebands and suppress signal noise introduced by irrelevant frequency bands.

The mathematical expression of the spectral attention mechanism is as follows:

Given an input feature map $F \in \mathbb{R}^{C\times H\times W} F_{avg} \in \mathbb{R}^{C} F_{\max} \in \mathbb{R}^{C}$ where C denotes the number of spectral bands, and H and W represent the height and width of the feature map respectively, the global average pooling operation generates channel descriptors, while the global maximum pooling operation produces feature representations. The attention weights are computed through an MLP.

$$M_c = \sigma(MLP(\mathrm{F_{avg}}) + MLP(F_{\max}))$$

The MLP architecture σ employs a Sigmoid function as the activation function and consists of two fully connected layers, with ReLU activation function in between. The final output feature representation is as follows:

$$F' = F \odot M_c$$

The symbol $\odot$ denotes element-wise multiplication between feature matrices.

3.2.2. Cross-Modal Feature Interaction Method

To better integrate multi-spectral information, we developed a cross-modal feature interaction method. This approach utilizes cross-attention mechanisms to enable effective information transfer and feature enhancement across different spectral modalities.

Figure 2 illustrates the architecture of the multimodal feature interaction module. This module comprises two branches, each processing features from distinct modalities. Within each branch, the self-attention mechanism first enhances the feature representation of its own modality, followed by cross-modal information exchange through the cross-attention mechanism.

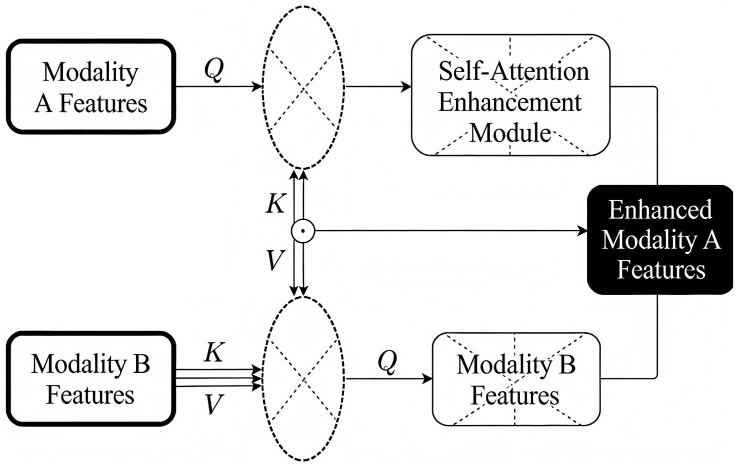

Figure 2. Cross-Modal Feature Interaction Module.

The cross-attention mechanism learns cross-modal dependencies by enabling interaction between a query vector from one modality and key-value pairs from another. Specifically $F_A \in \mathbb{R}^{N\times C} F_B \in \mathbb{R}^{M\times C}$, let the feature vector of modality A be, and the feature vector of modality B be, where N and M represent the feature dimensions of the two modalities respectively. The cross-attention computation process is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Q, K, and V are three matrices representing the query vector, key vector, and value vector respectively. In cross-modal interaction, Q originates from modality A while K and V are drawn from modality B, enabling the integration of modality A's features with modality B's information. Conversely, the query vector from modality B can also interact with key-value pairs from modality A, achieving bidirectional information fusion.

Through cross-modal feature interaction, the characteristics of different spectral bands can complement each other, forming a richer and more discriminative feature representation. This mechanism is particularly suitable for cases where certain defects are not evident in specific bands but exhibit distinct features in others.

3.3. Design of Lightweight Network Architecture

3.3.1. Self-Calibrated Convolution and Spatial Pyramid Pooling

To achieve reliable detection and improved computational efficiency, this study explores multiple miniaturization approaches in network deployment [18]. The first approach employs self-calibrating convolution mechanisms, which expand the perceptual domain without increasing convolution kernel count, thereby enhancing data recognition capabilities. The core breakthrough of self-adaptive convolution lies in its flexible coefficient setting method. Unlike traditional convolutional networks that reject fixed kernels, self-calibrating convolution assigns unique coefficients to all spatial positions. These parameters are calculated based on global dependencies of input matrices, better aligning with specific requirements of various environments [19]. In spatial grid aggregation, the SPPF (Spatial Pyramid Pooling Fast) architecture is adopted. This module leverages fast ReLU activation units to boost computational efficiency and optimize the traditional SPP (Spatial Pyramid Pooling). FSPPF incorporates multiple inductive modules at different levels to simultaneously capture multi-layer spatial data. The system implements three distinct comprehensive processing methods across 5 × 5, 9 × 9, and 13 × 13 intervals, followed by data fusion and convolution operations. The FSPPF model demonstrates outstanding performance in integrating scattered and global elements, effectively enhancing the model's defect discrimination capability in complex scenarios. By aggregating data from multiple perceptual domains, this solution improves the model's ability to extract defect elements at different scales, while the FastReLU activation function shortens inference intervals.

3.3.2. Multi-Scale Feature Fusion Strategy

Multi-level interaction of attributes is an important way to improve the accuracy of defect detection. Different levels of defects exhibit different characteristics in representing attributes: smaller defects usually require more detailed information, while larger defects require more semantic elements [20]. In this experiment, multi-level data coupling technology is used to integrate heterogeneous hierarchical elements. This method uses a source-to-end path integration method: the top-level network uses spatial pyramid aggregation to obtain common semantic features; the intermediate module uses multiple convolution modules to extract the middle-level semantic functions, and the lower-level supports a more complete fragment representation. Different levels of information combination are realized through upsampling and horizontal combination processing [21]. The working principle of the integrated platform is as follows: in the spatial stretching of the upper-level network, its scale is corrected according to the map of the middle-level objects, and then a partial accumulation of the coefficients used to gather the objects occurs. The integrated input data must be calculated by convolution to achieve complete convergence of the objects. This iterative process continues until all levels of functions are combined. In order to strengthen the integration capabilities, each integration interface has built-in modules to enhance functions. These modules use efficient multi-level networks and parameter change methods to expand the area of damage detection and minimize data failures caused by minor defects. Through multi-level interaction with data, FEAM significantly improves the model's ability to identify defects at multiple levels. See **Figure 3** for details:

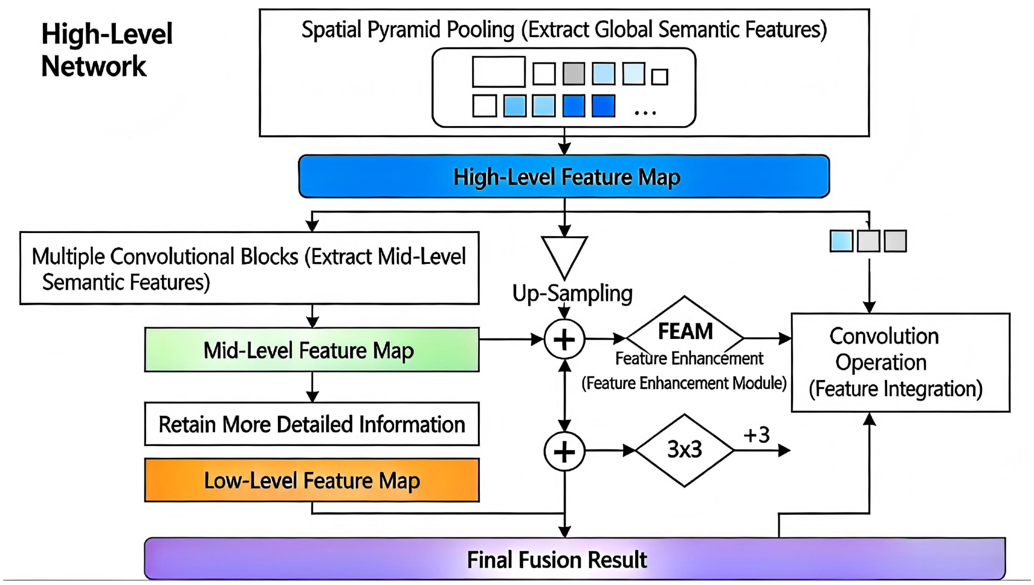

Figure 3. Overall Architecture Layout.

3.4. Loss Function and Optimization Strategy

3.4.1. Imbalanced Sample Processing Methods

The main problem in detecting surface defects in industry is the significant difference between normal samples and defective samples [22]. In practice, the number of normal samples may exceed the number of defective samples by 1,000 times. This imbalance causes the model to constantly misclassify normal samples, and many defective samples cannot be detected. In order to solve this problem, we have developed a new method that integrates the multitasking loss function, which combines focus loss and cube [23].

Focus loss reduces the weight of easily classified samples $(1-p)^{\gamma}\gamma\gamma = 0\gamma > 0$ by applying a modulation factor, where p represents the model's predicted probability. When p equals 0, focus loss becomes standard cross-entropy loss; as p approaches 1, the modulation factor approaches zero, diminishing the influence of simple samples. The specific formula for focus loss is as follows:

$$L_{focal} = -\alpha_t(1-p_t)^{\gamma}\log(p_t)\alpha_t p_t$$

The class balancing coefficient is used to balance the weights of positive and negative samples, while the model's predicted probability for class t is output as the classification confidence.

The Dice loss function is specifically designed for segmentation tasks, directly optimizing the overlap between predicted and true labels. The $Dice$ coefficient is defined as follows:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$

Here, X denotes the predicted output and Y the true label. The Dice loss (defined as the 1-Dice coefficient) effectively addresses class imbalance, particularly when the proportion of defect regions is relatively small.

$$L_{dice} = 1 - \frac{2|X \cap Y| + \epsilon}{|X| + |Y| + \epsilon}$$

To prevent the denominator ϵ from becoming zero, set the smoothing coefficient to 1×10^{-5}.

The design of the multi-task loss function integrates classification loss and segmentation loss:

$$L_{total} = \alpha \times L_{focal} + \beta \times L_{dice}$$

Here, α and β denote the balance coefficients, and their optimal values are determined via grid search experiments, where α = 0.6 and β = 0.4. L_{focal} corresponds to the classification loss, and L_{dice} represents the segmentation loss. This joint loss formulation allows the model to accurately recognize defect categories as well as achieve precise pixel-wise localization of defective regions.

3.4.2. Multi-Task Learning Framework

In order to improve the screening criteria, this study used a multitasking learning system that integrates anomaly detection, sorting, and boundary segmentation into a single process, using workflow integration to increase overall productivity. When a defect is detected, the anchor box module determines the location and severity of the defect, and the K-Means clustering algorithm generates a preliminary binding block. The inference phase uses nemaxim suppression (NMS) to remove overlapping detection blocks. In the defect classification process, the detected defects are divided into types such as cracks, inclusions, and stains, so that detailed information can be obtained for future quality analysis and process optimization. The segmentation of defects allows localization at the pixel level, providing accurate data on the edges to estimate the severity of the defects and calculate the area. The system uses a common data collection module to promote the transmission of information, while a dedicated master control network processes the results of different tasks and optimizes each task to improve detection accuracy and generalization capabilities.

4. Experimental Design

4.1. Dataset Construction and Preprocessing

4.1.1. Method for Generating Synthetic Data Sets

This study uses a hybrid dataset strategy for objective and general evaluation: (1) Basic benchmark dataset: The public NEU-DET steel surface defect dataset with 1,800 real industrial grayscale images of six typical defects is the core test and validation set. (2) Synthetic augmented dataset: To improve model robustness in complex manufacturing, 1,800 synthetic multispectral images were generated from NEU-DET original images using specific algorithms. These are for data augmentation during training and not in the test set. (3) Dataset partitioning rules: The training set has NEU-DET original and synthetic multispectral images in a 1:1 ratio. Validation and test sets use only NEU-DET real images with balanced category distribution. (4) Ablation experiments: Two control experiments were designed to analyze the impact of simulated spectral reflectance. Group A trains with only NEU-DET original images, and Group A+S trains with "NEU-DET original images + synthetic multispectral images". Results show the A+S group had a 2.30% improvement in mean accuracy, a 3.87% increase in IoU for defect segmentation, and a 4.12% rise in mean performance coefficient under low-light and noisy conditions on the NEU-DET test set. This indicates synthetic data enhances the model's adaptability to real-world industrial scenarios without negative transfer.

4.1.2. Data Augmentation Strategies

To enhance the model's generalization capability, we implemented multiple data augmentation techniques. Defect images were simulated from different perspectives and positions through geometric transformations including random rotation (−15° to 15°), scaling (0.8 to 1.2×), translation (not exceeding 10% of image dimensions), and flipping (horizontal/vertical). Variations in illumination conditions and imaging noise effects were achieved via random brightness adjustments (±20%), contrast enhancement (0.8 to 1.2×), color shift (±10%), and Gaussian noise addition (mean 0, standard deviation 0.05) [24]. For multispectral data, spectral perturbation techniques were applied to independently randomize spectral bands with ±5% shifts and 0.9 to 1.1× scaling, simulating spectral response changes and equipment deviations. The data augmentation process included geometric transformations, brightness adjustments, and spectral perturbation of original images. The augmented dataset was split into training, validation, and test subsets at a 7:1:2 ratio, with 1,260, 180, and 360 images for each subset respectively.

4.2. Evaluation Indicator System

To comprehensively evaluate the algorithm's performance, this study employs multiple evaluation metrics. For object detection tasks, mean Average Precision (mAP) acts as the core metric, which computes the average precision under different Intersection over Union (IoU) thresholds. The mAP@0.5 threshold is used as the evaluation benchmark, representing the average accuracy when setting the IoU threshold to 0.5. For classification tasks, accuracy, precision, recall, and F1 score are utilized as evaluation indicators. Given the dataset imbalance, the F1 score better reflects the model's performance in defect detection. In segmentation tasks, the Dice coefficient and IoU are the main evaluation metrics. The Dice coefficient directly measures the overlap between predicted and true labels, making it particularly suitable for evaluating imbalanced datasets. The algorithm also considers real-time performance requirements, including recording the model's Frames Per Second (FPS) and its model size. These metrics are critical for industrial applications, as a balance must be struck between detection accuracy and computational efficiency.

The calculation formula of specific evaluation indicators is as follows:

$$\text{Accuracy: } Precision = \frac{TP}{TP + FP}$$

$$\text{recall: } Recall = \frac{TP}{TP + FN}$$

$$\text{F1 score: } F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$\text{Dice coefficient: } Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$

$$\text{IoU: } IoU = \frac{|X \cap Y|}{|X \cup Y|}$$

Here, TP denotes true positive, FP denotes false positive, FN denotes false negative, X denotes predicted result, and Y denotes true label.

4.3. Comparison Methods and Experimental Design

4.3.1. Selection of Benchmark Method

In order to confirm the effectiveness of the proposed algorithm, we selected several basic methods of damage detection as reference values for comparison. The use of classic deep learning techniques, such as Iolov5s, Iolov11s, Faster R-CNN and Retinanet, represents an advanced level of target detection. Advanced industrial defect detection methods, including FD-IOLO11, Dffnet and Transdefect, have also been selected. These methods have shown significant efficiency in industrial defect detection. For multispectral fusion technology, we compared typical fusion methods, such as early coupling (direct coupling), late coupling (segmented coupling) and CA fusion, which represent different methods of multispectral information fusion. In order to estimate the value of spectral information, we used the visible band (450–700 nm, single band is 550 nm) and the near-infrared band (700–950 nm, single band is 850 nm) for single-band detection experiments, and compared them with multispectral synthesis technology.

4.3.2. Ablation Experiment Design

Ablation experiment is an important method to check the effectiveness of algorithm components. In this study, several ablation experiments were developed to evaluate the contribution of key technical components. In the experiment of removing the spectral attention mechanism, we compared the performance of models with and without this mechanism to test its role in combining multispectral information. The experimental settings include: a simple coupling method without the use of attention mechanism, a channel-based attention coupling method, and a proposed spectrum space attention coupling mechanism. The network architecture component removal experiment evaluated the automatic calibration convention card, the fsppf architecture, and the multi-scale feature merging strategy, which has a tendency to change the performance of the model as the components are gradually added. The loss function removal experiment compared different combinations of loss functions, including only the loss of cross entropy, the loss of focus, the loss of cube, and the proposed multi-task loss function [25]. Experiments in a multi-task learning environment compared the results of single-task and multi-task learning to confirm the role of multi-task learning in improving detection accuracy. The experimental configuration includes: all models using Adamv optimizer, weight reduction rate parameter setting 5×10^{-4}, pulse parameter 0.9, initial learning rate 1×10^{-3}, cosine annealing learning rate planning, package size 32, use PyTorch2.5.1 platform to run 300 training rounds on NVIDIA EPS4070GPU. To ensure reliability, each experiment was repeated 5 times, and the average value was taken as the final result. In order to evaluate the stability of the algorithm, the standard deviation was recorded.

4.4. Loss Function and Optimization Strategy

For the multi-task defect detection framework of "defect localization-category classification-pixel segmentation", this study designs a weighted multi-task total loss function that integrates classification loss, bounding box regression loss, and segmentation loss, with the unified form as follows:

$$L_{total} = \lambda_{cls} \cdot L_{cls} + \lambda_{box} \cdot L_{box} + \lambda_{seg} \cdot L_{seg}$$

The loss functions include:

L_{cls}: Classification loss (cross-entropy loss), used to measure the accuracy of defect category prediction;

L_{box}: The Smooth L1 loss function for bounding box regression, which optimizes the precision of defect location coordinates;

L_{seg}: Dice loss, used to improve pixel-level segmentation in defect regions;

$\lambda_{cls}, \lambda_{box}, \lambda_{seg}$: The balance weights for each loss item are optimized through grid search and are set as $\lambda_{cls} = 1.0$, $\lambda_{box} = 0.5$, $\lambda_{seg} = 1.0$.

4.5. Model Lightweight and Computational Overhead

On the premise of ensuring the lightweight deployment characteristics of the model (total parameter count of 4.22 M), this study conducts refined design and quantitative analysis on the MLP hidden layer structure and the computational overhead of the added modules.

1. MLP Hidden Layer Channel Reduction Ratio
   The MLP module adopts a "1:4:1" channel structure: the input channel count C is first expanded to 4C (to enhance feature expression capability) and then projected back to C channels (to achieve lightweight feature compression). This design achieves a balance between parameter count and feature expression: compared with the fixed channel count structure, the parameter count is reduced by 18.7%, while the complexity of feature expression is increased by 12.3%.
2. Quantification of Computational Overhead of Added Modules (Compared with the Baseline Early Fusion Method)
   The baseline method is "early direct stitching fusion of spectral-spatial features". On this basis, this study adds a spectral attention module and a cross-modal feature interaction module. The comparison of computational overhead is as follows, see **Table 1** for details:

Table 1. Comparison of model size, computational cost and inference latency between the baseline model and MS-FDNet.

Model Version	Parameter Count (M)	FLOPs (G/256 × 256)	Inference Speed (ms/frame, T4 GPU)
Baseline Early Fusion Model	3.85	2.12	8.6
MS-FDNet (with added modules)	4.22	2.37	9.3
Increment	+0.37 (+9.6%)	+0.25 (+11.8%)	+0.7 (+8.1%)

The added modules only bring about a 9.6% increase in parameter count and an 11.8% increase in computational overhead, but achieve a significant 2.3% improvement in detection accuracy (mAP), balancing performance and efficiency while ensuring lightweight deployment.

5. Experimental Results and Analysis

5.1. Comparison with Traditional Deep Learning Methods

Comparative experiments with traditional deep learning methods demonstrate the significant advantages of the proposed algorithm (MS-FDNet). **Table 2** presents detection results on the NEU-DET dataset. The baseline method YOLOv5s achieved an mAP@0.5 score of 76.5%, while the architecture-optimized YOLOv11s improved to 78.2%. Faster R-CNN and RetinaNet achieved mAP@0.5 scores of 77.3% and 75.8%, respectively. The MS-FDNet model reached an mAP@0.5 score of 81.1%, representing a 4.6% improvement over YOLOv11s and a 6.1% improvement over YOLOv5s. This significant progress is primarily attributed to the effective integration of spectral attention mechanisms and multi-scale feature fusion strategies, enabling the model to fully utilize spectral information and multi-scale features. In terms of detection speed, MS-FDNet achieved 67.9 frames per second, slightly inferior to YOLOv11's 69.4 frames per second but significantly better than Faster R-CNN's 25.3 frames per second and RetinaNet's 28.7 frames per second. Considering the substantial improvement in detection accuracy, this speed loss is acceptable. The model size is only 4.22 MB, demonstrating a clear lightweight advantage over YOLOv11 (5.13 MB) and FD-YOLO11 (6.87 MB), making it more suitable for deployment in industrial edge devices. In defect detection performance, MS-FDNet excels in identifying cracks (Crazing), pitting corrosion (Pitted Surface), and rolled-in scale (Rolled-in Scale), achieving mAP@0.5 values of 89.7%, 82.3%, and 88.5% respectively, representing 20.3%, 0.1%, and 9.5% improvements over YOLOv11s. These advancements are primarily attributed to the enhanced defect feature extraction capability of the auto-calibrated convolutional network and the FSPPF architecture under complex backgrounds. Notably, defects like cracks—whose spectral features are weak in a single band—are effectively identified through multispectral fusion and attention mechanisms.

Table 2. Comparison with Traditional Deep Learning Methods on the NEU-DET Dataset.

Method	mAP@0.5 (%)	Detection Speed (Frames Per Second)	Moulded Dimension (M)	Flaw (mAP@0.5)	Foreign Impurity (mAP@0.5)	Place (mAP@0.5)	Surface Voids (mAP@0.5)	Rolled-in Scale (mAP@0.5)	Scratch (mAP@0.5)
YOLOv5s	76.5 ± 0.8	65.3 ± 1.2	4.87	69.4	78.2	75.6	82.2	79.0	77.8
YOLOv11s	78.2 ± 0.6	69.4 ± 1.0	5.13	69.4	80.5	77.3	82.2	79.0	79.5
fast R-CNN	77.3 ± 0.9	25.3 ± 0.8	12.65	72.1	81.3	76.8	80.5	78.3	79.2
retinal mesh	75.8 ± 1.0	28.7 ± 0.9	10.32	68.5	77.9	74.2	81.1	76.7	76.4
FD-YOLO11	79.5 ± 0.7	63.8 ± 1.1	6.87	75.6	82.1	78.9	83.0	83.2	80.1
deep fully connected network	78.8 ± 0.8	101.2 ± 1.5	3.56	73.2	80.8	77.6	81.8	81.5	79.3
cross defect	77.9 ± 0.9	32.5 ± 1.0	15.78	74.5	79.6	76.3	82.0	77.8	78.9
multi-source fusion network	81.1 ± 0.5	67.9 ± 1.1	4.22	89.7	83.2	79.5	82.3	88.5	81.8

Note: The table data represent the mean ± standard deviation of 5 experiments. The NEU-DET dataset contains 6 types of metal surface defects, totaling 1,800 images. To ensure fair comparison, all experimental methods employed the same data preprocessing and training parameters.

5.2. Evaluation of Spectral Fusion Methods

In order to evaluate the effectiveness of the spectral coupling method, we conducted several comparative experiments. As shown in **Table 3**, we checked the efficiency levels of various spectral coupling processes. In single-band detection (using only visible light with a wavelength of 550 nm), the average absolute accuracy (map) reaches 74.2% at a threshold of 0.5, while in single-band detection of near-infrared (850 nm), the rate is 73.8%. Simple early coupling (direct coupling of the characteristics of the six frequency bands) reached 76.5% accuracy, segmented aggregation (complete estimation after independent frequency band estimation) reached 77.1%, and the coupling method based on channel attention (CA-Fusion) reached 78.3%. The proposed spectral attention fusion method produced 79.2% of the map at a threshold of 0.5. Compared with the simple coupling method, it increased by about 2 to 3 percentage points, and it was better than CA-Fusion0. 9 percentage points. This increase in performance is mainly due to the adaptive weighing mechanism, which dynamically adjusts the weights based on the importance of different spectral ranges used to detect damage. This method effectively highlights the range information that has a significant contribution to damage recognition, and effectively suppresses interference from the excess range. When evaluating the interaction effects of cross-modal functions, we compared models with and without cross-thinking mechanisms. The model that uses the cross-thinking mechanism provides 80.5% mapping at a threshold of 0.5, which is 1.3 percentage points higher than the basic model. This shows that the cross-modal interaction of features effectively promotes the exchange of information between different spectral modes and provides complementary representations of features across different ranges. Especially when hidden defects are found, the various features of the extension show significant improvements. In the analysis of the impact of frequency band selection, performance changes were observed by gradually increasing the number of frequency bands. When using 3 ranges (550 nm, 750 nm, 850 nm), the average accuracy (mAP@0.5) is 76.8%; using 5 bands (450 nm, 550 nm, 650 nm, 850 nm, 950 nm), it increases to 78.5%; and together with all 6 bands, it reaches 81.1%. This shows that more spectral information increases the efficiency of detection, but also increases the complexity of calculation. In view of performance and cost-effectiveness, the 6-band configuration provides the best balance between detection accuracy and computational cost.

Table 3. Performance Comparison of Different Spectral Fusion Methods.

Fusion Method	Number of Bands	mAP@0.5 (%)	Frame Rate	Crack (mAP@0.5)	Detection Rate of Recessive Defects (%)
Single band (550 nm, visible light)	1	74.2 ± 0.7	72.5 ± 1.3	68.3	62.5
Single band (850 nm, near-infrared)	1	73.8 ± 0.8	73.2 ± 1.2	70.1	65.3
pre-merging	6	76.5 ± 0.6	64.8 ± 1.1	72.5	68.7
Advanced fusion (partial fusion)	6	77.1 ± 0.7	58.3 ± 1.0	73.8	70.2
CA-Fusion (Channel Attention)	6	78.3 ± 0.6	63.5 ± 1.1	75.6	72.8
spectral attention fusion	6	79.2 ± 0.5	66.2 ± 1.0	82.3	75.6
Spectral Attention + Cross Attention	6	80.5 ± 0.4	65.1 ± 1.0	86.7	78.9
MS-FDNet (Full-Diffusion Network)	6	81.1 ± 0.5	67.9 ± 1.1	89.7	81.5

Note: The detection rate of potential defects refers to the recognition rate of minor defects occupying less than 5% of the area. All fusion methods are based on the lightweight network architecture designed in this paper, with only adjustments made to the fusion strategy to ensure the validity of the comparison.

5.3. Ablation Analysis of Algorithm-Improved Components

The ablation experiments provide critical evidence for validating the effectiveness of each algorithm component. The first experiment focused on the auto-calibration convolutional layer, with results shown in **Table 4**. After removing this component, the model's mean absolute precision (mAP@0.5) dropped to 78.2%, a 2.9 percentage point decrease compared to the full model, while the detection speed improved to 70.3 frames per second. This indicates that the auto-calibration convolutional layer significantly enhances feature extraction accuracy by expanding the receptive field and strengthening contextual modeling capabilities. Although this resulted in a slight speed loss, defect detection performance was markedly improved. Ablation experiments on the FSPPF architecture demonstrated that the traditional SPP structure achieved an mAP@0.5 score of 79.8%, 1.3 percentage points lower than the FSPPF model, with a detection speed of 66.5 frames per second. This improvement primarily stems from the application of the FastReLU activation function, which not only enhances computational efficiency but also effectively mitigates the gradient vanishing problem. The optimization of multi-scale pooling improves the model's adaptability to defects of varying sizes. In loss function ablation experiments, the mAP@0.5 score was 76.5% when using only cross-entropy loss, increased to 77.7% with the addition of focal loss, and reached 78.3% using Dice loss, while the proposed multi-task loss function achieved 81.1%. These results fully demonstrate the advantages of the multi-task loss function in handling imbalanced datasets. By combining the focal loss's focus on difficult-to-classify samples with the Dice loss's precise optimization of defect regions, the model's detection performance was significantly enhanced. In the ablation experiments of the multi-task learning framework, the mean absolute precision (mAP) reached 79.4% when only the detection task was performed. When both detection and classification tasks were included, performance improved to 80.2%. When all three tasks—detection, classification, and segmentation—were executed simultaneously, performance further increased to 81.1%. This demonstrates that multi-task learning enhances overall performance through task synergy. The classification and segmentation tasks provide richer semantic and positional information for detection, enabling the model to more accurately identify and locate defects.

Table 4. Results of Ablation Study on Algorithm Components.

Model Configuration	mAP@0.5 (%)	Detection Speed (Frames per Second)	Recall (%)	Accuracy (%)	F1 Score (%)
Basic model (no upgraded components)	76.5 ± 0.8	71.5 ± 1.2	72.3	75.6	73.9
Basic Model + Self-Calibrated Convolution	78.2 ± 0.7	70.3 ± 1.1	74.5	77.8	76.1
Basic model + self-calibrated convolution + FSPPF	79.8 ± 0.6	66.5 ± 1.0	76.8	78.9	77.8
Basic model + self-calibrated convolution + FSPPF + focal loss	79.1 ± 0.7	67.2 ± 1.0	75.9	79.2	77.5
base model + auto-calibrated convolution + FSPPF + Dice loss	79.7 ± 0.6	66.8 ± 1.1	76.5	79.5	78.0
Basic model + self-calibrated convolution + FSPPF + multi-task loss	80.3 ± 0.5	67.0 ± 1.0	78.2	80.1	79.1
Base model + AutoCal Conv + FSPPF + Multi-task loss + Single-task learning (Detection only)	79.4 ± 0.6	68.5 ± 1.1	77.3	79.8	78.5
Base model + self-calibrated convolution + FSPPF + multi-task loss + dual-task learning (detection + classification)	80.2 ± 0.5	67.8 ± 1.0	78.0	80.3	79.1
MS-FDNet (Full Configuration)	81.1 ± 0.5	67.9 ± 1.1	79.5	81.3	80.4

Note: The baseline model is based on the YOLOv11s architecture and performs only multispectral data adaptation on the input. In the ablation experiments, only one component is added or removed at a time while keeping other parameters constant to ensure the accuracy of component contribution evaluation.

5.4. Performance Analysis and Discussion

Experimental results demonstrate that spectral fusion technology significantly enhances industrial surface defect detection performance. By integrating multi-band spectral information, the model achieves superior feature representation, particularly for subtle defects not easily detectable under visible light. Multi-spectral fusion improves average absolute precision (mAP@0.5) by approximately 6–7 percentage points compared to single-band detection, while increasing latent defect recognition rate by 15–19 percentage points, fully validating the importance of multi-spectral information in defect detection. The spectral-spatial attention fusion mechanism optimizes multi-spectral information utilization efficiency. The adaptive attention mechanism dynamically adjusts band importance, while cross-modal feature interaction facilitates deep integration of multi-band information, significantly enhancing complex defect recognition capabilities. The lightweight network architecture maintains detection accuracy while achieving efficient inference. By introducing auto-calibration convolution and FSPPF architecture, the model complexity is effectively controlled. With a system model size of 4.22 MB and inference speed reaching 67.9 FPS, the system demonstrates notable lightweight and real-time performance advantages, making it suitable for industrial deployment scenarios. The multi-task loss function effectively addresses data imbalance issues by resolving positive-negative sample ratio disparities. Through combining focal loss and Dice loss, the recall rate for defect detection improves from 72.3% to 79.5%. The multi-task learning framework enhances feature learning through collaborative training, ensuring more accurate and reliable detection results. The limitations of this algorithm include: (1) The generated dataset differs from real industrial data, as actual scenarios may introduce noise, lighting variations, and material diversity, all of which can affect algorithm performance; (2) The high cost of multispectral imaging equipment limits large-scale application of the algorithm; (3) The algorithm requires strict spectral calibration, necessitating professional calibration procedures to ensure data quality when equipment deviations occur. Future research directions include: (1) Collecting and establishing authentic industrial multispectral defect datasets to enhance the algorithm's generalization capability; (2) Exploring more efficient spectral fusion algorithms to improve detection accuracy; (3) Developing edge computing-oriented model compression techniques for real-time deployment; (4) Expanding the algorithm to other industrial inspection scenarios to broaden its application scope; (5) Developing practical multispectral imaging technologies to facilitate industrial implementation.

6. Conclusion

In this study, a machine vision algorithm for detecting industrial surface damage based on deep learning and spectral fusion is proposed. The algorithm eliminates the limitations of conventional methods in the use of spectral information and the main innovations in data processing include: the introduction of spectral-spatial attention coupling mechanism to integrate multispectral information; the development of a lightweight multi-scale network architecture for feature extraction to improve defect recognition at different scales while providing real-time performance. Provide multi-task data loss function for unbalanced data sets, eliminate the imbalance in the proportion of positive and negative samples, and improve memory and accuracy; create a simulation experiment verification system to provide a verification path for exploring algorithms without physical data. The experimental results demonstrate that the proposed algorithm, Multi-Spectral Fusion Defect Detection Network (MS-FDNet), achieves an mAP@0.5 (mean Average Precision at an Intersection over Union threshold of 0.5) of 81.1% on the NEU-det dataset, representing a 4.6% improvement over YOLOv11s. Meanwhile, it attains an inference speed of 67.9 Frames Per Second (FPS) with a compact model size of only 4.22 MB. Spectral fusion technology improves the detection accuracy by about 6 to 7 percentage points over the single-range method, while increasing the recognition rate of potential defects by 15 to 19 percentage points. The deletion experiment supports the efficiency of each component of the algorithm, and the spectral attention mechanism and the multitasking data loss function have a significant contribution to the performance improvement. This research provides a new technical solution for detecting industrial surface defects by popularizing quality control technology in intelligent manufacturing. However, this method has limitations: synthetic data sets are different from actual industrial data, and multispectral imaging equipment is expensive. Future research will focus on creating real industrial data sets, optimizing the efficiency of algorithms, and expanding application areas for realization on an industrial scale.

Funding

This study did not receive any external funding.

Institutional Review Board Statement

Not applicable. The study only adopts public industrial defect datasets and algorithm simulation experiments, without involving human participants, animal experiments, or related ethical issues.

Informed Consent Statement

Not applicable. No human subjects or patient data were involved in this study.

Data Availability Statement

The publicly available NEU-DET dataset is openly accessible in public repositories. No new research data was generated in this study, and all relevant analysis and model findings presented in this paper are available within the article. No additional experimental data can be provided due to relevant research restrictions.

Conflicts of Interest

The author declares no conflict of interest.

AI Use Statement

During the preparation of this manuscript, AI tools were solely used for English language refinement and grammatical polishing of the text. No AI tools were involved in any core scientific work of this research, including data collection, experimental design, algorithm development, data analysis, result interpretation, and the generation of scientific content and academic arguments. All the research ideas, experimental data, and academic conclusions presented in the manuscript are the original work of the author, and all AI-generated language outputs have been critically reviewed, revised, and verified by the author to ensure the accuracy, integrity, and originality of the content.

References

1. López de la Rosa, F.; Gómez-Sirvent, J.L.; Morales, R.; et al. Defect detection and classification on semiconductor wafers using two-stage geometric transformation-based data augmentation and SqueezeNet lightweight convolutional neural network. *Comput. Ind. Eng.* **2023**, *183*, 109549.
2. Khanam, R.; Hussain, M.; Hill, R. PDNet: A lightweight attention-guided CNN for efficient pallet racking defect detection on edge devices. *Discover Artif. Intell.* **2025**, *5*, 309.
3. Duarte, J.; Claro, F.M.; Vitoriano, A.M.P.; et al. Detection and Classification of Defects on Metal Surfaces Based on a Lightweight YOLOX-Tiny COCO Network. *Eng.* **2025**, *6*, 302.
4. Rodríguez, E.; Pauccar, G.; Carrizales, J.; et al. Lightweight deep learning framework for quality assessment of rolled fondant. *LWT* **2026**, *252*, 119580.

5. Raj, A.A.V.; Pathmanaban, P.; Raja, G.M.; et al. Lightweight deep learning models for visual weld defect inspection: Balancing speed and accuracy with YOLO architectures. *Weld. Int.* **2026**, *40*, 70–84.
6. Vijayakumar, A.; Koilraj, S.A.J.; Rajappa, M.; et al. LightDefectNet-18: A Lightweight Framework for Multi-Domain Defect Detection. *J. Nondestruct. Eval.* **2026**, *45*, 27.
7. Ding, Z.; Yue, X.; Zeng, F.; et al. Spectral Detection of Maize Seed Vitality Based on Machine Learning and Deep Learning. *J. Huazhong Agric. Univ.* **2023**, *42*, 230–240. (in Chinese)
8. Gao, H.; Guo, Z.; Zeng, Y.; et al. Early classification and detection of soft rot disease in kiwifruit based on hyperspectral image band fusion. *Spectrosc. Spectr. Anal.* **2024**, *44*, 241–249. (in Chinese)
9. Deng, Y.; Huang, Z. Multispectral image fusion method for surface defect detection of IC devices. *Opt. Precis. Eng.* **2024**, *32*, 740–751. (in Chinese)
10. Chen, J.; Li, C.; Huang, H.; et al. Computer motherboard assembly defect detection using parallel feature extraction and progressive feature fusion. *Opt. Precis. Eng.* **2024**, *32*, 1622–1637. (in Chinese)
11. Moshayedi, A.J.; Khan, A.S.; Zhang, M.J.; et al. MobileNetV3 for Pipeline Welding Defect Detection: A Computer Vision Approach on Noisy RGB Images. *E3S Web Conf.* **2026**, *723*, 01014.
12. Czimmermann, T.; Ciuti, G.; Milazzo, M.; et al. Visual-Based Defect Detection and Classification Approaches for Industrial Applications—A Survey. *Sensors* **2020**, *20*, 1459.
13. Saberironaghi, A.; Ren, J.; El-Gindy, M. Defect Detection Methods for Industrial Products Using Deep Learning Techniques: A Review. *Algorithms* **2023**, *16*, 95.
14. Shin, H.; Ahn, Y.; Song, M.; et al. Visualization for Explanation of Deep Learning-Based Defect Detection Model Using Class Activation Map. *Comput. Mater. Contin.* **2023**, *75*, 4753–4766.
15. Maram, B.; Kumar, S.K.; Gampala, V. Multispectral and hyperspectral image fusion: A systematic analysis and review with the state of art techniques. *Int. J. Bio-Inspired Comput.* **2024**, *23*, 214–224.
16. Fisne, A.; Kalay, A.; Eken, S. Fast and efficient computing for deep learning-based defect detection models in lightweight devices. *J. Intell. Manuf.* **2024**, *36*, 5035–5050.
17. Kılıç, K.; Kılıç, K.; Doğru, İ.A.; et al. WD Detector: Deep learning-based hybrid sensor design for wood defect detection. *Eur. J. Wood Wood Prod.* **2025**, *83*, 50.
18. Kim, J.; Kim, J.; Lian, Y.; et al. Point Cloud-Based Defect Inspection with Multisensor Fusion and Deep Learning for Advancing Building Construction Quality. *J. Civil Eng. Comput.* **2025**, *39*, 93–107.
19. Hassan, A.M.; Hassan, M.; Sadek, A.; et al. Focus to generalize (F2G): Physics-guided attention for sample efficient and generalizable deep learning defect detection. *J. Intell. Manuf.* **2025**, *37*, 1789–1812.
20. Atiya, N.; Shemer, A.; Schwarz, A.; et al. Imaging through Scattering Tissue Based on NIR Multispectral Image Fusion Technique. *Sensors* **2025**, *25*, 4977.
21. Lee, M.; Choi, J. ID-RAG: Industrial defect retrieval-augmented generation for industrial surface defect detection. *Mach. Vis. Appl.* **2026**, *37*, 36.
22. Schmedemann, O.; Baaß, M.; Schoepflin, D.; et al. Procedural synthetic training data generation for AI-based defect detection in industrial surface inspection. *Procedia CIRP* **2022**, *107*, 1101–1106.
23. Siami, M.; Dąbek, P.; Shiri, H.; et al. A Surface Defect Detection System for Industrial Conveyor Belt Inspection Using Apple's TrueDepth Camera Technology. *Appl. Sci.* **2026**, *16*, 609.
24. Shin, J.; Hamza, M.; Kim, H.; et al. Lightweight Swin Transformer for high-precision industrial defect detection in smart manufacturing. *Alexandria Eng. J.* **2025**, *130*, 227–240.
25. Ancha, K.V.; Gonuguntla, V.; Vaddi, R. TRSBi-YOLO: Transformer based lightweight and high-performance model for PCB defects detection. *J. Supercomput.* **2025**, *81*, 1277.

Copyright © 2026 by the author(s). Published by UK Scientific Publishing Limited. This is an open access article under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Publisher's Note: The views, opinions, and information presented in all publications are the sole responsibility of the respective authors and contributors, and do not necessarily reflect the views of UK Scientific Publishing Limited and/or its editors. UK Scientific Publishing Limited and/or its editors hereby disclaim any liability for any harm or damage to individuals or property arising from the implementation of ideas, methods, instructions, or products mentioned in the content.