Proceedings of Engineering and Technology Innovation, vol. 31, 2025, pp. 45-57

Surface Defect Detection of Aluminum Plates Using Improved Faster R-CNN and ResNet-50

Xiaomei Ni[1], David Chua Sing Ngie[1,*], Wanzhen Wang[2], Miaomiao Xin[2], Na Li[2], Xiaole Han[2], Jialei Shi[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), Sarawak, Malaysia

[2]School of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, Jinan, China

Received 27 February 2025; received in revised form 05 July 2025; accepted 10 July 2025

DOI: https://doi.org/10.46604/peti.2025.14884

Abstract

This study aims to enhance the accuracy of surface defect detection in aluminum profiles. To address low recognition accuracy caused by irregular defect sizes and the coexistence of multiple defects, an inspection system integrating Faster Region-based Convolutional Neural Network (Faster R-CNN) and Residual Networks (ResNet-50) is proposed. After data enhancement and preprocessing of the aluminum profile image dataset from the Tianchi platform, ResNet-50 is used to extract deep features, and the Region Proposal Network (RPN) within Faster R-CNN is applied to generate candidate regions for classification and localization. Experimental evaluations demonstrate that the proposed model identifies all defect types with over 93% accuracy, while the error rate remains below 2%. Compared to the You Only Look Once Version 4 (YOLOv4) model, it exhibits greater performance in detecting surface defects. This advancement may lead to increased productivity and quality control in the manufacturing of aluminum profiles.

Keywords: deep learning, Faster R-CNN, feature fusion, ResNet-50, surface defect detection

1. Introduction

With the advancement of industrial modernization, the demand for sheet metal has increased, making the detection of metal surface defects increasingly critical. The rapid development of deep learning techniques has facilitated the widespread adoption of intelligent defect detection methods. While deep learning has propelled the advancement of intelligent defect detection, existing approaches still face challenges in effectively addressing the multi-scale coexistence of defects and the complex feature fusion necessary for aluminum plate inspection. During the production of aluminum plates, various surface defects may arise, which can negatively impact product quality, shorten service life, and reduce safety.

Traditional manual inspection methods used in aluminum manufacturing are often inefficient and exhibit inconsistent detection accuracy. Consequently, they struggle to meet modern industry demands for high precision and efficiency [1]. To overcome these challenges, researchers have developed defect detection technologies tailored to specific products, including microwave detection, nondestructive testing, and linear structured light scanning [2-6]. However, these techniques are effective only in site-specific applications. Therefore, it is necessary to develop a method that can adapt to diverse conditions and enhance the accuracy of defect classification in aluminum alloys. In recent years, with the rapid development and application of artificial intelligence, deep learning algorithms have been widely applied to object detection, particularly in

* Corresponding author. E-mail address: 21010377@siswa.unimas.my

industrial inspection [7-8]. These methods have significantly improved detection accuracy, productivity, and raw material utilization. Thus, developing advanced detection technologies capable of accurately identifying surface defects in aluminum materials can ensure a smooth production process and reliable product quality [9].

Accurate identification of defect features during image preprocessing is crucial for effective surface defect detection [10]. Researchers have leveraged Convolutional Neural Networks (CNNs) and Graph Convolutional Neural Networks (GCNs) to enhance the recognition rate of metal defects [11]. Faster Region-based Convolutional Neural Network (Faster R-CNN) has been applied to surface defect detection on rails, and experimental results have confirmed its feasibility [12]. Some studies have proposed an adaptive weight calculation method to enhance image quality, integrating it with Faster R-CNN and the You Only Look Once Version 5 (YOLOv5) algorithm for defect detection [13]. Some studies have improved defect recognition accuracy by modifying the residual block structure of Residual Networks (ResNet-50) and incorporating domain adaptation in the fully connected layer. These improvements have also significantly reduced training time [14]. Furthermore, some researchers have proposed a method that combines Deep Convolutional Generative Adversarial Network (DCGAN)-based data augmentation with ResNet-18 to improve defect detection accuracy in steel plates by generating realistic defect images [15].

For model optimization, CNN-based methods have been developed to detect defects in metal additive manufacturing processes, demonstrating strong robustness and detection capabilities against various defects such as cracks, gas pores, and lack of fusion [16]. Some researchers have successfully proposed texture quantization and attention mechanisms to optimize CNNs, increasing the mean Intersection over Union (mIoU) to 82.9% and enhancing robustness to complex textures [17]. Additionally, multilayer neural networks based on deep learning have been applied to detect defects in cold-rolled steel sheet processing [18].

In summary, these studies have played a pivotal role in advancing defect detection technologies for aluminum materials. Through rigorous research and analysis, they have significantly enhanced detection accuracy and reliability, benefiting the aluminum manufacturing industry.

Despite extensive research on aluminum plate defects, existing models and algorithms still face challenges in detection speed and accuracy [19]. However, they still have notable limitations:

(1) The irregular sizes of aluminum surface defects pose significant challenges for effective detection. The original algorithms operate under broad assumptions, necessitating further modifications to meet the precise requirements of aluminum surface defect analysis.
(2) The aluminum sheet dataset contains various defect types, with many images exhibiting multiple defects of different sizes. This complexity increases the likelihood of false positives and missed detections, as the algorithm may struggle to accurately differentiate and localize defects amid overlapping features.

To address these gaps, this study develops a surface quality monitoring model for aluminum profiles by integrating Faster R-CNN with ResNet-50. The approach involves preprocessing and segmenting an open-source dataset, followed by feature extraction, regression prediction, and classification. The results demonstrate high model accuracy. The contributions of this study are as follows:

(1) Through the residual block structure of ResNet-50, the cross-layer fusion of shallow edge features and deep semantic features is achieved, effectively solving the problem of the multi-scale defect coexistence in aluminum profiles. In the feature extraction layer, the output feature maps from different residual stages are used, each corresponding to a distinct receptive field. This enables the model to capture both the local texture of fine defects and the global semantics of large area defects.

(2) In the Region of Interest (ROI) pooling stage, the multi-scale feature maps output by ResNet-50 are combined with the candidate regions generated by the Region Proposal Network (RPN). Bilinear interpolation is applied to unify the feature sizes, thereby achieving precise feature alignment for defects of different sizes.

The subsequent sections provide detailed methods as follows: Section 2 introduces the integrated Faster R-CNN with the ResNet-50 model and dataset preprocessing. Section 3 outlines the experimental setup, including dataset selection, parameters, and evaluation metrics, as well as model training, describing optimization algorithms and overfitting handling. Finally, Section 4 summarizes the conclusions, discusses limitations, and proposes future directions.

2. Model and Method

This section first introduces the Faster R-CNN and ResNet-50 object detection system, which comprises a feature extraction layer, an RPN, an ROI pooling layer, and a fully connected layer. Subsequently, the evaluation metrics are presented in detail, including precision, recall, F1 score, Precision-Recall (PR) curve, Average Precision (AP), and mean average precision (mAP). Together, these metrics form a comprehensive evaluation framework.

2.1. Introduction of the model

This model integrates Faster R-CNN with ResNet-50 to develop an efficient target detection system. As an advanced object detection framework, Faster R-CNN enables fast and accurate identification and localization of multiple objects within an image [20]. However, Faster R-CNN alone does not always achieve the accuracy required for industrial surface defect detection. To enhance detection accuracy, this study employs ResNet-50 as the backbone feature extraction network. It is a widely used residual network known for its strong training performance. The Faster R-CNN model is further optimized in terms of feature extraction and candidate region localization. The 50-layer deep convolutional architecture and residual learning mechanism of ResNet-50 mitigate the gradient vanishing problem commonly encountered in deep networks.

The model comprises four main components, as illustrated in Fig. 1. The feature extraction layer processes input images through convolution, pooling, and nonlinear activation operations to generate high-dimensional feature maps containing key information. An RPN utilizes a fully convolutional network to propose potential object regions on the feature map, generating boundary positioning coordinates and class confidence scores. After normalization, these proposed regions were transformed by the ROI pooling layer into uniformly sized feature maps, preserving sufficient information for object identification and localization. Finally, the fully connected layer outputs the bounding box coordinates and classification probabilities, completing the transition from feature extraction to object detection [21].

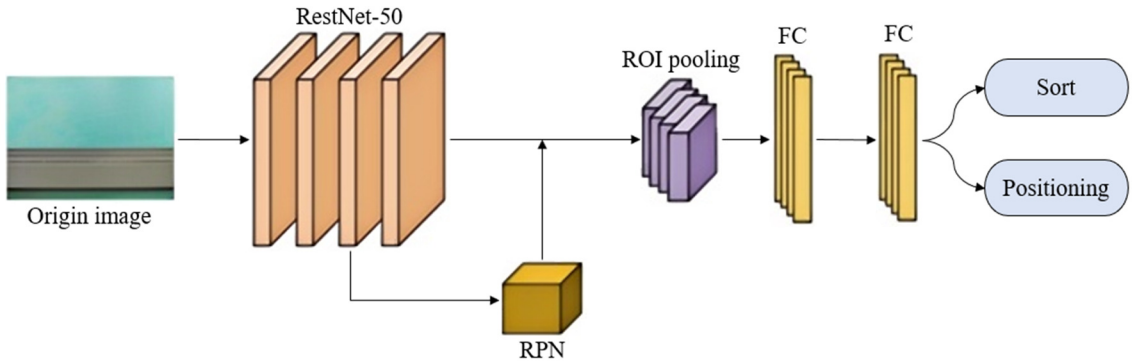


Fig. 1 Faster R-CNN and ResNet-50 target detection system

The integration of ResNet-50, particularly its residual learning structure, significantly enhances the model's ability to manage deep features, allowing it to extract more comprehensive and detailed image data. This structure introduces skip connections, enabling information to bypass multiple layers and facilitating deeper network training without performance

degradation. Fig. 2 illustrates the structure of the proposed model. Several data augmentation techniques are applied during the training phase to improve model resilience and generalization, including color dithering, random scaling, panning, and flipping. During validation, proportional scaling and center alignment are implemented to ensure compatibility with model input requirements. Additionally, image and bounding box data are preprocessed to comply with PyTorch specifications [22].

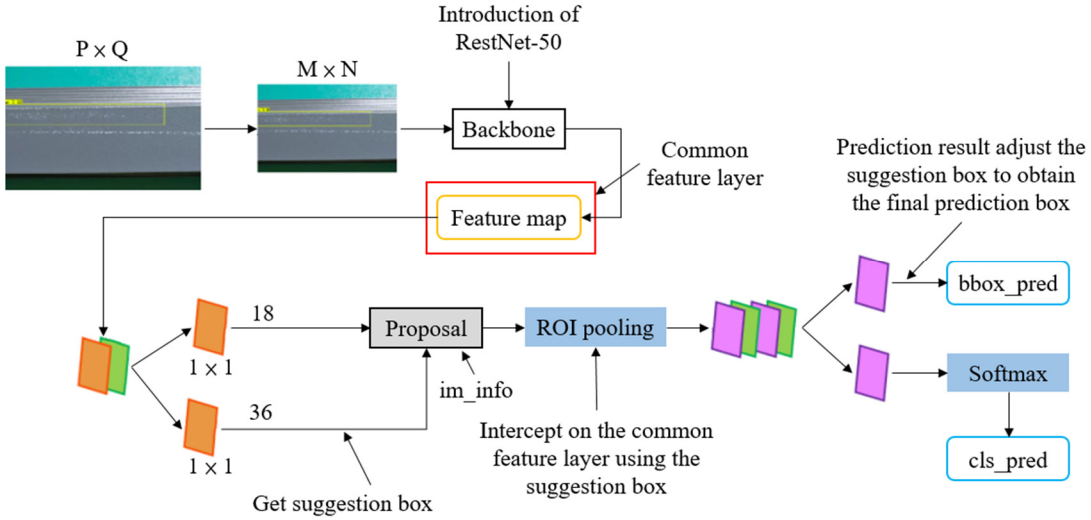


Fig. 2 Schematic diagram of model structure

2.2. *Evaluation index*

To evaluate the model's performance effectively, this study analyzes several key indicators for assessing the surface defect detection accuracy of aluminum profiles. These indicators include precision, recall, F1 score, the PR curve, AP, and mAP. The mAP is a widely accepted standard in target detection because it accounts for the model's performance across multiple categories and various detection thresholds. Furthermore, its calculation is based on the concepts of precision and recall. The mathematical expressions for these metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

True Positive (TP) refers to the number of defects correctly identified, while False Positive (FP) indicates the number of non-defects incorrectly classified as defects. Accuracy is measured by the proportion of samples that are both detected and labeled as defective by the model and subsequently verified to be correct.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

In the context of model evaluation, False Negative (FN) represents the number of defective samples that the model fails to detect. Recall measures the percentage of actual defects correctly identified by the model.

When there is a trade-off between precision and recall, the F1 score serves as the harmonic mean to balance these two metrics. A higher F1 score indicates better model performance.

$$\text{F1 score} = \frac{2 \times (\text{Prectision} \times \text{Recall})}{\text{Prectision} \times \text{Recall}} \tag{3}$$

The PR curve reflects variations in model accuracy under different recall rate thresholds and visualizes performance dynamics. At multiple recall levels, the average accuracy rate corresponds to the area under the PR curve. This metric quantifies the model's effectiveness in detecting defects within a single class. The mAP value, calculated by averaging the AP values for each class, provides a comprehensive assessment of the model's performance across multiple categories.

$$\mathrm{AP} = \int_0^1 P(R)\,dR \tag{4}$$

High mAP values signify strong model performance during evaluation. Additionally, a low log-average miss rate, which indicates fewer missed detections, serves as another hallmark of an effective model. These metrics collectively offer a robust framework for model evaluation.

3. Data Collection and Analysis

This section first describes the dataset, including its source, data division, defect conditions contained in the images, and the number of samples. Subsequently, the experimental software and hardware conditions are elaborated. Then, the entire data processing pipeline is presented, encompassing data preprocessing and feature extraction. Following the explanation of feature extraction, the training process is introduced, including all model parameter settings. Finally, the results are analyzed and discussed, with comparisons between different models.

(1) Introduction of experimental data

The data used in this study originates from the dataset of the "Tianchi Aluminum Surface Defect Dataset," which primarily collects image data for aluminum profile defect detection in actual production [23]. A total of 2,900 images are selected as samples, with the dataset split into training, testing, and validation sets at a ratio of 8:1:1. Each image contains one or more defects, representing a wide range of defect types encountered in aluminum production. Table 1 provides detailed information on the aluminum profile data, including definitions of defect types and corresponding descriptions.

Table 1 Specific information on aluminum profile data

Defect type (quantity)	Type description
Non-conductive (726)	Powder coating does not reach the aluminum plate.
Scratch (260)	Marks are caused by slight friction with other objects after surface treatment.
Dents (132)	Surface defects such as scuffs, scratches, and aluminum adhesion that have been pressed by the rollers.
Orange peel (183)	Delamination of plate and strip surfaces.
Leaky bottom (95)	Aluminum undertones reveal a lot.
Bruising (47)	A single stripe-shaped scar is caused by a sharp object (such as a plate corner, a metal chip, or a sharp object on equipment) coming into contact with the plate surface and sliding against it.
Pitting (336)	Because of the profile mold, the profile is concave throughout.
Convex powder (137)	Surface treatment powder spraying fails to evenly spray powder, resulting in a bunch of powder piles protruding.
Cracking coating (84)	The coating cracks during surface treatment, causing a phenomenon similar to the cracking of dry land.
Dirty spots (491)	During surface treatment, dust or dirt that is not properly wiped off results in more prominent particles in the coating.
Normal (815)	No defects.

(2) Experimental hardware and software

The experimental platform and environment, critical for ensuring the reliability and reproducibility of this study, are detailed in Table 2. The hardware configuration comprises a CPU and a GPU, while the software environment includes an operating system, a programming language, a compilation environment, and a deep learning framework. Together, they provide a comprehensive technical setup to support the implementation of the model.

Table 2 Experimental hardware and software

Flat-roofed building	Equipment parameters
CPU	AMD Ryzen 7 5800H
GPU	NIVIDA GeForce GTX 3060
Operating system	Windows 11
Programming language	Python 3.9
Compilation environment	PyCharm 2020.1
Deep learning framework	PyTorch 1.12.1

(3) Data preprocessing

During data preprocessing, several procedures are implemented. To ensure consistent input size, all images are resized to 600 × 600 pixels. Grayscale images are then automatically converted to RGB format to meet the model's input requirements. The standard image format adopted is ".jpg," maintaining coding coherence while enhancing both loading and processing speed. Additionally, category information and bounding box coordinates are recorded in an annotation file using the ".xml" format.

(4) Feature extraction

This model employs a pretrained ResNet-50 as the backbone network for feature extraction. ResNet-50, a 50-layer deep CNN, is capable of extracting intricate features from images. Within the Faster R-CNN framework, it processes input images through a series of convolution, batch normalization, and activation layers, efficiently capturing multilevel image features.

The output feature maps generated by ResNet-50 are subsequently fed into the RPN, which analyzes the feature maps to identify potential target regions. This step is critical for feature selection, as the performance of the RPN directly influences the effectiveness of subsequent ROI classifiers. By adjusting scale factors and aspect ratios, the model accommodates targets of varying sizes and shapes, thereby enhancing both the flexibility and efficiency of the feature selection process.

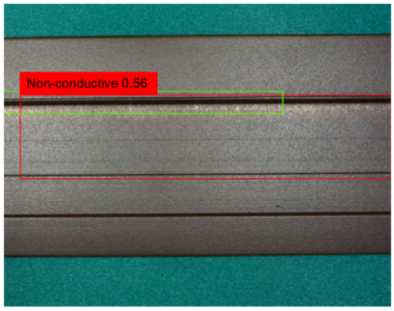
(5) Model training

In the network training experiments, ResNet-50 is used as the backbone of the CNN model to efficiently extract image features. The model begins with an initial convolutional layer employing a 7 × 7 convolutional kernel, 64 output channels, and a stride of 2 to extract features while effectively reducing dimensionality. Four residual layers progressively increase the number of channels, with each layer incorporating a different number of bottleneck blocks and adapting to different strides to further abstract and compress features. Finally, an average pooling layer and a fully connected layer map high-dimensional features to the classification output.

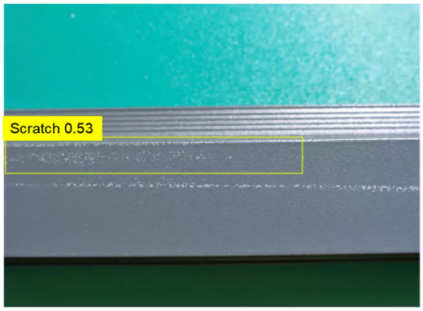
The initial learning rate (lr) is set to 0.0001, with a batch size of 8 and an epoch count of 200. Edge Box, the candidate box extractor, is configured by the parameter anchors_size, a list containing predefined sizes. This study defines anchors_size as [8, 16], enabling the candidate box extractor to utilize three anchor box sizes for object detection. During training, model performance variation is illustrated by plotting the number of training iterations (epochs) on the horizontal axis and the loss value on the vertical axis.

(6) Validation

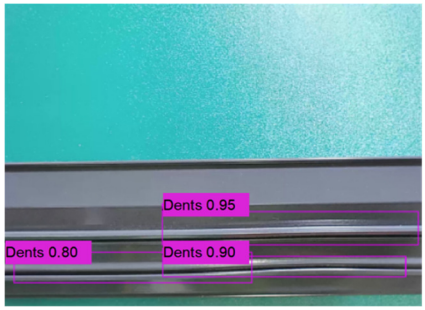
The validity of this model is demonstrated by its ability to accurately identify various defects. As shown in Fig. 3, the model achieves high accuracy in detecting both single defect types and multiple defects within a single aluminum plate. This result confirms the model's high recognition precision and accuracy, indicating its practical applicability.



(a) Non-conductive

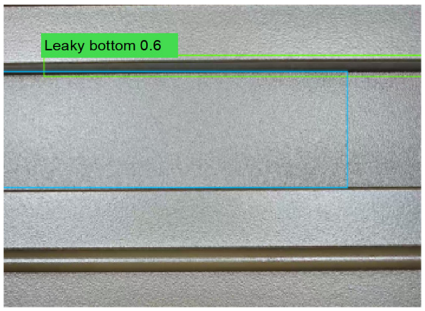


(b) Scratch

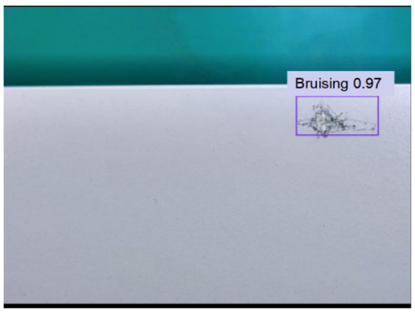


(c) Dents

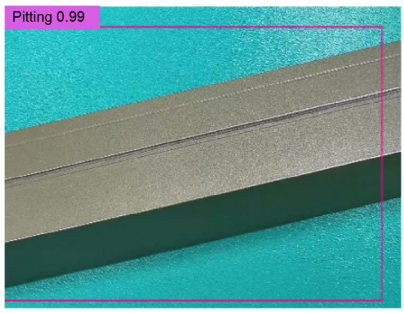
(d) Orange peel



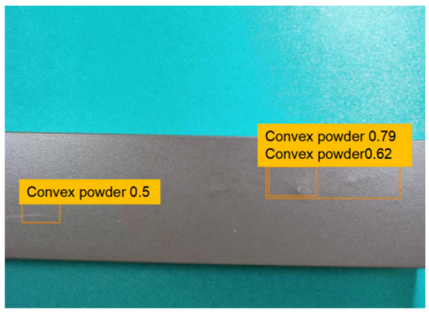
(e) Leaky bottom



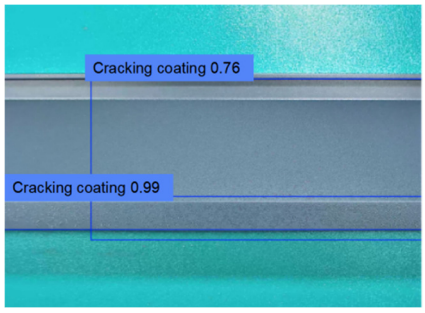
(f) Bruising



(g) Pitting

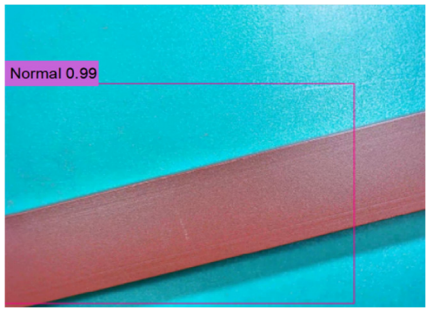


(h) Convex powder



(i) Cracking coating

(j) Dirty spots



(k) Normal

Fig. 3 Identification effect for different defect types

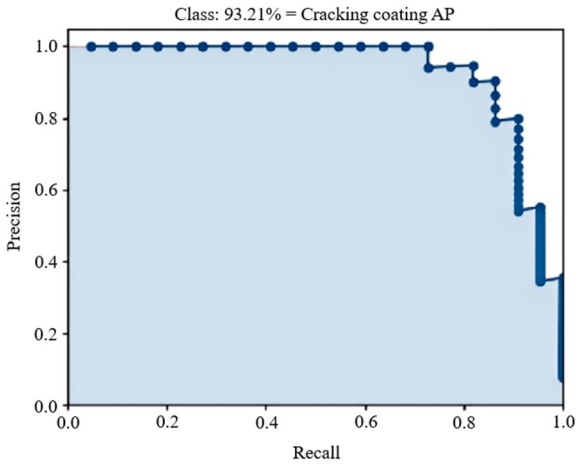
(7) Model evaluation

To comprehensively evaluate the proposed model's performance, an analysis is conducted using four key indicators: the PR curve, F1 score curve, accuracy curve, and recall curve. The model's performance in identifying specific defects, such as "cracking coating," "dents," and "non-conductive," is discussed in detail, while its performance on other defect types follows a similar trend.

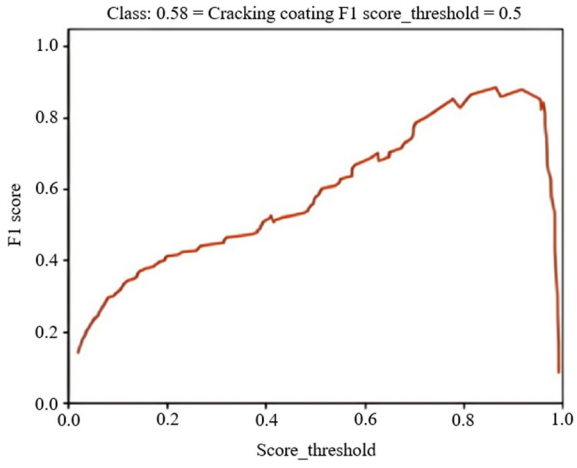
Performance of cracking coating

Fig. 4 illustrates the model's performance in recognizing "cracking coating" defects. The PR curve shown in Fig. 4(a) indicates that the model achieves an average accuracy of 93.21%, demonstrating strong predictive performance for aluminum plate "cracking coating" defects across multiple thresholds. However, as presented in Fig. 4(c), when the score threshold is set to 0.5, the model's prediction accuracy drops to 42%, suggesting a high misclassification rate at this specific threshold. The

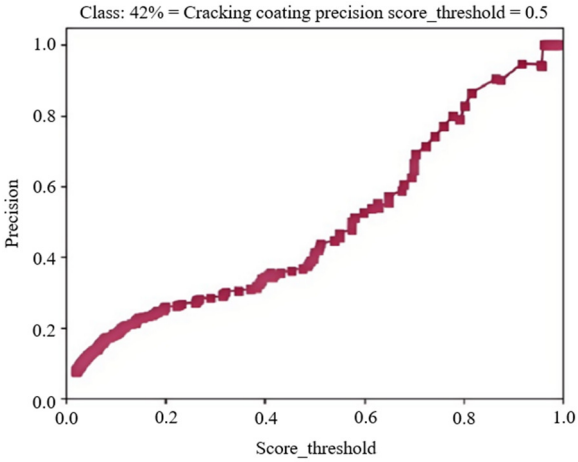
recall curve, as illustrated in Fig. 4(d), demonstrates that at the 0.5 score threshold, the model achieves a recall rate of 95.45%, indicating its effectiveness in detecting real aluminum layer cracking defects, which is crucial for practical applications. Fig. 4(b) shows that at the 0.5 threshold, the F1 score reaches 0.58, suggesting a relatively balanced performance between accuracy and recall. In summary, the model demonstrates high identification efficiency for "cracking coating" defects on aluminum plates, providing a strong theoretical foundation for practical defect detection.



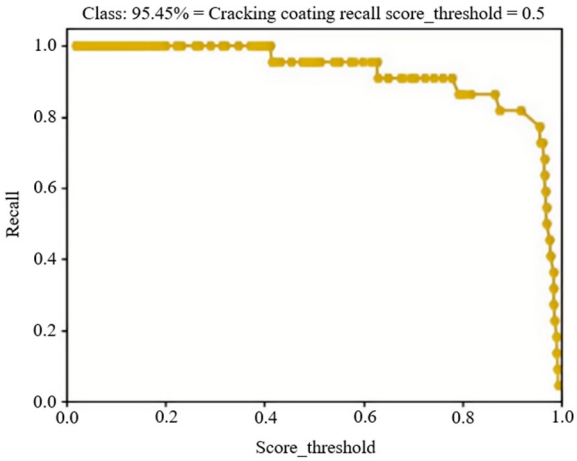
(a) Accuracy-recall curve



(b) F1 score curve



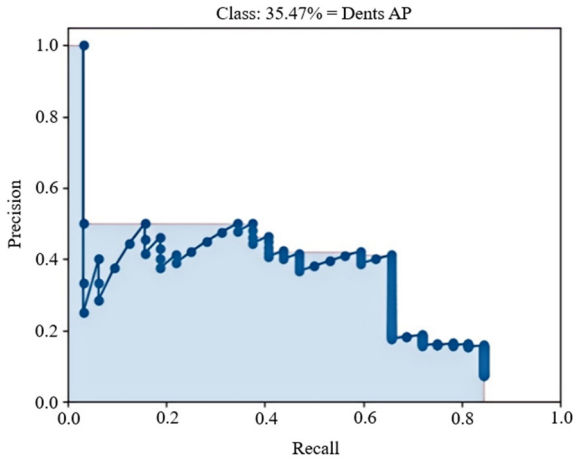
(c) Precision-Recall curve



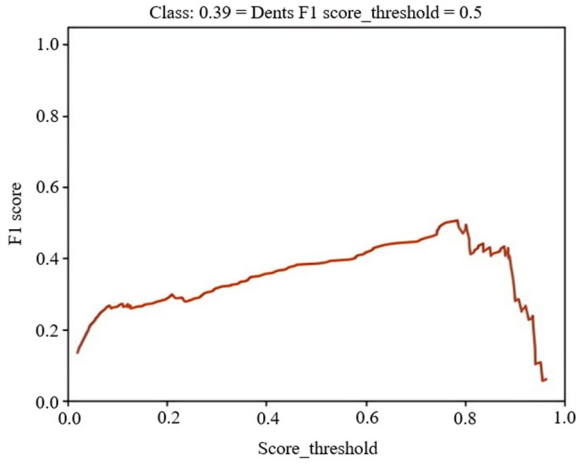
(d) Recall curve

Fig. 4 Model performance analysis of coating cracking defects

Performance of dents

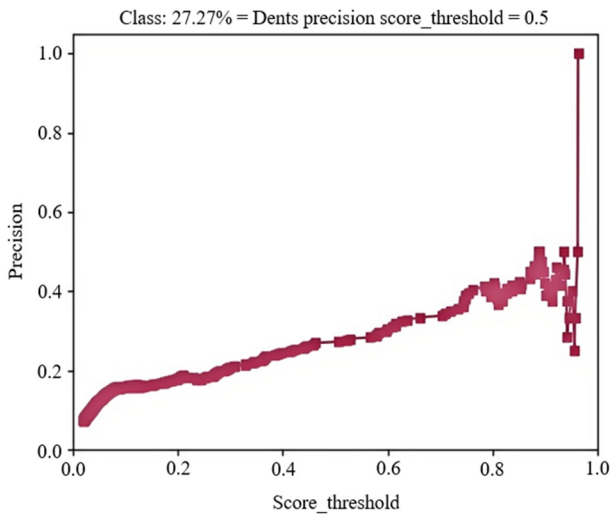


(a) Accuracy-recall curve

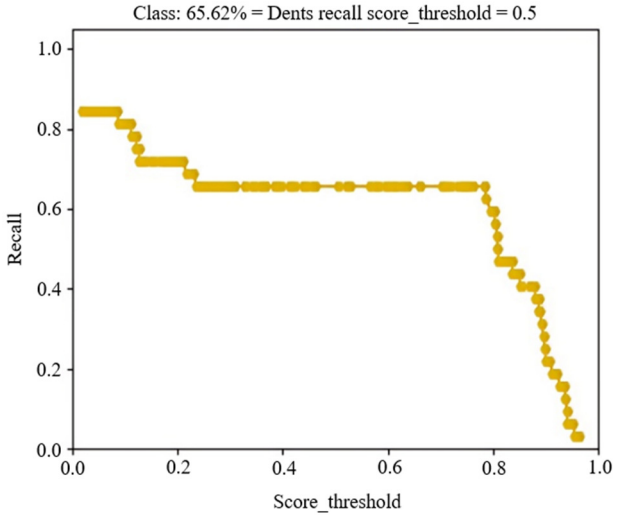


(b) F1 score curve

Fig. 5 Model performance analysis of dents defects



(c) Precision-Recall curve



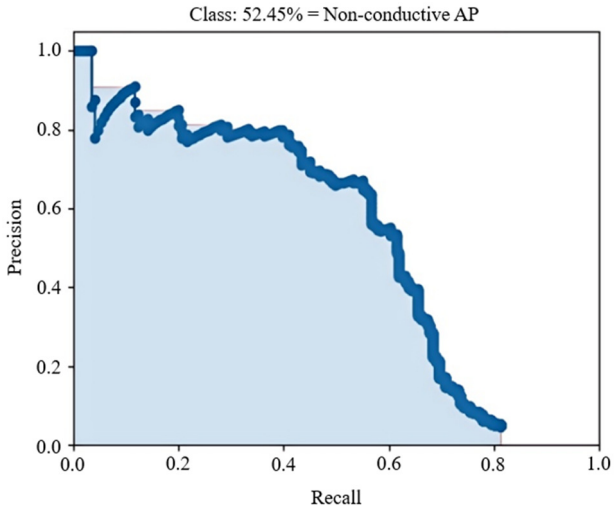
(d) Recall curve

Fig. 5 Model performance analysis of dents defects (continued)

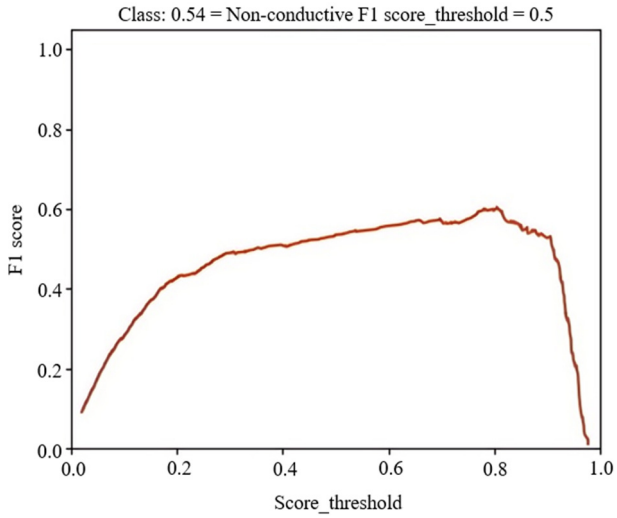
The model's ability to detect "dents" defects is further analyzed in Fig. 5. Fig. 5(a) shows that the model's AP in identifying "dents" defects is 35.47%, indicating room for improvement in accuracy. Fig. 5(b) illustrates the trade-off between precision and recall, revealing an F1 score of 0.39 at a threshold of 0.5, suggesting a moderate balance between these metrics. The precision curve in Fig. 5(c) indicates that at a 0.5 threshold, the model's precision is 27.27%, meaning that nearly one-third of the positive predictions at this threshold are actual defects. Fig. 5(d) shows that under the set threshold, the model achieves a recall rate of 65.62%, demonstrating effective capability in identifying "dents" defects.

Performance of non-conductive

The model's performance in identifying "non-conductive" defects is presented in Fig. 6. The accuracy-recall curve illustrates the model's accuracy variation across different recall levels. Fig. 6(a) shows that the model achieves an average accuracy of 52.45% for detecting "non-conductive" defects, indicating a relatively strong identification ability. As shown in Fig. 6(c), when the score threshold is set at 0.5, the model's accuracy is 47.32%, meaning that 47.32% of the samples predicted as defective are indeed defective. The recall curve in Fig. 6(d) shows that at the same threshold, the recall rate is 61.99%, indicating that 61.99% of actual defects are correctly identified. Fig. 6(b) demonstrates that the F1 score at a 0.5 threshold is 0.54, suggesting a relatively balanced performance between accuracy and recall. The model exhibits strong detection efficiency for "non-conductive" aluminum plate defects, offering a solid theoretical framework for practical defect detection applications.

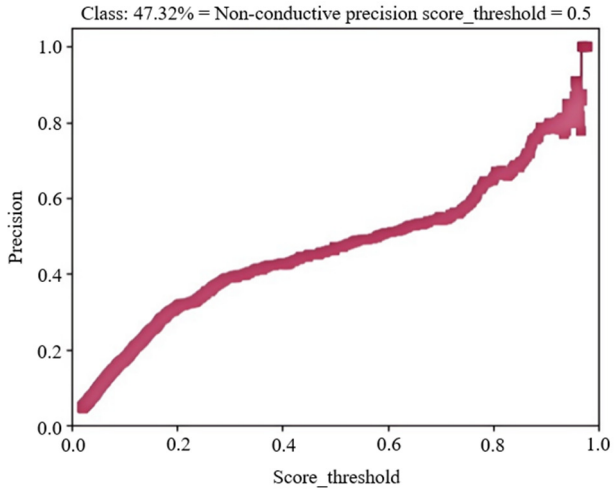


(a) Accuracy-recall curve

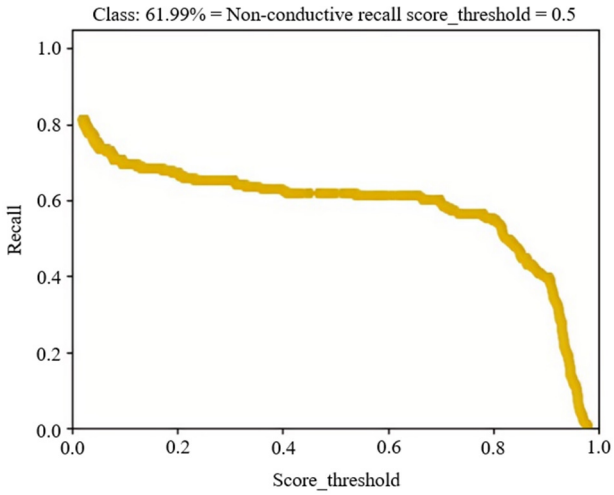


(b) F1 score curve

Fig. 6 Model performance analysis of non-conductive defects



(c) Precision-Recall curve



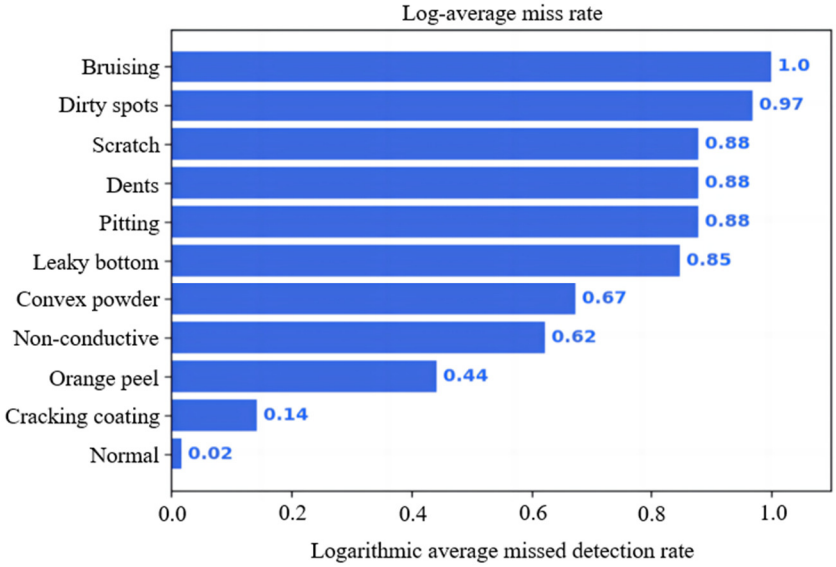
(d) Recall curve

Fig. 6 Model performance analysis of non-conductive defects (continued)

Overall comparison

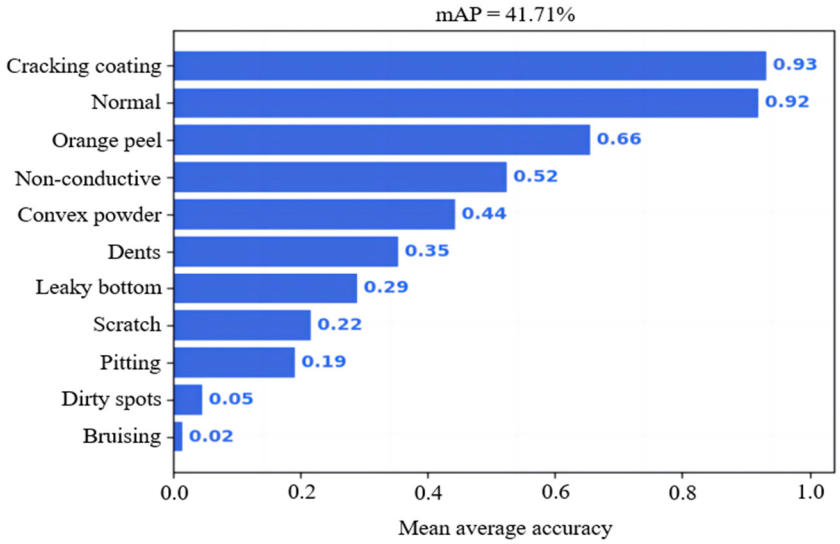
The quantitative performance of mAP and log-mean miss rates for different defect classes is analyzed in detail to assess the model's effectiveness in detecting aluminum plate defects. Fig. 7 presents the accuracy and miss rates for various defect types. As shown in Figs. 7(a) and 7(b), the model demonstrates strong reliability in detecting "cracking coating" defects and "normal" aluminum plates, with average accuracies of 0.93 and 0.92, respectively, while the log-mean miss rate is nearly zero. This result confirms the model's high efficiency in identifying these defect types. Overall, the model performs well in detecting various defects. However, the average precisions for detecting "dirty spot" and "bruising" defects are 0.02 and 0.05, respectively, indicating a significantly reduced ability to identify these defects effectively. The experimental data show that there are only 47 samples of "bruising." The small sample size prevents the model from comprehensively learning the features of this defect during the training, which reduces its detection precision.

On the other hand, dirty spots and bruising are small in size, and their features are gradually weakened or lost during the convolution and pooling process of the deep network. Although ResNet-50 has a strong feature extraction capability, the feature extraction focuses more on capturing obvious and large defects and is not sensitive enough to subtle defect features. This limitation makes it difficult to accurately extract and recognize dirty spots and bruising.



(a) Average missed rates for each category

Fig. 7 Overall performance evaluation of this model under each defect type



(b) The mean average accuracy of the overall class

Fig. 7 Overall performance evaluation of this model under each defect type (continued)

Comparison with other models

Deep learning-based object detection algorithms can be broadly categorized into two types. The first type includes two-stage algorithms, such as Faster R-CNN, which operate in two steps: (1) generating candidate regions likely to contain the target, and (2) refining and classifying these candidate regions. The second type comprises single-stage target detection algorithms, such as the You Only Look Once (YOLO) series, which generate complete detection results in a single forward pass, enabling real-time detection. To further validate the accuracy of the improved Faster R-CNN model, this study also trained the aluminum profile dataset using the You Only Look Once Version 4 (YOLOv4) model for 200 iterations, with the mAP values and overall results presented in Fig. 8.

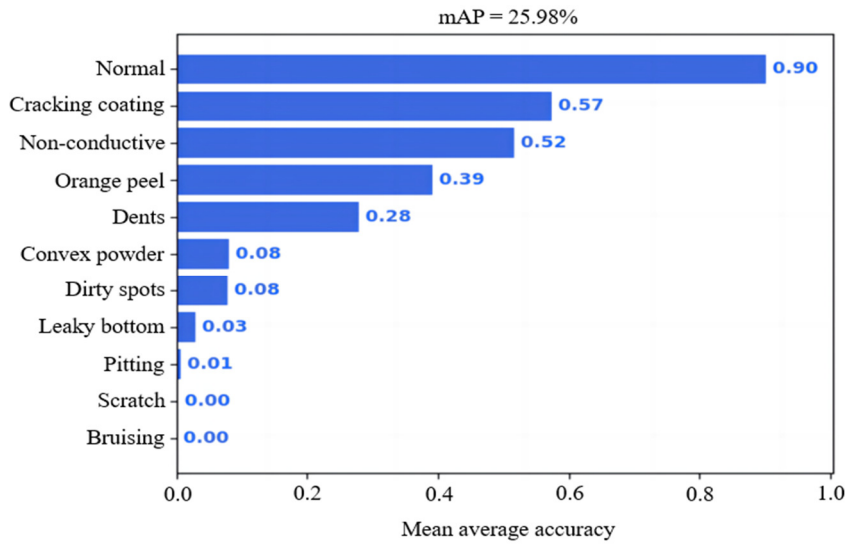


Fig. 8 Results of YOLOv4 detection

A comparison of the experimental results in Fig. 7(b) and Fig. 8 shows that the optimized Faster R-CNN model achieves higher AP values for most aluminum defect types than YOLOv4. Specifically, the total mAP value increases by 15.30% compared to YOLOv4. Notably, the mAP increases by 22% for "scratch," 7% for "dents," 27% for "orange peel," 26% for "leaky bottom," 18% for "pitting," 36% for "convex power," 36% for "cracking coating," and 2% for "normal." The proposed method demonstrates significantly higher detection accuracy than previous approaches, presenting considerable improvements across most defect categories. Extensive comparisons confirm that the Faster R-CNN model is highly effective, achieving superior performance in defect detection tasks.

To evaluate the performance of the improved algorithm, the frames per second (FPS) metric is compared with other target detection algorithms. The comparison results are shown in Table 3. The FPS of of proposed algorithm is 19.8 FPS, which is lower than that of YOLOv4. This is because Faster R-CNN is a two-stage detection algorithm with higher detection accuracy than a single-stage algorithm. However, the improved accuracy comes at the cost of slower detection speed. Compared with the original Faster R-CNN algorithm, the optimized model detection speed decreases from 20.5 FPS to 19.8 FPS, but it could meet real-time requirements for industrial applications.

Table 3 Comparison of the effect of the improved algorithm

Detector method	YOLOv4	Faster R-CNN	The proposed algorithm
FPS	34.6	20.5	19.8

4. Conclusion

This paper proposed a model that integrates Faster R-CNN with ResNet-50 for the surface defect detection of aluminum plates. The dataset used for training and testing was derived from the aluminum plate defect image dataset provided by the Tianchi Platform. Through the implementation of a cross-layer feature fusion mechanism and an optimized ROI pooling strategy, this model achieves high accuracy and stability in defect identification.

(1) This study utilizes the residual structure of ResNet-50 to achieve cross-layer fusion of shallow edge features and deep semantic features. By integrating multi-scale sensing fields from different residual stages, it accurately captures defects such as coating cracks, addressing the problem of detecting coexisting multi-scale defects. Experimental results show that the model performs well across most defect categories, especially in detecting “normal” and “cracked coatings”, with a missed detection rate of 0.14 and an accuracy of 93%.

(2) In the ROI pooling stage, the multi-scale feature maps from ResNet-50 are fused with candidate regions generated by the RPN. Bilinear interpolation is employed to accurately align features of defects with different sizes. Compared to traditional single-scale detection frameworks, the detection accuracy is significantly improved. Specifically, the total mAP of the Faster R-CNN model with ResNet-50 increased by 15.30% over the YOLOv4 model, demonstrating superior overall detection performance.

The current research has achieved progressive advancements in aluminum plate defect detection by integrating Faster R-CNN with ResNet-50. However, the ability of the model detect defects with a small number of samples decreases significantly. For instance, the mAPs of “bruising” and “dirty spots” are only 5% and 2%, respectively, indicating insufficient generalization capability in small sample scenarios. Thus, future research will focus on data augmentation and small sample learning, such as introducing generative adversarial networks to synthesize rare defect samples, combined with federated learning or transfer learning to alleviate the sample imbalance problem.

Acknowledgments

This work is supervised by professors from Universiti Malaysia Sarawak. This work is also supported by the research program of Qilu Institute of Technology (No.: QIT23NN038). This work is supported by Shandong Jinhui Material CO., LTD.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] B. Yang, Z. Liu, G. Duan, and J. Tan, “Mask2defect: A Prior Knowledge-Based Data Augmentation Method for Metal Surface Defect Inspection,” IEEE Transactions on Industrial Informatics, vol. 18, no. 10, pp. 6743-6755, 2022.

[2] J. Sun, P. Wang, Y. K. Luo, and W. Li, "Surface Defects Detection Based on Adaptive Multiscale Image Collection and Convolutional Neural Networks," IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 12, pp. 4787-4797, 2019.

[3] J. Cao, H. Wu, W. Wang, T. Qasim, and D. Wang, "A Visual Inspection and Classification Method for Camshaft Surface Defects Based on Defect Similarity Measurement," IEEE Access, vol. 12, pp. 74633-74648, 2024.

[4] Z. Guo and T. Sasayama, "Detection of Surface and Back-Surface Defects on Metal Plate via Rectangular Wave Eddy Current Testing Using Magnetoresistive Sensor," IEEE Transactions on Magnetics, vol. 59, no. 11, pp. 1-5, 2023.

[5] Q. Li, Z. Luo, H. Chen, and C. Li, "An Overview of Deeply Optimized Convolutional Neural Networks and Research in Surface Defect Classification of Workpieces," IEEE Access, vol. 10, pp. 26443-26462, 2022.

[6] S. She, X. Zheng, L. Xiong, T. Meng, Z. Zhang, Y. Shao, et al., "Thickness Measurement and Surface-Defect Detection for Metal Plate Using Pulsed Eddy Current Testing and Optimized Res2net Network," IEEE Transactions on Instrumentation and Measurement, vol. 73, pp. 1-13, 2024.

[7] K. Sun, Q. Wen, and H. Zhou, "Ganster R-CNN: Occluded Object Detection Network Based on Generative Adversarial Nets and Faster R-CNN," IEEE Access, vol. 10, pp. 105022-105030, 2022.

[8] X. Zheng, S. Zheng, Y. Kong, and J. Chen, "Recent Advances in Surface Defect Inspection of Industrial Products Using Deep Learning Techniques," The International Journal of Advanced Manufacturing Technology, vol. 113, no. 1-2, pp. 35-58, 2021.

[9] Y. S. Balcıoğlu, B. Sezen, C. C. Çerasi, and S. H. Huang, "Machine Design Automation Model for Metal Production Defect Recognition with Deep Graph Convolutional Neural Network," Electronics, vol. 12, no. 4, article no. 825, 2023.

[10] B. Li, F. Ren, H. Ni, X. Kang, S. Lv, and Z. Hao, "Classification Method of Surface Defects of Aluminum Profile Based on Transfer Learning," International Conference on Machine Learning and Intelligent Systems Engineering, pp. 1-5, 2022.

[11] C. Jia and F. Huang, "A Novel Fault Inspection Method of Steel Plate Surface," 4th International Conference on Artificial Intelligence and Advanced Manufacturing, pp. 66-73, 2022.

[12] J. Y. Choi and J. M. Han, "Deep Learning (Fast R-CNN)-Based Evaluation of Rail Surface Defects," Applied Sciences, vol. 14, no. 5, article no. 1874, 2024.

[13] L. Yang, X. Huang, Y. Ren, and Q. Han, "Study on Image Enhancement and Automatic Annotation of Steel Plate Surfaced Defect," Mechanical Science and Technology for Aerospace Engineering, vol. 44, no. 3, pp. 445-452, 2025. (In Chinese)

[14] X. Shi, S. Zhou, Y. Tai, J. Wang, S. Wu, J. Liu, et al., "An Improved Faster R-CNN for Steel Surface Defect Detection," IEEE 24th International Workshop on Multimedia Signal Processing, pp. 1-5, 2022.

[15] B. Jiang, P. Zhou, X. Sun, and T. Chai, "Intelligent Recognition of Steel Plate Surface Defect Based on Deep Convolutional GAN," Neural Computing and Applications, vol. 37, no. 14, pp. 8331-8345, 2025.

[16] M. Sun, F. Dong, Z. Huang, and J. Luo, "Adaptive Model Compression for Steel Plate Surface Defect Detection: An Expert Knowledge and Working Condition-Based Approach," Tsinghua Science and Technology. vol. 29, no. 6, pp. 1851-1871, 2024.

[17] C. Zhang, J. Cui, J. Wu, and X. Zhang, "Attention Mechanism and Texture Contextual Information for Steel Plate Defects Detection," Journal of Intelligent Manufacturing, vol. 35, no. 5, pp. 2193-2214, 2024.

[18] F. Selamet, S. Cakar, and M. Kotan, "Automatic Detection and Classification of Defective Areas on Metal Parts by Using Adaptive Fusion of Faster R-CNN and Shape from Shading," IEEE Access, vol. 10, pp. 126030-126038, 2022.

[19] L. Jiang, B. Yuan, Y. Wang, Y. Ma, J. Du, F. Wang, et al., "Ma-Yolo: A Method for Detecting Surface Defects of Aluminum Profiles with Attention Guidance," IEEE Access, vol. 11, pp. 71269-71286, 2023.

[20] P. Tan, X. Li, Z. Wu, J. Ding, J. Ma, Y. Chen, et al., "Multialgorithm Fusion Image Processing for High Speed Railway Dropper Failure–Defect Detection," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 7, pp. 4466-4478, 2021.

[21] L. Cui, X. Jiang, M. Xu, W. Li, P. Lv, and B. Zhou, "SDDNet: A Fast and Accurate Network for Surface Defect Detection," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-13, 2021.

[22] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun, "Infrared Thermal Imaging-Based Crack Detection Using Deep Learning," IEEE Access, vol. 7, pp. 182060-182077, 2019.

[23] K. Xiang, S. Li, M. Luan, Y. Yang, and H. He, "Aluminum Product Surface Defect Detection Method Based on Improved Faster RCNN," Chinese Journal of Scientific Instrument, vol. 42, no. 01, pp. 191-198, 2021. (In Chinese)