Machine Learning and Intelligent Systems
J.-L. Kim (Ed.)

doi:10.3233/FAIA241217

Advancements in Industrial Product Surface Defect Detection: From Traditional Methods to Modern Advanced Techniques

Qi QIAO[a,b] , Azlin AHMAD [a,1], Huiying HU [a] and Ke WANG[b]
[a] *College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia*
[b] *School of Computer and Communication, Jiangsu Vocational College of Electronics and Information, Huai an, Jiangsu, China*

Abstract. Surface defect detection plays a pivotal role in ensuring product quality in industrial production, as defects like cracks, scratches, and dents can compromise product performance and durability. Traditional detection methods, such as manual inspection and Non-Destructive Testing (NDT), are limited by inefficiency, reliance on human expertise, and susceptibility to errors, which restrict their application in large-scale production. With advancements in artificial intelligence, deep learning models, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have emerged as promising solutions for automated surface defect detection. This paper provides a comprehensive review of surface defect detection technologies, starting from traditional methods to modern deep learning-based techniques. The advantages and limitations of each approach are analyzed, highlighting key advancements in deep learning, including recent models like Faster R-CNN, Cascade R-CNN, and YOLOv4. Furthermore, challenges such as handling complex defects and improving detection accuracy in real-world industrial environments are discussed, along with potential directions for future research. Experimental evaluations using the Few Steels Classification (FSC) dataset demonstrate the effectiveness of modern detection methods in industrial applications, offering insights into enhancing defect detection systems.

Keywords. Surface Defect Detection, Deep Learning, Convolutional Neural Networks, Industrial Production

1. Introduction

In industrial production, industrial products surface defect detection is crucial because product quality often depends on the integrity of the surface. If defects such as cracks, scratches, or dents occur, they can significantly reduce the performance and durability of products, creating potential safety hazards in critical areas such as automotive and aerospace. Ensuring surface quality is therefore essential to guarantee product reliability. Traditional defect detection primarily relies on manual inspection and Non-Destructive Testing (NDT), such as ultrasonic testing and eddy current testing. While these methods are somewhat effective, they require significant manpower, have low efficiency, and are

[1] Corresponding Author: Azlin Ahmad, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Malaysia, E-mail: azlin@tmsk.uitm.edu.my

prone to human error, leading to the risk of misjudgment. These limitations greatly restrict the application of such detection methods in large-scale production [1].

In recent years, the rapid advancement of artificial intelligence technology [2] has sparked growing interest among researchers in exploring deep neural network models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [3]. CNNs excel at image recognition and classification tasks by effectively handling large datasets and minimizing manual intervention [4]. Meanwhile, RNNs perform strongly in sequential data processing, making them ideal for applications like natural language processing and time series forecasting [5]. As newer models are developed, the capabilities of these architectures and methods continue to improve.

The purpose of this research is to study the development of defect detection technology on the surface of industrial products and to understand the development from traditional manual detection methods to advanced deep learning-based methods. First, to understand the development of defect detection techniques and analyze the advantages and disadvantages of various methods; second, to understand the advantages of various models and algorithms used in current research, from traditional methods to deep learning methods in defect detection; and finally, to analyze the relevant methods through experiments on a specific dataset.

This review covers both traditional and deep learning-based defect detection methods to offer insights for enhancing defect detection in industrial production. Additionally, while deep learning applications in industrial products surface defect detection are relatively new, there is considerable potential for further development [6].

2. Traditional defect detection method

Industrial products surface defect detection has always been a critical process in the manufacturing industry, and traditional methods form the basis of quality control. These traditional methods include manual visual inspection, and NDT [7].

2.1. Manual Visual Inspection

Throughout the history of surface defect detection technology, manual inspection has remained a common and direct approach. Skilled inspectors visually assess industrial product surfaces for defects like scratches, dents, cracks, or corrosion. While this method is simple and widely employed, it is prone to subjectivity and can be easily influenced by human factors. In large-scale production settings, the fatigue resulting from extended work periods can significantly diminish defect detection accuracy [8].

2.2. NDT

NDT allows for the inspection of materials without causing damage or affecting their performance, utilizing techniques such as Ultrasonic Testing (UT), Eddy Current Testing (ECT), Magnetic Flux Leakage Testing (MFLT), and Infrared Testing (IRT) [9]. UT detects internal and surface defects by analyzing the propagation and reflection of ultrasonic waves within materials, with inspectors interpreting the reflected signals to determine the defect's location, size, and nature [10]. ECT, based on electromagnetic induction, identifies material defects by observing changes in the impedance of an induction coil [11]. MFLT is used to detect surface and near-surface flaws in ferromagnetic materials, where defects like cracks or corrosion cause magnetic flux to leak when the material is magnetized [12]. IRT identifies surface and subsurface defects by measuring temperature distribution or changes on the material's surface, utilizing infrared radiation to detect anomalies [13].

3. Modern Detection Methods

Early methods for detecting surface defects in products primarily relied on manual visual inspection and nondestructive testing technology. Although effective under certain conditions, these methods are subject to subjective factors and low efficiency due to their dependence on human operation. Consequently, with advancements in computer technology, machine learning has increasingly become a significant alternative, especially in large-scale production environments.

3.1. Simple Machine Learning Methods

Statistical pattern recognition technology in machine learning is one of the fundamental methods for detecting surface defects in products. Using a limited number of sample sets and a known statistical model or discriminant function, the model learns based on specific criteria. To improve classification accuracy, image preprocessing techniques are often employed to reduce noise, and feature extraction and selection technologies optimize the sample feature space to meet the requirements of the classification model. Finally, a classifier, such as a Bayesian classifier, decision tree, K-nearest neighbor method, or support vector machine, is used to identify product defects.

Pernkopf [14] proposed a method for detecting three-dimensional defects on the surface of steel blocks with oxide layers. The surface range data of the steel block were obtained using light sectioning technology, which addressed the problem of depth map recovery caused by vibrations of the steel block on the conveyor belt. Classification was performed using a Bayesian network classifier.

Aghdam et al. [15] introduced a method for classifying steel surface defects using decision trees combined with Principal Component Analysis and Bootstrap Aggregating. Local Binary Patterns are employed for feature extraction. The accuracy and speed of the decision tree classifier are enhanced by reducing dimensionality with PCA and improving performance through Bagging. Additionally, for multi-class classification tasks, the method incorporates a cascaded Support Vector Machine (SVM), further enhancing its real-time applicability in automatic surface inspection systems.

Li et al. [16] proposed a K-Means clustering image segmentation algorithm based on particle swarm optimization for detecting surface defects in automobile engine high-pressure oil circuit seals. The Speeded-Up Robust Features (SURF) algorithm was used to extract feature points from the seal image. Particle swarm optimization was then employed to optimize the initial clustering centers of the K-Means algorithm, thereby improving its clustering efficiency.

However, in real industrial production environments, defect detection is easily affected by factors such as product shape, position, lighting, and variations in industrial camera equipment. As a result, simple machine learning methods are only applicable in specific environments.

3.2. Methods Based on Deep Learning

With the advancement of deep learning technology, CNN-based classification networks have become crucial for surface defect detection. Surface defect classification using deep learning typically involves employing pre-trained networks such as VGG, ResNet, DenseNet, and SENet as backbone networks. A new network structure is then developed for specific detection tasks. An image of the product to be tested is input, and the network outputs its defect classification. Classification networks can be categorized into three methods based on their implementation: direct network classification, defect localization using the network, and feature extraction.

Masci et al. [17] used a maximum pooling CNN to classify steel surface defects. In seven defect classification tasks collected from actual production lines, this method achieved a 7% error rate. Compared with support vector machine (SVM) classifiers trained on common feature descriptors, this method performed at least twice as well.

Soukup et al. [18] used CNN to detect rail surface defects. They also applied regularization methods, including unsupervised hierarchical pre-training and data set augmentation. The study found that both unsupervised pre-training and data set augmentation significantly improved the classification performance.

3.3. Advanced methods in recent years

Ren et al. [19] proposed a Faster R-CNN classification network, which can improve efficiency by sharing full-image convolutional features, achieving almost zero-cost region proposal, and achieving the best object detection accuracy on multiple datasets through experiments.

Cai et al. [20] proposed a new detector architecture, Cascade R-CNN, which can extend R-CNN to solve common detection problems through a multi-stage approach. Each stage of the detector builds on the output of the previous stage to gradually improve the detection accuracy and alleviate the overfitting problem. The architecture improves the detection performance while only slightly increasing the computational requirements.

Li et al. [21] proposed an improved version of the YOLOv4 algorithm specifically for defect detection on steel strip surfaces. They integrated a convolutional block attention module into the backbone network and replaced the augmented path aggregation network with a design similar to that of sensory wild blocks.

Zhu et al. [22] developed AutoAssign, a new anchorless frame detection architecture that reduces the need for extensive manual adjustments. The method adaptively refines the label assignment strategy based on the features of different classes and instances. During training, a center-weighting and confidence-weighting module is combined to enable the assignment strategies for positive and negative samples to be tailored to each category's specific attributes and appearance.

Chobola et al. [23] proposed a method that combines migration learning with optimal migration mapping in the latent space. The method normalizes the feature space generated by the backbone network and uses the Sinkhorn algorithm to implement the optimal migration mapping.

Xiao et al. [24] use modules to embed the feature structure and align the distribution of new categories during inference. Effective migration classification is also enhanced by minimizing the Wasserstein distance.

Lu et al. [25] improved AutoAssig by proposing a new detection architecture, CA-AutoAssign, which has a streamlined channel-space adaptive feature pyramid network to mitigate the interference of complex backgrounds. The network is particularly effective in detecting defects of different sizes.

Li et al [26] proposed a novel feature map reconstruction network emphasizing foreground information. The network uses ridge regression to reconstruct query features from supporting features. It applies foreground weights to compute weighted distances, which are then used to predict the category distribution of the query image.

4. Experiments

4.1. Description of Few Steels Classification (FSC) Dataset

The FSC dataset [27] comprises 1,000 high-definition images organized into 20 categories, with each category containing 50 images of a specific defect type. The dataset includes 10 types of surface defects for cold-rolled steel and 10 types for hot-rolled steel. Examples of various types of surface defects in the FSC dataset are shown in Figure 1.

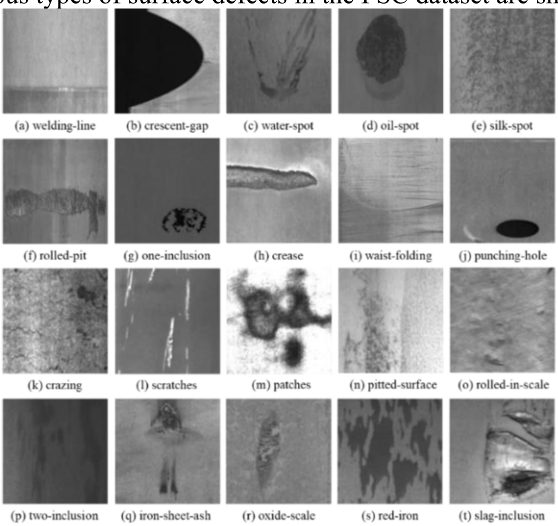

Figure 1. Examples of Each Type of Surface Defect in the FSC Dataset

4.2. Performance Evaluation

It is important to focus on critical metrics like accuracy and precision to effectively evaluate the proposed method's performance and identify its strengths and weaknesses.

Accuracy is a crucial measure of a model's classification ability, representing the ratio of correctly classified samples to total samples. Precisely, it reflects the proportion of True Positives (TP) and True Negatives (TN) among all samples. The calculation formula is shown in equation (1). A False Positive (FP) occurs when a negative sample is mistakenly identified as positive, while a False Negative (FN) refers to a positive sample being misclassified as unfavorable.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision quantifies the proportion of samples identified as positive by the model that are truly positive. Specifically, it is the TP to the total number of samples predicted as positive, including both TP and FP. Precision offers insight into the accuracy of the model's predictions, with its calculation formula presented in equation (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4.3. Experiment Setting and Results

4.3.1. Experimental Setting

In experimental Setting 1, both the training and test datasets are derived from the same source. The purpose of this experiment is to assess the model's defect recognition capabilities within that dataset. Only one of the cold-rolled steel or hot-rolled steel categories is selected, and ten defect samples are randomly selected as the training set (D_{Train}), and the remaining ten defect samples are used as the test set (D_{Test}).

In experimental setting 2, the training and test datasets are sourced from distinct datasets to assess the model's generalization capability across different domains. For instance, ten cold-rolled steel defect samples are selected as the training set (D_{Train}), and ten hot-rolled steel defect samples are designated as the test set (D_{Test}).

4.3.2. Performance in Same-Domain Setting (Experimental Setting 1)

Table 1 shows the specific results of setting 1. The CA-AutoAssign method achieved a classification accuracy of 75.76% in the 5way 1-shot and 82.36% in the 5way 5-shot. Compared with other methods, CA-AutoAssign improves the classification accuracy of 5way 1-shot by at least 2.07%. It improves the classification accuracy of 5way 5-shot by at least 4.23%. This suggests that the CA-AutoAssign method is highly effective in environments where the training and testing data share similar characteristics, reinforcing its robustness in controlled settings.

Table 1. Accuracy of recent advanced methods on the FSC dataset using experimental setting 1

Method	Backbone	5-way 1-shot	5-way 5-shot
Faster R-CNN [19]	ResNet-50	68.43%+0.11%	73.61%±0.13%
Cascade R-CNN [20]	ResNet-50	69.54%+0.12%	75.13%±0.11%
YOLOv4 [21]	CSPDarknet	71.36%+0.11%	76.21%±0.16%
AutoAssign [22]	CSPDarknet	73.26%+0.13%	78.13%±0.12%
PTNET [23]	WRN	71.28%±0.16%	73.68%±0.16%
GEDT [24]	WRN	73.69%±0.10%	75.69%±0.10%
CA-AutoAssign [25]	CSPDarknet	75.76%±0.13%	82.36%±0.13%
FRN [26]	ResNet-12	72.26%±0.12%	74.26%±0.12%

4.3.3. Performance in Cross-Domain Setting (Experimental Setting 2)

Table 2 shows the specific results of setting 2. The CA-AutoAssign method achieved a classification accuracy of 77.12% in the 5way 1-shot scenario and 85.86% in the 5way 5-shot scenario. Although the performance gains are more modest compared to the same-domain setting—1.43% in 5-way 1-shot and 0.74% in 5-way 5-shot—the results still affirm the model's ability to generalize across different domains. The ability to maintain high accuracy in cross-domain settings underscores the model's adaptability and robustness.

Table 2. Accuracy of recent advanced methods on the FSC dataset using experimental setting 2

Method	Backbone	5-way 1-shot	5-way 5-shot
Faster R-CNN [19]	ResNet-50	69.24%+0.12%	74.38%±0.13%
Cascade R-CNN [20]	ResNet-50	69.89%+0.11%	76.25%±0.11%
YOLOv4 [21]	CSPDarknet	73.17%+0.16%	77.36%±0.16%
AutoAssign [22]	CSPDarknet	74.27%+0.18%	79.38%±0.12%
PTNET [23]	WRN	73.68%±0.16%	77.98%±0.11%
GEDT [24]	WRN	75.69%±0.10%	82.00%±0.10%
CA-AutoAssign [25]	CSPDarknet	77.12%+0.21%	85.86%±0.13%
FRN [26]	ResNet-12	74.26%±0.12%	85.12%±0.11%

4.4. Comparative Analysis

The comparison between the two experimental setups reflects important aspects of the performance of each model. In the same-domain setup, this contributes to the model's accuracy due to the similarity between the training and test data. However, in cross-domain experiments, the accuracy of the models is more indicative of the potential of the models in real-world applications, where the models may need to process data from different sources. The relatively small performance gains in the cross-domain setup suggest that while CA-AutoAssign is effective, there is still room for further improvement, especially in improving generalization across different domains. Several challenges remain in the area of surface defect detection, particularly in terms of how to improve the ability of models to generalize across different production environments. Future research should focus on developing more adaptive algorithms.

5. Conclusion

In conclusion, this study provides a comprehensive overview of traditional and modern methods for surface defect detection, with a particular focus on deep learning techniques. The experiments conducted on the FSC dataset validate the effectiveness of advanced models such as CA-AutoAssign in identifying defects, even in challenging few-shot learning scenarios. With continued advancements in neural network architectures and adaptive learning techniques, surface defect detection is poised to become more accurate and efficient, significantly improving product quality in industrial production.

References

[1] Li Z, Wei X, Hassaballah M, Li Y, Jiang X. A deep learning model for steel surface defect detection. Complex & Intelligent Systems. 2024;10(1):885-97.

[2] Man Q, Zhang L, Cho Y. Efficient hair damage detection using SEM images based on convolutional neural network. Applied Sciences. 2021;11(16):7333.

[3] Lee J, Joo H, Lee J, Chee Y. Automatic classification of squat posture using inertial sensors: Deep learning approach. Sensors. 2020;20(2):361.

[4] Liu Y, Pu H, Sun D-W. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. Trends in Food Science & Technology. 2021;113:193-204.

[5] Hewamalage H, Bergmeir C, Bandara K. Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting. 2021;37(1):388-427.

[6] Li M-D, Huang Z-R, Shan Q-Y, Chen S-L, Zhang N, Hu H-T, Wang W. Performance and comparison of artificial intelligence and human experts in the detection and classification of colonic polyps. BMC gastroenterology. 2022;22(1):517.

[7] Czimmermann T, Ciuti G, Milazzo M, Chiurazzi M, Roccella S, Oddo CM, Dario P. Visual-based defect detection and classification approaches for industrial applications—a survey. Sensors. 2020;20(5):1459.

[8] Drury CG, Watson J. Good practices in visual inspection. Human factors in aviation maintenance-phase nine, progress report, FAA/Human Factors in Aviation Maintenance. 2002.

[9] Gupta M, Khan MA, Butola R, Singari RM. Advances in applications of Non-Destructive Testing (NDT): A review. Advances in Materials and Processing Technologies. 2022;8(2):2286-307.

[10] Honarvar F, Varvani-Farahani A. A review of ultrasonic testing applications in additive manufacturing: Defect evaluation, material characterization, and process control. Ultrasonics. 2020;108:106227.

[11] AbdAlla AN, Faraj MA, Samsuri F, Rifai D, Ali K, Al-Douri Y. Challenges in improving the performance of eddy current testing. Measurement and Control. 2019;52(1-2):46-64.

[12] Zhao D, Yang L, Liu Y, Geng H, Wang S, editors. A survey on pipeline magnetic flux leakage inspection data processing technology. 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022); 2022: SPIE.

[13] Pernkopf F. Detection of surface defects on raw steel blocks using Bayesian network classifiers. Pattern Analysis and Applications. 2004;7:333-42.

[14] Pernkopf F. Detection of surface defects on raw steel blocks using Bayesian network classifiers. Pattern Analysis and Applications. 2004;7:333-42.

[15] Aghdam SR, Amid E, Imani MF, editors. A fast method of steel surface defect detection using decision trees applied to LBP based features. 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA); 2012: IEEELi X, Zhu J, Shi H, Cong Z. Surface Defect Detection of Seals Based on K‐Means Clustering Algorithm and Particle Swarm Optimization. Scientific Programming. 2021;2021(1):3965247.

[16] Masci J, Meier U, Ciresan D, Schmidhuber J, Fricout G, editors. Steel defect classification with max-pooling convolutional neural networks. The 2012 international joint conference on neural networks (IJCNN); 2012: IEEE.

[17] Soukup D, Huber-Mörk R, editors. Convolutional neural networks for steel surface defect detection from photometric stereo images. International symposium on visual computing; 2014: Springer.

[18] Doshvarpassand S, Wu C, Wang X. An overview of corrosion defect characterization using active infrared thermography. Infrared physics & technology. 2019;96:366-89.

[19] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence. 2016;39(6):1137-49.

[20] Cai Z, Vasconcelos N, editors. Cascade r-cnn: Delving into high quality object detection. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.

[21] Li M, Wang H, Wan Z. Surface defect detection of steel strips based on improved YOLOv4. Computers and Electrical Engineering. 2022;102:108208.

[22] Chobola T, Vašata D, Kordík P, editors. Transfer learning based few-shot classification using optimal transport mapping from preprocessed latent space of backbone neural network. AAAI Workshop on Meta-Learning and MetaDL Challenge; 2021: PMLR.

[23] Zhu B, Wang J, Jiang Z, Zong F, Liu S, Li Z, Sun J. Autoassign: Differentiable label assignment for dense object detection. arXiv preprint arXiv:200703496. 2020.

[24] Xiao W, Song K, Liu J, Yan Y. Graph embedding and optimal transport for few-shot classification of metal surface defect. IEEE Transactions on Instrumentation and Measurement. 2022;71:1-10.

[25] Lu H, Fang M, Qiu Y, Xu W. An anchor-free defect detector for complex background based on pixelwise adaptive multiscale feature fusion. IEEE Transactions on Instrumentation and Measurement. 2022;72:1-12.

[26] Lu L, Cui X, Tan Z, Wu Y. MedOptnet: Meta-learning framework for few-shot medical image classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2023.

[27] Zhao W, Song K, Wang Y, Liang S, Yan Y. FaNet: Feature-aware network for few shot classification of strip steel surface defects. Measurement. 2023;208:112446.