Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37912

Multi-factor Analysis of Visual Inspection Technologies for Strip-Steel Surface Defects

Huinan Dong

School of Management, Harbin Institute of Technology, Harbin, China
2024110903@stu.hit.edu.cn

Abstract. Strip steel surface defect detection serves as a vital link for intelligent quality management in the steel industry. Surface defects can impair the mechanical properties of strip-steel products and bring potential economic risks during actual manufacturing, making effective defect inspection highly essential for industrial production. This paper conducts a comprehensive review on strip steel surface defect detection technologies. It sorts out the fundamental principles of four major technical categories, namely traditional detection, conventional machine-learning-based detection, deep-learning-driven detection, and generative data augmentation. The strengths and limitations of different representative methods are systematically compared. Combined with existing research based on public benchmark datasets such as NEU Surface Defect Detection Dataset (NEU-DET), this paper analyzes practical bottlenecks of existing approaches, including insufficient annotated defect samples, imbalanced sample categories, complex interference from industrial production environments, difficulties in multi-scale defect identification, and challenging model deployment requirements. Corresponding improvement suggestions are put forward from dataset construction, model robustness enhancement, multi-scale feature fusion, the fusion of Convolutional Neural Networks (CNN) and Transformer, as well as lightweight deployment. This review provides valuable references for subsequent research on intelligent strip steel quality inspection.

Keywords: strip-steel surface defect detection, multi-scale feature fusion, generative adversarial network, model lightweight deployment, sample imbalance

1. Introduction

As modern manufacturing develops toward high quality, high efficiency and intelligence, quality control for steel products attracts increasing attention. As an important product in the steel industry, strip steel is widely applied in automobile manufacturing, mechanical equipment, building construction, home appliances and other fields. Its surface quality directly affects the service performance and operational reliability of downstream products. However, during production processes such as rolling, cooling and transportation, defects including scratches, cracks, oxidation spots and inclusions easily occur on strip steel surfaces under the influence of process parameters, equipment conditions and environmental factors. These flaws not only spoil product appearances, they may also degrade material performance and increase production costs. Therefore, fast and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37912

accurate strip steel surface defect detection is of major significance for increasing production efficiency and guaranteeing product quality.

Traditional strip steel surface detection methods mainly contain manual visual inspection and non-destructive detection. Since manual inspection depends on inspectors' experience, it is vulnerable to subjective variables and has poor stability and efficiency, making it unable to meet the needs of contemporary high speed production lines. While magnetic detection, eddy current detection, and other technologies can detect partial defects, their detection techniques are relatively complex, and they hardly adapt to defect recognition tasks in complex industrial scenarios, their ability to identify multiple defects under complex backgrounds is still limited. Therefore, industrial production standards for automation and intelligent detection can be hardly met by traditional approaches.

In recent years, with the development of computer vision, machine learning and deep-learning technologies, intelligent strip steel surface defect detection has become a hot research topic. In order to achieve defect recognition through image acquisition, image processing, and feature extraction, early research primarily uses machine vision technology. Afterwards, defect classification tasks are subjected to machine learning techniques like random forests and support vector machines. Nevertheless, these methods depend on manually-designed features, and their adaptability toward complex environments and multi-scale defects is restricted. In recent years, deep-learning methods represented by CNN can automatically extract image features, and they achieve higher accuracy in defect classification, localization and segmentation. At present, models such as You Only Look Once (YOLO), Faster Region-based Convolutional Neural Network (R-CNN), U-Net and Vision Transformer are gradually used for strip steel surface defect detection, offering brand-new technical solutions for intelligent quality inspection on industrial production lines.

Though deep-learning-based strip steel defect-detection technologies have achieved phased research results, intelligent detection techniques still face practical application obstacles such as insufficient data, imbalanced defect samples, complex industrial environmental interference and difficulties in real-time model deployment. Therefore, how to further boost detection accuracy, inference real-time performance and generalization capability in industrial scenarios remains an important research direction. Based on the above research background, this paper carries out a review on strip steel surface defect-detection technologies; it systematically summarizes the development status of traditional detection methods, machine-learning methods and deep-learning methods, analyzes the advantages and drawbacks of different approaches, and discusses future development trends of strip steel surface defect-detection technologies. In the context of intelligent manufacturing, it offers references for steel quality control research.

2. Related work

Strip steel surface defect-detection technologies go through gradual development from manual empirical judgment and traditional image processing to machine learning and deep learning. According to defect-feature acquisition and recognition modes, existing research can be mainly divided into four categories: traditional detection, traditional machine learning, deep learning and generative data augmentation. Different methods show differences in feature-extraction modes, detection accuracy and computational complexity; their overall development trend shifts from manually-designed features toward automatic feature learning implemented by neural networks.

DOI: 10.54254/2755-2721/2026.37912

2.1. Traditional detection method

Traditional detection mainly includes manual visual inspection, physical non-destructive detection and manual image processing. Manual detection identifies defects by relying on operators' experience; it is vulnerable to visual fatigue and subjective bias and cannot guarantee detection efficiency and stability under high-speed continuous-production conditions. Non-destructive detection methods such as magnetic flux leakage testing, eddy-current testing and infrared testing obtain defect information by using steel's physical properties including electromagnetism and thermal characteristics. They have certain advantages in specific detection tasks, but detection results are easily affected by material properties, defect sizes and surface conditions, and they show limited capability for recognizing some tiny defects.

Gradient responses of images in horizontal and vertical directions are used by the Sobel operator to extract defect edge information; the following is a basic calculation formula:

$$G_x = S_x * I, \quad G_y = S_y * I \tag{1}$$

In formula (1), I denotes the input image; S_x and S_y represent Sobel convolution kernels in horizontal and vertical directions respectively; * stands for convolution operation; G_x and G_y are gradient responses in horizontal and vertical directions. By calculating gradients in the two directions, edges and gray-scale change regions inside images can be highlighted, which lays a foundation for edge extraction of strip steel surface defects.

With the development of machine vision, operators apply Sobel, Gabor, Local Binary Pattern (LBP) and other operators to strip steel defect detection. Sobel is mainly used for edge-information extraction; Gabor can describe textures of different scales and orientations; LBP owns the merit of simple calculation. Although these techniques rely on manually set filters and threshold values, they do not require complex network training. Manual extraction of features tends to drift when strip steel surfaces experience interference such as illumination variation, oil contamination, and dust. Because of this, even though conventional techniques have limited computing costs, their versatility in complex environments and multi-type defect detection are limited.

2.2. Traditional machine learning-based detection methods

Traditional machine learning generally follows the mode of "manual feature extraction + classifier / segmentation model". It extracts gray-scale, texture and geometric features by means of histogram of oriented gradients (HOG), gray-level co-occurrence matrix and other tools; then it completes classification or segmentation using hidden Markov models, low-rank sparse decomposition, Support Vector Machine (SVM), random forests and other approaches.

Xu et al. proposed the Context-adaptive Hidden Markov Tree (CAHMT) multi-scale hidden Markov tree model, and it reduces segmentation errors of tiny defects via hierarchical strategies [1]; Zhou et al. constructed the Double Low-rank and Sparse Decomposition (DLRSD) dual low-rank sparse decomposition saliency model, and it extracts defect foreground by combining the Otsu threshold algorithm [2]; Wang et al. built a guidance-template algorithm. It reconstructs backgrounds by using statistical features of normal strip-steel images so as to realize defect localization, and this algorithm achieves favorable detection efficiency and accuracy on corresponding datasets [3]. To a certain extent, the above-mentioned methods can improve the performance of traditional defect detection, yet model effects still depend on the quality of manually-extracted features. Defects such as scratches, pockmarks and plaques vary greatly in scale,

DOI: 10.54254/2755-2721/2026.37912

shape and texture; a single manual feature cannot cover all defect types. Meanwhile, changes in production environments will cause shifts of feature distributions, so traditional machine learning is gradually restricted by poor generalization capacity.

The support vector machine realizes sample classification by searching for hyperplanes with the maximum classification interval; its optimization objective is expressed as follows:

$$min_\{w,b,\xi\}\ 1/2\,||w||\hat{}\,2 + C\,\Sigma_\{i=1\}\hat{}\,\{n\}\,\xi_i \tag{2}$$

In formula (2), w denotes the weight vector of the classification hyperplane; b is the bias term; ξ_i represents the slack variable of the i-th sample; C is the penalty parameter; n means the quantity of training samples. This optimization objective balances classification intervals and classification errors, so the model maintains good generalization performance while keeping classification capability.

2.3. Deep learning-based detection methods

Deep learning enables neural networks to automatically learn texture, semantic and spatial features from images, and it lowers dependence on manual feature engineering. It has become an important method for strip steel surface defect detection at present. Among various models, one-stage object-detection networks of the YOLO series attract wide attention for their satisfactory detection speed. Gao et al. put forward YOLOv5s-DDC; it introduces Deformable Convolutional Networks v2 (DCNv2) deformable convolution and coordinate attention mechanism, and it reaches 80.4 % mean average precision (mAP)@0.5 on the NEU-DET dataset [4]. Yuan proposed FEB-YOLOv8; it adopts Partial Convolution (PConv), Efficient Multi-Scale Attention (EMA) attention and lightweight Bidirectional Feature Pyramid Network (BiFPN) to optimize feature extraction and multi-scale fusion [5]. Shen et al. designed YOLOv11-WDS; it adds Residual Feature Decomposition (WaveletRFD) wavelet down-sampling modules and re-parameterized structures, which improves detection performance and cuts down model parameters and computation load simultaneously [6].

Ren et al. put forward DDR-YOLO, which adopts Dual-Branch Dynamic Structure Convolution module (DualSPDG), Dynamic Attention Multi-Scale Fusion module (DAMF) and Restructurable-parameter-fusion large-kernel convolution attention aggregation block (RepLK-AA) modules to mitigate the difficulties brought by large-scale variations in steel strip defect sizes. It achieves mAP50 improvements of 3.6 % and 3.9 % on NEU-DET and GC10-DET datasets respectively, while maintaining competitive real-time inference performance [7].

Most above-mentioned YOLO-derived schemes mainly rely on convolutional operations to extract local defect features. Nevertheless, pure convolutional architectures are limited in capturing long-range contextual information for complex and irregular defects. To compensate for such deficiency, researchers have started to introduce Transformer components into detection frameworks.

With the application of Transformer in computer vision, its global-feature-modeling capability is also adopted for strip steel defect detection. Taking YOLOX as the basic framework, Du et al. introduced Swin-Tiny to construct AFF-Net [8]. AFF-Block and Foc-FPN are used to strengthen global-feature modeling and multi-scale fusion, and this network obtains 83.5 % mAP on the NEU-DET dataset. Liu et al. proposed CGTD-Net; it combines Swin Transformer with spatial-channel attention mechanisms to realize pixel-level defect segmentation, and it achieves 75.16 % mean Intersection over Union (mIoU) on segmentation tasks of NEU-DET [9]. Compared with object

DOI: 10.54254/2755-2721/2026.37912

detection, segmentation methods can further acquire information of defect contours and areas, but they usually consume more computing resources.

Aiming at real-time requirements of industrial online detection, some research focuses on lightweight networks. MINet extracts defect features by using multi-scale depth-wise convolution and feature interaction [10]. It only has 0.28 M parameters, and its Graphics Processing Unit (GPU) inference speed reaches 721 Frames Per Second (FPS). It maintains high detection speed while reducing model complexity.

Convolutional neural networks carry out local weighted calculation for input features through convolution kernels, so as to extract texture, edge and morphological features of defects layer by layer. The basic convolution computation is written as:

$$Y(i,j) = \Sigma_m \Sigma_n X(i+m, j+n) K(m,n) + b \tag{3}$$

In formula (3), X denotes the input feature map; K represents parameters of convolution kernels; b is the bias term; Y stands for the output feature map obtained from convolution calculation; i and j are spatial positions on output feature maps; m and n are position indexes inside convolution kernels. Convolution operations extract features of input data via local receptive fields, and multi-layer networks can generate more abstract defect-feature representations step-by-step.

2.4. Generative data-augmentation methods

Industrial defects usually suffer limited sample quantity and imbalanced category distribution, hence Generative Adversarial Networks (GAN) are used for training-data expansion. However, GAN is mainly applied for data augmentation; it cannot independently complete full defect-localization tasks. Besides, generated images may contain blurring artifacts, so GAN is generally combined with object-detection or segmentation networks.

Table 1. Comparison of different detection methods

Method Category	Typical Method	Feature Acquisition Mode	Advantages	Limitations
Traditional Detection	Sobel, Gabor, LBP	Manual design	Simple process, low computation cost	Insufficient robustness
Machine Learning	SVM	Manual features + classifier	Strong classification capability	High dependence on feature engineering
Deep Learning	CNN, YOLO	Automatic feature learning	High accuracy, high-level automation	High demand for data and computing resources
Generative Methods	GAN, CGAN	Data generation	Relieve data-shortage problems	Potential distribution deviation

3. Application status and limitations

3.1. Application status

Different detection methods show distinct characteristics in practical application. Traditional manual inspection and non-destructive detection possess solid technical foundations; they still have application value in specific detection and quality spot-checking scenarios, yet their automation level and complex-image-processing capacity are limited. Traditional machine learning can finish

DOI: 10.54254/2755-2721/2026.37912

basic defect classification and identification, but it relies on manually-designed features, so deep-learning-based schemes are gradually replacing its applicable scenarios.

At present, lightweight object-detection networks based on YOLO constitute an important research direction for online detection. Approaches including YOLOv5s-DDC, FEB-YOLOv8 and YOLOv11-WDS adopt attention mechanisms, deformable convolution, multi-scale fusion and lightweight structures. They strike a balance between detection accuracy and inference speed, and they display promising potential for online detection. Mixed Transformer models such as AFF-Net and CGTD-Net have powerful global-feature-modeling and defect-contour-expression capabilities; they fit scenarios requiring high detection accuracy and defect-region information, but their computation cost is relatively high [9]. Lightweight models such as MINet pay more attention to real-time performance and resource consumption [10]. GAN is mostly used for data expansion in the training phase, and it works together with object-detection and segmentation networks.

Public datasets such as NEU-DET and GC10-DET serve as important experimental foundations for relevant research; they provide unified conditions for performance comparison among different algorithms. Nevertheless, public datasets differ from real production environments in imaging conditions, defect distributions and background features. Therefore, experimental results obtained from laboratory datasets cannot fully reflect detection performance on real production lines.

To further compare application effects of different deep-learning-based strip steel surface-defect-detection approaches, this paper sorts out typical research methods, major improvement strategies and corresponding performance indicators, as shown in Table 2.

Table 2. Performance comparison of typical strip steel surface-defect-detection methods

Method	Basic Model	Main Improvement	Dataset	Evaluation Metrics
YOLOv5s-DDC	YOLOv5s	DCNv2 + CA	NEU-DET	mAP@0.5 = 80.4 %
FEB-YOLOv8	YOLOv8	PConv + EMA + BiFPN	NEU-DET	—
YOLOv11-WDS	YOLOv11	WaveletRFD + Rep	NEU-DET	—
AFF-Net	YOLOX + Swin	AFF-Block + Foc-FPN	NEU-DET	mAP = 83.5 %
CGTD-Net	Swin Transformer	Spatial-Channel Attention	NEU-DET	mIoU = 75.16 %
MINet	CNN	Multi-scale depth-wise convolution	—	0.28 M / 721 FPS

3.2. Limitations

First of all, insufficient data and imbalanced sample categories remain critical factors affecting the performance of detection models. In real industrial production, some defects occur at low frequencies, so plenty of real-world samples are hard to acquire, and obvious data-distribution gaps exist among different defect categories. GAN can supplement samples for minority categories to a certain degree, but distribution deviations may exist between generated data and real-world data; over-reliance on generated samples may degrade model generalization capability.

Secondly, model robustness under complex industrial environments still needs further improvement. Practical production conditions contain dust, illumination variation, mirror reflection, oxidation layers and other interferences. These factors will reduce contrast between defects and backgrounds, making tiny defects harder to identify. Attention mechanisms and deformable convolution can strengthen defect-feature extraction, yet image-distribution shifts caused by complex industrial environments may still bring about performance drops of models.

DOI: 10.54254/2755-2721/2026.37912

Thirdly, multi-scale defect detection brings practical difficulties. Defects including pockmarks, scratches and plaques differ greatly in size and morphology; small-size defects tend to lose detailed information in the network down-sampling process, while complex multi-scale fusion structures will increase computation volume. Hence, how to retain local information of tiny defects as well as overall features of large-size defects at the same time is an important research problem at present.

Fourthly, trade-offs exist between detection accuracy and real-time deployment. Transformer and multi-branch CNN models own strong feature-expression capability, but they usually have large parameter quantity and high computation load. Lightweight models can raise inference speed and lower deployment costs, yet they may lose partial fine-grained defect information. Accordingly, achieving a good balance among detection accuracy, model complexity and real-time performance is a key problem for further industrial application of strip steel surface defect detection.

On the whole, existing research has built a technical-development system of "traditional detection — traditional machine learning — deep learning — generative data augmentation". However, problems such as insufficient data resources, adaptability for complex environments, multi-scale feature representation and model deployment have not been completely solved. These problems lay a foundation for follow-up research on data augmentation, multi-scale feature fusion, model lightweight design and the fusion of CNN and Transformer.

To comprehensively evaluate the performance of defect-detection models, apart from detection accuracy, false-detection and missed-detection conditions should also be taken into consideration. Therefore, indicators such as precision, recall and F1-score are commonly adopted for evaluation.

$$Precision = TP / (TP + FP) \tag{4}$$

In formula (4), TP denotes true positives, namely samples with real defects correctly detected by the model; FP represents false positives, which refer to samples without real defects but misjudged as defective ones by the model. $Precision$ stands for the proportion of real defective samples among all samples predicted as defects by the model; higher precision values mean fewer false detections produced by the model.

Recall mainly reflects the model's detection capability for real defective samples; its calculation formula is shown below:

$$Recall = TP / (TP + FN) \tag{5}$$

In formula (5), TP is the number of true-positive samples; FN means false-negative samples, namely samples containing real defects yet failing to be detected by the model. $Recall$ represents the proportion of real defective samples successfully identified by the model; higher recall values correspond to fewer missed-detection cases.

F1-score can be adopted to synthesize precision and recall; its calculation formula is:

$$F1 = 2 \times Precision \times Recall / (Precision + Recall) \tag{6}$$

In formula (6), $Precision$ is precision value; $Recall$ is recall value. The F1-score is the harmonic mean of these two indicators; it can comprehensively reflect model performance in reducing false detection and missed detection. When both precision and recall maintain high values, the F1-score will increase correspondingly, so it is suitable for comparing comprehensive performance of different defect-detection models.

DOI: 10.54254/2755-2721/2026.37912

For defect-segmentation tasks, it is necessary to evaluate the overlap degree between predicted defect regions and real defect regions, and IoU is widely used for this purpose. Its calculation formula is:

$$IoU = |A \cap B| / |A \cup B| \tag{7}$$

In formula (7), A denotes defect regions predicted by the model; B stands for real defect regions; $A \cap B$ is the intersection of predicted regions and real regions; $A \cup B$ is their union set. IoU measures the overlap degree between predicted defect areas and real defect areas; higher IoU values indicate more accurate prediction results for defect positions and ranges.

Combined with previous analysis on existing research, current strip steel surface defect detection faces challenges in data resources, environmental conditions, feature expression and model deployment. Specific problems and corresponding improvement directions are listed in Table 3.

Table 3. Existing problems and improvement directions for strip steel surface defect detection

Existing Problems	Specific Manifestations	Improvement Directions
Insufficient data	Limited quantity of defect samples	Data augmentation, GAN
Category imbalance	Scarce samples of minority defects	Resampling, generative augmentation
Environmental interference	Dust, reflection, illumination variation	Attention mechanism, robust augmentation
Multi-scale defects	Easy loss of small-defect features	Multi-scale feature fusion
Insufficient real-time performance	Large computation volume of models	Lightweight design, pruning, quantization

4. Suggestions

Aiming at problems in the strip-steel-defect-detection field, including insufficient data, imbalanced sample categories, complex industrial environmental interference, difficulties in multi-scale defect identification and real-time model deployment, improvements can be carried out from multiple aspects such as data processing, feature extraction, model structure and practical deployment in the future. With continuous development of deep-learning and computer-vision technologies, simply stacking network modules and raising model complexity to pursue higher detection accuracy can hardly satisfy practical industrial demands. High mAP indicators obtained on public datasets cannot guarantee that algorithms can directly adapt to factory-site conditions. Hence, future research should pay more attention to the comprehensive balance among detection accuracy, robustness, real-time performance and industrial adaptability, and laboratory-dataset accuracy should not be taken as the sole evaluation criterion.

First of all, improve the construction system of industrial-defect datasets and optimize data-augmentation strategies. Current research generally relies on public datasets such as NEU-DET and GC10-DET, yet defect types, sample quantity and collection environments in public datasets still differ from real industrial production lines. Image-acquisition conditions of public datasets are relatively ideal, and defect samples are pre-screened, while images collected from industrial sites are mixed with various kinds of noise. Especially in real-production processes, some defects such as cracks and pockmarks occur at low frequencies, which will bring sample-quantity imbalance among different categories. Models tend to favor defect categories with sufficient samples, which weakens recognition capability for minority-category defects. Therefore, real-production-line data can be collected on a large scale in future work; samples covering different production batches, rolling processes and working conditions should be expanded so as to raise dataset diversity. On this basis,

DOI: 10.54254/2755-2721/2026.37912

traditional data-augmentation operations such as rotation, cropping, flipping, brightness adjustment and noise overlay, as well as generative models like GAN can be used to expand samples of minority categories. Nevertheless, excessive dependence on generated data should be avoided. Generated samples are prone to texture distortion and unreasonable morphology, so authenticity and effectiveness of generated samples need to be verified and screened. Distribution gaps between generated data and real-world data should be narrowed, preventing model-performance degradation caused by wrong-sample introduction.

Secondly, enhance model robustness in complex industrial environments. Practical production sites are easily affected by dust, illumination variation, mirror reflection, oxidation layers and other factors. These interferences will mix defect textures with background textures, reduce differences between defects and backgrounds, and further cause false detection or missed detection of models. Accordingly, in the model-training phase, disturbed samples with different illumination, noise, blurring and contrast variation can be added. Models are supposed to learn intrinsic essential features of defects rather than superficial interference features. In terms of network structures, attention mechanisms can be further adopted to guide models to focus on defect regions and suppress negative influences brought by invalid background information such as dust and reflection on detection results. Meanwhile, transfer-learning methods can be used; feature representations pre-trained from public datasets are adopted, and then models are fine-tuned with a small number of real industrial samples. In this way, models can quickly adapt to different production environments, and their cross-scene generalization capability and industrial-environment adaptability get improved.

Thirdly, strengthen multi-scale feature-extraction capability of models. Strip steel surface defects show obvious differences in size and morphology; tiny defects such as scratches and pockmarks as well as large-size defects such as plaques may appear in one single detection scene. Small-size defects are easy to disappear in the network down-sampling process, while blindly stacking multi-scale modules will greatly increase model computation overhead. Hence, feature-fusion structures can be further optimized in future research; multi-scale-fusion mechanisms such as FPN (Feature Pyramid Network) are adopted to fully fuse shallow texture details and deep semantic features, and redundant convolution units are streamlined to cut unnecessary network computation. For tiny defects, high-resolution feature layers can be reserved, so that models can make full use of edge and texture information of defects and improve detection capability for small-target defects.

Fourthly, explore fusion architectures of CNN and Transformer. Existing research indicates that CNN has advantages in extracting local texture and edge features, while Transformer possesses strong global-feature-modeling capability. For strip steel surface defects, scratches, pockmarks and other defects require attention to local-texture variation; defect recognition under complex backgrounds also needs context information in a larger scope to distinguish real defects from random noise. Reasonable combination of CNN and Transformer can make full use of both local and global features, and it improves model-recognition capability for different types of defects. However, Transformer structures usually consume large computing resources and bring higher parameter quantity and inference latency. For this reason, future research should not simply increase network complexity; lightweight fusion schemes should be designed for industrial-detection scenarios to control model parameter quantity and computation load while improving detection accuracy.

Finally, promote lightweight optimization of models and industrial-landing verification. Although some deep-learning models can achieve high detection accuracy at present, they have large parameter quantity and high computation load. Their direct deployment on industrial production lines is restricted by embedded-hardware performance and real-time requirements. Therefore,

DOI: 10.54254/2755-2721/2026.37912

lightweight networks, depth-wise separable convolution, model pruning, quantization, knowledge distillation and other methods can be adopted to reduce model complexity; edge-computing equipment is combined to raise online-detection speed. At the same time, model evaluation should not only focus on single detection-accuracy indicators; recall rate, F1-score, mAP, parameter quantity, inference speed and other indicators should be comprehensively considered. Besides, experimental verification across different devices, batches and production environments should be strengthened; complex field working conditions are fully simulated to test anti-interference ability and generalization performance of models, making models adapt to practical production conditions.

Overall, research priorities for future strip steel surface defect detection shift from single-accuracy optimization toward collaborative improvement of accuracy, robustness, real-time performance and industrial adaptability. By perfecting industrial-defect datasets, optimizing data-augmentation strategies, strengthening multi-scale feature modeling, fusing advantages of CNN and Transformer, and realizing iterative lightweight models, the engineering-application value of defect-detection algorithms can be greatly enhanced. It provides reliable technical support for intelligent quality inspection and automatic management of strip-steel production.

5. Conclusion

This paper systematically sorts out the development course and application status of strip steel surface-defect-detection technologies, and it analyzes research progress of traditional detection, machine learning, deep learning and generative data-augmentation methods. Traditional detection methods are easy to implement and own mature application foundations, yet they are easily affected by manual experience and environmental factors. Traditional machine learning realizes defect recognition relying on manually-extracted texture, gray-scale and morphological features, but it has strong dependence on feature design work. Deep learning can automatically extract defect features; it shows favorable performance in object detection, defect segmentation and classification tasks, and it has gradually become an important research direction for strip steel surface defect detection.

According to analysis of existing research, current strip steel surface defect detection still faces problems such as insufficient data, imbalanced sample categories, complex industrial-environment interference, multi-scale-defect-identification difficulties and real-time-model-deployment obstacles. Especially in real-production environments, dust, illumination variation, mirror reflection and other factors may affect detection results, and feature information of small-scale defects is easy to get lost in the model-feature-extraction process. Meanwhile, high-precision models usually have large parameter quantity and heavy computation load, which restricts the industrial-landing application of algorithms.

Aiming at these above-mentioned problems, future research can further strengthen the construction of industrial datasets and data augmentation to improve models' adaptability for different defect categories and complex environments. Multi-scale feature fusion and attention mechanisms can be used to enhance models' recognition capability for tiny defects and complex backgrounds. Advantages of CNN and Transformer can be combined to improve comprehensive-expression capability for local details and global information. In addition, attention should be paid to model lightweight design; pruning, quantization, knowledge distillation and other approaches are adopted to reduce computation costs and raise real-time-detection performance of models.

In general, strip steel surface defect detection is iteratively upgraded from traditional manual inspection toward automation and intelligence. Future research should guarantee detection accuracy, and further improve model robustness, real-time performance and industrial adaptability. Relevant

DOI: 10.54254/2755-2721/2026.37912

detection technologies should be promoted from laboratory-research stage to practical-production application, providing technical support for intelligent quality control in the steel industry.

References

[1] Xu, K., Song, M., Yang, C. L., & et al. (2013). Application of hidden Markov tree model in on-line detection of steel strip surface defects. *[J]., 2013, 49*(22), 34–40.

[2] Zhou, S., Wu, S., Liu, H., Lu, Y., & Hu, N. (2018). Double low-rank and sparse decomposition for surface defect segmentation of steel sheet. *Applied Sciences, 8*(9), 1628. doi: 10.3390/app8091628.

[3] Wang, H., Zhang, J., Tian, Y., Chen, H., Sun, H., & Liu, K. (2019). A simple guidance template-based defect detection method for strip steel surfaces. *IEEE Transactions on Industrial Informatics, 15*(5), 2798–2809. doi: 10.1109/TII.2018.2887145.

[4] Gao, W., & Zhou, Y. (2024). YOLv5s DDC: A rapid detection network for strip steel surface defects based on deformable convolution with coordinate attention and distribution shifting convolution. *2024 10th International Conference on Computer and Communications (ICCC)*, Chengdu, China, pp. 2073–2078. doi: 10.1109/ICCC62609.2024.10942311

[5] Yuan, J., Wan, F., Lei, G., & Xu, L. (2023). FEB YOLOv8: A steel surface defect detection algorithm based on improved YOLOv8s. *2023 3rd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Wuhan, China, pp. 629–633. doi: 10.1109/CEI60616.2023.10527815.

[6] Shen, J., Guan, S., Hou, L., & Dong, H. (2025). An improved surface defect detection algorithm for strip steel with YOLOv11. *2025 6th International Conference on Electronic Communication and Artificial Intelligence (ICECAI)*, 426–429. https://doi.org/10.1109/ICECAI66283.2025.11170574

[7] Ren, X., Liu, Z., Cheng, H., & Chen, L. (2025). DDR-YOLO: An efficient detection model based on YOLO11s for steel surface defect. *IEEE Access, 13*, 147202–147217. https://doi.org/10.1109/ACCESS.2025.3599978

[8] Du, Y., Chen, H., Fu, Y., Zhu, J., & Zeng, H. (2024). AFF Net: A strip steel surface defect detection network via adaptive focusing features. *IEEE Transactions on Instrumentation and Measurement, 73*, 1–14. https://doi.org/10.1109/TIM.2024.3398131

[9] Liu, H., Chen, C., Hu, R., Bin, J., Dong, H., & Liu, Z. (2024). CGTD Net: Channel wise global transformer based dual branch network for industrial strip steel surface defect detection. *IEEE Sensors Journal, 24*(4), 4863–4873. https://doi.org/10.1109/JSEN.2023.3346470

[10] Shen, K., Zhou, X., & Liu, Z. (2024). MINet: Multiscale interactive network for real-time salient object detection of strip steel surface defects. *IEEE Transactions on Industrial Informatics, 20*(5), 7842–7852. https://doi.org/10.1109/TII.2024.3366221