Journal of Electronic Research and Application, 2026, Volume 10, Issue 4
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Research on the Application of Computer Vision in Equipment Fault Diagnosis

Xiaoquan Zhu[1], Siyu Wang[1], Tong Zhang[2], Pengyuan Chen[3]

[1]China Construction Third Engineering Bureau Group (Shenzhen) Co., Ltd., Shenzhen 518109, Guangdong, China
[2]Faculty of Construction and Environment, The Hong Kong Polytechnic University, Hong Kong, China
[3]School of Artificial Intelligence, Jianghan University, Wuhan, China

Abstract: With the continuous improvement of industrial automation, rapid and accurate diagnosis of equipment faults is the key to ensuring production safety and efficiency. With the advantages of non-contact sensing, real-time processing and high-precision recognition, computer vision has broad application prospects in fault diagnosis. This technology integrates image acquisition, feature extraction and deep learning models to automatically identify and classify equipment faults such as appearance damage, motion abnormalities and thermal state changes. Multi-modal image fusion further improves fault positioning accuracy under complex working conditions. In scenarios such as mine electrical equipment, construction engineering inspection cold-chain storage and unmanned aerial vehicle (UAV) inspection, its detection performance is superior to traditional methods, providing strong technical support for building an intelligent equipment operation and maintenance system and promoting the in-depth integration of industrial Internet and intelligent manufacturing.

Keywords: Computer vision; Equipment fault diagnosis; Deep learning; Feature extraction; Image recognition

Online publication: May 21, 2026

1. Introduction

The stable operation of industrial equipment is an important guarantee for modern production. The concealment and suddenness of faults are likely to cause economic losses and safety accidents. Traditional diagnosis relies on manual inspection, which is inefficient, subjective and difficult to adapt to complex working conditions. Artificial intelligence technology has promoted innovation in the field of equipment diagnosis, and computer vision, which senses equipment status through images, has become a mainstream technical path. Relying on high-definition imaging and neural network models, it can monitor the appearance, motion and thermal characteristics of equipment without shutdown, realizing early warning and accurate positioning of faults. This technology has been applied in power, mining, agriculture, aviation and other fields, with continuously improving engineering capabilities, and has important research and practical value.

2. Core principles and method system of computer vision technology

2.1. Image acquisition conditions and technical specifications for preprocessing

In equipment fault diagnosis, image quality directly determines the effectiveness of subsequent analysis. Industrial sites are generally plagued by strong noise, low illumination, high dust and other interferences, resulting in low signal-to-noise ratio of original images [1]. Therefore, establishing standardized acquisition standards is a basic link in the entire diagnosis process. At the hardware level, industrial cameras, infrared thermal imagers or depth cameras are selected according to different detection objects, combined with reasonable lighting and protective packaging to ensure stable image acquisition. At the software level, original images are sequentially subjected to denoising, enhancement and normalization processing to make the input data meet the quality specifications required for model training. Experiments show that standardized preprocessing improves the convergence speed in the feature extraction stage by about 25% on average, and the model recognition accuracy by 8%–12% (**Table 1**).

Table 1. Comparison of common image preprocessing methods

Method	Applicable Scenarios	Processing Effect	Computational Complexity
Gaussian Filtering	Gaussian noise suppression	Smooths images and preserves edges	Low
Histogram Equalization	Low-contrast image enhancement	Stretches dynamic range	Low
Morphological Processing	Target extraction after binarization	Fills holes and removes burrs	Medium
Adaptive Threshold Segmentation	Uneven illumination scenarios	Enhances local contrast	Medium
Bilateral Filtering	Detail-preserving denoising	Removes noise while preserving edges	High

The standardized design of the preprocessing process not only improves the usability of single-frame images, but also provides a unified data format foundation for cross-scenario model migration, creating favorable conditions for subsequent feature extraction and model training.

2.2. Multi-scale feature extraction and representation learning methods

Feature extraction is the core link in the computer vision fault diagnosis process. Traditional methods rely on manually designed descriptors (such as SIFT, HOG), which have limited adaptability to illumination changes and complex industrial backgrounds, restricting recognition accuracy in real scenarios. Convolutional Neural Networks (CNNs) automatically learn hierarchical features from low-level textures to high-level semantics through multi-layer convolution operations. The convolution feature extraction process can be expressed as:

$$F^{(l)}=\sigma\left(W^{(l)}*F^{(l-1)}+b^{(l)}\right) \tag{1}$$

Where $F^{(l)}$ is the feature map of the l-th layer, $W^{(l)}$ is the convolution kernel weight, $b^{(l)}$ is the bias term, and $\sigma(\cdot)$ is the nonlinear activation function.

This mechanism endows the model with natural adaptability to translation invariance, performing prominently in tasks such as surface defect and deformation recognition. Multi-scale feature fusion strategies

(such as Feature Pyramid Network, FPN) further enhance the perception ability for fault targets of different sizes, which is particularly critical in small target detection scenarios such as chain links and tiny cracks. Studies show that after introducing FPN, the mAP of small target detection is improved by about 7.3% compared with the single-scale baseline, effectively guaranteeing the generalization ability of the model under complex working conditions (**Figure 1**).

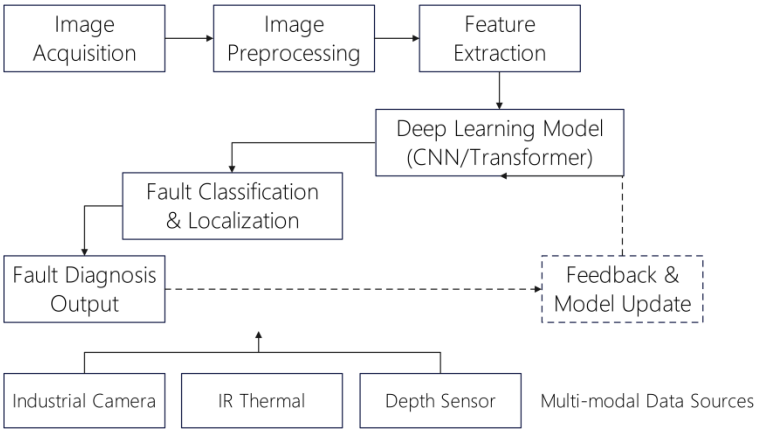

Figure 1. Computer vision fault diagnosis system architecture.

2.3. Construction of deep learning models for fault recognition

On the basis of feature extraction, the selection of model architecture determines the final efficiency of fault recognition. Single-stage detection models represented by YOLOv8 are widely used in real-time fault detection tasks due to their fast inference speed and flexible deployment. The single-frame inference delay can be compressed to within 22 milliseconds, meeting the real-time requirements of industrial online detection. The introduction of Transformer architecture further enhances the model's ability to model global context information, thus improving accuracy in fault positioning tasks with occlusion and complex backgrounds. In the model training process, data augmentation (rotation, cropping, brightness perturbation) and transfer learning strategies effectively alleviate the problem of insufficient industrial annotated samples, enabling the model to have good robustness even under small-sample conditions. It is worth noting that model design must balance accuracy and inference efficiency. Excessive pursuit of network depth often leads to difficulties in edge deployment, so lightweight design has become an important consideration for engineering implementation [2].

3. Key technical implementation of computer vision in equipment fault detection

3.1. Visual anomaly recognition for equipment appearance damage

Equipment appearance damage is the most direct perception object in fault diagnosis, covering cracks, corrosion, deformation, wear and other morphological types. Based on image classification and target detection technology, the system can automatically annotate surface images of equipment and judge fault categories, replacing the traditional manual visual inspection process [3]. Studies show that the detection model with ResNet-50 as the backbone network achieves a mean Average Precision (mAP@0.5) of 92.3% in metal surface crack recognition tasks, with a detection speed of 30 frames per second, meeting the real-

time requirements of online detection. Compared with manual detection, the vision system has significant advantages in missed detection rate and false alarm rate, eliminating the uncertainty caused by human subjective judgment and greatly improving the consistency and repeatability of detection results. For fine damages such as surface micro-cracks, introducing an attention mechanism to amplify the weight of high-response areas can effectively improve the perception sensitivity to weak fault features, further raising detection accuracy to over 95%, thus having strong engineering applicability in precision equipment appearance quality inspection scenarios [4].

3.2. Motion state analysis and dynamic fault perception

For continuously moving mechanical equipment, static images are difficult to capture fault characteristics caused by abnormal speed, excessive vibration or offset motion trajectory. Therefore, introducing video sequence analysis and optical flow estimation technology has become a key path. Based on the two-stream network architecture, the system can simultaneously model the spatial features and temporal dynamics of equipment motion, effectively identifying dynamic faults such as chain slack, conveyor belt offset and unbalanced rotating components. In the quantitative evaluation of target positioning accuracy, Intersection over Union (IoU) is the core indicator to measure the coincidence degree between the detection box and the ground truth box, defined as:

$$\mathrm{IoU}=\frac{\left|B_p \cap B_{gt}\right|}{\left|B_p \cup B_{gt}\right|} \tag{2}$$

Where B_p is the prediction box and B_{gt} is the ground truth annotation box.

Experiments show that the two-stream network achieves an accuracy of 94.7% in conveyor belt offset detection, an increase of about 11% compared with single-frame static detection, and the fault response time is compressed to the second level. In addition, the abnormal gradient distribution of the optical flow field can also be used as a precursor signal of equipment motion abnormality, enabling the system to trigger an early warning mechanism before obvious deformation of the equipment, striving for a time window for active intervention by maintenance personnel [5].

3.3. Multi-modal image fusion and accurate fault positioning

A single image modality often has information blind spots under complex working conditions, leading to inaccurate fault positioning. Fusing visible light images with infrared thermal images, 3D point clouds and other multi-modal data can obtain equipment status information from different perception dimensions, thus significantly improving positioning accuracy [6]. Among the two main strategies of image-level fusion and feature-level fusion, feature-level fusion is preferred in engineering practice because it can retain the deep semantic information of each modality and has higher tolerance for resolution differences between modalities. Studies show that the infrared-visible fusion scheme improves the recall rate by 18.6% compared with the single-modal scheme in the automatic detection of dead poultry in caged environments, reducing the positioning error to the pixel level; in high-dust mine scenarios, the fusion strategy of 3D point clouds and grayscale images achieves a 3D positioning accuracy of 1 mm for chain deformation detection (**Table 2**).

Table 2. Performance comparison of multi-modal image fusion schemes

Fusion Method	Input Modalities	Typical Applications	Positioning Accuracy Improvement	Deployment Complexity
Image-level Fusion	Visible light + Infrared	Surface temperature anomaly detection	+9.2%	Low
Feature-level Fusion	Visible light + Infrared	Defect positioning in complex backgrounds	+18.6%	Medium
Decision-level Fusion	Visible light + 3D Point Cloud	3D deformation measurement	+23.4%	High
Cross-modal Attention	RGB + Depth Map	Precision parts inspection	+27.1%	High

The selection of a multi-modal fusion strategy must be comprehensively weighed according to on-site deployment conditions and accuracy requirements. The system complexity should be controlled as much as possible on the premise of ensuring detection performance, making the scheme feasible for practical implementation [7].

4. System performance verification and engineering application in typical scenarios

4.1. Experimental dataset construction and evaluation index design

The objective evaluation of system performance relies on high-quality datasets and a complete evaluation index system. In terms of dataset construction, industrial fault samples have the inherent characteristics of unbalanced category distribution and high annotation cost [8]. Data augmentation (rotation, cropping, contrast adjustment) and Generative Adversarial Networks (GAN) are usually used to expand minority samples to ensure the completeness of category coverage in the training set. Taking the chain wear detection task as an example, the original positive-negative sample ratio is about 1:8, which is adjusted to 1:2 after GAN amplification, and the model recall rate is increased by about 9.4%. In terms of evaluation index design, precision, recall, mAP and F1-score are core indicators to measure fault detection performance. For real-time detection scenarios, efficiency dimensions such as inference delay and frame rate must also be considered (**Table 3**).

Table 3. Core evaluation indicators of fault detection system

Indicator	Brief Formula	Focus Dimension	Applicable Scenarios
Precision	TP / (TP + FP)	False alarm control	Scenarios with high false alarm cost
Recall	TP / (TP + FN)	Missed detection control	Scenarios with high missed detection cost
mAP@0.5	Mean AP of all categories	Comprehensive detection accuracy	Multi-category target detection
F1-score	Harmonic mean of P and R	Balance between precision and recall	Unbalanced category scenarios
Inference Delay (ms)	Single-frame processing time	Real-time performance	Online detection deployment

The joint evaluation of multi-dimensional indicators can comprehensively reflect the comprehensive performance of the system in terms of accuracy, efficiency and robustness, avoiding potential performance shortcomings being covered by a single indicator, thus providing a clear improvement direction for

subsequent optimization and iteration.

4.2. Comparative analysis and optimization of fault diagnosis accuracy

Through systematic comparative experiments, the diagnostic performance differences of different algorithm architectures on the same fault dataset can be objectively evaluated, thereby guiding the targeted optimization of the model [9]. In the chain wear detection task, the YOLOv8 model achieves an mAP@0.5 of 93.8%, an increase of about 21% compared with traditional machine vision methods, and the inference speed remains above 45 frames per second, thus having practical feasibility for online deployment. Aiming at the problems of false detection and missed detection, after introducing the channel attention mechanism and multi-scale loss function optimization strategy, the false detection rate is reduced from 4.2% to 1.8%, the missed detection rate from 3.7% to 1.5%, and the comprehensive F1-score is increased to 96.5%. The continuous strengthening of model generalization ability depends on the strict control of data cleaning quality and annotation consistency, which directly affect the stability after final deployment. In addition, under the condition of limited edge computing power, transferring the capabilities of large models to lightweight networks through knowledge distillation reduces the inference delay to within 18 milliseconds, while keeping the accuracy loss within 1.3%, thus achieving collaborative optimization of accuracy and efficiency.

4.3. Deployment practice in construction engineering, mining, agriculture, UAV inspection and other scenarios

The aforementioned technical system has achieved mature engineering implementation in multiple industries, showing significant application benefits [10]. In the field of construction engineering inspection, the Low-Altitude Technology Industry Division of China Construction Third Engineering Bureau Group independently developed a Building Engineering Inspection Management System. The core functions of this system are based on artificial intelligence and computer vision recognition technology, enabling automated inspection task execution, data collection and analysis, hazard identification and closed-loop management, while integrating safety control and team collaboration capabilities. Leveraging automated task execution, AI defect recognition, closed-loop hazard management, and GIS visualization, the system significantly improves the efficiency and accuracy of traditional inspections, especially in high-risk, high-cost, and hard-to-reach scenarios. In the construction industry, safety inspections of elevated machinery and work platforms have long relied on manual labor, which is characterized by high danger, high repetition, heavy workload, and susceptibility to omissions, leading to recurring safety accidents on active construction sites. Intelligent and automated inspection therefore represents a critical means of overcoming these limitations. Through self-developed flight control software, the system enables autonomous UAV path planning and “one-click” launch of automated inspection missions from the ground. For key structural components such as tower crane standard sections, sleeves, jibs, and counterbalance arms, the UAV performs reciprocating automated flight and image capture from top to bottom. Using high-fidelity image compression technology, images captured across various scenarios are uploaded to the platform in real time for storage and subsequent review. The platform integrates AI algorithms capable of automatically identifying and flagging ten categories of safety hazards, including tower crane pin retraction, missing bolts, and structural corrosion. Single-tower inspection time has been reduced from a manual 0.5–3 hours to 15 minutes, improving inspection efficiency by 50%, achieving an accuracy rate of 98%, and increasing overall work efficiency by a factor of ten, with significantly enhanced safety [11,12].

In the field of equipment operation and maintenance, the 3D intelligent visual inspection system for coal mine scraper conveyors has increased overall work efficiency by more than 40 times. In the agricultural field, a vision detection system based on infrared-visible fusion has realized automatic identification of dead poultry in caged environments, significantly reducing the risk of disease spread, with a detection recall rate superior to manual inspection. In the field of UAV inspection, the collaborative deployment of lightweight vision models and edge computing devices has transformed fault identification of transmission lines and wind turbine blades from manual periodic inspection to real-time automatic early warning, compressing the inspection cycle from days to hours. The engineering practices in the above multiple scenarios verify the wide applicability and deployment reliability of the computer vision fault diagnosis system, laying a solid foundation for the large-scale promotion of the intelligent operation and maintenance system.

5. Conclusion

Computer vision provides an efficient, intelligent and scalable technical path for equipment fault diagnosis, and has formed a complete technical closed loop from image acquisition and preprocessing, deep learning fault feature extraction and recognition to multi-modal fusion engineering implementation. This technology shows good adaptability and reliability in scenarios such as construction engineering inspection mine electrical equipment monitoring, cold-chain storage and UAV inspection, verifying the feasibility of industrial implementation. At present, problems such as illumination changes, occlusion interference, difficulties in small target detection and insufficient cross-scenario generalization still limit the improvement of system performance. In the future, with the development of lightweight networks, interpretable models and edge computing, fault diagnosis based on computer vision will develop towards higher accuracy, stronger real-time performance and wider application scope, providing solid support for the comprehensive construction of industrial intelligent operation and maintenance systems.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang Q, 2026, Design of Intelligent Fault Diagnosis System for Electrical Equipment in Coal Mines based on Computer Vision and Deep Learning. International Journal of System Assurance Engineering and Management, 2026(prepublish): 1–11.

[2] Lakhan M, Oskar G, A.P. W, et al., 2026, Detection of Glacier Calving Events from Time-Lapse Images Using Computer Vision and a Neural Network. Journal of Glaciology, 2026(72): e5.

[3] Jiang T, Mu H, Liang L, et al., 2026, Robust Detection of Dead Broilers in Caged Environments Using Infrared-Visible Image Fusion and Computer Vision. Measurement, 258(PE): 119502–119502.

[4] Feature Extraction and Image Processing for Computer Vision, Elsevier Ltd, October 11, 2025.

[5] Harikrishna K, Sen B, Nithin A, et al., 2025, Advanced Computer Vision Algorithm for Extraction of Microstructural Features from BSE Images of Powder Metallurgical Microstructures. The International Journal of Advanced Manufacturing Technology, 140(7–8): 1–19.

[6] Liao D, Bi R, Zheng Y, et al., 2025, LCW-YOLO: An Explainable Computer Vision Model for Small Object Detection in Drone Images. Applied Sciences, 15(17): 9730–9730.

[7] Najafi J, Mirzakuchaki S, Shamaghdari S, 2025, Autonomous Drone Detection and Classification Using Computer Vision and Prony Algorithm-Based Frequency Feature Extraction. Journal of Intelligent & Robotic Systems, 111(1): 8–8.

[8] Junfeng J, Tian G, Weichuan Z, et al., 2023, Image Feature Information Extraction for Interest Point Detection: A Comprehensive Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 4694–4712.

[9] Cedric C, Thomas F, 2023, Computer Vision and Internet Meme Genealogy: An Evaluation of Image Feature Matching as a Technique for Pattern Detection. Communication Methods and Measures, 17(1): 17–39.

[10] Feng Y, 2023, An Intelligent Detection Method of Local Feature Points in Computer Vision Image. International Journal of Information and Communication Technology, 23(3): 266–277.

[11] Xi C, Yali M, ShuHui L, 2023, Vision based Defect Detection Technologies in Civil Structures: A Review Study. Journal of Optics, 53 (2): 1456–1461.

[12] Fang C, Yu S, Li Y, et al., 2024, Deep Learning-Based Computer Vision Health Monitoring for Civil Engineering. Journal of Tongji University (Natural Science), 52(2): 213–222.