The Development History and Frontier Paradigms of Industrial Visual Quality Inspection Empowered by Deep Learning

Ruiyang Tang *

School of Resources and Safety Engineering, Central South University, Changsha, Hunan Province, China

Abstract. Industrial visual quality inspection technology has achieved breakthroughs in four technological levels: manual inspection, traditional machine vision, deep learning, and multimodal intelligence. It has gradually shed human dependence and moved towards intelligent, autonomous decision-making. Traditional methods, constrained by human subjectivity and rigid rules, are unable to cope with the complexity of industrial scenarios. A deep learning solution combining convolutional neural networks with an object detection framework enables automatic defect identification and classification, making defect detection more accurate and faster. Multimodal solutions utilize composite data inputs, including visual features and acoustic cues, to optimize system fault tolerance. Generative AI effectively compensates for the lack of rare defect samples. Lightweight models and edge-cloud collaboration facilitate the implementation of technology in real-time environments. Industrial quality inspection systems are forming a tightly integrated intelligent closed-loop system of "perception-understanding-action," achieving a systematic upgrade from passive response to active intervention. By summarizing the technological evolution and key innovations, this paper seeks to promote deeper integration of AI and industrial inspection, paving the way for more adaptive and autonomous quality control systems.

1 Introduction

In the stage of industrial intelligent transformation, the manufacturing industry is currently in a major transformation stage of intelligence and digitalization, which has prompted a comprehensive upgrade of the production operation model. In the product manufacturing stage, quality maintenance has always been the primary demand of the production link. The core basis for production line value assessment is product quality parameters. Today, industrial quality inspection methods have significantly enhanced the efficiency of defective product detection, which is the main way and safety barrier to promote production efficiency and stable output.

* Corresponding author: 8210242429@csu.edu.cn

Manual inspection methods suffer from inherent limitations such as subjective judgment, high physical exertion, and low efficiency. In today's high-volume, high-standard, and digital production environment, these methods have become a core obstacle to modern manufacturing processes. Early visual systems based on rule-based engines, which employed manually designed features (such as HOG and SIFT) combined with classifiers, were able to accomplish basic automated quality inspection tasks. However, these approaches presented a high barrier to entry for industrial products and were out of touch with production realities, such as those involving variable backgrounds, rare defects, and high-speed requirements. In response to these technological demands, the rise of deep learning, integrated with data-driven concepts, has overcome the bottlenecks of existing technologies and promoted a comprehensive upgrade of industrial quality inspection systems. Advances in technologies such as CNNs, Transformers, and ViT have driven the rapid development of deep learning applications in industrial visual quality inspection. Technologies such as large-scale lightweight models and self-supervised learning can be used to address the challenges of data annotation and the high hardware requirements of edge devices in industry. Combining recent academic research with industry application experience, this paper comprehensively outlines the development path of deep learning-enabled industrial visual quality inspection methods, explains the underlying logic behind technological iteration, and analyzes future technological development paradigms from the perspective of practical needs.

2 The Technological Evolution of Industrial Visual Quality Inspection

The evolution of industrial visual quality inspection is inseparable from the technical support of artificial intelligence. It has gone through four technical stages: manual, traditional vision, deep learning, and multimodal, completing a comprehensive transition from human intervention to intelligent automation. Each stage of development has effectively aligned innovative technologies with the core demands of the industry. The phased evolution of industrial visual quality inspection technology and its significant characteristics are summarized in Table 1:

Table 1: Comparison of industrial vision quality inspection technology evolution.

Stage	Time Period	Technical Characteristics	Core Methods/Technologies	Technical Limitations	Typical Applications
Manual Visual Inspection	Before the 1970s	Relies on human eye observation and simple tools	Manual visual inspection + calipers/micrometers	Miss detection rate of 15%-20%, high subjectivity, easy fatigue	Random inspection of mechanical dimensions, fabric defect sorting
Traditional Machine Vision	1970s-2010s	Optical imaging + rule-based algorithms	Canny Edge Detection, Otsu Segmentation, Template Matching	Can only handle preset defects, poor flexibility	Metal plate inspection, electronic component positioning
Deep Learning AI Quality Inspection	2010s-2020s	Data-driven, self-supervised feature learning	CNN, YOLO, Few-Shot Learning	Relies on annotated data, high computing power	Semiconductor wafer inspection, automotive body weld inspection

				requirements	
Multimodal Intelligent Quality Inspection	2020s - Present	Cross-modal fusion, cognitive intelligence	GPT-4V + YOLOv8, Stable Diffusion	Complex system, high integration cost	PCB inspection, 3D measurement of curved workpieces

2.1 Manual Visual Inspection Stage

In the initial stages of industrial quality inspection, work was entirely manual and conducted using simple inspection tools. Product defect determination criteria were subjective, resulting in significant fluctuations in product yield rates during this phase. Technically, the core evaluation criterion for manual visual inspection was subjective judgment, with basic measurement work primarily supported by a few simple tools, such as calipers. During this quality inspection phase, given the gradual decline in human function, missed inspections due to fatigue accounted for 15%-20% of defective products. Inspection efficiency was limited by the inspector's reaction speed and subjective judgment. Emotional states, overly subjective judgments, and a lack of experience often resulted in inconsistent quality inspection standards, reducing the controllability of product quality. Conventional manufacturing, the primary vehicle for quality inspection activities at this stage, primarily involves tasks such as measuring bearing bores and sorting out broken threads and stains in fabric. These activities barely meet the requirements for low-precision, low-volume quality inspections, directly hindering the transition to high-quality, high-efficiency, large-scale production models.

2.2 Traditional Machine Vision Stage

With the iterative upgrades of optical imaging technology and machine learning algorithms, industrial quality inspection has completed the transition to automation, breaking through the rate limit of manual inspection. The inspection focus has shifted from defect elimination to quality control. Technological breakthroughs at the hardware level are the starting point of development in this stage. Industrial cameras have achieved breakthroughs in resolution of tens of millions of pixels, and the upgrade of CMOS image sensor photosensitivity has achieved a dual improvement in industrial imaging clarity and low-light adaptability. By adopting telecentric lenses and linear scan lenses, higher-precision image acquisition is achieved, so that the inspection quality of workpieces such as metal sheets is reliably guaranteed. The collaborative control of motion controllers and PLCs realizes the coordinated operation of quality inspection and production lines, and achieves an automated closed loop for inspection and sorting.

In addition to technological breakthroughs at the hardware level, the widespread use of algorithmic detection technology has become the core driving force for the development of quality inspection. OpenCV, Halcon and other algorithms with manual feature annotation are used to complete fast and high-precision identification of workpieces. The Canny operator is used to complete edge detection, and the Otsu method is combined to perform threshold segmentation. Template matching and collaborative Blob analysis are used for defect judgment to achieve preliminary processing of feature extraction. In the detection implementation stage, sub-pixel positioning is used to optimize the measurement accuracy to ±0.1μm [1], thereby meeting the measurement requirements of high-precision parts. The integration of the two stages leads traditional industrial quality inspection technology into

the automation stage. Figure 1 shows the defects of PCB, and Figure 2 is an intuitive presentation of fabric defects.

Fig. 1. Circuit board manufacturing defects [2]

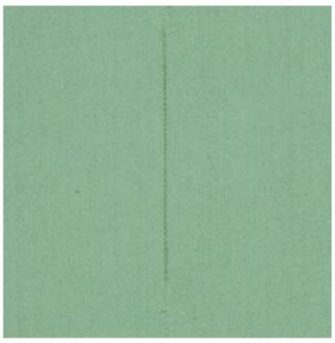

Fig. 2. Fabric defects [3]

2.3 Deep Learning AI Quality Inspection Stage

Traditional vision technology, used in industrial quality inspection, often exposes technical flaws due to its rule-driven parameter system. Its detection capabilities only cover pre-calibrated defect types. Changes to production line configurations, such as those encountered with interfering textures (such as wood surfaces) or submillimeter defects (such as chip scratches), can significantly increase the rate of false positives, necessitating engineers to recalibrate the detection logic. The rapid rise and maturity of deep learning technology have significantly revolutionized the existing industrial quality inspection model, transitioning from a rule-based approach to a new, data-driven one. This breakthrough in autonomous feature learning has substantially broken through the technical limitations and application boundaries of traditional machine vision. During the rapid expansion of deep learning applications, from 2012 to 2016, convolutional neural network technology made rapid progress [4]. LeNet became the first model to successfully classify and detect surface defects in electronic components. AlexNet, with its hierarchical feature extraction architecture, significantly surpassed existing methods in its ability to discriminate industrial defects [5-6]. Deep learning technology has penetrated the industrial quality inspection industry, driving a leap forward in the overall technological advancement of the industry. In 2017, with the help of models such as the YOLO series and Faster R-CNN, defect location and classification could be completed in one go. YOLOv5 uses CSPDarknet as its basic architecture, and the detection speed of photovoltaic module EL images exceeds 30 frames/second [7]. ResNet with a residual structure significantly alleviates the gradient vanishing dilemma of deep

networks. Figure 3 fully illustrates that ResNet has achieved a recall rate of more than 92% for cold-rolled steel strip crack detection [8]. In 2020, in order to reduce the requirements for large-scale labeled data in the training stage, the use of small sample learning methods can significantly reduce the dependence on labeled data. Its semi-supervised framework has extremely low requirements for labeled samples, achieving an effect of reducing the false alarm rate of electronic component detection by more than 40% compared with traditional methods [9].

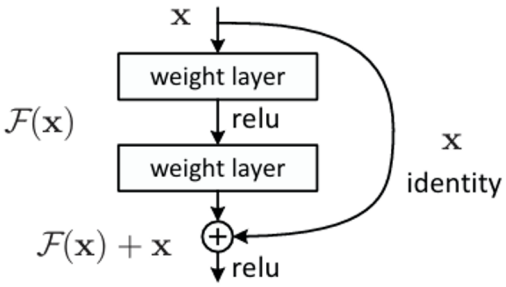

Fig. 3. Residual learning: a building block [10].

The development of lightweight architectures such as MobileNet and SqueezeNet has paved the way for the implementation of deep learning models. Their detection accuracy has been greatly improved compared to traditional image technology. The fast and efficient detection capabilities based on deep learning have made it a core technology path for industrial quality inspection [11].

2.4 Multimodal Intelligent Quality Inspection Stage

The industrial quality inspection sector is entering a period of innovative development characterized by "cross-modal integration + intelligent cognition," enabling intelligent integration and upgrades across the entire quality inspection process. The key focus lies in overcoming the cognitive limitations of traditional single-modal vision and comprehensively leveraging multi-channel data and knowledge assets to improve quality management. The current core technology framework focuses on three key areas: multimodal fusion for cross-modal collaboration to enhance inspection robustness, generative AI to address the long-tail sample gap, and an integrated inspection and execution system driven by embodied intelligence. Multimodal technology enables correlation and analysis of image and text processing data. For example, combining GPT-4v with YOLOv8 enables joint analysis of PCB inspection and design drawing text, reducing false positives caused by camera angles. Stable Diffusion within generative AI can efficiently generate simulated defect data such as metal corrosion and glass scratches, expanding the dataset size and thereby improving model generalization and, consequently, the model's discriminative capabilities. Embodied intelligence technology can significantly reduce measurement errors caused by manual labor and traditional technologies and effectively cope with environmental complexity. Collaborative innovation at the hardware level effectively supports technological implementation. FPGAs are used to improve the inference speed of lightweight edge models, achieving inference latency of less than 50 milliseconds to ensure real-time quality inspection. Structured light scanning technology enables 3D vision systems to achieve high-precision measurement of three-dimensional defects with an accuracy of ±0.01mm, making them particularly suitable for quality inspection of curved workpieces such as automotive wheels [12-13]. Hybrid computing architectures (GPUs paired with AI chips, with heterogeneous

computing platforms represented by NVIDIA Jetson) can effectively support the training and deployment of large models, paving the way for their practical application. Figure 4 systematically summarizes the hierarchical relationship between AI technology, embodied intelligence theory, and its robotic applications.

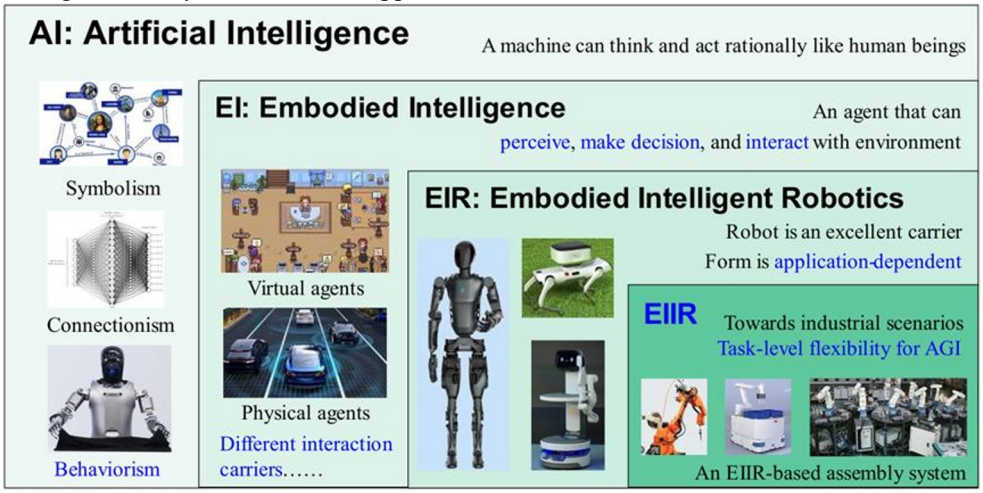

Fig. 4. Connections involving AI, and notably EIIR [14].

3 Deep Learning Drives the Future Paradigm of Industrial Visual Quality Inspection

3.1 Multimodal Fusion Detection System

Exploration of the sample space is limited to the scope of visual images, making it difficult to fully grasp the key characteristics of the sample and limiting the complete description of complex industrial defects. With the continuous evolution of multimodal fusion technology, the detection framework that combines multimodal data such as image, sound, and force feedback has become a cutting-edge development direction. With the advancement of sensing technology, parameters such as temperature and vibration will be integrated into the multimodal model system to form a more systematic quality inspection framework. The key technologies of multimodal fusion detection are detailed in Table 2.

Table 2: Core technologies of multimodal fusion detection system.

Modality Type	Sensors/Data Sources	Fusion Methods	Application Scenarios
Visual	Industrial cameras, 3D structured light	Image + point cloud fusion	Defect detection of curved workpieces
Acoustic	Microphone arrays, voiceprint sensors	Acoustics + vibration analysis	Abnormal noise detection of bearings
Force	Torque sensors, pressure sensors	Force feedback + visual positioning	Assembly quality monitoring
Temperature	Infrared thermal imagers	Thermal imaging + visual images	Welding quality evaluation
Modality Type	Sensors/Data Sources	Fusion Methods	Application Scenarios

3.2 Lightweight and Edge-Cloud Collaborative Architecture

The rigid requirements of the product quality inspection process for low latency and elastic deployment drive the upgrade of models to lightweight and edge-cloud linkage architectures. Given the limited computing power of edge nodes, it is difficult to achieve real-time reasoning of large models due to computing bottlenecks. The use of multimodal methods to integrate various sensor data to achieve an overall analysis of workpieces is also limited. Implementers apply knowledge distillation and quantization technology to ensure that the model output quality does not decrease significantly, and achieve a 70% reduction in model size to match the computing power boundaries of edge nodes [15], significantly reduce implementation costs, and combine cloud-based model reasoning to improve the reliability of edge equipment reasoning results in industrial quality inspection scenarios. Precision medical wire production lines now adopt an edge-cloud collaborative operation model: when the edge intelligent vision device detects a deviation, it immediately sends a signal to the control module, correcting production parameters in real time and completing millisecond-level process corrections. Relying on data accumulation, the cloud implements iterative model upgrades, building a closed-loop mechanism of "edge real-time processing-cloud continuous optimization". To comprehensively enhance the automation and intelligence capabilities of industrial production and improve production quality and efficiency, it is necessary to fully unleash the characteristics and value of each layer based on lightweight models and edge-cloud collaborative solutions to maintain production stability. Table 3 presents the main deployment modes of the edge-cloud collaborative architecture.

Table 3: Typical configuration modes of edge-cloud collaborative systems.

Hierarchy	Function	Hardware Examples	Task Type
Edge End	Real-time inference, control feedback	FPGA, Jetson Nano	Defect detection, process correction
Edge Side	Data aggregation, lightweight model update	Industrial gateway, industrial computer	Multi-device collaboration, model fine-tuning
Cloud End	Model training, data storage	GPU cluster, cloud server	Large model training, historical analysis

3.3 Generative AI-Enabled Solutions for Defect Detection and Migration

Defect types in industrial production are extremely complex, and defect data exhibits a long-tail distribution. Low-data-volume scenarios present significant data collection challenges and high costs. Therefore, using generative AI to generate supplementary data is an effective approach to addressing the shortage of rare defects. Automated generation of defect samples using GANs and diffusion models can significantly enhance dataset diversity. Using diffusion models to generate scarce samples has become a core strategy for addressing data scarcity. When conducting quality inspections on automotive parts, generative models can reproduce sparse defect patterns such as surface flaws, significantly enhancing the model's ability to detect new types of defects. Table 4 summarizes the key application scenarios and results of generative AI in defect detection.

Table 4: Technologies using generative AI for defect detection.

Generative Model	Generated Defect Types	Training Data Requirements
GAN	Scratches, dents, stains	Hundreds of normal samples

Diffusion Model	Metal corrosion, glass cracks	A small number of defect samples + synthetic data
VAEs	Texture anomalies, color deviations	No annotation required, self-supervised learning

3.4 "Collection-Optimization-Deduction" Closed-Loop Intelligent Model

Driven by major breakthroughs in deep learning and an overall leap in automation, industrial quality inspection models are expanding from point-based defect identification to linear quality control chains. A closed-loop "detection-optimization-prediction" system will become the industry's mainstream model. This system organically integrates multi-source data models and knowledge bases, rapidly collects data through an edge-cloud linkage framework, and instantly generates risk grading maps. Intelligently implements control based on output data, enabling autonomous regulation throughout the entire production cycle. Virtual Technology's DLIA system relies on manual re-inspection to achieve a closed-loop feedback loop, enabling model self-upgrades and a self-evolutionary closed-loop of "iterative enhancement and increased refinement." Its technology has already been implemented in major industrial sectors such as 3C electronics and new energy. This closed-loop operation transforms quality inspection from a marginal link into a core decision-making unit driving manufacturing optimization, driving the manufacturing sector's upgrade from delayed detection to proactive prevention.

4 Conclusions

This study fully presents the development trajectory of industrial visual quality inspection technology from the manual judgment stage, to traditional machine vision relying on rule bases, and then into the intelligent quality inspection stage driven by deep learning, and gradually transitioned to a new paradigm of multimodal intelligent fusion. Traditional methods have technical limitations when faced with complex defects and dynamic scenes due to the non-adaptability of their rule systems; the revolutionary progress of deep learning models, using convolutional neural networks combined with detection algorithms, has achieved a qualitative leap in performance indicators and developed into a core solution for modern quality inspection. People are in a stage where the dual characteristics of "integration" and "cognition" are highlighted. Multimodal technology combines heterogeneous data such as images, sounds and text, breaking free from the constraints of a single visual dimension; generative AI technology has successfully addressed the limitations of a shortage of long-tail defect samples; with the help of model optimization and edge-cloud linkage mechanisms, it has successfully resolved the problems of insufficient computing power and high costs in industrial AI practice. Industrial visual quality inspection will gradually form an intelligent system with a closed loop of "detection-analysis-decision-making-optimization" throughout the entire process. Its mechanism has evolved from passive quality control to intelligent decision-making nodes that autonomously control process flow and achieve predictive maintenance, continuously driving the manufacturing industry towards the trinity of intelligence, flexibility and high quality.

References

1. Guo, J., Yang, Y., & Chen, X. (2024). Research on sub-pixel accuracy flange disk dimension measurement based on machine vision. Signal, Image and Video Processing, 18(8), 6185–6195

2. Raihan, F. I., & Ce, W. (2017, November). PCB defect detection using OpenCV with image subtraction method. In 2017 International Conference on Information Management and Technology (ICIMTech) (pp. 204–209). IEEE.
3. Jia, X. (2010, July). Fabric defect detection based on open source computer vision library OpenCV. In 2010 2nd International Conference on Signal Processing Systems (Vol. 1, pp. V1-342). IEEE.
4. Jha, S. B., & Babiceanu, R. F. (2023). Deep CNN-based visual defect detection: Survey of current literature. Computers in Industry, 148, 103911.
5. Wan, L., Chen, Y., Li, H., & Li, C. (2020). Rolling-element bearing fault diagnosis using improved LeNet-5 network. Sensors, 20(6), 1693
6. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.
7. Wang, Z., Shao, W., & Hao, Y. (2023). Surface defect detection based on improved YOLOv5. Modular Machine Tool & Automatic Manufacturing Technique, 4, 177–182.
8. Konovalenko, I., Maruschak, P., & Brevus, V. (2022). Steel surface defect detection using an ensemble of deep residual neural networks. Journal of Computing and Information Science in Engineering, 22(1), 014501
9. Wang, A., Zhang, Y., Wu, H., Liu, X., & Zhou, Y. (2020). Few-shot learning based balanced distribution adaptation for heterogeneous defect prediction. IEEE Access, 8, 32989–33001
10. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778).
11. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobilenetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510–4520).
12. Wei, L., & Peng, L. (2021, April). An efficient OpenCL-based FPGA accelerator for MobileNet. In Journal of Physics: Conference Series (Vol. 1883, No. 1, p. 012086). IOP Publishing.
13. Yu, X. Y., Zhang, F. B., Yuan, H. J., & Lang, Z. G. (2005). A structured-light-based on-line 3D measurement system and software development. Key Engineering Materials, 295, 399–404.
14. Zhang, C., Zhang, C., Xu, Z., Xie, Q., Hou, J., Feng, P., & Zeng, L. (2025). Embodied intelligent industrial robotics: Concepts and techniques. arXiv preprint arXiv:2505.09305.
15. Tanama, C., Peng, K., Marinov, Z., Stiefelhagen, R., & Roitberg, A. (2023, October). Quantized distillation: Optimizing driver activity recognition models for resource-constrained environments. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 5479–5486). IEEE.