Proceedings of the CONF-MLA 2026 Symposium: Explainable Computing, Modeling & Data Science in Complex Systems
DOI: 10.54254/2755-2721/2026.GF37194

Smart Manufacturing System Architecture with Intelligent Convergence: Technological Evolution and Implementation Pathways

Xuanyi Chen

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
cxuanyi2024@outlook.com

Abstract. As the global manufacturing industry accelerates its digital and intelligent transformation, the deep integration of Artificial Intelligence (AI) with the Internet of Things (IoT) is emerging as a central driver behind the evolution of smart manufacturing systems. This paper provides a comprehensive review of the technical architecture, core technologies, and typical application scenarios of smart manufacturing within the context of this integration. From a technological perspective, it examines the transition from convolutional neural networks to vision transformers in industrial quality inspection, and explores how large language models enhance manufacturing interaction and test automation by serving as industrial Copilots, knowledge base question-answering systems, and code generators. The study also highlights the foundational role of IoT in data acquisition, real-time transmission, and intelligent analytics. In terms of applications, it demonstrates productivity improvements and modeling innovations across areas such as predictive maintenance, intelligent quality inspection, production scheduling, and smart warehousing. Despite notable progress, several challenges persist, including data silos, insufficient interpretability of models, difficulties in integrating operational technology with information technology, and a lack of interdisciplinary talent. Future research should prioritize strengthening data infrastructure, optimizing model architectures, advancing system integration and ecosystem development, and facilitating the shift of smart manufacturing from specialized tools to universal platforms, thereby offering stronger support for the intelligent advancement of the manufacturing industry.

Keywords: Artificial intelligence, Internet of Things, vision transformer, large language model, digital transformation

1. Introduction

With the continuous transformation of production methods, the global manufacturing industry faces significant challenges in industrial transformation and upgrading: production costs are under increasing pressure, industry regulations are becoming more stringent, and factors such as production efficiency and environmental issues impose higher demands on operational profitability

© 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.GF37194

[1]. To consolidate and enhance international competitiveness, various countries have introduced strategic policies related to manufacturing to drive development: China launched "Made in China 2025", the United States released the "Advanced Manufacturing Strategy", Germany formulated the "Industry 2030 Strategy", and Japan introduced the "Connected Industries Strategy".

Currently, knowledge discovery and service applications in manufacturing still primarily rely on general big data and cloud computing technologies. A few leading enterprises at the high end of the industrial chain, leveraging their own advantages, integrate data from various internal management platforms and intelligent equipment resources to carry out limited-scope knowledge services such as internal management, decision support, and solutions [2]. During the digital transformation process, manufacturing enterprises face a new set of contemporary challenges, such as how to cope with the high cost and complexity of digital transformation, how to fully leverage the driving role of data elements, how to adapt to the rapid iteration of digital technologies, and how to re-evaluate and adjust existing business processes. Under the dual pressure of inherent difficulties and new challenges in the digital era, manufacturing enterprises urgently need to seek corresponding solutions and governance paths [3]. In recent years, under the dual influence of artificial intelligence and the Internet of Things, the manufacturing industry has presented new development modalities. Traditional manufacturing is shifting from labor-intensive to intelligent, standardized, and normalized production, making production processes more streamlined. For instance, in areas such as predictive equipment maintenance, intelligent quality inspection and appearance, intelligent production scheduling and planning, and smart warehousing and logistics, the IoT serves as the connection foundation for the perception system, while AI forms the decision-making and optimization engine. Together, they constitute the closed-loop technical system of "comprehensive perception — real-time transmission — intelligent decision-making — precise execution".

Take machine vision online inspection as an example. The online monitoring method based on machine vision principles refers to a non-contact, automatic detection method for monitoring particle information.

To get the particle size and distribution info through getting and processing image info. Online monitoring of based on the principle of machine vision compared to others is that it is safer, more efficient, accurate, precise, and informative [4]. In traditional inspections, there is only manual visual checks or some simple mechanical sensors and single electromagnetic induction use many people's manpower for production. Though it helped with the use of manpower, hard to get around the issue of product quality deviation due to eye fatigue and also data truthfulness had great dispute. The technology of Machine Vision Online Inspection Based On Ai Is Getting Stronger. The Technology is made up of Industrial Cameras and Smart Algorithms. In modern smart manufacturing it has been widely used on complex production lines and can quickly obtain the image information of the product, taking many high clarity product photos in a very short period. And the high-speed, continues, non-contact inspection can run around-the-clock, it will speed up more than a few times. Also, getting inspection info at real time along with like MES system would allow you to do production quality trend analysis,predictive maintenance ,intelligent decisions. Nowadays it has been widely used in some complex situation like electronic assemble, car manufacture, semiconductor, pharmaceuticals, logistics, security, infrastructure inspection, also extend to the temperature, face recognition and pass attendance. Although the first cost is quite large, but in the future can greatly save labor costs, rework rate, scrap rate and other expenses.

This paper mainly looks into the developments in AI, IoT technology for the manufacturing industry. Introducing some intelligent technology applications in order to achieve better production,

DOI: 10.54254/2755-2721/2026.GF37194

so as to guide manufacturing enterprises to carry out the transformation from conventional production to intelligent digital production, so that the industry can become stronger.

2. Core intelligent technologies

2.1. Computer vision algorithms

2.1.1. Convolutional neural networks (CNN)

CNN is a very good deep model that has the ability to automatically learn features. Local connection and share weights by using the convolution kernel over the input, then it can learn some local spatio-temporal correlations. Such a network consists of convolutional layer ,activation layer,pooling layer and fully-connection layer,convolution extracts features at different levels with convolution layer,non-linearity increases by adding activation layers,also pooling decrease the size of data so it will be more translation invariant and fully connect integrates the features from front end for classification or regressing [5].

In the field of deep learning, it's a very fast time. From one side, CNN has performed excellently in fields like computer vision, self-driving and medical image analysis and become an important drive to technology. From another side,CNN begin to be researched for solving communication problems such as code design,signal detection and channel estimation, it has a bright prospect of cross field application. Currently, most of the high-performing CNN models still relies on the server side system which has very strong computing power. The CNN has solid backing on doing calculations with data, looking into pictures and making stuff.

Analysis of core algorithms involves two types: standard convolution and depth-wise separable convolution. Standard convolution is the most fundamental and computationally intensive operation in CNN, as shown in Figure 1. Its process can be described as follows: a set of convolution kernels with dimensions $k_c \times k_l \times k_l$ performs feature extraction on an input feature map with dimensions $I_c \times I_N \times I_N$. Within each window, the convolution kernel performs element-wise multiplication and accumulation (Multiply-Accumulate, MAC) with the corresponding input feature map data, ultimately generating one pixel of the output feature map. It can be seen that the core computation mode of standard convolution is high-density MAC operations within a sliding window data stream [6].

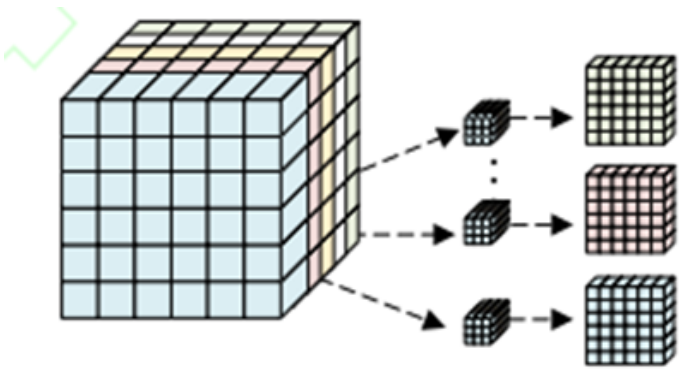

Figure 1. Schematic diagram of conventional convolution [6]

To build lightweight network models, depth-wise separable convolution was proposed to replace standard convolution. Its computation process is shown in Figure 2. It decomposes standard

DOI: 10.54254/2755-2721/2026.GF37194

convolution into two serial steps: (1) Depth-wise convolution, where each input channel is filtered independently using a single convolution kernel. This process is similar to standard convolution, but the relationship between convolution kernels and input channels is one-to-one, and its core computation remains MAC operations within a sliding window. (2) Pointwise convolution, which uses a 1×1 convolution kernel to perform channel fusion on the output of the depthwise convolution. This is computationally equivalent to a standard 1×1 convolution, and its essence is also high-density MAC operations [6].

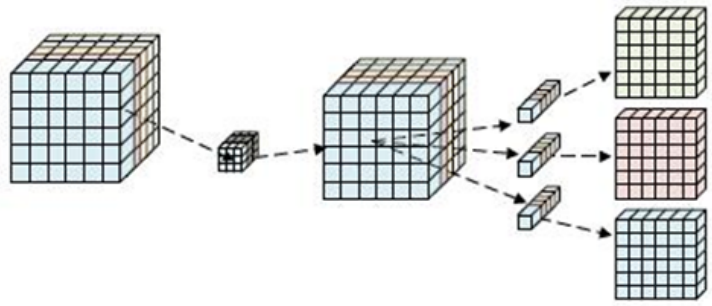

Figure 2. Schematic diagram of depth-separable convolution [6]

2.1.2. VIT network

VIT stands for Vision Transformer. It was first put forward with great success in the domain of natural language processing [7]. Figure 3 shows the whole framework about the VIT model.

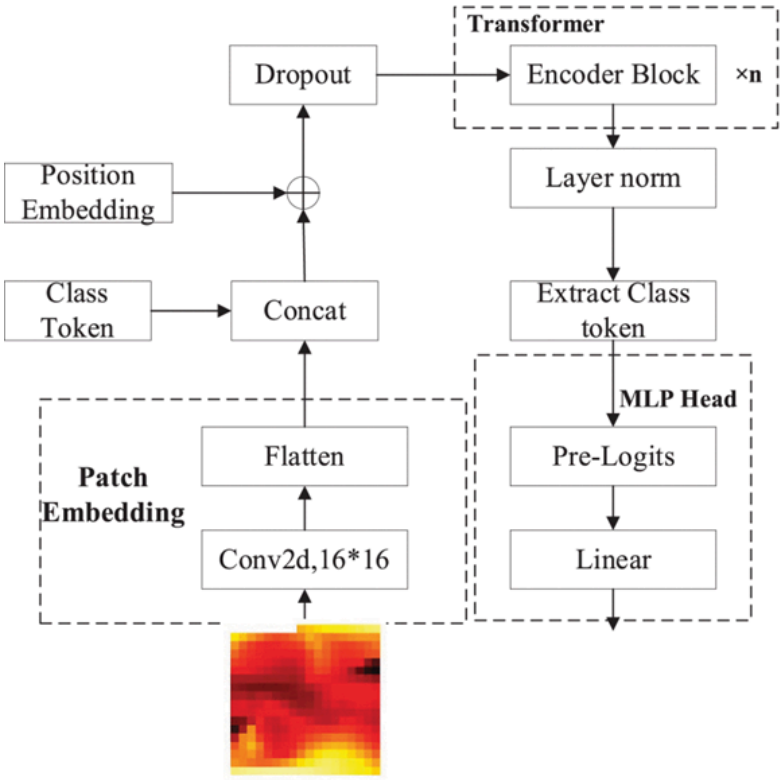

Figure 3. Overall framework of the ViT model [7]

DOI: 10.54254/2755-2721/2026.GF37194

With Attention-SEfls's help,Vision Transformer (VI) could know on the picture about long-dependency among some places. Tasks like image classification and object detection, it does much better than regular CNN when it comes to understanding the big picture [8]. In order to get more balanced result and efficiency of the model, researches put forward lots of lightweight VIT architectures. Among which hybrids combining CNN with Transformers have become main stream as they want to combine the good local feature extraction ability from CNN and the global modelling advantage from transformers [8].

The application of vision trans-former(VIT) in manufacturing has become an important technique very quickly, especially for the field like automated visual inspection and quality supervision. It is important in this field for industrial production. Based on the idea of transformer, the big plans for vision inspection have broken through from traditional CNN. When the amount of model parameters is greater than 1billion, the phenomenon of intelligence emergence will occur and it can have human-like learning capabilities. It provides three benefits, one is few shot learning that just require the sample size of 1-5 prompt samples per type of defect to finish the whole model construction process, second cross-domain generalizability, it can reuse a single foundation model for different industries instead of building intermediary industry models; Third, fast adaptation ability to respond flexibly to changes in production line parameters and processes [9].

2.2. Natural language processing algorithms

Large scale Language Models (LLMs) based on transformers can be seen as foundations of many different NLP tasks as well as Vision problems because they are scalable and model strong [10]. Transformer was first developed to do Machine Translation; LLMs are ultra large scale language models that have Transformers at the heart of them and were trained on huge data sets. To put it simply, Transformer is just like skeleton and engine of a car, and LLMs are the ultimate extension of transformer in scale and capacity. LLMs refer to NLP model which is Deep Learning based earlier trained with huge amount of info taught to find out hidden rules / relationships among different part in that particular piece of information itself and add detail around them so that transformers can grasp the long term relationship by focusing on one nearby feature at any instance rather than just one alone [11].

3. Application scenarios

3.1. Smart manufacturing scenarios for computer vision applications

The basis of smart manufacturing is the connectivity between all sorts of factories. To form a wholeness for all-round information exchange, data communication must extend on many different levels and dimensions such as device, cloud, user. In the big architecture of smart manufacturing, Data Communication System is like a bridge [12]: Nowadays, it's moving towards fully digitalized coverage, so there is plenty of digital management. But it's the algorithm bottleneck of quality testing that is still the big brake on raising production. When traditional visual recognition methods were used for tasks like inspecting metal components and electronic parts, they would be influenced by things like the light and any noise which resulted in a lot of errors [13]. In the 1960s, Roberts proposed the concept of the "blocks world" visual environment. By the 1970s, formal machine vision systems began to emerge [14]. A three-dimensional vision system developed by the Euclid Laboratory in Italy possessed wireless programming capability, demonstrating unique advantages in terms of convenience. The American company Cognex launched the DataMan 7500, a mobile

DOI: 10.54254/2755-2721/2026.GF37194

product capable of using barcodes to track machine positions in real-time, achieving 100% scanning accuracy [15]. For example, on an industrial motor production line, it is necessary to inspect whether the rotor surface has scratches or pits. Traditional methods rely on manual visual inspection, where each person can only inspect a few hundred units per day and is prone to missing defects due to fatigue. After introducing an object detection model based on PaddleDetection, the system can make judgments within milliseconds and push abnormal images in real-time to quality inspectors for review.

3.2. Smart manufacturing scenarios for large language model applications

As the architecture of industrial software systems becomes more modularized, it is getting more complicated inside. The traditional manual design -test is hard when dealing with system which are very uncoupled or processes keep changing. Conversely, with more complex systems, this paper starts running into trouble with diverse interface definitions, closely coupled behaviors, path explosions due to combinations, etc., and thus smarter/newer/better tests is what need. In the last few years, large language models have demonstrated powerful capabilities in natural language understanding and structured generation, opening up new opportunities for automating industrial testing tasks [16].

Large language models, represented by mainstream models like GPT-4 and GLM, leverage their structured understanding of natural language requirements, API interface definitions, and system documentation, combined with prompting and few-shot learning, to automatically generate test cases for multiple scenarios, establishing a new paradigm for efficient testing of industrial software systems. Convert some into an executable test case like interface, Component , Process can save a lot of manpower for testing, but it increases the range and speed of cover tests. At the level of specifics implementation in this context is large language model embedding into manufacturing system mostly take 4 forms: Industrial Copilot serves as a unified operational interface for querying and analyzing production data using natural language; knowledge base question-answering systems enable intelligent retrieval and summarization of content such as equipment manuals and process documents; code and configuration generators automatically generate scripts, recipes, and other engineering configurations; multimodal defect analysis combines with vision models to identify, classify, and reason about the root causes of defect images. These four forms collectively constitute the primary implementation paradigm for empowering smart manufacturing with large language models [16].

3.3. Internet of things technology

With the swift advance of technology, Internet of things has become a new information technology and it changes the landscape for all kinds of industry in a rapid manner. To get away from the limitations of traditional detection gadgets means delivering it at any time and looking far, smartly examining information [17]. In the field of chemical production detection instrumentation, the use of precise sensors is important in making sure data gathering is accurate and reliable. New materials sensor which depend new material sciences and physics are much better. Take MEMS sensors as an example, it would use micromachining process and make the mechanical components and electronic circuits be placed in a very small chip, so basically pretty tiny weightless thing taking much less energy: With regard topressure. MEMS Pressure S ensor takes advantageof silicons Piezoresist ive eff ect todetect even the slightest chang esin pre s su reu p tor m is; way b e fore traditiona lpress uresen so r s: It will give an exact measure of some pressure parameters during chemical producing

DOI: 10.54254/2755-2721/2026.GF37194

process giving very good supporting data for making operation and optimizing producing process [17].

4. Challenges and prospects

Although many emerging technologies have been integrated into the smart manufacturing industry so far, due to technological limitations, many industries still need to continuously accelerate transformation and upgrading. The implementation process faces four main technical challenges: Data aspects include data silos, inconsistent formats, poor quality, high annotation costs, and demanding real-time requirements; model and algorithm aspects face the contradiction between black-box characteristics and the need for interpretability, insufficient generalization capability, and risks potentially caused by large model hallucinations; system integration aspects face difficulties in OT and IT integration, challenges in retrofitting legacy equipment, and increasing pressure on network security; organization and talent aspects face a shortage of interdisciplinary talent and difficulties in adapting organizational structures to intelligent transformation.

As for improvement on future development, it is possible to do some work about technology integration, system architecture, application model and ecosystem & infrastructure of the industry production.

5. Conclusion

Smart manufacture is the main direction of manufacture transformation and upgrading, it's speeding up its transition from information to intelligence. Its essence has become a multi dimensional convergence system based on data with ai being its brain , IoT as nervous system and system architecture to carry it off. In the smart manufacturing system,Computer vision technology especially VIT which has global modeling ability and few shot learning capability, is pushing industrial visual inspection towards general large model from specific model which can address the quality inspection problem of multi variety small batch production scenario. In the manufacturing system of industrial Copilot or other form such as an industrial knowledge base Q&A, automatic code generation etc., LLMs are changing human-computer interaction model and improving production operation efficiency, system testing automation. IoT Technology offers basic help in feeling all aspects of making and controlling them at once. They set up the loop of smart make in terms of "full circle grab---send right away--smarts for making choices--carry out precisely".

The application of intelligent technology has been constantly evolving on the production line; but putting smart manufacturing theory into practical production is still full of problems, like data silos, bad models' explanation ability, hard to unite OT/IT together and no interdisciplinary personnel. As for the future, it is believed that the development of intelligent manufacturing should focus more on strengthening the foundation of industrial data; developing lightweight and intelligent models; speeding up the intelligent transformation of old equipment; cultivating cross-domain talents combining "manufacturing+AI+data"; building a cooperative industrial system open to all. And this will bring about a great leap of Smart Manufacturing from "specialized tool" to "common platform", give continuous impetus for manufacturing industry going all-round, high-quality and high-efficiency.

References

[1] 2013 Hanover European Machine Tool Exhibition focuses on "intelligent production" – Global enterprises provide solutions for manufacturing challenges. *Industrial Engineering, 16*, 136.

DOI: 10.54254/2755-2721/2026.GF37194

[2] Sun, L., & Jiao, W. (2017). Research on knowledge discovery and knowledge service construction driven by industrial big data. *Information Studies: Theory & Application, 40*, 86-89+104.
[3] Liu, Y. (2024). *Research on modular governance of manufacturing enterprise digital transformation from a paradox perspective*. Jilin University.
[4] Wang, N., Lin, C., Wang, X., et al. (2026). Status and progress of monitoring technology for median diameter of suspended particles in reservoir water injection. *Tianjin Science & Technology, 53*, 15-17+22.
[5] Liu, K., Wang, H., Han, R., et al. (2025). Fault diagnosis method for metro bogie based on improved graph convolutional network. *Journal of East China Jiaotong University, 1*, 1-12.
[6] Sun, K., & Chen, Y. (2025). Design and implementation of vector processor for edge-end CNN acceleration. *Journal of Beijing University of Aeronautics and Astronautics, 1*, 1-14.
[7] Sun, H., Lu, D., Cai, B., & Lan, T. (2022). Railway environment scene recognition for GNSS positioning based on deep learning and ViT model. *International Conference on Electromagnetics in Advanced Applications (ICEAA)*, 167–171.
[8] Li, Z., Wang, X., Ma, J., et al. (2026). Lightweight ViT integrating orientation and frequency domain awareness. *Computer Engineering*, 1–14.
[9] Large vision inspection model based on transformer – reshaping the billion-level industrial vision market. *China Machine Vision Network*.
[10] Elayat, O., Gaudet, V., & Elmasry, M. (2025). Efficient pruning and acceleration of encoder-based LLM transformers on eFPGAs. *International Conference on Microelectronics (ICM)*, 1–6.
[11] Jiang, N., Fan, C., Zhang, R., et al. (2023). Model-based zero-trust network security architecture. *Journal of Chinese Computer Systems, 44*, 1819–1826.
[12] Shen, H., Cui, L., & Zhang, T. (2022). Research on 5G network deployment scheme for smart manufacturing. *Telecommunications Science, 38*, 158–163.
[13] Zhao, X. (2026). Research on improvement of computer vision recognition algorithms in smart manufacturing quality inspection for digital management. *Information Recording Materials, 27*, 191-193.
[14] Ouyang, Z., & Xiao, X. (2018). Application of machine vision in smart manufacturing. *Big Data Time, 3*, 4.
[15] Han, L., Ma, F., & Shi, Y. (2025). Application and technical analysis of machine vision in smart manufacturing. *Die & Mould Manufacture, 25*, 16-18.
[16] Luo, J. (2025). *Method for generating and verifying test cases for industrial software components based on large language models*. Guangdong University of Technology.
[17] Zhang, D. (2026). Mechanism of IoT technology improving performance of chemical production detection instruments. *China Instrumentation, 3*, 56-60.