

Journal of Electronic Research and Application

Editor-in-Chief

Joselito A. Dolot

Lyceum of the Philippines University-Batangas, Philippines

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

(619 649 400)

Level 10

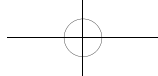
50 Clarence Street

SYDNEY NSW 2000

Copyright © 2026. Bio-Byword Scientific Publishing Pty Ltd.

Complimentary Copy





ISSN (ONLINE): 2208-3510

ISSN (PRINT): 2208-3502



Journal of Electronic Research and Application

Focus and Scope

Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication

About Publisher

Bio-Byword Scientific Publishing is a fast-growing, peer-reviewed and open access journal publisher, which is located in Sydney, Australia. As a dependable and credible corporation, it promotes and serves a broad range of subject areas for the benefit of humanity. By informing and educating a global community of scholars, practitioners, researchers and students, it endeavors to be the world's leading independent academic and professional publisher. To realize it, it keeps creative and innovative to meet the range of the authors' needs and publish the best of their work.

By cooperating with University of Sydney, University of New South Wales and other world-famous universities, Bio-Byword Scientific Publishing has established a huge publishing system based on hundreds of academic programs, and with a variety of journals in the subjects of medicine, construction, education and electronics.

Publisher Headquarter

BIO-BYWORD SCIENTIFIC PUBLISHING PTY LTD

Level 10

50 Clarence Street

Sydney NSW 2000

Website: www.bbwpublisher.com

Email: info@bbwpublisher.com

Table of Contents

- 1 Design and Simulation of a Miniaturized Printed Wideband Antenna with a Gradual Radiator and Defected Ground**
Qi Wang
- 7 Structural and Control System Design of an Automatic Cooking Robot**
Juanjuan Huang, Zhe Huang, Haitao Sang, Qiwen Liang, Runqiang Lu*
- 14 Architectural Design of RT-DETR-L for PCB Surface Defect Detection: A Systematic Comparison of Attention Mechanisms, Backbone Replacement, and Cross-Dataset Generalization**
Tao Huang
- 21 Finite-Element Modeling and Analytical Validation of a Coaxial Cylindrical Capacitor Using Maxwell 2D**
Qi Wang
- 27 Analysis of the Application of AI Visual Inspection Technology in Precision Mechanical Manufacturing Automation**
Tianwei Zheng
- 33 Research on the Controllable Growth Mechanism and Structural Characterization Methods of Single-Walled Carbon Nanotubes**
Jingchun Zhang, Wenqing Shen, Zhikun Chen
- 40 Research on Position-Sensorless Control of PMSM Based on ELO and NLESO-PLL**
Wei Liu, Jiawei Cao
- 53 Research on Multi-feature Fusion Method for Software Defect Prediction Based on Industrial Internet of Things**
Jinpeng Xu

- 59 Research on the Key Role of Data Assets in Operation and Maintenance Retrospective Analysis Under Rail Transit Network Security Control**
Jianyong Huang
- 67 YOLO11-SPMB: Application of an Improved Instance Segmentation Network for Tire Defect Detection**
Haixu You, Jun Xing
- 76 Structure- and Frequency-aware Domain Adaptation for SAR Raft Aquaculture Extraction**
Yijie Wu, Shan Jiang
- 84 Research on Development and Performance of Intelligent Protective Mask Based on Flexible Heating and Piezoresistive Sensing Technology**
Ziyi Dang, Miao Wang, Xiangyi Hou, Mingze Sun, Yin He
- 90 Research on Real-Time Intelligent Rendering Method for Immersive Interactive Virtual Reality Scenes**
Wenzhen Wang
- 97 Exploration of Digital Twin-Driven Heterogeneous Terminal Linkage and Intelligent Control Technology in the Internet of Things**
Tao Zhang
- 104 A Vessel Segmentation Method Based on Prior-Guided False Positive Suppression Gating**
Jintao Zhang
- 118 Design and Simulation of Wireless Charging Circuit for Electric Vehicles**
Zijian Lu
- 125 Strategic Analysis of Dynamic Detection and Tracking Methods for Autonomous Vehicles**
Chenhe Kang
- 131 Distributed Flexible Job Shop Scheduling Based on an Improved Genetic Algorithm**
Wenwen Chen, Bowen Du
- 137 Virtual Simulation Modeling and Validation of Traffic Measurement Equipment Based on Multi-Physical Field Coupling**
Lin Zhou, Jing Wang, Nan Wang, Chi Zhang, Zhongtian Jin, Lucheng Wang
- 143 Structural Design and Optimization of Tesla Valve Based on Bezier Curves**
Mingtao Gong

154 CVaR-Based Optimal Scheduling of an Integrated Energy System with Battery Storage, Hydrogen Storage, and Demand Response

Delong Yan

165 A Brief Analysis of the Development of the Wire and Cable Industry and Quality Test Technology in Shaanxi Province in the Age of Artificial Intelligence

Beini An

Design and Simulation of a Miniaturized Printed Wideband Antenna with a Gradual Radiator and Defected Ground

Qi Wang

School of Electronic and Information Engineering, Liaoning Technical University (Huludao Campus), Huludao 125105, Liaoning Province, China

*Corresponding author: Qi Wang, wangqi545@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To improve the integration capability and impedance matching of printed antennas for wideband wireless terminals, a miniaturized planar antenna with a gradual radiator and a defected partial ground is designed and verified by full-wave simulation. The antenna is implemented on an FR4 substrate of 31 mm × 18 mm × 1 mm and fed by a 50 Ohm microstrip line. A gradual, chamfered radiator with a top arc and T-shaped loading is combined with a partial ground containing a central slot and tapered edge, so that the forward and return current paths can be tuned simultaneously. The Driven Modal solver in Ansys HFSS is used to evaluate the reflection coefficient, field, and current distributions, and far-field radiation. The simulated S11 curve exhibits resonances near 3.3 GHz, 4.4–4.6 GHz, 7.7–8.0 GHz, and 13.0 GHz; the deepest resonance reaches about -35 dB, and the typical peak GainTotal is about 3.2 dBi. A parameter sweep of La from 0.6 mm to 0.8 mm shows little influence on the main resonances. The results indicate that the proposed structure provides a compact and feasible basis for multi-band or wideband terminal antennas.

Keywords: Printed antenna; Ultra-wideband antenna; Defected ground structure; HFSS; Impedance matching

Online publication: Jul 7, 2026

1. Introduction

Antennas are key interfaces in wireless communication and electromagnetic sensing systems because they convert guided waves in circuits into radiated waves in free space. Their impedance matching, bandwidth, and radiation characteristics directly influence link budget, coverage, and system integration. With the continuous miniaturization of terminals, a compact antenna must maintain acceptable matching and stable radiation within a limited printed-circuit-board area^[1,2].

Printed planar antennas, including microstrip-fed monopole and slot-loaded forms, are widely adopted because of their low profile, low cost, and compatibility with standard PCB processes^[3]. However, their

resonant behavior is highly sensitive to the geometry of the radiator, feeding transition, ground plane, and slot defects [4]. Small changes in these structures may alter the surface-current path, shift resonant points, and introduce multi-mode coupling [5]. Therefore, full-wave modeling and targeted parameter optimization are essential in practical design [6].

Ultra-wideband (UWB) applications require an antenna to maintain good impedance matching over a broad spectral range while remaining tolerant of narrowband interference [7,8]. Reported solutions usually employ gradual radiators, loaded branches, slots, and defective ground structures (DGS) to form multiple resonances or controllable notched bands [9]. Based on this design concept, this paper presents a miniaturized printed wideband antenna and analyzes its matching and radiation mechanisms by HFSS simulation [10].

2. Antenna structure and simulation method

2.1. Antenna configuration

The proposed antenna adopts a double-sided printed structure on an FR4 substrate. The overall size is 31 mm x 18 mm x 1 mm, which is suitable for compact terminal integration. The top layer contains a gradual and chamfered radiator connected to a 50 Ohm microstrip feed. A top arc and T-shaped loading are introduced to extend the effective current path and to generate additional resonant modes. The bottom layer uses a partial ground plane with a central slot and tapered upper edge, which modifies the return-current path and equivalent capacitance/inductance, thereby improving impedance transition over a wider frequency range. See **Figure 1** and **Table 1**.

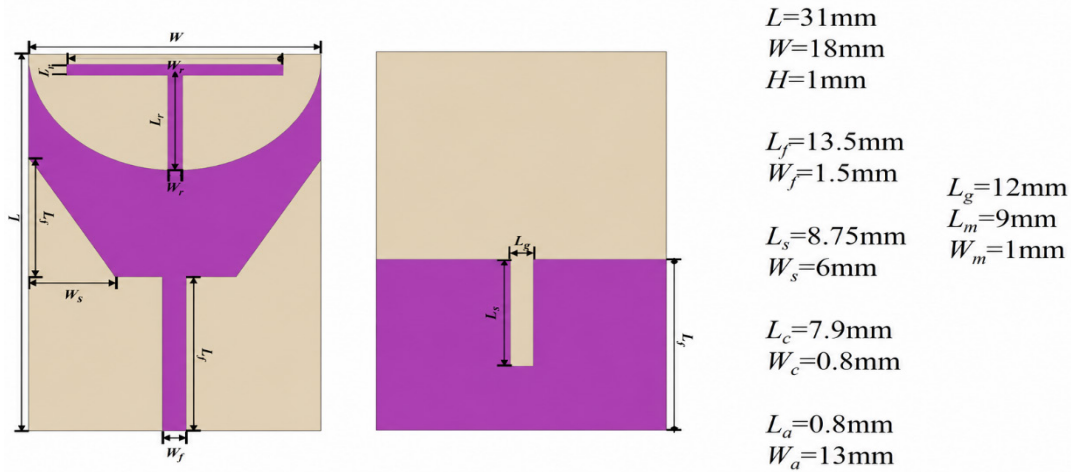


Figure 1. Geometry and main dimensions of the miniaturized printed antenna.

Table 1. Main design parameters and simulated performance indicators

Category	Parameter or output	Value / description
Substrate	Material and size	FR4, 31 mm × 18 mm × 1 mm
Feeding	Reference impedance	50 Ohm microstrip feed
Radiator	Geometry tuning	Gradual/chamfered radiator with top arc and T-shaped loading

Ground	DGS treatment	Partial ground with tapered edge and central slot
Solver	Simulation platform	Ansys HFSS, Driven Modal frequency-domain analysis
Matching	Main resonances	3.3, 4.4–4.6, 7.7–8.0, and 13.0 GHz
Radiation	Typical peak GainTotal	About 3.2 dBi at the selected operating frequency
Sensitivity	La sweep	0.6–0.8 mm with 0.05 mm step; main response nearly unchanged

2.2. Simulation setup

The model is built in Ansys HFSS with parameterized substrate, radiator, feed, and ground features. The Driven Modal solution is employed for frequency-domain analysis, and a radiation air box is assigned around the antenna. The reflection coefficient S11 is used as the primary matching index, and $S_{11} \leq -10$ dB is adopted as the engineering criterion for a usable operating band. The initial electrical size is estimated from a quarter-wavelength current-path assumption, while final resonance positions are determined by full-wave simulation because the actual current flows along the radiator edge, loading branch, and ground-return path.

3. Simulation results and discussion

3.1. Impedance matching characteristics

The simulated S11 response shows a clear multi-resonant behavior, as shown in Figure 2. The deepest resonance occurs near 3.3 GHz with a return loss of about -35 dB, indicating excellent impedance matching around the lowest operating mode. Additional resonant dips appear near 4.4–4.6 GHz, 7.7–8.0 GHz, and 13.0 GHz, with approximate return-loss levels of -25 dB, -20 dB, and -15 dB, respectively. These dips confirm that the combined radiator loading and defective ground introduce several effective current paths rather than a single narrow-band resonance. See **Figure 2**.

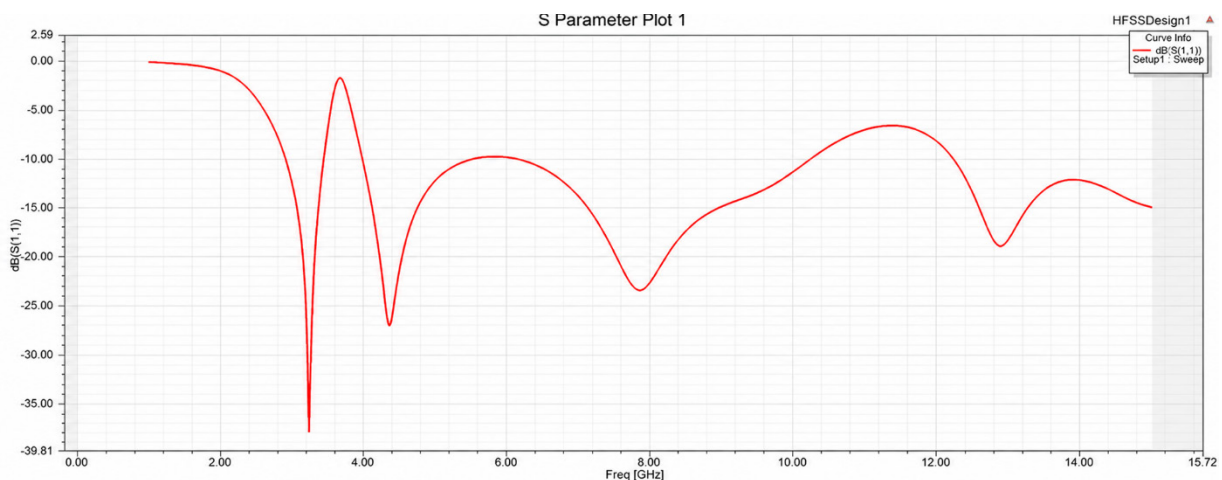


Figure 2. Simulated S11 curve of the proposed antenna.

A further sweep of parameter La from 0.6 mm to 0.8 mm with a step of 0.05 mm indicates that the main resonant positions and depths are almost unchanged. This result suggests that La is not the dominant tuning parameter in the present configuration. In contrast, the low-frequency resonance is expected to be more strongly associated with the total path length from the feed to the radiator edge and the ground return,

whereas higher-frequency modes are more sensitive to the local top-loading, slot, and DGS features.

3.2. Radiation behavior and mechanism

Far-field post-processing yields a typical peak GainTotal of about 3.2 dBi, and the two-dimensional E-plane/H-plane patterns show that the antenna exhibits identifiable main radiation regions. Although the design is not intended for high-gain directional operation, the positive gain indicates that the matched input power can be effectively radiated into free space. Surface-current and electric-field distributions further show that strong current appears along the microstrip feed, the feed-radiator transition, and the upper loading region. This concentration is consistent with the edge-radiation mechanism of printed planar antennas.

The multi-resonant response can therefore be interpreted from the coexistence of several current paths. The low-frequency mode is mainly governed by the global path consisting of feed, radiator, and ground return, while the middle and high-frequency modes are produced by local loading and slot perturbations. The present structure provides a compact platform for multi-band operation; however, if a continuous UWB passband or explicit notched-band suppression is required, the radiator loading, ground-slot length, and slot position should be optimized more systematically.

3.3. Mode attribution and parameter sensitivity

To further clarify the electromagnetic mechanism, the simulated resonances can be associated with different equivalent current paths rather than treated only as isolated numerical dips. The resonance near 3.3 GHz is primarily determined by the global path from the microstrip feed to the gradual radiator edge and then to the ground return. This mode is responsible for the lowest usable band and is therefore strongly related to the overall electrical length of the printed structure. The 4.4–4.6 GHz resonance is influenced by the feed-radiator transition and the upper arc, which introduce a smoother impedance transition and additional capacitive coupling. The 7.7–8.0 GHz and 13.0 GHz responses are more closely associated with local perturbations caused by the T-shaped loading, central ground slot and tapered ground edge.

The parameter sweep of L_a from 0.6 mm to 0.8 mm, with a step of 0.05 mm, shows that the major resonant positions and return-loss depths remain nearly unchanged. This indicates that L_a is not the primary control variable in the present geometry and that the antenna response has a certain tolerance to small dimensional variations of this parameter. From an engineering perspective, such robustness is useful in PCB fabrication because small etching deviations or alignment errors are unlikely to disrupt the main matching behavior. By contrast, parameters directly related to the total current path, such as the length of the top loading strip, the gap between the radiator and the partial ground, and the length and width of the ground slot, are expected to have a stronger influence on impedance matching and should be emphasized in subsequent optimization. See **Table 2**.

Table 2. Physical interpretation of the simulated resonant modes

Mode	Frequency range	Dominant structural factor	Design implication
M1	about 3.3 GHz	Global feed-radiator-ground current path	Sets the lowest operating mode and compact-size limit
M2	4.4–4.6 GHz	Feed transition and top arc coupling	Improves local matching after the first resonance
M3	7.7–8.0 GHz	T-shaped loading and slot perturbation	Provides an additional middle/high-frequency matching point

3.4. Optimization strategy and engineering applicability

Although the proposed antenna achieves compact multi-resonant matching, the S11 response still does not form a fully continuous UWB passband. Adjacent resonances can be merged by jointly tuning the radiator contour, ground-slot length and position, microstrip transition width, and feed-region coupling. For practical implementation, FR4 loss, copper thickness, SMA connector, soldering pad, and finite-ground effects may reduce gain or shift the input impedance; therefore, prototype fabrication and vector-network-analyzer measurements are necessary for final validation. Overall, the workflow of current-path estimation, HFSS simulation, and field/current analysis is repeatable and can support future bandwidth extension, notch-band design, and radiation-stability optimization.

4. Conclusion

A compact printed wideband antenna based on a gradual radiator and a defected partial ground has been designed and simulated. The antenna occupies only $31 \text{ mm} \times 18 \text{ mm} \times 1 \text{ mm}$ on an FR4 substrate and is excited by a 50 Ohm microstrip feed. HFSS results show resonant points near 3.3 GHz, 4.4–4.6 GHz, 7.7–8.0 GHz, and 13.0 GHz, with the best simulated return loss reaching about -35 dB. The typical peak GainTotal is about 3.2 dBi, and the current distribution verifies that the feed transition, radiator edge, and top loading form the main radiation and tuning regions. The study demonstrates that simultaneous control of the forward current path and ground-return path is effective for compact multi-band matching. Future work should focus on merging adjacent resonances to obtain a more continuous UWB passband and on introducing controllable notch structures for interference suppression.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Xiang L, Wei K, Zhang X, et al., 2025, A Miniaturized and Wideband Millimeter-Wave $\pm 45^\circ$ Differentially Fed Dual-Polarized Arrowhead-Shaped Patch Antenna Array. *IEEE Antennas and Wireless Propagation Letters*, 24(12): 4760–4764.
- [2] Wang B, Zhou Z, Deng C, et al., 2025, A Compact D-Band Multilayer Horn Antenna with Thin 3-D-Printed Lens for Subterahertz Applications. *IEEE Antennas and Wireless Propagation Letters*, 24(9): 2919–2923.
- [3] Zhou X, Zhang G, Tam K, et al., 2026, Folded-Patch-Based Miniaturized Quasi-Omnidirectional Filtering Antennas. *IEEE Transactions on Antennas and Propagation*, 74(1): 298–308.
- [4] Kumar P, Rao P, Saxena S, 2025, Design and Analysis of 0.5–2.0 GHz RF Frontend Subsystems for IEMI Applications. *IEEE Transactions on Electromagnetic Compatibility*, 67(1): 237–246.
- [5] Mazzinghi A, Barras D, Danev B, et al., 2024, Miniaturized UWB Dual-Port Antenna for Localization Applications. *IEEE Antennas and Wireless Propagation Letters*, 23(3): 1080–1084.

- [6] Salimitorkamani M, Mehranpour M, Odabasi H, 2024, A Miniaturized Wideband Sinuous Antenna for Microwave Brain Imaging Systems. *IEEE Transactions on Antennas and Propagation*, 72(3): 2228–2240.
- [7] Shukoor M, Dey S, 2024, Wideband Reconfigurable Multifunctional Absorber/Reflector with Bandpass/Bandstop Filtering and Band-Notch Absorption for RCS and EMI Shielding. *IEEE Transactions on Electromagnetic Compatibility*, 66(1): 153–160.
- [8] Askarzadeh R, Farahbakhsh A, Zarifi D, et al., 2025, Wideband High-Efficiency Slot Array Antenna Based on Gap Waveguide Single-Layer Feeding Network. *IEEE Antennas and Wireless Propagation Letters*, 24(2): 519–523.
- [9] Li D, Deng C, Fu Z, et al., 2025, Wideband Electrically Small Vivaldi Antenna Using Non-Foster Matching. *IEEE Antennas and Wireless Propagation Letters*, 24(8): 2672–2676.
- [10] Cheng S, Chen S, Huang W, 2025, Low-Profile MIMO Trapezoidal Patch Antenna for 5G Wideband Mobile Antenna Application. *IEEE Antennas and Wireless Propagation Letters*, 24(3): 696–700.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Structural and Control System Design of an Automatic Cooking Robot

Juanjuan Huang¹, Zhe Huang², Haitao Sang^{1*}, Qiwen Liang¹, Runqiang Lu³

¹School of Electronics and Electrical Engineering Lingnan Normal University Zhanjiang, 524048 China

²School of Computer Science and Technology Harbin University of Science and Technology, 150080, China

³Guangdong HUACO Electrical Appliance Group Co., Ltd., Zhanjiang, 524400 China

**Corresponding author:* Haitao Sang, sanght@lingnan.edu.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the rapid development of electronics, household kitchen appliances are gradually becoming intelligent. To improve the wellbeing of office workers and address challenges such as limited time for cooking rice, this paper proposes the structure and control system of a cooking robot. The system uses an STM32 microcontroller as its core and integrates IoT and other new technologies to enable interaction among the smartphone, cloud, and the cooking robot. It supports remote control and monitoring via a mobile app, and ondevice monitoring of the operating status or the cooking curve through a touch screen. Experimental tests show that the robot can automatically complete a series of operations, including rice metering, rice washing, and rice cooking, making the system more userfriendly and intelligent.

Keywords: Cooking robot; Internet of things (IoT); STM32; Control system; Cloud platform

Online publication: Jul 7, 2026

1. Introduction

In recent years, within the smallappliance manufacturing industry, rice cookers have enjoyed a strong sales advantage. To expand the market and meet diverse needs, ricecooker manufacturers have continuously innovated and upgraded, resulting in significant improvements in performance and operability. Temperature control during cooking has evolved from the previously used magnetic temperature limiter to microcomputer control, and the heating method has expanded from traditional resistive heating to electromagnetic heating^[1]. Market surveys indicate that commonly used highend rice cookers adopt IH (induction heating), enabling functions such as rice cooking, soup making, and steaming; however, there is still no household cooking robot on the market that integrates automatic rice dispensing, automatic rice washing, automatic rice cooking, and remote operation. To further raise the level of intelligence in rice cookers and move China's brands closer to the ranks of worldclass brands, this paper proposes a controlsystem design for a cooking robot that automatically performs rice fetching, rice washing, rice dispensing, water addition, and cooking, making rice preparation

convenient, fast, and within easy reach, thereby improving quality of life ^[2].

2. Overall structural design of the cooking robot

2.1. Rice storage & metering-dispensing mechanism and working principle

The rice storage and metering unit consists of: rice storage hopper (1), remaining quantity detection sensor (2), rice inlet of the metering device (3), spherical rice metering cup (4), mounting bushing for the metering device (5), mounting bracket (6), rice outlet (7), coupling (8), stepper motor (9), and rotating shaft of the spherical metering device (10), as shown in **Figure 1**.

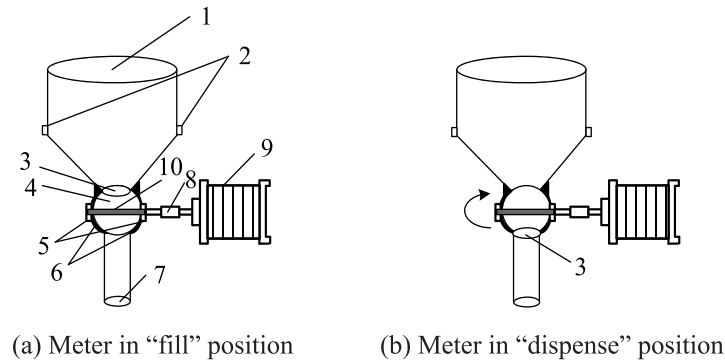


Figure 1. Structure of the rice storage and metering unit.

The storage hopper adopts a funnel shape to realize first in first out (FIFO), solving issues such as stale rice accumulation, moisture, and insect infestation in traditional bins. A spherical metering cup (4) is installed at the lower end of the hopper (1). The cup is spherical with a rice inlet (3) at the top. According to the requirement, the spherical cup (4) is designed with a fixed volume—for example, the maximum amount per fill is 3 liang (150 g) per cup. A rotating shaft (10), aligned horizontally with the inlet (3), is fixed to the spherical cup (4). Supported by the mounting bushing (5) and bracket (6), the entire metering device is mounted beneath the hopper (1).

During metering, the inlet faces vertically upward so rice flows smoothly from the hopper into the spherical cup (**Figure 1a**). When dispensing is required, the stepper motor drives the cup, via the coupling, to rotate clockwise by 180° (as indicated in **Figure 1b**), turning the inlet vertically downward (**Figure 1b**) so that one cup (3 liang) of rice exits through the outlet into the washing bucket. After dispensing, the motor rotates the cup another 180° clockwise to return to the fill position. If more rice is needed, the action repeats. A rice level sensor is installed at one third of the hopper capacity. Comprising a photoemitter and photoreceiver, it detects "sufficient" when light is blocked by rice above the 1/3 level, and when the level falls below 1/3, the photoreceiver senses light from the emitter; the MCU then issues a refill reminder. The hopper level can also be monitored via a mobile app to implement low level alerts.

2.2. Ricewashing and water dosing mechanism & working principle

In the ricewashing and water dosing unit, a three-dimensional hydrojet washing method is used to simulate agitator stirring, removing debris from the rice while better preserving nutrients. This 3D hydrojet approach eliminates the need for a mechanical agitator and motor, reducing the washer's size and simplifying both

mechanics and control. Coordinated with the sprinkler rings to actuate the pushdown rice valve, which prevents rice adhesion and improves reliability. The overall structure is shown in **Figure 2** ^[3].

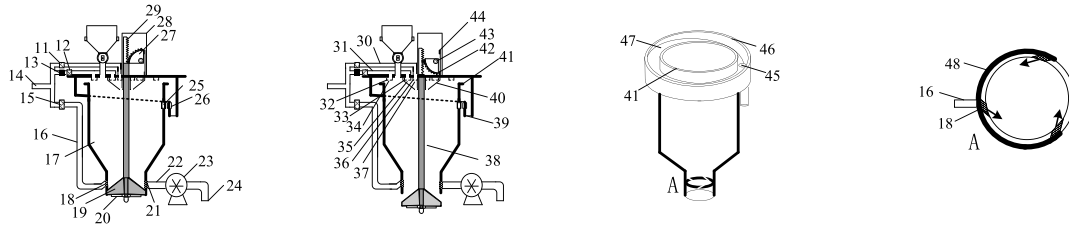


Figure 2. Overall structure of the washing unit. (a) Rice valve closed (b) Rice valve open (c) Structure of the washing bucket (d) Structure at location A.

The washing unit mainly comprises: solenoid valve for floatingdebris removal (11), watervolume solenoid valve (12), flow sensor (13), water inlet port (14), washing solenoid valve (15), washing water inlet pipe (16), washing bucket (17), filter screen (18), ricedischarging valve (19), valve fixing gasket (20), drain filter mesh (21), drain pipe (22), selfpriming drain pump (23), drain outlet (24), bucket waterlevel sensor (25), overflow detection sensor (26), lift cam (27), protective cover (28), mainshaft lift control rack (29), inlet pipe for the first spray ring (30), inlet pipe for the second and third spray rings (31), third spray ring (32), second spray ring (33), spray holes of the third ring (34), spray holes of the second ring (35), first spray ring (36), spray holes of the first ring (37), pushpull rod (38), lower drain outlet of the bucket (39), water collector (40), bucket overflow port (41), “downpush inplace” switch for the rice valve (42), controlmotor main shaft (43), “uppull inplace” switch for the rice valve (44), upper drain outlet of the bucket (45), outer waterblocking ring of the drain channel (46), annular drain channel (47), and annular washer (48). The height of the bucket’s overflow port is lower than that of the outer waterblocking ring.

When a washing command is received, the unit operates as follows:

2.2.1. Automatic water addition and rice loading

Upon receiving the command, a geared motor rotates the lift cam clockwise to drive the pushpull rod upward; as it rises, the rice valve is pulled up. When the cam edge triggers the “uppull inplace” switch, lifting stops, and the rice outlet is closed. The MCU energizes the watervolume relay to open the watervolume solenoid; water passes the flow sensor and is sprayed into the bucket via the second and third spray rings. When the preset volume is reached, the solenoid is deenergized to stop inflow. To prevent adhesion to the bucket wall, water is added before rice so the grains enter water and are wetted and loosened. After the first water addition, the storage/dispensing unit adds rice according to the set quantity.

2.2.2. Automatic washing and removal of floaters

After rice loading, the MCU energizes the washing solenoid; water entering the annular washer moves upward in a helical direction, washing the rice in 3D. When the level reaches the bucketlevel sensor, spraying stops to avoid overflow. Rice settles while floaters remain on the surface. The MCU then opens the floaterremoval solenoid; water from the first spray ring is sprayed downward 360° around the pushpull rod, raising the water level. When it exceeds the overflow port, the flow carries floaters into the annular drain channel. Because this channel is circumferentially sloped and equipped with an overflow sensor at the

discharge end, once overflow is detected, the MCU delays around 5 s, then closes the solenoid and stops inflow. Next, the drain pump empties the bucket, completing the first wash. The second wash repeats the spraying until the level sensor is reached, then drains again. The unit then proceeds to water dosing and rice discharge^[4].

2.2.3. Automatic water dosing and rice discharge

The flow sensor, containing a rotor/impeller and a Hall sensor, is in series with the inlet solenoid to measure throughflow. With the inlet energized (valve open), the impeller's magnet rotates at a speed proportional to flow; the Hall sensor outputs pulses counted by the control circuit to estimate volume^[5]. This method achieves high dosing accuracy, especially for small portions (e.g., 150 g rice with 250 mL water).

Typically, the ricetowater ratio is 1:1.2–1.4. Experiments show that, for small volumes, adhesion at the rice piston/outlet can occur. To avoid sticking and ensure smooth discharge into the cooker's inner pot, the MCU first doses twothirds of the required water to prewet the rice surface, then rapidly drives the rice valve downward while spraying the remaining onethird via the second and third spray rings: 360° onto the bucket's inner wall and downward along the pushpull rod onto the top of the rice valve. This disperses the grains evenly into the inner pot without needing manual leveling and prevents adhesion.

2.2.4. Automatic selfcleaning

The MCU raises the pushpull rod to close the rice valve and enter cleaning. Water from the three bottom inlets rises in a swirling motion to flush the bucket, rice valve, and rod. When the level exceeds the overflow port and flows to the floater discharge, the overflow sensor signals the MCU, which delays around 5 s, then stops inflow and runs the drain pump to empty the bucket, completing cleaning. If the unit will be idle after cooking, the MCU lowers the pushpull rod to standby and opens the rice valve to ventilate and prevent odors.

2.3. Cooking mechanism

A cooking mechanism is arranged beneath the washing and dosing unit. After washing and dosing are completed, the riceandwater mixture is discharged into the robot's inner pot. The control circuit then starts the cooking program, putting the robot into the cooking state.

3. Control system design of the cooking robot

The cooking robot provides the following functions: automatic rice fetching, automatic water addition, automatic rice washing, automatic water dosing, automatic cooking, and remote monitoring/control via a mobile app. A block diagram of the control system is shown in **Figure 3**. The hardware consists of an STM32 core module, WiFi circuitry, ricequantity metering and control circuits, waterquantity metering and control circuits, drain control circuitry, washing control circuitry, cooking control circuitry, temperaturesensing circuitry, and a touchscreen display/input module^[6]. The core board adopts the STM32F103C8T6, featuring 64 K Flash, 20 K RAM, up to 72 MHz clock, 32 generalpurpose I/Os, and a 12bit synchronous ADC (two groups, ten channels). It carries an onboard 24C04 EEPROM and supports oneclick downloading via USBtserial, meeting the system requirements^[7]. During operation, the MCU receives commands for rice quantity and cooking mode via the touch screen or WiFi module; the LCDs display the rice quantity, cooking mode, and process information. The flow sensors, watervolume relay, and drain relay in the washing/

discharging/drainage mechanisms control washing water and drainage. After washing, the ricedischarging mechanism meters and sends rice into the cooker's inner pot. Under the control of the cooking unit—using lid and bottom temperature sensors and the heating main control circuit—the robot performs automatic cooking.

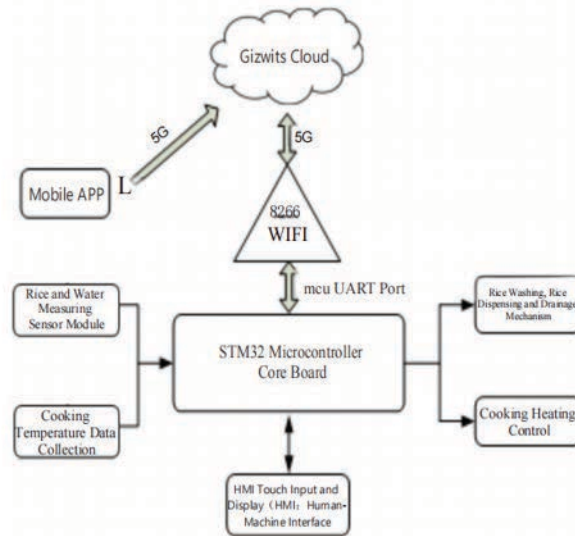


Figure 3. Block diagram of the cooking robot control system.

3.1. Mobile-app remote control and monitoring

For appbased control, the system uses an ESP8266 WiFi module in conjunction with the OneNET cloud platform^[8]. Once the device is registered on OneNET, the robot connects via the WiFi module using the EDP protocol specified by OneNET: it composes a data packet and sends a connection request to the cloud. Upon success, the platform shows the device as online, and the mobile client can query the device status by calling methods from the OneNET API with the corresponding device ID. Since the robot is programmed to send a connectionrequest packet at powerup, the app indicates online whenever the device is powered; if power is lost, it shows offline. Only when the device is online are the app's full functions enabled^[9].

During cooking, the app can monitor parameters in real time: lid (top) temperature, bottom temperature, relay actuation states, and the stages of the cooking process, such as preheating, water absorption, heating, boiling, simmering, and keepwarm.

3.2. Touch-screen HMI functions

For human–machine interaction, a serial HMI touch screen is used^[10,11]. The HMI display module communicates with the STM32 MCU via UART. The display interface is created with PCside configuration software, while the MCU only needs to exchange commands over the serial link to display the robot's operating status in real time and to handle ricedischarge/cooking control, status display, and alerts. The touchscreen interface supports: ricequantity display, ricequantity categories, cookingprocess display, time display, and parameter adjustment.

4. Workflow of the cooking robot

When cooking is needed, the user can control the system remotely via the mobile app or operate it through the robot's touch screen. After setting the rice quantity and issuing the command, the robot starts operating according to the workflow shown in **Figure 4**.

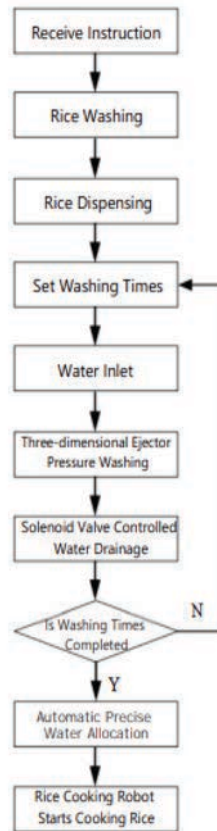


Figure 4. System workflow diagram.

5. System testing

After system debugging, tests were conducted on rice quantity, waterdosing volume, and ricewashing time. Method: start the washing process; use a stopwatch to record the time from the start to completion (as described); open the lid and pour the water from the inner pot into a measuring cup; record the values. The results are shown in **Table 1**.

Table 1. Tests of rice mass, water volume, and washing time

Rice mass (g)	Measured water (mL)	Designed water (mL)	Washing time (min)
150	245	250	3
300	415	400	6
450	525	550	9
600	715	700	12
750	850	850	15

6. Conclusion

Future household kitchen appliances should evolve toward greater intelligence, ease of use, style, and energy efficiency. Given its smart, energysaving, and ecofriendly characteristics, the ricecooking robot is destined to become a promising direction. Based on the design and experiments, this paper proposes a structural and controlsystem scheme that implements water storage, rice storage, automatic rice discharge, automatic rice washing, automatic water addition, automatic water dosing, and automatic cooking. Leveraging IoT technology enables interaction among the smartphone, the cloud, and the robot. With a userfriendly interface, it provides heating curves and selectable taste/texture options, and also allows remote cooking control and parameter tuning via the app, as well as realtime monitoring of bottom/lid temperatures and cooking time during the cooking stages, thereby enhancing the robot's intelligence.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Peng D, 2019, Design and Implementation of the Main Functional Control Software of an AH Rice Cooker Based on Microwave Heating Technology. thesis, University of Electronic Science and Technology of China: 1–5.
- [2] Zeng G, Kang J, 2025, Noise Analysis and Noise Reduction Research of the IH Rice Cooker. *Home Appliance Technology*, 2025(02): 36–41.
- [3] Liang Q, Long S, Chen B, et al., 2022, A Multifunctional Rice Washing Device: CN113384177B.
- [4] Wang H, Chen B, Long S, et al., 2022, A Fully Automatic Rice Washing Method and System: CN113384176B.
- [5] Huang C, Luo Y, Huang Y, 2024, Design and Implementation of a Remote-Control System for an Intelligent Rice Cooker. *Mechanical Engineer*, 2024(11): 1–5 + 13.
- [6] Fu W, Wang H, 2022, Design and Implementation of a Temperature and Humidity Acquisition System Based on STM32 Microcontroller and DHT11 Sensor. *Light Sources and Lighting*, 2022(03): 119–121.
- [7] Liu L, Wang J, Zhao G, 2020, Fundamentals and Practice of Embedded Systems. *Electronics Industry Press*: 22–49.
- [8] Liu F, Shen S, Deng S, 2022, Design and Implementation of a Data Acquisition and Monitoring System Based on OneNET. *Electronic Production*, 30(17): 47–49.
- [9] Liang Q, Long S, Deng J, et al., 2019, Design of an IoT Rice Cooker App. *Internet of Things Technologies*, 9(01): 52–55.
- [10] Bai L, Zhu J, Lei H, et al., 2021, Application of HMI Serial Touch Screens in Microcontroller System Development. *Journal of Henan Institute of Technology*, 29(02): 64–66.
- [11] He J, Zhong L, Huang C, 2023, Design and Implementation of Automatic Feeding and Aging Line for Electric Rice Cookers. *Mechanical and Electrical Engineering Technology*, 52(12): 107–110.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Architectural Design of RT-DETR-L for PCB Surface Defect Detection: A Systematic Comparison of Attention Mechanisms, Backbone Replacement, and Cross-Dataset Generalization

Tao Huang*

School of Electronics and Information Engineering, Liaoning Technical University, Huludap 125105, Liaoning, China

*Corresponding author: Tao Huang, phamvanlinh65116@gmail.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Based on RT-DETR-L, this paper systematically compares five attention mechanisms (SE, CBAM, CA, ECA, and EMA) at the P3/P4/P5 outputs of the feature-pyramid neck under identical training conditions, and evaluates FasterNet backbone replacement and a P2 small-object detection head as complementary improvements. Experiments reveal a performance gap of up to 4.29 percentage points (CA: 93.04% to EMA: 97.33% in mAP_{50}), indicating that the choice of attention mechanism has a substantial impact on RT-DETR-type PCB detectors. EMA achieves the best mAP_{50} (97.33%) and the highest $mAP_{50:95}$ (56.45%); ECA offers a competitive trade-off without increasing GFLOPs (96.69%); CA performs worst (93.04%), a 3.34 pp drop below the baseline, tentatively attributed to an architectural conflict with the AIFI encoder. FasterNet backbone replacement trades accuracy for efficiency (31% fewer parameters, 40% lower GFLOPs); and, when trained from scratch on the second dataset DeepPCB, the EMA variant again yields the largest gain (mAP_{50} 89.33%, +4.83 pp over the baseline), showing that the improvement is not specific to a single dataset.

Keywords: PCB defect detection; RT-DETR; Attention mechanism comparison; EMA; FasterNet; Cross-dataset generalization

Online publication: Jul 7, 2026

1. Introduction

Surface defects in printed circuit board (PCB) manufacturing—open circuit, short, spur, spurious copper, missing hole, and mouse bite—lead to product failure and increased recall costs if left undetected^[1]. Deep learning has greatly advanced PCB defect detection: the YOLO series attains strong accuracy at high throughput, and various PCB-oriented YOLO variants further adapt it to manufacturing constraints. The Transformer-based RT-DETR overcomes the NMS bottleneck via bipartite matching and has drawn wide

attention in PCB detection: HSA-RTDETR integrates a FasterNet-EMA backbone and reaches 96.9% mAP_{50} , while other works improve small-object localization by introducing shallow P2 features or achieve light-weighting by drastically reducing parameters^[2,3].

However, two research gaps remain: (1) a systematic comparison of different attention mechanisms at the neck (after the backbone, before AIFI) of the RT-DETR-L feature pyramid is still absent in PCB detection; (2) systematic evaluation on more than one PCB dataset for RT-DETR-type detectors is largely missing.

The main contributions of this paper are:

the first controlled comparison of five attention mechanisms (SE, CBAM, CA, ECA, EMA) at the P3/P4/P5 neck outputs of RT-DETR-L, revealing a 4.29 pp performance gap;

- (1) Quantifying the effect of neck-level EMA insertion, achieving the best mAP_{50} (97.33%) and $mAP_{50:95}$ (56.45%);
- (2) Faithfully reporting the accuracy–efficiency trade-off of FasterNet backbone replacement (with the FasterNet+P2 configuration evaluated on DeepPCB);
- (3) Providing a multi-variant comparison on a second dataset (DeepPCB), where the EMA variant again yields the largest gain (89.33%), showing the improvement is not specific to one dataset.

2. Proposed method

2.1. RT-DETR-L baseline architecture

RT-DETR-L adopts an HGNetv2 backbone, an AIFI Transformer encoder, a RepC3 FPN neck, and an RTDETRDecoder for end-to-end detection^[2]. With 640×640 input, the backbone outputs three-level features P3 (80^2), P4 (40^2), and P5 (20^2); the AIFI encoder applies 8-head self-attention (embedding dimension 256, feed-forward dimension 1024) to the highest-level feature; the decoder requires no NMS thanks to bipartite matching. Because backbone replacement makes pretrained weights incompatible, all modified models are trained from scratch (pretrained=False)..

2.2. Neck-level EMA Multi-scale Attention (Main Contribution)

The EMA module is inserted at the P3/P4/P5 outputs, immediately before the input projection layer of the AIFI encoder. In simplified form, EMA can be summarized as two parallel branches operating on $\mathbf{X} \in R^{C \times H \times W}$ (the full module additionally applies grouped cross-spatial learning, see^[4]):

$$\mathbf{f}_c(\mathbf{X}) = \sigma(\mathbf{W}_1 \cdot GAP_{1D}(\mathbf{X})) \otimes \mathbf{X}, \mathbf{f}_s(\mathbf{X}) = \sigma(\mathbf{W}_3 * \mathbf{X}) \otimes \mathbf{X}, \mathbf{Y} = \mathbf{f}_c + \mathbf{f}_s$$

Unlike SE, EMA does not compress the channel dimension; compared with CBAM and CA, its parallel design avoids the sequential bottleneck. Our neck-level insertion differs from the backbone-integrated scheme of HSA-RTDETR, acting on post-backbone features before multi-scale fusion^[3, 5–7].

2.3. FasterNet Backbone Replacement and P2 Detection Head

HGNetv2 is replaced with FasterNet, whose core partial convolution (PConv) applies standard convoutin only to $C/4$ channels: $PConv(\mathbf{X}) = [Conv(\mathbf{X}_{1:C/4}); \mathbf{X}_{C/4+1:C}]$, greatly reducing computation^[8]. By leaving the remaining $3C/4$ channels untouched, PConv also avoids much of the memory-access overhead that bottlenecks depthwise-separable designs, which is the main reason FasterNet attains higher realized

throughput than its FLOPs alone would suggest. On top of the FasterNet backbone, an additional P2 feature map (160×160 , stride 4) is introduced as a fourth-scale detection head to improve the localization of small targets such as spurs and missing holes, following the established idea of introducing shallow features in RT-DETR to improve small-object localization. The P2 variant is evaluated only on DeepPCB and is not independently ablated on pcb-cropped. **Figure 1** contrasts the baseline (a) with the proposed EMA-enhanced architecture (b).

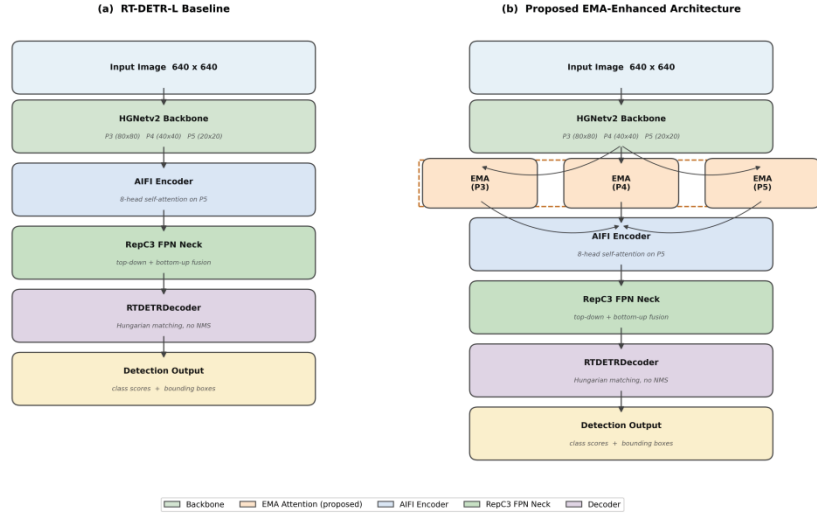


Figure 1. Architecture comparison. **(a)** RT-DETR-L baseline (HGNetv2 backbone); **(b)** the proposed EMA-enhanced architecture: EMA modules (orange) inserted at the P3/P4/P5 outputs, followed by the AIFI encoder → RepC3 FPN → RTDETRDecoder to produce detections.

3. Experiments

3.1. Datasets and implementation details

3.1.1. Pcb-cropped (main benchmark)

From the Peking University PCB defect image library, 693 original images and six defect classes (open circuit, short, mouse bite, spur, missing hole, spurious copper). The split is first performed at the original-image level in an 8:1:1 ratio. 640×640 Sliding-window cropping is applied independently to avoid data leakage fully, yielding 4 379 cropped images (3 507 train / 439 val / 433 test) and 5 266 annotated instances.

3.1.2. DeepPCB

1 000 training and 500 test images at (binarized CCD image pairs, with a class definition different from pcb-cropped), used only as a second dataset to assess whether the improvement generalizes beyond pcb-cropped ^[9] 640×640 .

3.1.3. Training configuration

AdamW optimizer; lr=1e-4; weight decay 1e-4; cosine schedule; 2 warmup epochs; batch 16; 100 epochs (early stopping patience=30); AMP=True; seed 42. FPS is measured on an NVIDIA RTX 4090 D with batch=1 (50 warmup + 300 timed runs).

3.1.4. Evaluation metrics

We report mAP_{50} and $\text{mAP}_{50:95}$ (COCO-style, averaged over IoU thresholds 0.50:0.05:0.95), together with precision and recall at the default confidence threshold; parameter count, GFLOPs, and FPS serve as efficiency indicators. All reported accuracy numbers are obtained on the held-out test split, never on the validation split used for early stopping, so that no model selection leaks into the test measurement.

3.2. Ablation study

Table [tab:ablation] progressively adds improvement modules starting from the RT-DETR-L baseline.

The EMA-only variant achieves the highest mAP_{50} (97.33%, + 0.95 pp) and $\text{mAP}_{50:95}$ (56.45%, + 1.75 pp). The FasterNet variant drops 1.07 pp in accuracy but reduces parameters by 31% (32.82 $\rightarrow$ 22.78 M), GFLOPs by 40% (108.0 $\rightarrow$ 65.3), and raises FPS to 11.3. The combined FasterNet+EMA variant (95.64%) falls below both the EMA-only configuration and the baseline, indicating that the restriction of PConv on channel diversity weakens the discriminative feature variance that EMA can exploit.

Figure 2 shows the training process of the proposed EMA variant: all losses converge smoothly, mAP_{50} and $\text{mAP}_{50:95}$ stabilize after about 40 epochs with no sign of overfitting, confirming convergence stability under from-scratch training (pretrained = False).

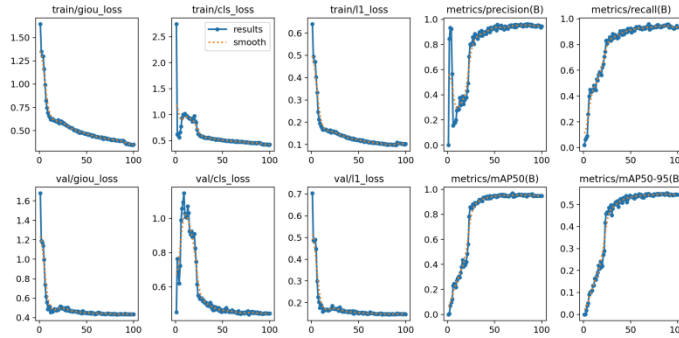


Figure 2. Training curves of the proposed EMA variant on pcb-cropped (100 epochs). Top row: training losses, precision, and recall; bottom row: validation losses, mAP_{50} , and $\text{mAP}_{50:95}$.

3.3. Attention mechanism comparison

All five attention mechanisms are integrated at the P3/P4/P5 neck outputs of the original HGNetv2 backbone of RT-DETR-L under identical training conditions.

The mAP_{50} ranking is EMA (97.33%) > ECA (96.69%) > None (96.38%) > CBAM (95.55%) > SE (95.43%) > CA (93.04%), with a maximum gap of **4.29 pp**. CA shows a notable drop (−3.34 pp); we tentatively attribute this to a conflict between its horizontal–vertical coordinate-decomposed encoding and the holistic spatial query mechanism of the AIFI encoder, with the exact mechanism awaiting further verification via feature-map visualization and similar means.

Two broader patterns emerge. First, only EMA and ECA surpass the no-attention baseline, whereas SE, CBAM, and CA all degrade it. This suggests that, at the neck of a DETR-style detector, attention that compresses or sequentially reshapes the channel description—SE squeezes channels, CBAM applies channel-

then-spatial attention in sequence, and CA factorizes space into two 1-D directions—can disturb the feature distribution that the downstream AIFI encoder expects. Second, the two mechanisms share a common trait: they preserve the full channel description. ECA does so through a parameter-free 1-D local cross-channel interaction, leaving GFLOPs identical to the baseline, while EMA does so through parallel channel and multi-scale spatial branches without channel reduction. EMA’s extra multi-scale spatial modeling explains why it edges out ECA (+0.64 pp mAP₅₀) at the cost of 11.5 more GFLOPs, making ECA the preferable choice when compute is constrained and EMA the preferable choice when accuracy is paramount^[10].

3.4. Comparison with state-of-the-art methods

[†] Cited from the original paper, using PKU-PCB original images (no sliding-window cropping); FPS measured on different hardware, not directly comparable, provided for method reference only.

The proposed EMA variant achieves the highest mAP₅₀ (97.33%) and mAP_{50:95} (56.45%) under the unified protocol, slightly above YOLOv8l (97.32%/55.69%). The end-to-end FPS of the YOLO series (76.7–112.7) is markedly higher than that of the RT-DETR series (8.7–11.3), mainly because NMS post-processing is highly engineering-optimized; the FasterNet variant (11.3 FPS) has the highest throughput among the RT-DETR family.

3.5. Cross-dataset generalization: DeepPCB

Table 1. DeepPCB cross-dataset generalization results (500 test images; all three models trained from scratch on the DeepPCB training set)

Method	Params (M)	GFLOPs	FPS	mAP ₅₀ (%)	mAP _{50:95} (%)
RT-DETR-L (baseline)	32.82	108.0	8.5	84.50	57.55
RT-DETR-L+EMA (ours)	33.68	119.5	8.8	89.33	61.08
FasterNet+P2	26.03	152.0	9.4	85.66	58.89
EMA vs. baseline	+0.86	+11.5	+0.3	+4.83	+3.53

When all three models are trained from scratch on DeepPCB, the EMA variant reaches 89.33% mAP₅₀ (+ 4.83 pp) and improves mAP_{50:95} by 3.53 pp. Because DeepPCB differs from pcb-cropped in imaging modality and class definition, this indicates that the gain from neck-level EMA attention is not specific to a single dataset but holds on a second, distinct dataset. Unlike most existing RT-DETR-type PCB studies that report results on only one dataset, our multi-variant comparison across two datasets shows that the choice of attention mechanism contributes robustly to detection quality. Note that FasterNet+P2 has the highest GFLOPs (152.0) in this table because the high-resolution P2 head (160×160) adds substantially more computation than the FasterNet backbone saves.

3.6. Qualitative results and error analysis

Figure 3 shows the normalized confusion matrix of the proposed EMA variant on the pcb-cropped test set: the diagonal recall of all six classes is no lower than 0.96 (short reaches 1.00), indicating strong inter-class discrimination; the main errors come from the background column (spurious_copper missed at 0.32, open_circuit at 0.18), reflecting that confusing tiny, low-contrast defects with copper-foil texture remains a core

challenge in PCB detection.

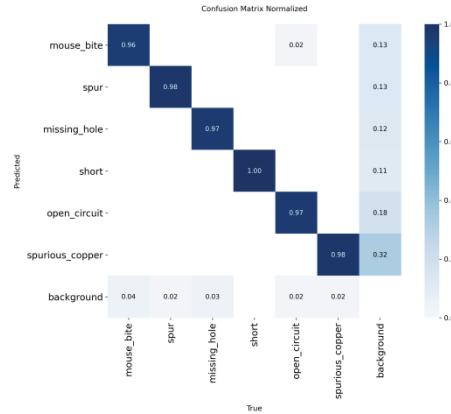


Figure 3. Normalized confusion matrix of the proposed EMA variant on the PCB-cropped test set. The diagonal gives per-class recall; the rightmost column gives the proportion missed as background.

4. Conclusion

Based on RT-DETR-L, this paper systematically compares five neck attention mechanisms and reveals a 4.29 pp performance gap: EMA achieves the best mAP_{50} (97.33%) and $\text{mAP}_{50:95}$ (56.45%), both surpassing the baseline; CA drops 3.34 pp, tentatively attributed to an architectural conflict with the AIFI encoder; FasterNet trades 1.07 pp of accuracy for 31% fewer parameters and 40% lower GFLOPs; and the EMA gain also holds on the second dataset DeepPCB (89.33%, 4.83 pp over the baseline). Future work will study the FasterNet–EMA negative interaction and knowledge distillation for edge deployment.

Data availability

pcb-cropped is available from the author upon reasonable request; DeepPCB is publicly available (<https://github.com/tangsanli5201/DeepPCB>).

Author contributions

T.H. conceived the study, designed and performed the experiments, analyzed the data, and wrote the manuscript.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Ling Q, Isa N, 2023, Printed Circuit Board Defect Detection Methods Based on Image Processing, Machine Learning and Deep Learning: A Survey. IEEE Access, 11: 15921–15944.

- [2] Zhao Y, Lv W, Xu S, et al., 2024, DETRs Beat YOLOs on Real-Time Object Detection. CVPR.
- [3] Wang Y, Wu B, Zhang L, et al., 2025, Enhanced PCB Defect Detection Via HSA-RTDETR on RT-DETR. Scientific Reports, 15: 31783.
- [4] Ouyang D, He S, Zhang G, et al., 2023, Efficient Multi-Scale Attention Module with Cross-Spatial Learning. ICASSP.
- [5] Hu J, Shen L, Sun G, 2018, Squeeze-and-Excitation Networks. CVPR: 7132–7141.
- [6] Woo S, Park J, Lee J, et al., 2018, CBAM: Convolutional Block Attention Module. ECCV: 3–19.
- [7] Hou Q, Zhou D, Feng J, 2021, Coordinate Attention for Efficient Mobile Network Design. CVPR: 13713–13722.
- [8] Chen J, Kao S, He H, et al., 2023, Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. CVPR: 12021–12031.
- [9] Tang S, He F, Huang X, et al., 2019, On-Line PCB Defect Detector on a New PCB Defect Dataset. arXiv: 1902.06197.
- [10] Wang Q, Wu B, Zhu P, et al., 2020, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. CVPR: 11534–11542.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Finite-Element Modeling and Analytical Validation of a Coaxial Cylindrical Capacitor Using Maxwell 2D

Qi Wang*

School of Electronic and Information Engineering, Liaoning Technical University (Huludao Campus), Huludao 125105, Liaoning, China

*Corresponding author: Qi Wang, wangqi545@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To improve the reliability of capacitance prediction for compact coaxial components, a cylindrical capacitor with copper conductors and an air dielectric is modeled and verified by Ansys Maxwell 2D electrostatic simulation. The inner conductor radius is 0.6 mm, the outer conductor inner radius is 1.0 mm, and the outer conductor wall thickness is 0.2 mm. A 1 V potential is assigned to the inner conductor, while the outer conductor is grounded. Based on Gauss's law, the analytical per-unit-length capacitance is calculated to be 108.9 pF/m. The finite-element model yields 107.4 pF/m, corresponding to a relative error of 1.38%. The simulated electric-field distribution is radially symmetric and reaches its maximum near the inner conductor, which agrees with the theoretical relation $E(r) \propto 1/r$. The results show that Maxwell 2D can accurately reproduce the electrostatic behavior of coaxial capacitors and provide an efficient workflow for capacitance extraction, field visualization, and preliminary structural optimization.

Keywords: Coaxial cylindrical capacitor; Capacitance extraction; Maxwell 2D; Finite-element method; Electric-field distribution

Online publication: Jul 7, 2026

1. Introduction

Capacitors are fundamental passive components in electronic systems and are widely used for filtering, coupling, decoupling, energy storage, and signal integrity control ^[1]. With radio-frequency modules, medical electronics, and compact sensing devices advancing toward miniaturization and high reliability, accurate prediction of capacitance and local electric-field concentration before fabrication is increasingly important ^[2]. Coaxial cylindrical capacitors are representative structures owing to their compact geometry, uniform field distribution, and direct relationship among capacitance, conductor radius, and dielectric properties ^[3–5].

Traditional capacitor design often relies on repeated prototyping or empirical adjustment. At the

same time, numerical electromagnetic simulation can evaluate capacitance, visualize field concentration, and analyze geometric influence more efficiently at the design stage ^[6]. For coaxial structures, analytical electrostatic theory provides a clear reference solution ^[7]. Therefore, comparison between analytical calculations and finite-element simulations is an effective way to verify model accuracy ^[7-9]. In this study, a copper-air coaxial capacitor is established in Ansys Maxwell 2D. The capacitance and electric-field distribution are extracted from the electrostatic solution, compared with the theoretical value derived from Gauss's law, and discussed in terms of mesh size, boundary representation, and numerical convergence ^[10].

2. Physical model and analytical formulation

2.1. Coaxial capacitor configuration

The investigated structure consists of two coaxial cylindrical conductors separated by air. The inner solid conductor is used as the high-potential electrode, and the outer annular conductor serves as the grounded shield. The inner conductor radius is $a = 0.6$ mm, the outer conductor inner radius is $b = 1.0$ mm, and the outer conductor wall thickness is 0.2 mm. In the two-dimensional electrostatic model, the cross section represents a uniform, infinitely long coaxial structure, so the extracted capacitance is interpreted as per-unit-length capacitance. See **Figure 1**.

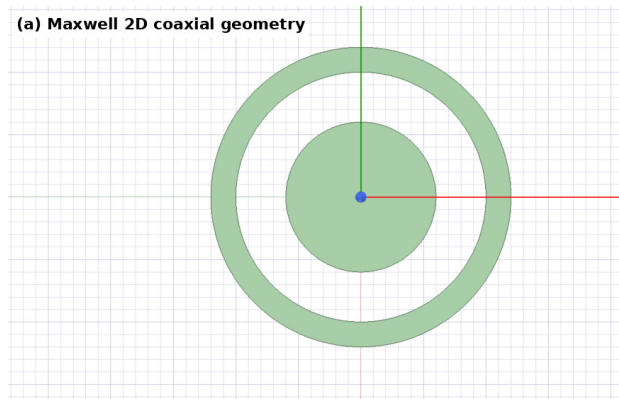


Figure 1. Two-dimensional coaxial capacitor geometry used in the Maxwell electrostatic model.

In the region $a < r < b$ The dielectric is air, and there is no free-volume charge. By applying Gauss's law to a coaxial cylindrical surface, the radial electric field can be expressed as $E(r) = \lambda / (2\pi\epsilon r)$ where λ is the line charge density and ϵ is the permittivity of the medium. Integrating the electric field between the two conductors gives the potential difference:

$$U = \int_a^b E(r) dr = \frac{\lambda \ln(b/a)}{2\pi\epsilon}$$

Therefore, the per-unit-length capacitance is obtained as:

$$C' = \frac{\lambda}{U} = \frac{2\pi\epsilon}{\ln(b/a)}$$

For air, ϵ is approximately ϵ_0 . Substituting $a = 0.6$ mm and $b = 1.0$ mm gives $\ln(b/a) = \ln(1.6667) = 0.5108$,

and the analytical result is $1.089 \times 10^{-10} \text{ F/m}$ namely, 108.9 pF/m. This value serves as the benchmark for validating the Maxwell 2D simulation results. See **Table 1**.

Table 1. Model parameters and theoretical evaluation basis

Item	Value	Role in the model
Inner conductor radius a	0.6 mm	Defines the high-potential central electrode
Outer conductor inner radius b	1.0 mm	Determines the air-gap radius for radial field formation
Outer conductor wall thickness	0.2 mm	Represents the grounded copper shield
Dielectric medium	Air, $\epsilon_r \approx 1$	Stores electrostatic energy between the electrodes
Excitation and boundary	1 V inner; 0 V outer	Forms the electrostatic potential difference
Analytical reference	$C' = 108.9 \text{ pF/m}$	Used to verify the finite-element result

3. Maxwell 2D finite-element simulation method

3.1. Modeling and boundary assignment

The finite-element model is built in Ansys Maxwell 2D under an electrostatic solution type. First, the concentric circular regions are drawn according to the specified radii. Copper is assigned to the inner and outer conductors, and the space between them is set as air. The outer conductor surface is defined as the ground reference at 0 V, and a 1 V excitation is applied to the inner conductor. This boundary setting reproduces the ideal coaxial capacitor condition and enables capacitance extraction from the electrostatic matrix.

After material and excitation assignments, the capacitance matrix is computed using the solver. The simulation focuses on two outputs: the capacitance at the voltage terminal and the spatial distribution of the electric field. The former provides a quantitative index for comparison with theory, whereas the latter gives physical evidence of whether the field is mainly radial and concentrated near the smaller-radius electrode. See **Figure 2**.

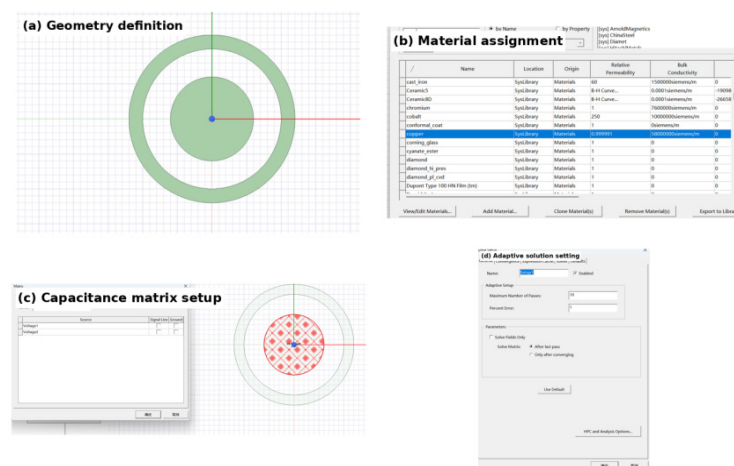


Figure 2. Maxwell 2D modeling procedure: geometry definition, material assignment, capacitance-matrix setup, and adaptive solution setting.

3.2. Numerical evaluation indicators

The relative error between the simulated capacitance and the analytical value is used as the primary validation metric, and an error below 5% is considered acceptable for preliminary engineering evaluation. The electric-field distribution is also examined qualitatively. For a correct coaxial electrostatic solution, the field should be primarily radial, strongest near the inner conductor, and decrease gradually with increasing radius.

Since the numerical solution is obtained on a finite mesh, a slight deviation from the closed-form result is unavoidable. Possible error sources include circular-boundary discretization, finite computational-region treatment, solver residuals, and the ideal assumptions used in the analytical model.

4. Results and discussion

4.1. Capacitance extraction and electric-field distribution

The Maxwell 2D simulation gives a capacitance of 107.4 pF/m. As shown in **Figure 3**, the electric field is mainly radial, starting from the inner conductor and ending on the grounded outer conductor. The field intensity is highest near the inner conductor and decreases outward, which agrees with the theoretical relation $E(r) \propto 1/r$. This confirms that the finite-element model accurately describes the electrostatic behavior of the coaxial capacitor.

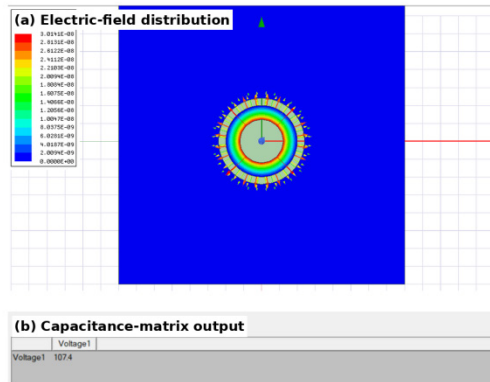


Figure 3. Simulation outputs: electric-field distribution and capacitance-matrix result showing 107.4 pF/m.

4.2. Accuracy verification and error analysis

Compared with the analytical reference of 108.9 pF/m, the relative error of the simulation is calculated as

$$\delta = \frac{|108.9 - 107.4|}{108.9} \times 100\% = 1.38\%$$

The error is well below the 5% design criterion, indicating that the model is sufficiently accurate for the training-stage design and preliminary engineering evaluation. The close agreement also indicates that the adopted cross-sectional model preserves the coaxial symmetry and that the air dielectric has been correctly assigned. See **Table 2**.

Table 2. Comparison between finite-element simulation and analytical calculation

Index	Simulation result	Analytical reference	Evaluation
Capacitance C'	107.4 pF/m	108.9 pF/m	Relative error: 1.38%
Field pattern	Radial distribution: strongest near the inner	$E(r) \propto 1/r$	Consistent with theory
Accuracy requirement	Error below 5%	Design criterion	Satisfied

The remaining difference between theory and simulation is mainly attributable to mesh approximation and numerical truncation. Discrete finite elements represent a circular boundary, so the exact geometry cannot be reproduced without sufficiently fine meshing. In addition, the solver stops after meeting its convergence criterion rather than reaching an exact analytical solution. For higher-precision design, mesh convergence testing, local refinement near conductor edges, and comparison under different boundary extents should be performed.

From an engineering perspective, the verified model can be used for rapid parametric studies. Increasing the dielectric constant would proportionally increase capacitance, while changing the radius ratio b/a would directly alter the logarithmic denominator in the capacitance expression. Therefore, the modeling workflow is suitable for evaluating coupling and decoupling capacitors in high-frequency modules, compact electronic packages, and other structures where electric-field concentration must be controlled.

5. Conclusion

A coaxial cylindrical capacitor with copper conductors and an air dielectric was modeled and analyzed using Ansys Maxwell 2D. The inner conductor radius was 0.6 mm, the outer conductor inner radius was 1.0 mm, and the grounded outer wall thickness was 0.2 mm. With a 1 V excitation applied to the inner conductor, the simulated per-unit-length capacitance was 107.4 pF/m. The theoretical value derived from Gauss law was 108.9 pF/m, yielding a relative error of 1.38%. The electric-field distribution was predominantly radial, and the strongest field occurred near the inner conductor, consistent with electrostatic theory. These results verify the accuracy of the finite-element model and demonstrate that Maxwell 2D is effective for predicting capacitance, visualizing fields, and early-stage optimization of cylindrical capacitor structures. Future work should include mesh convergence tests, dielectric material comparison, finite-length end-effect analysis, and experimental measurements.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Osorio L, Hall S, Saavedra R, et al., 2025, Finite Element Modeling for Biophysical Models of Nervous System Stimulation: Best Practices for Multiscale Adaptive Meshing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 33: 298–309.
- [2] Chan M, Chiang K, 2025, Study on the Applicability of Multi-Point Constraints for 3D WLP Finite Element Model. 2025 26th International Conference on Thermal, Mechanical and Multi-Physics Simulation and

Experiments in Microelectronics and Microsystems (EuroSimE), Utrecht, Netherlands: 1–5.

- [3] Jiang C, Deng F, Wang Y, et al., 2026, Experimental Verification of Finite Element Modeling for ICT-Type DC High-Voltage Power Supply. *IEEE Transactions on Applied Superconductivity*, 36(3): 1–6.
- [4] Sharma R, Soni P, Mohale V, 2026, Development and Electrical Characterization of Compact High-Voltage Film-Foil Cylindrical Pulse Capacitor. *IEEE Transactions on Plasma Science*.
- [5] Fujita K, Kuramoto N, 2024, Finite-Element Simulations of the Effects of Shielding and Guarding of a Capacitor Using Coaxial Cylindrical Electrodes for Electrostatic Force Balance Measurements. 2024 Conference on Precision Electromagnetic Measurements (CPEM), Denver, USA: 1–2.
- [6] Fujita K, Kaneko N, Domae A, et al., 2025, Effects of Shielding and Guarding Electrodes on Electrostatic-Force Balance Measurements in a Coaxial Cylindrical Capacitor. *IEEE Transactions on Instrumentation and Measurement*, 74: 1–11.
- [7] Alavi S, Sumaiya O, Silva D, et al., 2026, Benchmarking Machine Learning Methods Against Finite Element and Empirical Models for Predicting Solder Joint Characteristic Life. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 16(5): 1047–1061.
- [8] Elkin S, Haider M, Roth T, 2026, Finite-Element Time-Domain Model for Lossy Josephson Traveling-Wave Parametric Amplifiers. *IEEE Microwave and Wireless Technology Letters*, 36(4): 489–492.
- [9] Zhao L, Zhao B, 2026, Co-Simulation and Performance Analysis of Tidal Current Energy PMSG with RMxprt & Maxwell 2D. 2026 6th International Conference on Advances in Electrical, Electronics and Computing Technology (EECT), Hangzhou, China: 1–5.
- [10] Zhao Y, Zhang Y, 2025, Frequency Response Function-Based Finite Element Model Updating Using Principal Component Analysis and Extreme Learning Machine. 2025 5th International Conference on Machine Manufacturing and Intelligent Control (ICMMIC), Tianshui, China: 22–25.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Analysis of the Application of AI Visual Inspection Technology in Precision Mechanical Manufacturing Automation

Tianwei Zheng

Monash University, Melbourne 3800, Australia

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In recent years, driven by the Industry 5.0 wave, precision mechanical manufacturing is accelerating its evolution toward ultra-precision, intelligence, and unmanned operation, posing unprecedented, stringent requirements for precision control, efficiency improvement, and cost optimization in the manufacturing process. As a core carrier of in-depth integration of artificial intelligence and industrial automation, AI visual inspection technology, with its advantages of non-contact inspection, real-time response, and adaptive learning, has gradually replaced traditional manual inspection and conventional machine visual inspection. It has become a key technical support for addressing pain points such as micro-defect identification, adaptation to complex working conditions, and full-process quality control in precision mechanical manufacturing. In this regard, this paper first analyzes the application scenarios of AI visual inspection technology in precision mechanical manufacturing automation, and then expounds the development trends of its applications in this field, to provide a reference for relevant researchers.

Keywords: AI visual inspection technology; Precision mechanical manufacturing; Automation; Application

Online publication: Jul 23, 2026

1. Application scenarios of AI visual inspection technology in precision mechanical manufacturing automation

1.1. Dynamic adaptive inspection in precision machining process

In precision mechanical machining, tool wear, workpiece deformation, deviations in cutting parameters, and other factors can reduce the dimensional accuracy of parts and components and cause surface defects. Most traditional inspection methods rely on offline inspection after machining is complete, which cannot detect abnormalities in the machining process in a timely manner, can easily lead to batch waste, and increase manufacturing costs. The application of AI visual inspection technology realizes dynamic adaptive inspection in the machining process, breaks through the traditional closed loop of “machining–inspection–rework”, and establishes an automatic closed-loop mode of real-time inspection–timely adjustment–precise machining^[1].

Compared with traditional offline inspection, the new AI visual inspection technology can be embedded in the core parts of processing equipment, such as the spindle and turret. It employs high-frame-rate industrial cameras and adaptive light sources to capture dynamic images during tool cutting, and then uses deep learning algorithms to instantly identify and quantitatively evaluate tool wear, workpiece surface roughness, dimensional deviation, and other indicators. Meanwhile, this technology possesses strong adaptive learning capability. It can automatically adjust inspection parameters and identification models based on the characteristics of different processing materials and technologies, without manual recalibration, making it suitable for multi-type, small-batch precision machining. Taking the precision grinding of aero-engine blades as an example, AI visual inspection technology can monitor micro-scratches, edge chipping, and other defects on the blade surface in real time, as well as the wear of grinding tools. Once abnormalities are detected, feedback signals will be sent to the numerical control system to automatically adjust parameters such as grinding speed and feed rate, prevent defects from spreading further, and ensure that blade machining accuracy meets micron-level requirements ^[2].

1.2. Intelligent positioning and error compensation in precision assembly links

Precision mechanical assembly is a key link determining product performance and reliability. In fields such as aerospace and medical devices, where parts and components require high assembly accuracy, even tiny assembly errors can lead to product failure. In the past, assembly positioning was mostly done manually based on experience, resulting in low efficiency and a high risk of assembly errors caused by human operator mistakes, which did not meet the requirements of automated assembly. The application of AI visual inspection technology transforms positioning, inspection, and correction in precision assembly from manual operation to intelligent positioning, real-time detection, and dynamic adjustment. The assembly process has changed from manual assistance to full automation, greatly improving assembly efficiency and accuracy.

By adopting multi-camera collaborative acquisition and 3D visual modeling technologies, AI visual inspection technology can quickly obtain the three-dimensional spatial coordinates and surface features of parts to be assembled. It uses deep learning algorithms to accurately compare with preset assembly standard models, automatically identify assembly abnormalities such as part position and angle deviations, calculate compensation in real time, and control assembly robots to achieve accurate positioning and assembly. Different from traditional 2D visual positioning, AI visual inspection technology based on the integration of structured light and laser triangulation can accurately acquire the 3D contour information of parts, effectively solving the problems of missing Z-axis direction information and susceptibility to light interference in 2D visual inspection, and meeting the assembly requirements of parts with complex structures. In the assembly of micro precision gearboxes, the adopted AI visual inspection system can simultaneously scan the positions of gears, bearings, shafts and other components, automatically detect the meshing clearance between gears and the assembly correctness of bearings, and then adjust the moving mode of assembly robots according to the above information to ensure the gear meshing accuracy meets design standards, with the assembly efficiency increased by more than 5 times ^[3].

1.3. Non-contact precision inspection under extreme working conditions

The manufacturing and inspection processes for precision mechanical products must be carried out under extreme working conditions, such as high temperatures, high pressures, intense radiation, and high vacuum.

Traditional contact inspection methods easily damage part surfaces, and inspection equipment can hardly operate reliably under extreme working conditions, thereby compromising inspection accuracy. Thanks to the characteristics of non-contact inspection, AI visual inspection technology employs high-temperature- and radiation-resistant industrial cameras, along with adaptive imaging technology, to inspect precision parts under extreme working conditions, thereby breaking the application boundaries of traditional inspection technology ^[4].

For example, in the pouring process of precision castings, high-temperature resistant industrial cameras and infrared imaging technology can penetrate high-temperature smoke, observe the casting pouring situation in real time, accurately locate defects such as air holes, cracks, and shrinkage cavities on the casting surface, and monitor the dimensional changes of castings, providing data support for timely adjustment of pouring technology. Compared with traditional inspection methods, this non-contact inspection method neither damages inspection equipment due to high temperatures nor enables real-time monitoring of the pouring process, thereby preventing casting defects. During inspections of precision parts in nuclear reactors, radiation-resistant cameras, combined with remote control, are used to examine surface defects and the parts' dimensional accuracy, thereby improving inspection efficiency and accuracy while ensuring operator safety ^[5].

1.4. Predictive inspection and operation & maintenance of full life cycle

Traditional inspection in precision mechanical manufacturing primarily focuses on the machining and assembly stages, with insufficient attention to condition monitoring and early fault warning during product service. Products are prone to failure due to hidden defects in the service process, which increase operation and maintenance costs and pose safety risks. AI visual inspection technology can enable predictive inspection and operation & maintenance across the full life cycle of precision mechanical products, break the previous thinking that emphasizes manufacturing while ignoring operation & maintenance, and establish an automatic control system that integrates manufacturing, inspection, and operation & maintenance. During product service, the AI visual inspection system can be embedded in key equipment components. It uses real-time acquired image data of the product operation process, combined with deep learning algorithms and big data analysis methods, to detect hidden flaws on the product surface, monitor product operation parameter indicators, establish a product service condition prediction model, so as to send fault early warning signals in advance and provide strong data support for operation and maintenance decision-making. For example, miniature industrial cameras are installed inside aero-engines during service to monitor wear, cracks, and other conditions on the blade surface in real time. Combined with parameters such as vibration and temperature during blade operation, the remaining service life of blades can be predicted, enabling maintenance and replacement in advance and preventing safety accidents caused by engine failures ^[6].

1.5. Composite inspection application of multimodal fusion

At present, the structure of precision mechanical products is becoming increasingly complex, and single-modal visual inspection can no longer meet the requirements of multi-type and multi-dimensional inspection. In recent years, multimodal fused AI visual inspection technology has gradually become popular. By integrating multiple inspection modalities such as visible light, infrared, X-rays, and lasers, it enables multi-angle, all-round inspection of precision mechanical parts, breaks the constraints of single-modal inspection, and improves the comprehensiveness and accuracy of inspection. Meanwhile, multimodal-fused AI visual

inspection technology can meet the inspection requirements for parts with complex materials. For example, single-modal inspection has difficulty identifying surface and internal defects in precision parts made of special materials such as carbon fiber composites and ceramics, whereas multimodal fused inspection can effectively overcome this difficulty and achieve accurate inspection. When inspecting precision parts made of carbon fiber composites, multimodal AI visual inspection technology can detect internal delamination, pores, and other defects. The composite inspection application of multimodal fusion can broaden the scope of AI visual inspection technology, improve the comprehensiveness and accuracy of inspection, and provide a new solution for quality control of complex precision mechanical products^[7].

2. Development trends of AI visual inspection technology in precision mechanical manufacturing automation

2.1. Continuous breakthrough of ultra-precision inspection technology

As precision mechanical manufacturing moves toward nanometer-level accuracy, AI visual inspection technology will continue to make breakthroughs in ultra-precision inspection, continuously improving accuracy and resolution. On the one hand, high-resolution imaging technology will continue to advance, and new nanometer-level industrial cameras and laser imaging equipment will gradually become more widely adopted, enabling accurate detection of nanometer-level micro-defects and dimensional deviations. On the other hand, deep learning algorithms will be continuously optimized, and new algorithms based on the fusion of Transformer and CNN will be gradually applied to achieve accurate feature extraction and the identification of complex-structured parts, thereby improving the model's generalization ability and detection accuracy^[8].

At the same time, integrating quantum imaging, terahertz imaging, and other technologies with AI-based visual inspection can overcome the limitations of traditional imaging, accurately detect internal defects in precision mechanical parts, and enable comprehensive inspection without damaging the parts, thereby providing technical support for ultra-precision mechanical manufacturing. In addition, AI visual inspection technology can penetrate the product surface to accurately detect internal micro-cracks, inclusions, and other defects, achieving nanometer-level inspection accuracy. It is widely used in high-end precision mechanical manufacturing fields such as aerospace and microelectronics.

2.2. In-depth and collaborative development of technology integration

AI visual inspection technology will be deeply integrated with emerging technologies such as edge computing, 5G, digital twin, and blockchain, thereby continuously improving the coordination and compatibility of these technologies and establishing a full-process, intelligent, and integrated automatic management and control system for precision mechanical manufacturing. First, it is deeply integrated with edge computing to enable real-time processing of inspection data, reduce data transmission delays, improve real-time inspection performance, and closely match the demands of high-speed automated production. Deep integration with 5G technology enables high-speed data transmission and sharing among multiple devices and systems, breaks down data silos, and achieves full-process collaborative management and control^[9].

Second, the deep integration of AI visual inspection technology and digital twin technology will form a digital twin model of precision mechanical products, realizing real-time synchronization of inspection data and product digital models, simulating the manufacturing and service process of products, so as to

improve machining, assembly, and operation & maintenance strategies and achieve precise manufacturing and precise operation & maintenance. Meanwhile, through in-depth integration with blockchain technology, inspection data can be traced throughout the process and cannot be tampered with at will, thereby improving its authenticity and reliability and effectively controlling product quality. In addition, AI visual inspection technology is closely integrated with robots, enabling seamless collaboration between inspection and operations. Installing visual inspection devices on robots can automatically perform operations such as parts inspection, sorting, and rework, enabling full-process automation of precision mechanical manufacturing, reducing errors caused by manual intervention, improving the stability and efficiency of complex operational links, and further releasing the production capacity potential of precision manufacturing.

2.3. Mult scenario adaptive and intelligent upgrade

AI visual inspection systems will have a wider adaptive range, automatically adapting to various processing methods, product varieties, and working environments. They can automatically adjust inspection parameters and actively adapt to models without human intervention, realizing “one-click inspection”. Taking the iteration of precision mechanical products as an example, the AI visual inspection system can automatically identify the characteristics of new products, quickly adjust inspection models and parameters without manual recalibration and training, realize accurate inspection of new products, and quickly respond to the customized production demand of multi-variety and small-batch production in the precision manufacturing field.

AI visual inspection technology is gradually developing toward higher intelligence. By adopting reinforcement learning, deep learning, and other technologies, the inspection model can self-optimize and upgrade, adjust its identification strategies in response to changes in inspection data, and continuously improve inspection accuracy and speed. In addition, AI visual inspection systems will have stronger independent decision-making capabilities and can automatically determine rework and parameter adjustments based on inspection results, thereby integrating inspection and disposal, reducing human intervention, and promoting the continuous transformation of precision mechanical manufacturing toward automation^[10].

3. Conclusion

In a word, with the outstanding advantages of non-contact, high precision, and intelligence, AI visual inspection technology has been deeply integrated into the whole process of machining, assembly, operation, and maintenance of precision mechanical manufacturing automation. It has shown significant technical value in multiple typical application scenarios, effectively addressing the pain points of traditional inspection technology, such as insufficient precision, poor timeliness, and limited applicability, and providing core technical support for quality control and the automation upgrade of precision mechanical manufacturing. In the future, with the continuous breakthrough of ultra-precision imaging technology and deep learning algorithms as well as the in-depth integration with various emerging technologies, AI visual inspection technology will further upgrade toward adaptiveness and intelligence, continuously expand its application boundary in the field of high-end precision manufacturing, provide stronger power for the development of precision mechanical manufacturing automation, promote the high-end equipment manufacturing industry to develop toward higher precision and higher efficiency, and better support the high-quality development of China's high-end manufacturing industry.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Yin T, Fang G, Liu G, 2025, Application of AI Visual Inspection Technology in Refrigerator Production Equipment. *Information & Computer*, 37(13): 8–10.
- [2] Zhang Y, Tian M, Lu T, 2025, Application of AI Visual Inspection Technology in Quality Management of Wood-based Panels. *China Wood-based Panels*, 32(5): 32–36.
- [3] Fan X, Chen Z, Sai Y, et al., 2025, Research and Application of General Assembly Detection Technology for Complex Products Based on Visual Inspection. *Quality and Reliability*, 2025(2): 6–14.
- [4] Qiu Y, 2024, Optimization and Application of High-adaptability Visual Inspection Technology. *Modern Manufacturing Technology and Equipment*, 60(12): 119–122.
- [5] Wang Y, Yan K, Hu X, et al., 2024, Research and Application of Machine Vision Inspection Technology for Integrated Large Die-castings. In: *Proceedings of the 31st Annual Conference of China Society of Automotive Engineers (5)*. Chongqing Changan Automobile Co., Ltd. Manufacturing Center: 184–187.
- [6] Lin Q, 2024, Key Technology and Equipment of Intelligent Visual Inspection for Complex Parts. Zhejiang: QiJing Machinery Co., Ltd.
- [7] Zheng G, 2024, Research on Mechanical Design Manufacturing Automation Technology and Precision Machining Technology. *China Strategic Emerging Industry*, 2024(20): 160–162.
- [8] Yue J, Fan J, 2023, Application Analysis of Precision Inspection Technology in the Field of Mechanical Machining and Manufacturing. *Mold Manufacturing*, 23(11): 133–135.
- [9] He J, 2022, Application Analysis of Precision Inspection Technology in the Field of Mechanical Machining and Manufacturing. *China Metal Bulletin*, 2022(1): 58–60.
- [10] Ni J, 2021, Insights on Application Analysis of Precision Inspection Technology in the Field of Mechanical Machining and Manufacturing. *China Equipment Engineering*, 2021(24): 185–186.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on the Controllable Growth Mechanism and Structural Characterization Methods of Single-Walled Carbon Nanotubes

Jingchun Zhang, Wenqing Shen, Zhikun Chen

Ningbo Xiwo New Material Technology Co., Ltd., Ningbo 315300, Zhejiang, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: As a typical one-dimensional carbon nanomaterial, single-walled carbon nanotubes (SWCNTs) possess a unique electronic structure, excellent mechanical properties, and outstanding thermal and electrical conductivity, showing important application value in cutting-edge fields such as nanoelectronic devices and composite materials. Structure determines properties, which is the core foundation for the application of SWCNTs. Their physicochemical properties are significantly affected by microstructural parameters such as diameter and chirality. Therefore, in-depth analysis of the controllable growth mechanism and research on precise structural characterization methods are of great significance for promoting the large-scale application of SWCNTs.

Keywords: Single-walled carbon nanotubes; Controllable growth mechanism; Structural characterization methods; Research progress

Online publication: Jul 23, 2026

1. Research the value of controllable growth and structural characterization of SWCNTs

1.1. Improving the growth theory system of low-dimensional carbon materials

SWCNTs are one-dimensional tubular nanostructures formed by rolling up graphene sheets. Their growth involves a series of reactions, including carbon source cracking, tube wall extension, and growth termination, which together constitute a typical vapor-liquid-solid (VLS) or vapor-solid-solid (VSS) interfacial catalytic reaction system^[1]. Traditional theories of carbon material growth cannot fully explain the microscopic mechanisms at the nanoscale. Research on the controllable growth mechanism effectively fills the gaps in the basic theory of low-dimensional carbon materials. It improves the theoretical framework of nanocatalysis and low-dimensional material growth.

1.2. Unlocking the industrial application potential of SWCNTs

The physicochemical properties of SWCNTs are highly structure-dependent. Metallic SWCNTs exhibit excellent electrical conductivity suitable for high-performance conductive materials, while semiconducting SWCNTs possess superior on/off ratios and carrier mobility. Controllable growth technology enables precise regulation of diameter, chirality, and other parameters, producing high-purity samples with uniform structures that accurately match the requirements of different scenarios^[2]. For example, single-chirality semiconducting SWCNTs can be used to construct ultra-high-speed, low-power transistors in the chip field; high-crystallinity SWCNTs greatly enhance the rate performance and cycle life of lithium batteries and supercapacitors in the energy field. Moreover, precise structural characterization methods significantly improve industrial quality control, enabling rapid detection of structural parameters and ensuring stability in large-scale production.

1.3. Promoting independent innovation of high-end nanomaterials

Low-dimensional carbon materials are a strategic focus of global competition in the new material sector. As a core representative material, the controllable preparation and precise characterization technologies of SWCNTs reflect a country's core competitiveness in nanoscience and technology. Currently, developed countries such as the United States, Japan, and South Korea hold leading positions in the controllable growth of chiral SWCNTs and the high-precision characterization of SWCNTs, gradually monopolizing core technologies and equipment^[3]. Research on the controllable growth mechanism and structural characterization methods for SWCNTs helps overcome the bottleneck in precise structural regulation, establish a characterization technology system with independent intellectual property rights, and enhance China's international competitiveness in carbon nanomaterials. Furthermore, the development of SWCNT-controllable technologies drives the coordinated upgrading of upstream and downstream high-end industries, such as electronic information and new energy, thereby supporting the advancement of China's high-end manufacturing.

2. Research the status of the controllable growth mechanism and structural characterization of SWCNTs

2.1. Research status of controllable growth mechanism

2.1.1. Development and limitations of classical growth models

At present, chemical vapor deposition (CVD) is the most common method for preparing SWCNTs, and the classical VLS and VSS models primarily explain its growth mechanism. The early VLS model proposed that transition-metal catalysts (Fe, Co, Ni, etc.) melt into liquid droplets at high temperatures; carbon-source gases crack on the catalyst surface to form active carbon atoms, which dissolve and diffuse into the droplets. Upon saturation, carbon atoms precipitate and assemble into SWCNTs. This model provides a preliminary explanation of the growth process but fails to adequately account for chiral selectivity and precise diameter regulation^[4]. The VSS model applies to high-melting-point solid catalysts such as W, Mo alloys, and metal carbides. Active carbon atoms adsorb, migrate, and nucleate on the solid catalyst surface, and lattice symmetry matching between catalyst crystal planes and SWCNT walls enables directional structural regulation. In recent years, the screw dislocation theory has improved our understanding of growth kinetics, stating that the SWCNT growth rate correlates positively with the number of active sites at the open end. However, follow-up studies found that this theory applies only to high-etchant-concentration systems, with

limited universality.

2.1.2. Research progress of structure-controllable growth

For diameter control, the academic community has reached a consensus: catalyst particle size is the core factor determining SWCNT diameter. By regulating catalyst particle size, linear controllable adjustment of diameter in the range of 0.8–3 nm can be effectively achieved. For chiral control, no complete breakthroughs have yet been achieved ^[5]. International research teams have improved chiral selectivity through substrate induction and electric field regulation, but 100% chiral purity remains unattainable.

2.1.3. Core existing problems

Currently, challenges in the controllable growth mechanism of SWCNTs in China include: an unclear atomic-scale interaction mechanism at the catalyst-SWCNT interface; unquantified thermodynamic and kinetic thresholds for chiral-selective nucleation; and complex coupling effects among temperature, atmosphere, catalyst state, and other parameters during growth, with no unified theoretical model ^[6].

2.2. Research status of structural characterization methods

SWCNT structural characterization mainly focuses on core parameters such as diameter, chirality, and length. Existing technologies fall into three categories: microscopic imaging, spectral analysis, and electron diffraction. The principles, application scenarios, advantages, and limitations of each technique are summarized in **Table 1**.

Table 1. Comparison of mainstream structural characterization methods for SWCNTs

Characterization method	Core principle	Measurable structural parameters	Core advantages	Main limitations
High-Resolution Transmission Electron Microscopy (HRTEM)	Electron beam penetrates the sample; atomic scattering forms lattice diffraction images	Diameter, tube wall defects, atomic arrangement	Atomic-level resolution, direct observation of microstructure	Complex sample preparation, electron beam damage, not suitable for batch testing
Scanning Tunneling Microscopy (STM)	Detects surface atomic morphology based on the quantum tunneling effect	Chirality index, atomic arrangement, and diameter	Ultra-high chiral accuracy, strong surface structure resolution	Only applicable to conductive samples, slow testing, and a small characterization range
Raman Spectroscopy	Inelastic scattering between photons and carbon atoms; characteristic peaks correspond to structural information	Diameter (RBM peak), defect density (ID/IG), chirality	Fast, non-destructive, batch characterization, easy operation	Limited chiral accuracy, difficult to distinguish similar chiralities
UV-Vis-NIR Absorption Spectroscopy	An electronic transition absorbs specific wavelengths that correspond to the electronic structure.	Diameter, metallic/semiconducting ratio, purity	Non-destructive, fast, quantitative analysis of the conductivity type	Cannot directly measure chirality, highly sensitive to high-purity samples
Nanobeam Electron Diffraction (NBED)	Electron beam diffraction forms characteristic patterns; lattice structure analysis.	Chirality index, lattice symmetry	High chiral accuracy, reliable results	Expensive equipment, complex testing, single-tube characterization only

2.2.1. Application progress of mainstream characterization technologies

Microscopic imaging is the core method for atomic-level structural characterization. HRTEM clearly observes tube wall defects and diameter uniformity, and can be combined with NBED for precise chirality index measurement. STM directly analyzes chirality through atomic-scale imaging, serving as an important standard for chiral characterization, but its low efficiency limits its use to basic laboratory mechanistic research^[7]. Spectral analysis is the preferred choice for industrial characterization. Raman spectroscopy quickly determines diameter distribution via the inverse relationship between radial breathing mode (RBM) peaks and diameter. It quantitatively evaluates defect density using the ID/IG ratio, making it the most widely used rapid characterization method. UV-Vis-NIR spectroscopy efficiently distinguishes metallic and semiconducting SWCNTs, enabling quantitative analysis of conductivity-type ratios for rapid batch-sample screening.

2.2.2. Existing problems in the characterization field

Current structural characterization technologies face a trade-off between precision and efficiency. High-precision methods (STM, NBED) are slow and costly, failing to meet the needs of online detection in large-scale production. Meanwhile, rapid methods (Raman and absorption spectroscopy) lack precision, making it difficult to measure chirality accurately and to trace defects^[8]. Additionally, China lacks unified characterization standards, leading to large deviations in test results across laboratories and equipment, with poor data comparability and interoperability.

3. Realization paths for controllable growth and structural characterization of SWCNTs

This section analyzes the controllable growth mechanism from three perspectives: multi-scale theoretical simulation, targeted catalyst optimization, and process parameter regulation. See **Table 2**, which shows the correlation between growth and characterization.

Table 2. Corresponding characterization table

Core chapter	Core technical points	Core functions
Construction of a multi-scale theoretical simulation system	Combine DFT, MD, MLFF to build a multi-scale simulation system; quantify temperature, atmosphere, etc.	Analyze elementary reactions, establish theoretical models, and provide comprehensive theoretical guidance.
Targeted design and optimization of catalysts	Design intermetallic compounds and supported catalysts; regulate crystal planes and particle sizes	Improve chiral selectivity, inhibit catalyst agglomeration, and provide catalytic support.
Multi-parameter synergistic regulation of growth processes	Optimize CVD processes; regulate carbon source/etchant concentrations; develop in-situ real-time technologies.	Reduce defects, improve sample uniformity, and realize closed-loop regulation.

3.1. In-depth analysis paths for controllable growth mechanism

3.1.1. Construction of a multi-scale theoretical simulation system

To achieve controllable growth of SWCNTs, a multi-scale growth simulation system ranging from the atomic to the macroscale can be established by combining density functional theory (DFT) and molecular dynamics (MD). DFT calculations can precisely analyze elementary reaction mechanisms, such as carbon-

source cracking and carbon-atom migration. MD simulations reveal atomic-scale interactions at the catalyst-SWCNT interface, uncovering crystal-plane matching and chiral-selective growth rules^[9]. Furthermore, introducing machine-learning force fields (MLFFs) improves simulation efficiency, enabling large-scale, long-time growth-process simulations. This quantifies the effects of temperature, atmosphere, and other parameters on growth behavior, establishing a precise thermodynamic-kinetic coupled theoretical model to guide controllable growth.

3.1.2. Targeted design and optimization of catalysts

Break through the limitations of traditional transition metal catalysts and develop novel catalysts with precise structures. For example, prepare monodisperse intermetallic compounds with uniform size and regulate catalyst crystal planes and particle sizes to match the lattice symmetry of the target SWCNTs, thereby greatly improving chiral selectivity^[10]. Meanwhile, develop functional supported catalysts, systematically optimize carrier-catalyst interactions, suppress high-temperature agglomeration, and establish more stable active growth sites.

3.1.3. Multi-parameter synergistic regulation of growth processes

Based on the carbon concentration regulation kinetic model and screw dislocation theory, systematically optimize CVD growth processes. Precisely control carbon source supply rate and etchant concentration: achieve chiral-selective growth using active site differences in high-etchant systems, and terminate non-target structure growth by regulating carbon source concentration in low-etchant systems^[11]. Optimize reaction temperature and heating rate to balance nucleation thermodynamics and growth kinetics, effectively reducing defect formation. Additionally, develop in situ, real-time growth technologies for dynamic monitoring and closed-loop regulation of growth parameters, thereby significantly improving the structural uniformity of samples.

3.2. Construction paths for precise structural characterization methods

3.2.1. Integration of high-precision and high-efficiency characterization technologies

Establish a two-stage characterization system of “rapid screening + precise calibration”. The first stage uses rapid methods such as Raman and UV-Vis-NIR spectroscopy for preliminary screening of diameter, defect density, and other parameters, quickly eliminating unqualified samples. The second stage employs high-precision techniques (HRTEM-NBED, STM) to accurately measure chirality and atomic structure of high-quality samples, balancing efficiency and precision^[12]. Meanwhile, develop multi-technology-combined characterization platforms that integrate spectroscopy, imaging, and diffraction to enable simultaneous multi-parameter characterization on a single device, thereby simplifying testing processes and improving data reliability.

3.2.2. Intelligent characterization and establishment of standard systems

Introduce machine learning to assist in the in-depth analysis of characterization data. Train deep learning models to automatically identify characteristic peaks in Raman and absorption spectra, rapidly inverting diameter, chirality, and other parameters^[13]. Machine learning algorithms process electron microscopy images to automatically extract key information such as atomic arrangement and chirality index, reducing human error and improving data processing efficiency.

3.2.3. Development of in-situ and online characterization technologies

Targeting research on growth mechanisms and industrial quality control, with a focus on developing in situ and real-time characterization technologies. Use in-situ TEM and in-situ Raman spectroscopy to observe atomic migration, nucleation, and extension during SWCNT growth, verifying mechanisms and revising theoretical models^[14]. Meanwhile, design online characterization modules integrated into CVD equipment to enable real-time monitoring and feedback-based regulation of structural parameters during growth, ensuring structural stability in large-scale production.

3.3. Collaborative development paths of mechanism research and characterization technology

Controllable growth mechanism and structural characterization research are mutually supportive. Precise characterization obtains structural evolution data during growth, which is used to invert mechanisms, revise theoretical models, and provide experimental evidence for growth parameter optimization^[15]. Furthermore, build an industry-university-research collaborative innovation platform that integrates the theoretical advantages of universities and research institutes with the industrial experience of enterprises, effectively transforming laboratory-precise, controllable growth technologies into industrial mass production.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Lu X, Li Z, Dong W, et al., 2025, Progress in Scanning Tunneling Microscopy of Carbon Nanotubes. *Scientia Sinica Chimica*, 55(12): 3373–3392.
- [2] Han C, Wu Q, Xu G, et al., 2025, Preparation of Single-Walled Carbon Nanotubes Using Industrial Fly Ash Supported Ruthenium Catalyst. *Shandong Chemical Industry*, 54(22): 1–5.
- [3] Chi X, He M, 2025, Preparation of Single-Walled Carbon Nanotubes Using MgO Supported Fe-Pb Bimetallic Catalyst. *Shandong Chemical Industry*, 54(22): 49–53.
- [4] Zuo H, Ye T, 2025, Preparation of Large-Area High-Density Horizontal Arrays of Single-Walled Carbon Nanotubes. *Liaoning Chemical Industry*, 54(10): 1682–1685.
- [5] Wu Q, Zhang X, An H, et al., 2025, Selective Growth of (5,4) Single-Walled Carbon Nanotubes. *Science Bulletin*, 70(19): 3116–3119.
- [6] Yang Z, Wang T, Fang H, et al., 2025, Adsorption Characteristics of Single-Walled Carbon Nanotubes for Quinolones in Farm Sewage. *China Animal Health Inspection*, 42(09): 111–117.
- [7] Yang L, Liao J, Ruan H, 2025, Diameter Regulation of Single-Walled Carbon Nanotubes Grown by Novel MgO Supported Y-Fe Bimetallic Catalyst. *Contemporary Chemical Research*, 2025(18): 53–55.
- [8] Luo F, 2025, Modification of Single-Walled Carbon Nanotube Films and Their Terahertz Metasurface Properties, thesis, Xi'an University of Technology.
- [9] Cheng F, 2025, Study on Diameter Regulation of Single-Walled Carbon Nanotubes, thesis, Qingdao University of Science and Technology.
- [10] Fang C, 2025, Application of Single-Walled Carbon Nanotubes in Silicon-Based Anodes of Lithium-Ion Batteries,

thesis, Hubei University.

- [11] Zhu L, 2025, Optimization of Growth Conditions for Single-Walled Carbon Nanotubes Based on Iron Catalyst. *Tianjin Chemical Industry*, 39(S1): 79–83.
- [12] Xiong Z, 2025, Interface Regulation and Thermoelectric Enhancement Mechanism of Polyethyleneimine/Single-Walled Carbon Nanotube Composites, thesis, Wuhan Institute of Technology.
- [13] Wan Y, Zhang J, Han G, et al., 2025, Separation of Semiconducting Single-Walled Carbon Nanotubes by Thiophene-Fluorene Containing Polymers. *Journal of Shanxi University (Natural Science Edition)*, 48(05): 997–1008.
- [14] Duan H, Ge Y, Chen H, et al., 2025, Preparation of Rosin-Based Small Molecule Gels and Dispersion of Single-Walled Carbon Nanotubes. *Biomass Chemical Engineering*, 59(03): 17–24.
- [15] Yang H, 2025, Confined Assembly and Catalytic Properties of Multi-Walled Carbon Nanotubes-Molybdenum-Based Polyoxometalates, thesis, Inner Mongolia University.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on Position-Sensorless Control of PMSM Based on ELO and NLESO-PLL

Wei Liu, Jiawei Cao*

College of Electrical Engineering and Control, Jilin University of Chemical Technology, Jilin 132022, Jilin, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In the context of the position and speed information of the rotor being difficult to directly obtain in the position sensorless control of permanent magnet synchronous motors, this paper establishes a PMSM position sensorless vector control system model and designs a speed and position estimation structure consisting of an extended Luenberger observer and NLESO-PLL. This method first uses the extended Luenberger observer to estimate the motor's back electromotive force, and then estimates its components in the α and β coordinate systems from the stator voltage and current signals. Then, the estimated back electromotive force is input into NLESO-PLL, and the rotor position, angular velocity, and mechanical speed are estimated through nonlinear error feedback and the expansion state observation loop. Compared with the traditional PLL, NLESO-PLL can enhance dynamic tracking capability and anti-disturbance performance under conditions of sudden speed changes and load disturbances. Using MATLAB/Simulink, a PMSM position sensorless vector control simulation model is built, and speed step, load disturbance, and position estimation experiments are conducted under different operating conditions. The simulation results show that the proposed method can better track the actual speed and rotor position in the medium-to-high-speed operating range and maintain good estimation stability under speed changes and load disturbances, providing a feasible solution for position-sensorless control of PMSM at medium and high speeds.

Keywords: Permanent magnet synchronous motor; Position sensorless control; Extended Luenberger observer; Nonlinear extended state observer; Phase-locked loop

Online publication: Jul 23, 2026

1. Introduction

Permanent magnet synchronous motors offer high power density, high operating efficiency, fast torque response, and good speed regulation. They are widely used in electric vehicles, industrial servo systems, robots, and numerical control equipment. In high-performance vector control systems, rotor position and speed information are important bases for coordinate transformation, current decoupling control, and speed closed-loop regulation. Traditional control systems usually obtain rotor position information from

encoders, rotary transformers, or Hall sensors. Still, mechanical sensors increase system cost and installation complexity and may reduce system reliability in high-temperature, vibration, dust, and strong electromagnetic interference environments. Therefore, the position-sensorless control method that estimates rotor position and speed from measurable signals, such as motor terminal voltage and current, has high research significance ^[1].

At present, the position sensorless control methods for permanent magnet synchronous motors mainly include the back electromotive force estimation method, the flux linkage observation method, the sliding mode observer method, the model reference adaptive method, the Kalman filter method, and the high-frequency signal injection method, etc. Among them, the back electromotive force-based methods utilize the back electromotive force information generated during the motor operation to estimate the rotor position. These methods have relatively clear structures, are easy to integrate with the motor mathematical model and vector control system, and are well-suited to the medium- and high-speed operating range ^[2]. In back electromotive force-based estimation methods, the extended Luenberger observer can uniformly observe the stator current and back electromotive force as extended states, and correct the state estimate through current-observation-error feedback ^[3]. It has the characteristics of a clear model form, a simple structure, and convenient engineering implementation. However, the observed reverse electromotive force still requires position extraction to determine the rotor angle and speed. The traditional phase-locked loop is simple in structure. Still, its dynamic response speed and anti-interference capability are easily affected by changes in speed, load disturbances, or fluctuations in reverse electromotive force estimation. This can lead to overshoot, lag, or steady-state fluctuations in the estimated speed. Especially when the motor operating state changes rapidly, it is difficult to achieve both speed tracking and smooth estimation simultaneously. Therefore, it is necessary to improve the back-end phase-locked loop structure by incorporating reverse electromotive force observations to enhance the dynamic performance and anti-interference capability of rotor position and speed estimation.

To address the aforementioned issues, this paper establishes a vector control system model for permanent magnet synchronous motors without position sensors. It employs an extended Luenberger observer to estimate the component of the reverse electromotive force in the two-phase stationary coordinate system. It combines this with a nonlinear expanded-state observer phase-locked loop to extract the rotor position and speed. This method uses the estimated reverse electromotive force output by the front-end extended Luenberger observer as input, constructs the phase-locked adjustment quantity via nonlinear error feedback, and compensates for disturbances and dynamic changes using the expanded state observer loop, thereby improving the estimation performance of the traditional PLL under variable-speed and disturbance conditions.

2. Mathematical model and control system of PMSM

To facilitate subsequent observer design, this paper selects the surface-mounted permanent magnet synchronous motor as the research object. Based on the equivalent circuit of the motor and under the condition of ignoring the effects of magnetic saturation, iron loss, and higher-order harmonics, the stator voltage equation of the PMSM can be established in the three-phase stationary coordinate system:

$$\begin{bmatrix} u_a \\ u_b \\ u_c \end{bmatrix} = \begin{bmatrix} R_s & 0 & 0 \\ 0 & R_s & 0 \\ 0 & 0 & R_s \end{bmatrix} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \psi_a \\ \psi_b \\ \psi_c \end{bmatrix} \quad (1)$$

In the formula, u_a, u_b, u_c For the three-phase voltage of the stator, i_a, i_b, i_c For the three-phase current of the stator, R_s For stator phase resistance, ψ_a, ψ_b, ψ_c The three-phase stator flux components for surface-mounted permanent magnet synchronous motors are usually approximated. $L_d = L_q = L_s$. After the Clarke transformation, the three-phase stationary coordinate system can be converted into a two-phase stationary coordinate system. At this point, the stator voltage equation of the PMSM in the $\alpha\beta$ coordinate system can be expressed as:

$$\begin{cases} u_\alpha = R_s i_\alpha + L_s \frac{di_\alpha}{dt} + e_\alpha \\ u_\beta = R_s i_\beta + L_s \frac{di_\beta}{dt} + e_\beta \end{cases} \quad (2)$$

Among them, u_a, u_b For the stator voltage components in the two-phase stationary coordinate system, i_a, i_b For the stator current component; e_a, e_b For the anti-electromotive force component. The relationship between the reverse electromotive force and the rotor position and rotational speed is:

$$\begin{cases} e_\alpha = -\omega_e \psi_f \sin \theta_e \\ e_\beta = \omega_e \psi_f \cos \theta_e \end{cases} \quad (3)$$

The component of the reverse electromotive force contains information about the rotor position and speed, so it is possible to obtain this information through the analysis of e_a, e_b . The estimation is obtained indirectly to determine the rotor state. To facilitate the subsequent expansion of the design of the Longberg observer, Equation (2-2) is rewritten as the current state equation:

$$\begin{cases} \frac{di_\alpha}{dt} = -\frac{R_s}{L_s} i_\alpha - \frac{1}{L_s} e_\alpha + \frac{1}{L_s} u_\alpha \\ \frac{di_\beta}{dt} = -\frac{R_s}{L_s} i_\beta - \frac{1}{L_s} e_\beta + \frac{1}{L_s} u_\beta \end{cases} \quad (4)$$

3. PMSM position sensorless observer

3.1. Basic principles of the Loberg observer

In the motor control system, not all state quantities can be directly measured by sensors. To obtain the system's internal state from limited measurable signals, a state observer is usually required. The Luenberger observer is a state-estimation method based on the system's state-space model [3]. Its core idea is to dynamically reconstruct the unmeasurable state by using the system input, measurable output, and output estimation error. For general linear systems, their state space expression can be written as:

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases} \quad (5)$$

In the formula, x . As a system state variable, u . For system input, y . For system output, A, B, C . They are, respectively, the state matrix, the input matrix, and the output matrix. When the system status x When it is not possible to measure everything directly, the following form of Lorch observer can be constructed;

$$\begin{cases} \dot{\hat{x}} = Ax + Bu + L(y - \hat{y}) \\ \hat{y} = Cx \end{cases} \quad (6)$$

In the formula, $\hat{x}$ For the estimated values of the state $\hat{y}$ To output the estimated value, L , for the observer gain matrix, $y - \hat{y}$ To output the observation error. Compared with the original system model, the Lorchberg observer incorporates an error feedback term. $L(y - \hat{y})$ This method can continuously correct state estimates based on the deviation between actual and estimated outputs, thereby improving state estimation accuracy.

Define the state estimation error as:

$$\tilde{x} = x - \hat{x} \quad (7)$$

The observable error dynamic equation can be obtained:

$$\dot{\tilde{x}} = (A - LC)\tilde{x} \quad (8)$$

From Equation (3-4), it can be seen that the error convergence performance of the observer is mainly determined by the eigenvalues of the matrix $A - LC$. By reasonably selecting the observer gain matrix L , the error system can be made stable, and the estimated state can gradually converge to the vicinity of the actual state. Therefore, the Lorchberger observer has the advantages of a clear structure, clear physical meaning, and ease of engineering implementation.

3.2. Extension of the basic principles of the Lorenz observer

In position-independent sensorless control of a PMSM, the rotor position and speed information are contained in the back electromotive force. Still, the back electromotive force cannot be measured directly. Therefore, the back electromotive force can be regarded as an extended state, and together with the stator current, it constitutes the extended state variables:

$$x = \begin{bmatrix} i_\alpha & i_\beta & e_\alpha & e_\beta \end{bmatrix}^T \quad (9)$$

The input variables and output variables are selected as follows:

$$u = \begin{bmatrix} u_\alpha & u_\beta \end{bmatrix}^T, \quad y = \begin{bmatrix} i_\alpha & i_\beta \end{bmatrix}^T \quad (10)$$

By combining the voltage equation of PMSM in the two-phase stationary coordinate system and the dynamic relationship of the back electromotive force, the extended state equation can be obtained:

$$\begin{cases} \frac{di_\alpha}{dt} = -\frac{R_s}{L_s}i_\alpha - \frac{1}{L_s}e_\alpha + \frac{1}{L_s}u_\alpha \\ \frac{di_\beta}{dt} = -\frac{R_s}{L_s}i_\beta - \frac{1}{L_s}e_\beta + \frac{1}{L_s}u_\beta \\ \frac{de_\alpha}{dt} = -\omega_e e_\beta \\ \frac{de_\beta}{dt} = \omega_e e_\alpha \end{cases} \quad (11)$$

When expressing Equation (3-7) in matrix form, the system matrix can be represented as:

$$A = \begin{bmatrix} -\frac{R_s}{L_s} & 0 & -\frac{1}{L_s} & 0 \\ 0 & -\frac{R_s}{L_s} & 0 & -\frac{1}{L_s} \\ 0 & 0 & 0 & -\omega_e \\ 0 & 0 & \omega_e & 0 \end{bmatrix} \quad (12)$$

$$B = \begin{bmatrix} \frac{1}{L_s} & 0 \\ 0 & \frac{1}{L_s} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (13)$$

Based on the above extended state model, an extended Luenberger observer for PMSM can be constructed. Let the state variables of the observer be:

$$\hat{x} = [\hat{i}_\alpha \quad \hat{i}_\beta \quad \hat{e}_\alpha \quad \hat{e}_\beta]^T \quad (14)$$

Then the specific expression of the observer is:

$$\begin{cases} \frac{d\hat{i}_\alpha}{dt} = -\frac{R_s}{L_s}\hat{i}_\alpha - \frac{1}{L_s}\hat{e}_\alpha + \frac{1}{L_s}u_\alpha + k_i(i_\alpha - \hat{i}_\alpha) \\ \frac{d\hat{i}_\beta}{dt} = -\frac{R_s}{L_s}\hat{i}_\beta - \frac{1}{L_s}\hat{e}_\beta + \frac{1}{L_s}u_\beta + k_i(i_\beta - \hat{i}_\beta) \\ \frac{d\hat{e}_\alpha}{dt} = -\hat{\omega}_e\hat{e}_\beta + k_e(i_\alpha - \hat{i}_\alpha) \\ \frac{d\hat{e}_\beta}{dt} = \hat{\omega}_e\hat{e}_\alpha + k_e(i_\beta - \hat{i}_\beta) \end{cases} \quad (15)$$

In the formula, $\hat{i}_\alpha, \hat{i}_\beta$ As an estimate of the stator current. $\hat{e}_\alpha, \hat{e}_\beta$ For the estimated value of the reverse electromotive force. k_i As the feedback gain for the current error. k_e For the feedback gain of the anti-electromotive force error, in this paper, the extended Luenberger observer adopts a fixed gain structure, and the simulation parameters are taken as $k_i=1000$ 、 $k_e=10000$. Due to the actual angular electric velocity ω_e It is difficult to obtain directly. In the observer, the estimated angular velocity is used. $\hat{\omega}_e$ Participate in the dynamic calculation of the reverse electromotive force. See **Figure 1**.

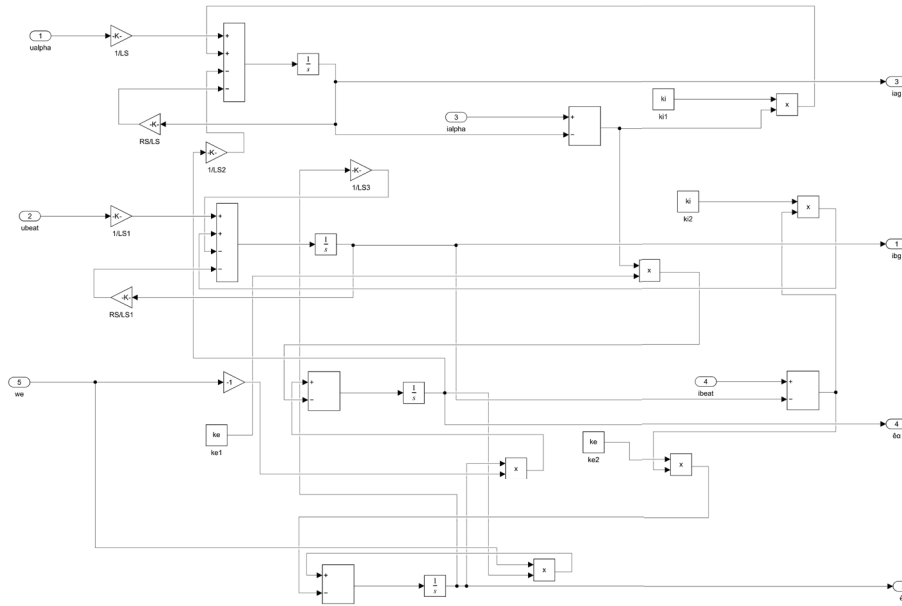


Figure 1. Extended Loberg observer simulation model.

4. Position and speed extraction method based on NLESO-PLL

4.1. Principle of traditional PLL (Phase-Locked Loop)

In the PMSM position-sensorless control system, the extended Luenberger observer can estimate the armature-voltage component in the two-phase stationary coordinate system. However, the armature voltage itself cannot directly serve as the rotor position and speed signals required for vector control. Therefore, an additional position-extraction section needs to be designed to extract the rotor position angle and speed from the estimated armature-voltage component. The phase-locked loop is simple in structure, easy to implement, and a widely used method for extracting the armature voltage position in a PMSM. According to the relationship of the PMSM armature voltage, the estimated armature voltage can be expressed as:

$$\begin{cases} \hat{e}_\alpha = -\hat{\omega}_e \psi_f \sin \hat{\theta}_e \\ \hat{e}_\beta = \hat{\omega}_e \psi_f \cos \hat{\theta}_e \end{cases} \quad (16)$$

A traditional PLL typically constructs the phase error signal based on the estimated reverse electromotive force:

$$\varepsilon = -\hat{e}_\alpha \cos \hat{\theta}_e - e_\beta \sin \theta_e \quad (17)$$

When the estimated angle matches the actual rotor angle, the phase-locked error approaches zero; when there is a deviation between the two, the regulator corrects the error signal to estimate the angular velocity, and the estimated position angle is obtained by integrating the estimated angular velocity. The traditional PI-type PLL can be expressed as:

$$\begin{cases} \dot{\xi} = \varepsilon \\ \hat{\omega}_e = K_p \varepsilon + K_i \xi \\ \dot{\hat{\theta}}_e = \hat{\omega}_e \end{cases} \quad (18)$$

In the formula, K_p and K_i represent the proportional and integral coefficients of the PLL, and ξ is the error integral term. The traditional PLL structure is relatively simple, but its dynamic performance mainly depends on the PI parameters. When the motor experiences sudden changes in speed or load, fixed PI parameters struggle to achieve both fast tracking and stable smoothness simultaneously. This often leads to problems such as speed estimation lag, overshoot, or increased fluctuations.

4.2. NLESO-PLL design

To enhance the tracking capability and anti-disturbance performance of the position and speed extraction section under dynamic conditions, this paper introduces a nonlinear expanded state observation approach based on the traditional PLL and constructs NLESO-PLL^[4]. This design draws on ADRC's extended-state disturbance estimation idea and adaptive NLESO-based PMSM sensorless speed-control research^[5]. Recent ESO-based PLL sensorless-control research also provides a comparable implementation reference^[6]. Reviews of PMSM position-sensorless control technology also identify observer-based compensation and PLL-based extraction as common technical routes^[7]. This method uses the reverse electromotive force estimated by the front-end extended Luenberger observer as the input, and the phase-locked error is still constructed using Equation (4-2). To mitigate the significant impact of fluctuations in the estimated reverse electromotive force on the phase-locked loop, amplitude limiting or filtering can be applied to the error signal in the actual implementation, thereby yielding the error input for the NLESO-PLL. The state equation of NLESO-PLL in this paper is designed as $\hat{e}_a, e$

$$\begin{cases} \dot{\hat{\theta}}_e = \hat{\omega}_e + \beta_1 fal(\varepsilon) \\ \dot{\hat{\omega}}_e = \hat{z} + \beta_2 fal(\varepsilon) \\ \dot{\hat{z}} = \beta_3 fal(\varepsilon) - \lambda_z z \end{cases} \quad (19)$$

In the formula, To estimate the rotor electrical angle, To estimate the angular velocity of electricity, To expand the state estimator, it is used to characterize the comprehensive disturbance and dynamic compensation terms during the phaselocking process. For the observer's gain, To expand the state attenuation coefficient. The NLESO-PLL not only utilizes the phase-locked error to correct the position and velocity estimation, but also introduces the expanded state to compensate for the effects caused by velocity jumps, load disturbances, and fluctuations in the estimated back electromotive force. Compared with the traditional PI-type PLL, this structure can enhance the dynamic regulation capability and reduce the estimation fluctuations under the influence of disturbances. $\hat{\theta}_e, \hat{\omega}_e, \hat{z}, \beta_1, \beta_2, \beta_3, \lambda_z$

The nonlinear error-feedback function in NLESO-PLL is employed. fal Function is defined as:

$$fal(\varepsilon, \alpha, \delta) = \begin{cases} \frac{\varepsilon}{\delta^{1-\alpha}}, & |\varepsilon| \leq \delta \\ |\varepsilon|^\alpha sign(\varepsilon), & |\varepsilon| > \delta \end{cases} \quad (20)$$

In the formula, $\text{fal}(\cdot)$ is the nonlinear feedback function, α is the nonlinear factor, and δ is the width of the linear interval. When the error is relatively small, the function is approximately linear, which helps reduce steady-state jitter; when the error is large, the function shows nonlinear feedback characteristics and improves dynamic correction. After this function is introduced into the PLL structure, NLESO-PLL can improve tracking performance under dynamic speed variation and load disturbance.

Output $\hat{\omega}_e$ Estimated angular electric speed. To obtain the mechanical speed, use the relationship between the electrical angular speed and the mechanical speed. Conduct the conversion:

$$\hat{n}_{rpm} = \frac{60}{2\pi p_n} \hat{\omega}_e \quad (21)$$

In the formula, p_n represents the number of pole pairs of the motor. $\hat{n}_{rpm}$ To estimate the rotational speed of the machine, use r/min.

In conclusion, the position and speed estimation process adopted in this paper is as follows: First, the extended Luenberger observer estimates the armature reaction voltage component from the stator voltage and current signals. $\hat{e}_\alpha, \hat{e}_\beta$ Then, NLESO-PLL constructs the phase-locked error based on the estimated rotor flux vector; finally, the estimated position is obtained through the nonlinear extended state observation structure. $\hat{\theta}_e$ Estimate the angular electric velocity $\hat{\omega}_e$. And estimating the rotational speed of the machinery $\hat{n}_{rpm}$. See **Figures 2 and 3**.

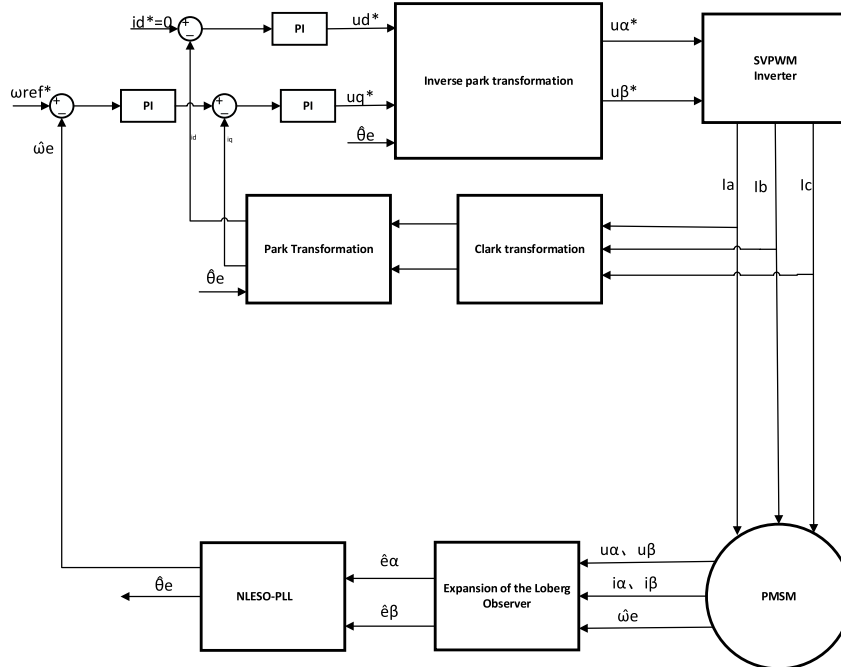


Figure 2. Structure diagram of PMSM position sensorless control system based on ELO and NLESO-PLL.

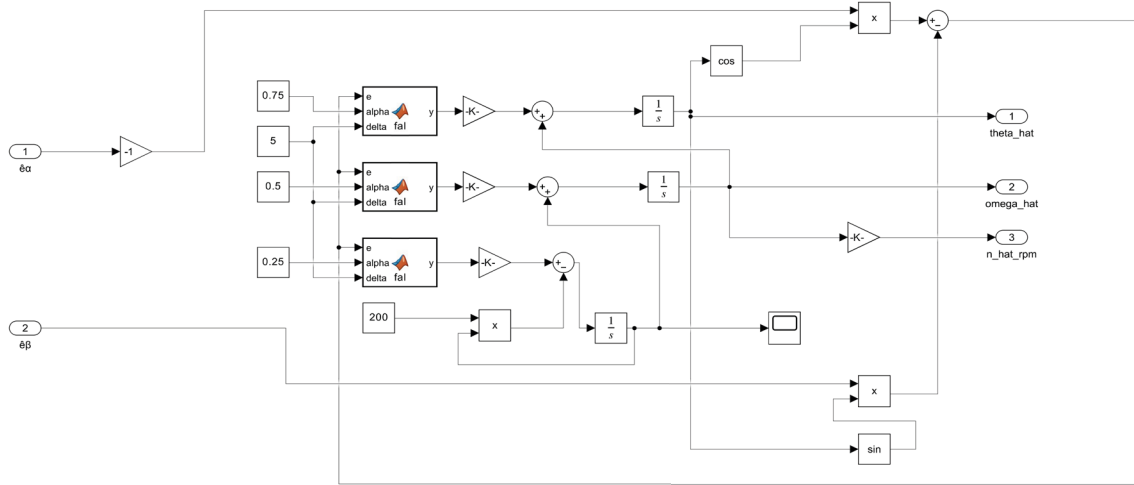


Figure 3. Simulation model of the NLESO-PLL structure.

5. Simulation results and analysis

5.1. Simulation model and parameter settings

To verify the effectiveness of the extended Lorch observer and the NLESO-PLL position- and speed-extraction method proposed in this paper, a PMSM position-sensorless vector-control simulation model was built in MATLAB/Simulink. The system uses the extended Lorch observer as the front-end back-electromotive-force observation link and outputs the estimated back electromotive force, following the same observer-based modeling direction as adaptive Luenberger-observer schemes^[8]. The backend uses NLESO–PLL to extract the position and speed of the reverse electromotive force, and outputs the estimated rotor position. Estimate the electrical angular velocity. And estimate the mechanical rotational speed.. $\hat{e}_\alpha, \hat{e}_\beta, \hat{\theta}_e, \hat{\omega}_e, \hat{n}_{rpm}$

This paper presents a comprehensive test that combines speed step and load disturbance. The simulation duration is set to 10 seconds. The given rotational speeds are 700 r/min, 1500 r/min, 2500 r/min, and -500 r/min, and a $T = 6$ N.m operation is applied at 3.5 seconds. The load is reduced to $T = 3$ N.m at 6.5 seconds to verify the proposed method's estimation performance under medium-to-high-speed shifting and disturbance conditions.

5.2. Comprehensive test of speed and load

Figure 4 presents the comparison results of the actual rotational speed and the observed rotational speed under the combined conditions of speed step and load disturbance.

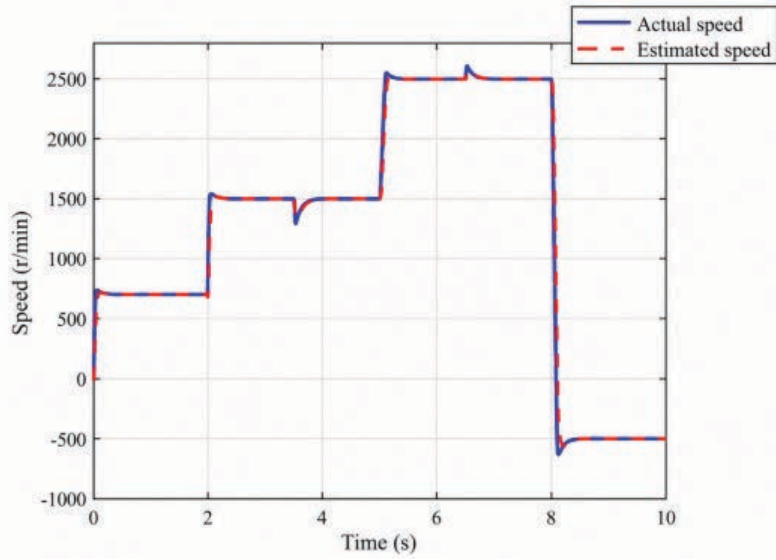


Figure 4. Comprehensive test results of speed step change and load disturbance.

As shown in **Figure 4**, under the operating conditions of 700 r/min, 1500 r/min, 2500 r/min, and -500 r/min, the observed speed closely tracks the actual speed, and the overall trend is consistent. When the given speed changes, the estimated speed dynamically adjusts to the actual speed and reaches a stable state after a brief transition. This result indicates that the extended Lorchberg observer can provide an effective estimate of the back electromotive force for the rear-end phase-locked loop, and the NLESO-PLL can extract speed from this information, which is consistent with improved PLL-based position and speed extraction methods^[9].

To further examine the dynamic tracking effect during speed variation, a detailed analysis was conducted of the speed-step stage in the comprehensive test. **Figure 5** presents the rotational speed response curve during the speed step stage.

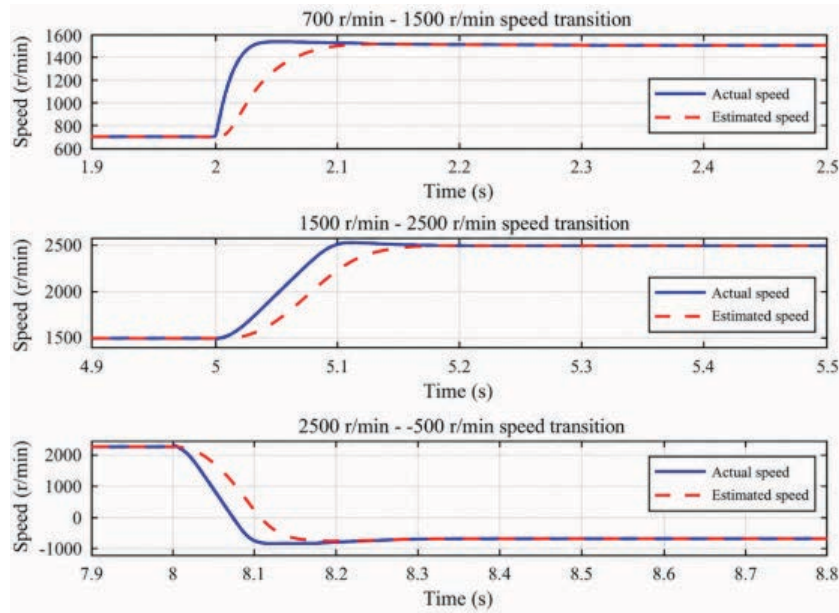


Figure 5. Local plot of rotational speed response during the speed step-up stage.

As shown in **Figure 5**, when the rotational speed is changed from 700 r/min to 1500 r/min and then to 2500 r/min, it reaches a steady state within approximately 0.1–0.2 seconds. The observed rotational speed can quickly follow changes in the actual rotational speed, and no obvious divergence occurred during the dynamic process. When the motor is switched from forward to reverse at -500 r/min, the observed rotational speed completes the direction change within 0.1–0.2 seconds. Due to significant changes in operating state during reverse switching, the estimation results may briefly fluctuate but will subsequently converge again. This indicates that the proposed method has good dynamic tracking performance under sudden speed changes.

To verify the anti-disturbance performance of the proposed method, load disturbances were applied near the 3.5-second and 6.5-second steady-state points, and a local amplification analysis was conducted during the disturbance stage. **Figure 6** presents the speed response curve during the load disturbance stage.

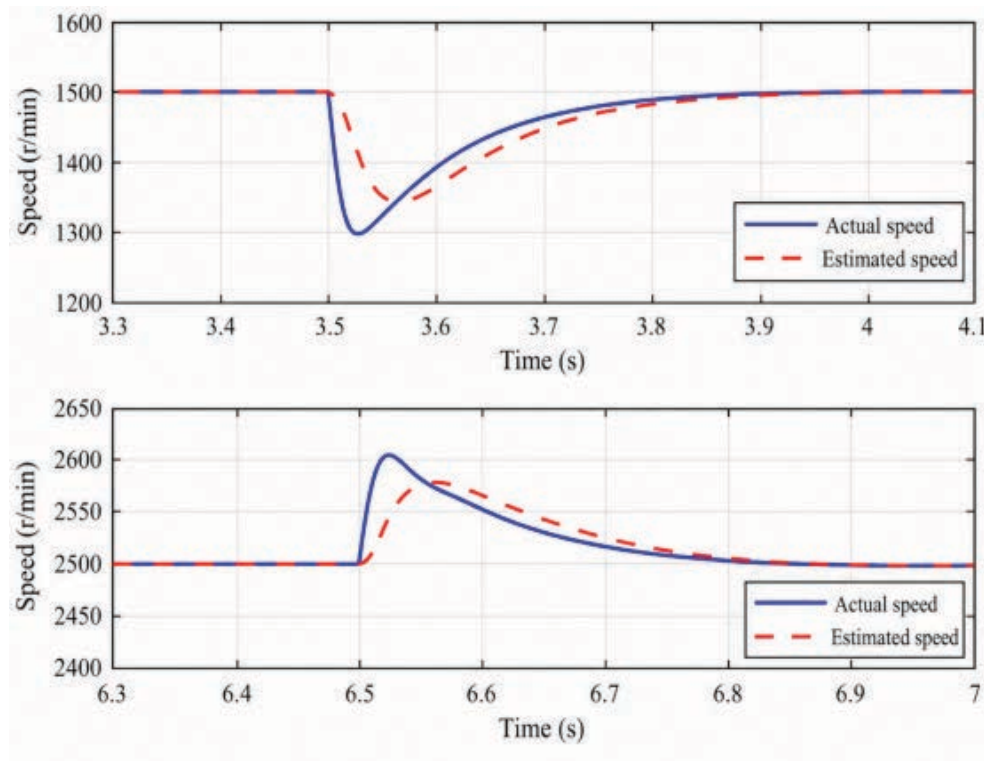


Figure 6. Local plot of rotational speed response during the load disturbance stage.

As shown in **Figure 6**, after the load disturbance is applied, the actual speed shows a short-term fluctuation, and the observed speed changes accordingly. It takes approximately 0.3 to 0.4 seconds to re-track the actual speed. The expanded state in NLESO-PLL can play a certain compensatory role in the dynamic changes caused by the disturbance, preventing the estimated speed from continuously deviating or experiencing significant oscillations; this disturbance-compensation idea is also reflected in LESO phase-lead correction research^[10]. This result indicates that the proposed method has good estimation stability and anti-disturbance capability under medium- and high-speed load-disturbance conditions.

To verify the effectiveness of the position estimation, four typical rotational speed points of 700 r/min, 1500 r/min, 2500 r/min, and -500 r/min were selected, and the actual rotor position and the estimated rotor

position were compared. **Figure 7** shows the positions at different rotational speed points. Estimated result:

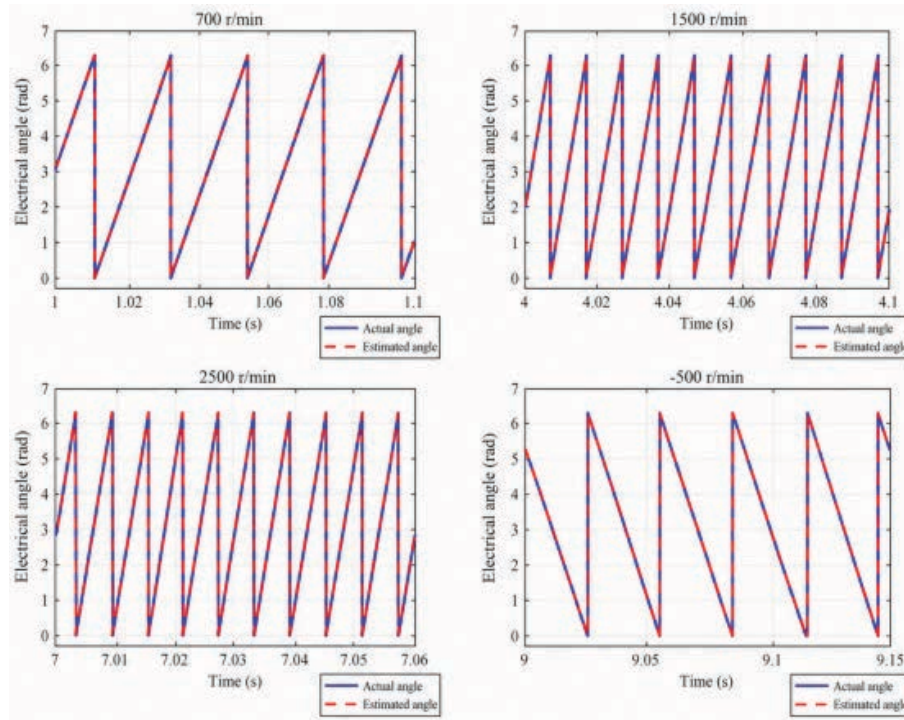


Figure 7. Comparison of rotor positions at different rotational speeds.

As shown in **Figure 7**, under both forward and reverse operating conditions, the estimated position closely follows the actual rotor position, and the two exhibit essentially the same trend.

6. Conclusion

This paper extends the application of the Longberg observer and NLESO-PLL to the position-sensorless vector control system for permanent magnet synchronous motors. Based on MATLAB/Simulink, a simulation model for PMSM position sensorless control was built. The simulation results show that the proposed method effectively achieves speed tracking and rotor position estimation under varying operating conditions, including different speeds and loads. During speed step changes and load disturbances, the observed speed can follow the actual speed changes, and the estimation results are overall stable. The extended Longberg observer can effectively estimate the armature reaction voltage component. At the same time, NLESO-PLL can extract the rotor position and speed from the armature reaction voltage information, thereby improving the system's estimation stability under dynamic changes and disturbances. Overall, the ELO + NLESO-PLL method adopted in this paper is suitable for high-speed, position-sensorless control of PMSMs and can serve as a reference for designing position-sensorless control strategies for permanent magnet synchronous motors.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Wang G, Valla M, Solsona J, 2020, Position Sensorless Permanent Magnet Synchronous Machine Drives: A Review. *IEEE Transactions on Industrial Electronics*, 67(7): 5830–5842.
- [2] Chen Z, Tomita M, Doki S, Okuma S, 2003, An Extended Electromotive Force Model for Sensorless Control of Interior Permanent-Magnet Synchronous Motors. *IEEE Transactions on Industrial Electronics*, 50(2): 288–295.
- [3] Luenberger D, 1971, An Introduction to Observers. *IEEE Transactions on Automatic Control*, 16(6): 596–602.
- [4] Han J, 2009, From PID to Active Disturbance Rejection Control. *IEEE Transactions on Industrial Electronics*, 56(3): 900–906.
- [5] Gan M, Wang C, 2015, An Adaptive Nonlinear Extended State Observer for the Sensorless Speed Control of a PMSM. *Mathematical Problems in Engineering*, 2015: 807615.
- [6] Zhang X, Zhu J, Zheng Z, et al., 2025, Sensorless Control of PMSM Based on Fifth-Order Generalized Integral Flux Observer and Extended State Observer-Based Phase-Locked Loop. *IEEE Transactions on Power Electronics*, 40(4): 5080–5093.
- [7] Liu J, Xiao F, Shen Y, et al., 2017, Review on Position-Sensorless Control Technology of Permanent Magnet Synchronous Motors. *Transactions of the China Electrotechnical Society*, 32(16): 76–88.
- [8] Xiao P, Zhao S, Qiu X, 2023, Research on Position-Sensorless PMSM Based on Adaptive Lobel Observer. *Micro and Mini Motors*, 51(6): 45–50.
- [9] Wu X, Chen S, Li J, et al., 2024, Position-Sensorless Control of Permanent Magnet Synchronous Motors Based on Improved Orthogonal Phase-Locked Loop. *Transactions of the China Electrotechnical Society*, 39(2): 475–486.
- [10] Ma Y, Zhao C, 2025, Research on Position-Sensorless Control of PMSM Based on Phase Lead Correction LESO. *Electrical Machines and Control Applications*, 52(7): 778–787.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on Multi-feature Fusion Method for Software Defect Prediction Based on Industrial Internet of Things

Jinpeng Xu

The Third Research Institute of the Ministry of Public Security, Shanghai 200031, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the rapid development of Industrial Internet of Things technology, the scale of its software is expanding, the complexity of the system continues to rise, and the problem of software defects has become increasingly prominent. Software defect prediction technology can locate potential defects in advance and improve software reliability. However, most of the traditional software defect prediction methods rely on a single code metric feature, which makes it difficult to fully characterize the complex characteristics of Industrial Internet of Things software in the semantic information, program structure, and software evolution process, resulting in limited prediction performance. In view of the above problems, this paper focuses on the research of multi-feature fusion in software defect prediction in Industrial Internet of Things scenarios, focusing on the analysis of the role of different types of features in defect prediction and their fusion mechanism. Firstly, the features of code metrics, semantic features, and structure features involved in Industrial Internet of Things software defect prediction are analyzed. Secondly, the influence of different feature fusion methods on prediction performance is studied, including feature concatenation, weighted feature concatenation, attention fusion, and gating fusion. The research results have a certain reference value for Industrial Internet of Things software quality assurance and intelligent defect analysis.

Keywords: Industrial internet of things; Software defect prediction; Multi-feature fusion

Online publication: Jul 23, 2026

1. Introduction

Software defect prediction is a method to predict potential defect modules by analyzing historical software data, which can help developers improve the efficiency of software testing and repair, thereby reducing the cost of software maintenance. Traditional software defect prediction methods mainly rely on static code metrics such as lines of code, cyclomatic complexity, and coupling degree, and use machine learning models such as random forest and support vector machine to classify and predict defects^[1].

In recent years, with the development of deep learning and program analysis technology, researchers

have begun to pay attention to source code semantic information and program structure information. For example, the abstract syntax tree is used to obtain the syntax characteristics of the code ^[2]. The control flow graph is used to describe the program execution logic, and the program dependence graph is used to reflect the data dependence relationship ^[3]. At the same time, a large number of commit records and modification histories generated in the process of software version evolution can also reflect the law of software stability and defect evolution, which can improve the prediction performance of the model ^[4].

However, relying only on a single code metric feature has limited performance in Industrial Internet of Things defect prediction. Multiple features integrate the advantages of different features, represent the source code in multiple dimensions, and improve the model performance ^[5,6]. Therefore, this paper focuses on the research of multi-feature fusion for software defect prediction in Industrial Internet of Things scenarios, focusing on the analysis of the characteristics of different types of features and their fusion strategies, and the discussion of the impact of multi-feature fusion on the performance of defect prediction, which provides a reference for Industrial Internet of Things software quality assurance.

2. Multi-feature fusion

2.1. Multi-feature

(1) Traditional feature

Through the static analysis of the software source code, extract its structure, scale, complexity, software dynamic behavior, and other quantitative indicators to form features, focusing on the dynamic data of the software product itself and the development process ^[7]. Its core logic is: the complexity of the code structure, size, design rationality, and the dynamic behavior in the development process are closely related to the probability of the existence of defects. The more complex the code, the more chaotic the structure, and the more frequent the code changes, the easier it is to hide design defects and logic errors.

(2) Semantic feature

Semantic features mainly analyze the lexical information, context, and program logic meaning in the source code to vectorize the semantic information of the code. Unlike traditional features, which focus on code size and complexity, semantic features have more advantages in identifying similar code fragments ^[8]. Semantic features usually use deep learning techniques to extract semantic information from the sequence of nodes in the abstract syntax tree. Semantic features can effectively mine the hidden defect information, which is difficult to characterize by traditional metric features, so as to enhance the recognition ability of the model for complex defect patterns.

(3) Structural feature

The overall structural behavior of the software program is modeled mainly by analyzing the control relationship, data dependency, and module call relationship within the program. Software defects are not only related to the content of the code itself, but also closely related to the execution process of the program, the degree of module coupling, and the data transmission path ^[9]. Complex control branches, deep loop nesting, exception jumps, and highly coupled call structures in programs are often more likely to lead to logic errors and potential defects. By constructing a control flow graph, a program dependence graph, and a function call graph, the interaction between program behavior and modules can be described more comprehensively, which enhances the model's ability to perceive the complexity of program logic and the characteristics of defect propagation.

2.2. Feature fusion method

Because different types of features can describe software characteristics from different perspectives, multi-feature fusion has gradually become an important research direction of software defect prediction. Traditional code metrics features can reflect the complexity and scale information of software, semantic features can describe the functional logic of the program, and structural features can describe the control relationship and dependence relationship of the program. Different features are strongly complementary to each other, and it is often difficult to fully describe the defect patterns in complex software systems only by relying on a single feature. Therefore, it is of great significance to study the fusion method of multiple features for improving the performance of software defect prediction^[10].

Currently, feature fusion methods in software defect prediction mainly include feature splicing, feature parameter splicing, attention fusion, and gating fusion, as shown in **Figure 1**.

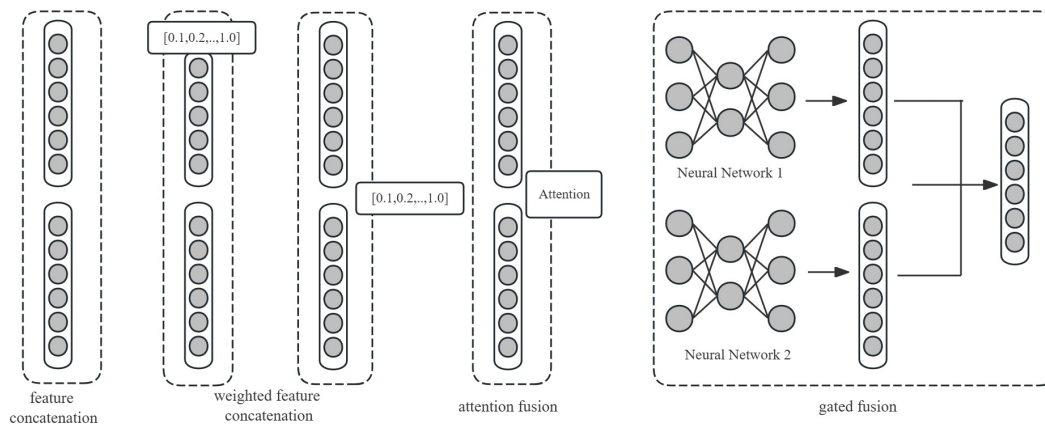


Figure 1. Framework diagram of the multi-feature fusion method.

2.2.1. Feature concatenation

Feature concatenation is a common multi-source feature fusion method, whose core idea is to directly connect feature vectors from different sources according to their dimensions to form a unified high-dimensional feature representation. The feature concatenation method is widely used in traditional machine learning and deep learning models because of its simple implementation, low computational overhead, and easy scalability.

However, this method also has some limitations. Due to the great differences in data distribution, dimension scale, and semantic space of different feature sources, direct stitching can easily lead to feature redundancy and noise accumulation. When the feature dimension is too high, it may also lead to “dimension disaster” and reduce the generalization ability of the model. In addition, by default, different features have the same importance, and the importance of key features can not be learned dynamically, so the feature expression ability is insufficient in complex Industrial Internet of Things software scenarios.

2.2.2. Attention fusion

Attention fusion dynamically learns the importance weights of different features by introducing an attention mechanism, so as to enhance the contribution ability of key features to the prediction results of the model.

Compared with the traditional feature stitching method, attention fusion can adaptively adjust the influence degree of different features according to the feature distribution of the current input sample, so it has a stronger feature expression ability.

Different software modules have different dependencies on various features in the process of defect prediction. For example, for complex control logic modules, structural features may be more important; For code fragments with dangerous function calls, semantic characteristics can be even more critical. Therefore, the importance of different features needs to be dynamically evaluated. Attention fusion can highlight the key defect features, reduce the interference of irrelevant features, and improve the recognition ability of the model for complex defect patterns.

However, there are also some problems with attention fusion. When the training data size is small, the attention mechanism may find it difficult to learn stable weights, resulting in unstable model training. In addition, the attention model has many parameters, and its computational overhead is relatively high in resource-constrained environments such as Industrial Internet of Things edge devices.

2.2.3. Weighted feature concatenation

Weighted feature concatenation is a further improvement based on the traditional feature concatenation method. Its core idea is to introduce the weight adjustment parameter in the feature fusion stage, and give different importance to different types of features or feature dimensions. The weight parameter is usually defined between 0 and 1 to indicate the importance of the corresponding feature to the final fusion representation. The weight can be set according to domain prior knowledge, experimental experience, or model training results, so as to realize the adjustment of different feature contribution degrees.

By assigning different weights or dimensions to different features, weighted feature concatenation can highlight more important feature information for defect prediction and improve the recognition ability of the model for key defect patterns. However, this method still has some limitations. Because the weight parameters are usually adjusted by human experience or experiments, it is difficult to obtain the optimal weight combination automatically. When there are more feature types, the parameter search space will also increase.

2.2.4. Gated fusion

Gated fusion is a dynamic feature fusion method whose core idea is to dynamically control the contribution of different features to the final representation through gating units, so as to achieve more refined feature selection and information fusion.

Unlike the attention mechanism, which mainly emphasizes “which features to pay attention to”, the gating mechanism pays more attention to “how much feature information to retain”. It is essentially similar to the gating structure in recurrent neural networks, which dynamically filters different features by generating gating weights between 0 and 1 through the Sigmoid function.

2.3. Construction of software defect prediction model based on multi-feature fusion

The construction of a software defect prediction model mainly includes data preprocessing, feature extraction, feature fusion, model training, and performance evaluation. Firstly, in the data preprocessing stage, the source code, defect tags, and version data are cleaned and standardized. Subsequently, in the feature extraction stage, multi-source information such as traditional features, semantic features, and structural features is extracted

from the software^[11].

After feature extraction, multi-source features are fused by feature splicing, weighted fusion, or an attention mechanism to enhance the model's ability to represent complex defect patterns. Subsequently, the fused features are input into models such as random forest, support vector machine, convolutional neural network, or long and short-term memory network for training, and the mapping relationship between features and defect labels is learned.

Finally, the performance of the model is evaluated by using the test data set, and the commonly used evaluation indicators include Precision, Recall, F1-score, and AUC. Through the comparative analysis of different models and feature fusion methods, the improvement effect of multi-source feature fusion on software defect prediction performance can be further verified.

3. Data set and evaluation index

In the research on Industrial Internet of Things software defect prediction, typical data sets highly consistent with the research theme are selected, as shown in **Table 1**. **Table 1** covers key indicators such as dataset name, source address, and language involved. **Table 2** shows the performance evaluation indicators commonly used in software defect prediction, which provides a clear and convenient reference for relevant researchers.

Table 1. Common data sets

Serial No.	Dataset	Language is involved	Address
1	PROMISE	Java, C, C++	http://promise.site.uottawa.ca/SERepository/datasets-page.html
2	KAMEI	Java, C, C++	https://research.cs.queensu.ca/~kamei/jittse/jit.zip
3	SIR	Java, PHP, C#, C, C++	https://sir.csc.ncsu.edu/portal
4	NASA	Java, PHP, C#, C, C++	https://github.com/klainfo/NASADefectDataset

Table 2. Evaluation index

Serial No.	Indicator name	Value range	Central role
1	Precision	[0,1]	Measure the actual positive proportion of the predicted positive samples, and reflect the accuracy of the prediction results.
2	Recall	[0,1]	Measure the proportion of the actual positive sample that is successfully predicted, reflecting the ability to check all.
3	F1	[0,1]	The harmonic mean of precision and recall is used to comprehensively evaluate the overall classification effect of the model.
4	MCC	[-1,1]	Adapt to the scene of unbalanced positive and negative samples to comprehensively evaluate classification performance.
5	AUC	[0,1]	Evaluate the overall discrimination ability and generalization performance of the model, based on the ROC curve calculation.

4. Concluding remarks

In the Industrial Internet of Things, the construction of a software defect prediction model is an important means to achieve software quality assurance, and multi-source feature fusion can comprehensively depict the potential defect patterns in software from different perspectives. By integrating traditional features, semantic

features, and structural features, the proposed model can not only enhance the ability to represent complex program behaviors but also improve the defect recognition effect of the model in complex Industrial Internet of Things scenarios. With the development of deep learning, graph neural networks, and pre-training models, software defect prediction models will gradually develop in the direction of intelligence, multi-modal, and high robustness, which will provide more effective technical support for the safe and stable operation of Industrial Internet of Things software.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Deng T, Deng Y, 2025, Review of Development and Application of Software Defect Prediction Technology in Industrial Internet of Things Environment. *Computer Science*, 52(S2): 739–749.
- [2] Guan X, Zhang H, 2025, Research on Software Defect Prediction Technology Based on Deep Learning. *Network Security and Informatization*, 2025(6): 56–58.
- [3] Phan A, Le Nguyen M, Bui L, 2017, Convolutional Neural Networks over Control Flow Graphs for Software Defect Prediction. *Proceedings of the IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, Boston, MA, USA: IEEE, 45–52.
- [4] Choudhary G, Kumar S, Kumar K, et al., 2018, Empirical Analysis of Change Metrics for Software Fault Prediction. *Computers & Electrical Engineering*, 67: 15–24.
- [5] Yang M, Yang S, Wong W, 2024, Multi-Objective Software Defect Prediction via Multi-Source Uncertain Information Fusion and Multi-Task Multi-View Learning. *IEEE Transactions on Software Engineering*, 50(8): 2054–2076.
- [6] Siachos I, Kanakaris N, Karacapilidis N, 2025, Software Bug Prediction Using Graph Neural Networks and Graph-Based Text Representations. *Expert Systems with Applications*, 259: 125290.
- [7] Xu J, Guo X, Wang R, et al., 2023, Aggregation Model for Software Defect Prediction Based on Data Enhancement by GAN. *Computer Science*, 50(12): 24–31.
- [8] Shen J, Liu N, Sun H, et al., 2025, Lightweight Semantic Feature Extraction Model with Direction Awareness for Aerial Traffic Object Detection. *IEEE Transactions on Intelligent Transportation Systems*, 27(4): 4569–4586.
- [9] Liu H, Li Z, Zhang H, et al., 2024, CFG2AT: Control Flow Graph and Graph Attention Network-Based Software Defect Prediction. *IEEE Transactions on Reliability*, 74(3): 3412–3426.
- [10] Qiu S, E B, He J, et al., 2025, Survey of Software Defect Prediction Features. *Neural Computing and Applications*, 37(4): 2113–2144.
- [11] Li J, Zhu Y, Yu Q, et al., 2025, Software Defect Prediction Based on Gated Fusion of Pre-Trained Models. *Proceedings of the 25th International Conference on Software Quality, Reliability, and Security Companion (QRS-C)*, IEEE, 204–213.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on the Key Role of Data Assets in Operation and Maintenance Retrospective Analysis Under Rail Transit Network Security Control

Jianyong Huang

Shenzhen Yongda Electronic Information Co., Ltd., Shenzhen 518063, Guangdong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This paper takes the national railway ticketing system as the research object, and studies the role of data assets in the operation and maintenance retrospective analysis under the network security control situation. Comprehensively sort out the overall scheme of network security operation and maintenance of the railway passenger ticket system, and focus on the significance of data assets with asset accounts as the core in operation and maintenance management, including important links such as asset modeling, status monitoring, log correlation, and fault tracing. Based on this premise, this paper studies the data collection, correlation analysis, and retrospective analysis technology for security operation and maintenance, and explains the supporting significance of the data asset entity model and relationship model to improve the efficiency of fault location and security event analysis. Research and summarize the practical experience of data asset management, operation, and maintenance, and provide a reference for data asset management of other major information infrastructures in network security operation and maintenance.

Keywords: Railway ticketing system; Network security; Data assets; Operation and maintenance traceback

Online publication: Jul 23, 2026

1. Introduction

The national comprehensive three-dimensional transportation network planning outline issued in 2021 focuses on the main significance of transportation infrastructure safety. As an important operation system in the field of rail transit, the network security guarantee ability of the railway passenger ticket system is very important. This study focuses on the network security operation and maintenance of the railway passenger ticket system, focusing on the analysis of the basic framework of network security, the role of data assets in operation and maintenance management, as well as the asset account and dynamic management. Under the circumstances of policy guidance and continuous expansion of system scale, studying the role of data assets

in operation and maintenance retrospective analysis during network security management and control is not only conducive to summarizing the experience of data asset management of railway ticketing system, but also provides reference for operation and maintenance management of large information systems of the same type, and provides ideas for the integrated development of cross domain data governance framework.

2. Basic theories of rail transit network security and data assets

2.1. Basic framework of rail transit network security

The basic network security framework of the railway passenger ticket system includes many important aspects. From the perspective of system structure, the system forms a multi-layer structure of network layer, system layer, and application layer from the center to the station. Each layer depends on the others and operates cooperatively to ensure the stability of the ticket business. The main technical elements include access control, data encryption, intrusion monitoring, and prevention technologies. Access control restricts illegal access by setting rules and permissions to ensure the rational use of resources by authorized subjects; Encryption means to protect ticketing transactions and system data to avoid data leakage and tampering; Intrusion monitoring and prevention methods detect and block abnormal access in time by continuously monitoring network traffic and system activities ^[1]. During the implementation of management and control, a relatively complete network security protection scheme for the railway passenger ticket system has been gradually built by formulating a unified security strategy, clarifying the operation and maintenance responsibilities, strengthening the cultivation of personnel's security awareness, and combining with regular security assessment and detection.

2.2. Core status of data assets in operation and maintenance management

Data assets occupy a core position in the operation and maintenance management of the railway passenger ticket system. The data assets mentioned in this paper, in addition to the operation data of servers, networks, and security devices, also involve the information of operation and maintenance objects such as business system solutions, node software, and hardware configuration, service and process status, and related logs ^[2]. During system operation and maintenance, these data assets together form a comprehensive description of the system operation. On the one hand, the operation data of nodes and processes can reflect the real-time status of the business system and provide support for fault detection and early warning; On the other hand, transaction data and log information can be used to analyze business exceptions and security events, and help the operation and maintenance personnel find the root cause of the problem. Asset accounts and configuration baseline data provide a unified reference standard for operation and maintenance operations to ensure the standardization and controllability of operations. It can be seen that data assets are involved in all aspects of operation and maintenance management of the railway passenger ticket system, which is the main basis for ensuring safe operation and maintenance, and carrying out retrospective analysis.

3. Data processing technology in operation and maintenance retrospective analysis

3.1. Multi-source data collection and fusion technology

To carry out operation and maintenance retrospective analysis under the network security control of the railway passenger ticket system, it is necessary to effectively collect and integrate multi-source operation

and maintenance data. The system operation includes multiple nodes such as bureau-level servers, station terminals, and network and security equipment, with diverse data sources and obvious structural differences. In order to support retrospective analysis, it is necessary to build a collection system including asset static account, operation status, log, and alarm information, and obtain relevant data through agents, probes, and log interfaces. For heterogeneous data, we use data cleaning to eliminate invalid information, use standardized means to unify the data form, and establish the corresponding relationship among assets, states, and events through correlation analysis technology ^[3]. Through the integration and processing of multi-source data, it can provide relatively complete data support for the retrospective analysis of operation and maintenance, and strengthen the effectiveness of network security control and operation and maintenance analysis.

3.2. Intelligent retrospective analysis algorithm model

In the operation and maintenance retrospective analysis of the railway passenger ticket system, the intelligent analysis model realizes anomaly identification and fault tracing through the study of operation and maintenance data assets. The model takes the asset status, network traffic, and log data as inputs to refine the typical characteristics of system operation to identify abnormal actions that deviate from the normal baseline. In the process of troubleshooting, according to the business scheme and asset association status, reverse reasoning is carried out for performance degradation or business interruption to determine the possible fault node. For example, through the comprehensive analysis of equipment parameters, process status, and network connection changes, and the combination of decision tree, neural network, and other methods, a retrospective analysis model is built to assist the operation and maintenance decision ^[4]. This model can improve the efficiency of fault location and provide technical support for the safe and stable operation of the railway passenger ticket system through the accumulation of historical data and continuous optimization analysis.

4. Data asset management requirements under the security control of the rail transit network

4.1 Analysis of data asset characteristics

4.1.1. Types and standards of data assets

There are clear and strict requirements for data asset management in the operation and maintenance retrospective analysis of the urban rail transit ticketing system. In the organizational structure and the collection and management process of data assets such as personnel, business systems, network environment (traditional network, distributed network, cloud service), products, and physical devices, nodes, etc., it is necessary to follow the network security and operation and maintenance specifications uniformly. The above data assets show the characteristics of multi-faceted, multi-faceted, and multi-source heterogeneity. The management objects include business, system, network, equipment, nodes, and other aspects. Its accuracy directly affects the effectiveness of operation and maintenance retrospective analysis and security event positioning. At the same time, data assets should maintain a high degree of integrity, focusing on recording the baseline information of hardware, system software, application software, middleware, database under the node, as well as services, processes, ports, network connections, dependent resource files, and other operations, so as to provide a reliable basis for fault tracing and safety audit. Through unified and standardized modeling and association management and control of data asset entities and their access,

Association and log characteristics, the consistency and traceability of data asset baselines can be ensured, providing support for automated security operation and maintenance and retrospective analysis ^[5].

4.1.2. Construction of the operation and maintenance retrospective data asset relationship chain

In the asset management of the railway passenger ticket system, data assets show obvious characteristics of hierarchy, relevance, and complexity. From the bureau level core to the end of the station, the data assets are structured, including different levels and types of organizational units, business systems, network conditions, equipment, and software nodes. Its association characteristics are mainly reflected in organizational affiliation, physical connection, association, network topology association, and service invocation Association. The complex characteristics are derived from the diversity of data assets, the richness of attributes, and the dynamic change of operation status.

In order to support operation and maintenance backtracking and security incident tracing, a complete data asset relationship chain should be built on the premise of clearly defining the entity attributes of data assets. By establishing an all-factor correlation model from the organization and business system to the supporting software and hardware, and including the network connection and service dependency, we can clearly describe the dependence and influence path between data assets ^[6]. In case of failure or a safety event, it can carry out rapid backtracking along the relationship chain and accurately locate the root cause and impact area of the problem. This requires the asset ledger to have excellent structural modeling and association management capabilities, ensure the accuracy, integrity, and traceability of data asset information, and provide a reliable data basis for the safe and stable operation of the passenger ticket system and the operation and maintenance retrospective analysis.

4.2. Operation and maintenance monitoring data governance system

4.2.1. Quality control mechanism for operation and maintenance monitoring data

In the operation and maintenance of the railway passenger ticket system, a data asset dynamic monitoring and governance system is established for security control and retrospective analysis. The data quality control mechanism of the system is very important. From the perspective of management requirements, real-time and quasi-real-time data collection mechanisms should be designed. Because the operation status of equipment, process, network connection, and other data assets in the ticketing system directly reflects the business and security situation, any missed acquisition, MIS acquisition, or inaccurate status may affect the accuracy of operation and maintenance judgment and retrospective analysis, so it is necessary to ensure the timeliness and consistency of status data through standardized acquisition methods. According to the defined data asset entities and baselines, it is necessary to formulate status determination and verification criteria, set clear normal and abnormal determination conditions for different types of assets, and identify abnormal conditions that deviate from the baseline through rule constraints. Only through reliable collection and a unified determination mechanism can the dynamic status data be credible and provide effective data support for real-time monitoring, anomaly identification, and operation and maintenance retrospective analysis ^[7].

4.2.2. Multi-party collaborative operation and maintenance data sharing platform

In the operation and maintenance of the railway passenger ticket system, it is necessary to build an operation and maintenance data collaborative analysis platform that can support data asset governance and operation and maintenance retrospective analysis, so as to realize the unified management and correlation analysis of

asset static accounts, dynamic state data, logs, and event data. The platform needs to carry out standardized access and associated identification of data from different operation and maintenance support subjects in the data aggregation link to ensure the consistency and associativity of data assets. In the field of data storage and processing, it is necessary to have the ability to efficiently process status data and log data, and to support context analysis based on the relationship between data assets. The association between static asset attributes and dynamic operation changes can support the whole process retrospective analysis of fault and safety events according to the timeline, so as to improve the efficiency of problem location and impact analysis. This collaborative analysis platform can effectively improve the utilization efficiency of operation and maintenance data assets, support the safe operation and maintenance, and rapid recovery of complex systems, and meet the main role of operation and maintenance collaboration and correlation analysis in improving efficiency highlighted in ^[8].

5. Application of data assets in project management

5.1. Security data-driven decision model

5.1.1. Risk assessment data modeling

In the operation and maintenance management of network security control of the railway passenger ticket system, the risk assessment data modeling establishes a dynamic risk assessment matrix by integrating asset vulnerability information, Threat Intelligence, and operation status data. The asset vulnerability information reflects the known vulnerabilities and configuration weaknesses of servers and applications, and the health data includes performance indicators, abnormal alarms, and other instant information. Integrate the two with external threat intelligence, and give weights to different data according to the importance of assets, vulnerability severity, threat activity, and other factors, and establish a matrix ^[8]. For example, the high-risk vulnerability data of the core business server is provided with a high weight. This matrix can dynamically show the system risk situation, assist the manager to accurately assess the risk, and timely adjust the security strategy or patch planning. The dynamic risk assessment matrix established based on the risk assessment data has become the main basis for safety data-driven decision-making, providing core support for the retrospective analysis of ticket system safety operation and maintenance, and improving the overall operation and maintenance management efficiency ^[9].

5.1.2. Design of an intelligent warning system

In the operation and maintenance traceability analysis of network security management and control of the railway passenger ticket system, the design of the intelligent early warning system is the main step. Through data assets, the business anomaly early warning decision tree model can be developed through the analysis of asset status and log characteristics. Extract main features from data assets such as performance benchmarks, process performance, and network traffic patterns, and conduct in-depth analysis of these data using data mining and machine learning algorithms ^[10]. According to the analysis results, a decision tree model is built. The nodes of the model set judgment conditions according to different data characteristics. Once the real-time data reaches the specific conditions, the system can send out an early warning in time. For example, by monitoring the abnormal exit of the core process, the deviation of transaction response time from the baseline, and the increase in the number of abnormal network connections, the decision tree can quickly identify the abnormal trend, provide accurate early warning messages for the operation and maintenance

workers, so as to take countermeasures in advance, prevent business interruption or decline in service quality, and ensure the safe and stable operation of the railway passenger ticket system.

5.2. Digital control of the detection process

5.2.1. Data integration of mobile detection terminals

In the application practice of data assets for network security control of the railway passenger ticket system, the data integration generated by operation and maintenance terminals and tools is very important for the digital control of the detection process. Through the specific technical scheme, it can realize the real-time upload of inspection results, vulnerability scanning reports, configuration verification data, etc., and synchronize with the central platform. The collected data can be quickly integrated by developing a data interface suitable for a variety of operation and maintenance tools and terminals. Ensure that the data is transmitted to the operation and maintenance data platform in a stable and efficient mode through the Railway internal network or a safe channel. A data encryption algorithm is adopted to ensure the security and integrity of data during transmission. A data integration module is built on the platform side to carry out standardized processing and classified storage of the uploaded data, so that the relevant data assets can be quickly and accurately called during the subsequent operation and maintenance retrospective analysis, providing strong support for the operation and maintenance decision, and further improving the efficiency and quality of the safe operation and maintenance of the railway passenger ticket system.

5.2.2. Topology visualization monitoring system

In the operation and maintenance management of the railway passenger ticket system, the topology visualization monitoring system is the main way of digital control of the detection process, which can visually display and analyze the data assets. Based on data assets, this system maps the real-time status and alarm messages of network equipment, servers, terminals, and other assets to the network topology and service topology model. By collecting the online status, link traffic, business service operation status, and other data of the equipment, accurate search and dynamic display can be realized in the visual interface, so that the operation and maintenance personnel can intuitively grasp the overall operation situation of the system, and improve the monitoring and control efficiency. At the same time, by virtue of the data asset relationship model, it can carry out the analysis of fault impact fields and the identification of abnormal situations. In case of failure, it can carry out rapid backtracking analysis according to data assets and topology association, determine the failed node, affected links, and related businesses, provide comprehensive and intuitive data support for operation and maintenance decisions, and indeed improve the efficiency and accuracy of backtracking analysis of the railway passenger ticket system operation and maintenance.

5.3. Project performance data analysis

5.3.1. Quantitative evaluation of detection efficiency

In the operation and maintenance of network security management and control of the railway passenger ticket system, the quantitative evaluation of operation and maintenance efficiency is very important for the application of data assets. By building a KPI system including fault discovery timeliness (mttd), fault repair timeliness (MTTR), and other aspects, the scientific quantification of operation and maintenance efficiency can be achieved. For example, MTTR can accurately measure the length of time from failure to business recovery. A shorter repair time represents a higher operation and maintenance efficiency. Data assets

accurately record the occurrence, response, processing, recovery time, and other relevant data of each event. Through in-depth analysis, we can gain insight into the long-term links in the operation and maintenance process and further carry out targeted optimization. In terms of compliance rate, the actual configuration will be compared with the compliance status of the security baseline, and the relevant data of configuration verification will be recorded through data assets to analyze the causes of deviation, so as to carry out timely rectification, improve the quality and efficiency of operation and maintenance, ensure the stable operation of the ticket system under the established SLA standards, and fully reflect the value of data assets in the quantitative evaluation of operation and maintenance efficiency.

5.3.2. Path of data value conversion

In the operation and maintenance work of network security management and control of the railway passenger ticket system, it is analyzed that the operation and maintenance data has the main way of data value conversion in fault tracing and preventive maintenance. In the fault processing stage, the deep correlation analysis of logs, monitoring, and topology data in data assets can quickly locate the root cause, reduce the duration of business interruption, further avoid the economic losses caused by long-term shutdown, and realize the transformation of economic value in business continuity guarantee. In the preventive maintenance phase, data assets can help identify performance deterioration trends and potential risks, analyze their causes, transform passive response into active intervention, reduce the probability of major failures and the consumed emergency resources and costs, and help improve the overall reliability of the system, save potential losses for enterprises, create more economic benefits, and ultimately promote the optimization of operation and maintenance costs and benefit improvement of the passenger ticket system in the whole life cycle.

6. Conclusion

Through the research on the data assets of the railway passenger ticket system in the context of rail transit network security control in the retrospective analysis of operation and maintenance, we can conclude the data asset governance concept based on the entity model and relationship model of data assets, with the asset Ledger as the core, supported by status monitoring and log analysis, and with the purpose of retrospective analysis. This idea focuses on asset modeling at all levels, from organization, system, network, equipment, to process and log, providing clear and verifiable data support for operation and maintenance backtracking. The relevant practice has important enlightenment value for the operation and maintenance management of other large-scale major information infrastructures, which is conducive to the establishment of a data management system for the full life cycle of assets, improving the ability of asset state perception, dependency identification, and problem location, and further improving the accuracy and efficiency of fault and safety event retrospective analysis.

From a more macro perspective, the asset data governance scheme for operation, maintenance, and security has strong versatility. Although there are differences in business models and technical conditions among various industries, they have common features in the core logic of asset identification, relationship establishment, baseline control, and retrospective analysis. Future research can further focus on the standardization of data asset models and the establishment of automated operation and maintenance capabilities, explore data asset governance solutions with reusability and scalability, and promote the value of data assets in ensuring safe operation and maintenance and stable business operation.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Jiang C, 2023, Research on Data Asset Value Evaluation with and Without Transaction Scenarios, thesis, Tianjin University of Technology.
- [2] Zhou Y, 2022, Research on Data Asset Value Evaluation Driven by the Digital Economy, thesis, Capital University of Economics and Business.
- [3] Liu G, 2023, Research on Data Asset Value Measurement of Power Grid Enterprises in Typical Application Scenarios, thesis, North China Electric Power University (Beijing).
- [4] Zhang Y, 2023, Research on Data Asset Value Evaluation of Manufacturing Enterprises from the Perspective of Big Data—A Case Study of Enterprise A, thesis, Hebei University of Economics and Business.
- [5] Qin X, 2022, Research on the Comparison and Optimization of Data Asset Value Evaluation Methods, thesis, Capital University of Economics and Business.
- [6] Wang F, 2023, Research on Intelligent Management and Control System for Operation and Maintenance of Rail Transit Power Supply Equipment. *Electric Power Equipment Management*, 2023(24): 186–188.
- [7] Chen X, 2021, Design of Data Asset Management and Control System for Radio and Television Supervision Platform Based on Hierarchical Classification. *Radio & Television Network*, 28(4): 43–45.
- [8] Fan X, 2022, Research on the Accounting of Data Assets in the Context of the Digital Economy. *Investment and Entrepreneurship*, 33(10): 111–113.
- [9] Zhou Y, Zhang W, 2022, Research on Data Asset Management System of Data Middle Platform. *Wireless Internet Technology*, 19(1): 30–31.
- [10] Xiang B, Shao Y, Lin J, et al., 2024, Research on Data Asset Security Risks and Prevention in Government Digital Transformation. *Computer Knowledge and Technology*, 20(9): 72–74 + 79.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

YOLO11-SPMB: Application of an Improved Instance Segmentation Network for Tire Defect Detection

Haixu You, Jun Xing*

School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Tire quality is vital for vehicle safety, but finding surface defects automatically is hard. Current systems often fail to spot tiny flaws hidden inside complex tread patterns. Also, because defects vary in size and blend into the dark rubber, standard models usually output broken or messy segmentation masks. To solve this, we built YOLO11-SPMB, a new instance segmentation network based on YOLO11. We designed it specifically for tire inspection. First, we added a Multi-Scale Adaptive Feature Fusion (MS-AFF) module to the network. It filters out the repeating noise from normal treads so tiny defects can stand out. Second, we created a Boundary Enhancement and Imbalance Calibration (BE-IC) strategy. It uses a boundary Dice loss to keep the mask edges smooth and whole. At the same time, it adjusts class weights to help the model catch rare but major damage, fixing the data imbalance problem. Tests on our custom dataset show the model works very well. YOLO11-SPMB achieves a mAP@0.5:0.95 of 0.702, a mIoU of 0.782, and a recall of 0.928. It also runs at 60 frames per second (FPS), making it fast and accurate enough for real factory production lines.

Keywords: Tire defect detection; Instance segmentation; YOLOv11

Online publication: Jul 23, 2026

1. Introduction

Tires are the vital parts that keep a vehicle on the road, so their build quality is key to safety. In the world of industrial checks, using deep learning to automate this task has become a common practice. Many have already used smart networks to find flaws. For instance, deep learning models helped detect defects in weld surfaces, while YOLO algorithms with extra attention features were used to find flaws in metal ^[1,2]. At the same time, finding defects in ultrasonic images using neural networks and PCB flaw detection models worked much better than the old manual ways ^[3,4]. As factories began to demand faster and more precise checks, researchers built decoupled YOLO structures to inspect bridge surfaces, improved Faster R-CNN layouts for 3D printed parts, and updated YOLOv3-Tiny to spot surface flaws more quickly ^[5-7].

Even with these gains, checking tires remains uniquely difficult due to their material and shape. Since tires are made of black rubber, they absorb most light, and their molded tread patterns create heavy background noise that often hides tiny cracks. Just drawing boxes around defects isn't enough to understand these odd-shaped problems; so, researchers started using pixel-level networks. Better instance segmentation tools were made to get clearer results^[8]. To keep track of exact borders under tricky light, methods based on shape contours were built to check copper tubes, while multi-scale Faster R-CNN frameworks were used to better pin down micro-defects on steel^[9,10]. These spatial tracking tools were also used in bigger areas, like finding flaws on aircraft using modern recognition techniques^[11]. Big reviews of surface defect detection have confirmed that modern automated checks depend heavily on how well a network can pick out features on the fly^[12]. With this in mind, smarter networks were born; for example, better YOLOv8 setups made it easier to spot small flaws on PCBs, and the YOLOv8-LSD algorithm cut down the math needed to check steel surfaces on fast lines^[13,14].

Recently, the YOLO series has changed quickly to handle the quick pace and messy nature of factory work. YOLOv9 made feature extraction better by using programmable gradient information to stop data loss^[15]. Soon after, YOLOv10 removed the delay from non-maximum suppression (NMS) and was used to spot small items in satellite photos, as well as helping drones find people in rescue missions^[16,17]. In the newest generation, YOLOv11 has taken another step forward in seeing details. New studies have successfully used improved YOLOv11 models to find subtle defects in electronic items^[18]. Also, lightweight versions of YOLOv11 that focus on small parts were made for insulator checks, and highly updated YOLOv11 algorithms were used to pull out anomalies from printed circuit boards^[19,20].

Although today's YOLO models work well, they often figure out mask losses using simple binary cross-entropy, which lets the flat background wash out the thin boundary lines. This often leads to messy or broken edges when spotting cracks on dark tires. To solve this, this paper brings forward an improved single-stage instance segmentation framework called YOLO11-SPMB. By rebuilding the neck of the network with a multi-scale adaptive feature fusion module, the network can filter out tread noise on the fly.

We also added a Boundary Enhancement and Imbalance Calibration (BE-IC) step. This puts strict rules on the defect shapes. It stops the masks from splitting into pieces when the contrast is poor. At the same time, it adjusts the class weights to help find rare flaws.

2. Proposed method

2.1. Overall architecture of YOLO11-SPMB

Inspecting tire surfaces is hard. The background patterns are noisy, and the defect edges are difficult to see. We created YOLO11-SPMB to handle this. It is a single-stage instance segmentation model. We started with the standard YOLO11 and changed its structure. These updates boost accuracy but keep the model lightweight. **Figure 1** shows the new layout. It has a backbone for feature extraction, a custom neck, and an updated prediction head. We put the Multi-Scale Adaptive Feature Fusion (MS-AFF) module inside the neck. This module blocks tread noise before mixing the features. We also added a Boundary Enhancement and Imbalance Calibration (BE-IC) step at the head. BE-IC applies geometric rules and updates sample weights. This helps the network draw better shapes for irregular flaws while still running fast.

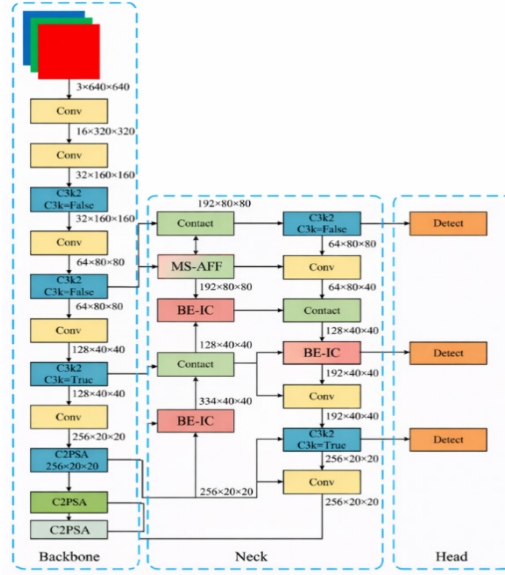


Figure 1. YOLOv11 overall network structure diagram
图 1. YOLOv11 整体网络结构图

Figure 1. YOLOv11-SPMB network architecture diagram.

2.2. Multi-scale adaptive feature fusion (MS-AFF)

Tire tread patterns easily hide small cracks. Standard models do not filter this background noise. Instead, they pass it directly into deeper layers. We built the MS-AFF module to solve this problem. As **Figure 2** shows, it filters the feature maps using channel and spatial data.

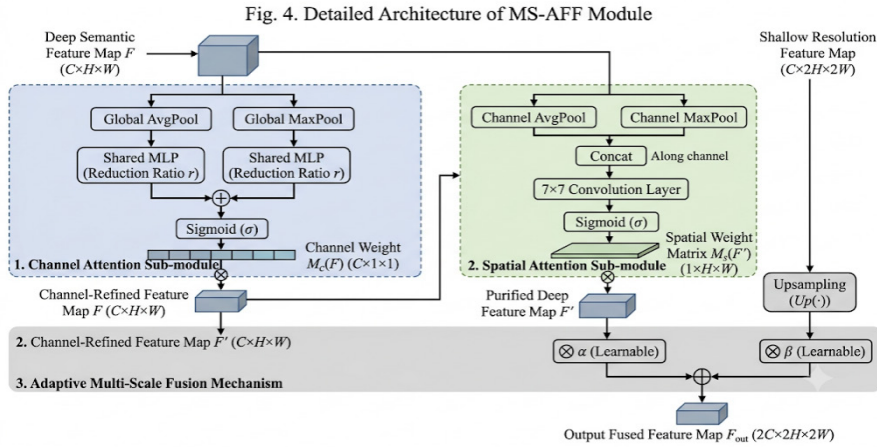


Figure 2. Schematic diagram of MS-AFF module structure.

The backbone extracts a feature map, $F \in \mathbb{R}^{C \times H \times W}$. The channel attention module then calculates a 1D weight vector. This vector highlights the channels that actually contain defect features. This is achieved by utilizing both global average pooling and global max pooling, followed by a shared multi-layer perceptron (MLP):

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

where σ denotes the Sigmoid activation function, the channel-refined feature map is obtained by element-wise multiplication:

$$F' = M_c(F) \otimes F \quad (2)$$

Subsequently, to precisely locate spatial anomalies and suppress interference from uniform tread patterns, a spatial attention mechanism is applied. The channel-refined feature map F' is first aggregated along the channel dimension and then processed by a large-kernel convolution operation to capture broader spatial receptive fields:

The fully purified feature map is thus represented as

$$F'' = M_s(F') \otimes F' \quad (3)$$

When merging the purified deep semantic features F''_{deep} with the high-resolution shallow features $F''_{shallow}$, MS-AFF employs learnable scalar weights α and β to adaptively balance the fusion proportion:

$$F_{out} = \alpha \times F''_{deep} \oplus \beta \cdot \text{Up}(F''_{shallow}) \quad (4)$$

Here, $\oplus$ stands for concatenation, and $\text{Up}(\cdot)$ refers to spatial upsampling. This adaptive method helps the network pay more attention to unusual pattern changes. By doing this, the model can better tell the defects apart from the busy tread background.

2.3. Boundary enhancement and imbalance calibration (BE-IC)

Tire defect detection is not just about background noise; it also struggles with the shapes of the defects. Since both the defects and the tire rubber are black, they blend (as shown in **Figure 3**). This makes it hard for standard loss functions to see the edges, often leading to broken or jagged mask lines. Also, serious damage is rare compared to small surface marks. This creates a data imbalance where the model sees fewer important defects than common ones. Our BE-IC strategy works at the prediction head to fix both of these problems.

Fig. 6. Overall Architecture of the Integrated BE-IC Strategy for Segmentation Head

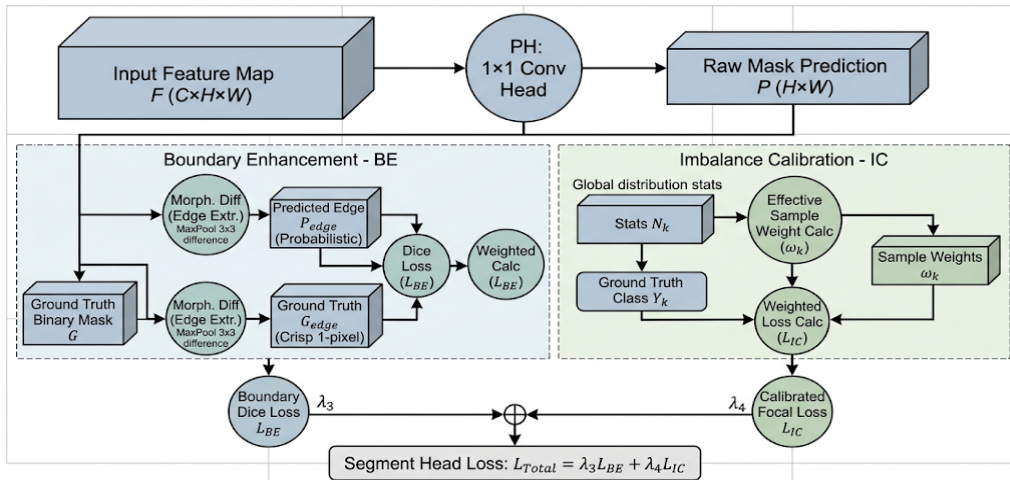


Figure 3. BE-IC module schematic diagram.

To make sure the defect shapes stay whole, the Boundary Enhancement (BE) part uses a specific geometric rule. We get the sharp, single-pixel ground-truth boundary G_{edge} and the predicted boundary P_{edge} by using a max-pooling difference method:

$$G_{edge} = \text{MaxPool}_{3 \times 3}(G) - G \quad (5)$$

where G denotes the binary ground-truth mask. To penalize boundary misalignment, a boundary-specific Dice loss L_{BE} is formulated as follows:

$$L_{BE} = 1 - \frac{2 \sum_{i,j} (P_{edge}^{i,j} \cdot G_{edge}^{i,j}) + \epsilon}{\sum_{i,j} (P_{edge}^{i,j})^2 + \sum_{i,j} (G_{edge}^{i,j})^2 + \epsilon} \quad (6)$$

where ϵ is a smoothing constant introduced to prevent division by zero. This loss continuously drives steep gradient propagation toward low-contrast contour regions.

Simultaneously, the Imbalance Calibration (IC) strategy mitigates the long-tail bias based on the effective sample volume theory. Let N_k denote the total number of samples in the category k . The effective sample weighting factor ω_k is computed to suppress the dominant gradient influence of the majority classes:

$$\omega_k = \frac{1 - \gamma}{1 - \gamma^{N_k}} \quad (7)$$

where $\gamma \in [0, 1)$ controls the attenuation rate. This dynamic weighting factor ω_k is incorporated into a modified Focal Loss framework to construct the calibrated classification objective L_{IC} :

$$L_{IC} = - \sum_{k=1}^K \omega_k \cdot y_k \cdot (1 - p_k)^{\hat{\alpha}} \log fo(p_k) \quad (8)$$

where y_k and p_k denote the ground-truth label and the predicted probability, respectively, and $\hat{\alpha}$ is the focusing parameter.

2.4. Unified joint loss function

To establish an end-to-end multi-objective optimization framework, the standard YOLO loss is comprehensively reconstructed. The total joint loss function L_{Total} . The proposed YOLO11-SPMB is defined as a weighted aggregation of bounding box regression, mask segmentation, boundary enhancement, and calibrated classification:

$$L_{Total} = \lambda_1 L_{CIoU} + \lambda_2 L_{mask_bce} + \lambda_3 L_{BE} + \lambda_4 L_{IC} \quad (9)$$

where L_{CIoU} denotes the Complete Intersection over Union loss for robust spatial bounding box localization, and L_{mask_bce} represents the standard binary cross-entropy loss for interior mask regions. The hyperparameter weights $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ are empirically optimized via grid search to ensure stable gradient convergence across competing spatial and semantic objectives.

3. Experiments

3.1. Dataset and implementation details

To evaluate the YOLO11-SPMB network, experiments were performed on a custom tire defect dataset named TireDefect-Self. The original images were collected using a 3D laser scanner to avoid ambient light interference, and the 3D point cloud data was converted into 2D height grayscale images. The dataset

includes three common tire defect types: cracks, scratches, and impurities or bulges. Industrial datasets usually present data imbalance where certain defects are rare. To prevent model overfitting caused by this imbalance, data augmentation methods such as geometric flipping, random noise injection, Mosaic, and MixUp were applied. The final augmented dataset contains 4,000 images. We sorted the dataset by tire IDs to avoid data leaks. This gave us 3,200 images for training, 400 for validation, and 400 for testing.

We ran all tests on an Ubuntu 20.04 machine. It had an Intel Xeon CPU and one RTX 4090 GPU (24GB). We wrote the code in PyTorch 2.0 and used CUDA 11.8. We resized all input images to 640×640 pixels. We trained the network for 300 epochs with a batch size of 32. We used the AdamW optimizer. The starting learning rate was 0.001, and we lowered it step by step using cosine annealing. For testing, we set the NMS IoU limit to 0.45 and the confidence limit to 0.25.

3.2. Evaluation metrics

We measure both bounding box detection and mask quality using five metrics: $mAP@0.5$, $mAP@0.5:0.95$, *Recall*, *mIoU*, and *FPS*. Each metric handles a specific task. The $mAP@0.5$ score checks if the model finds the general area of a defect using a loose 50% overlap rule. The $mAP@0.5:0.95$ score measures box accuracy strictly across multiple overlap levels from 50% to 95%. Recall counts the percentage of real defects the model successfully catches to monitor missed flaws. The *mIoU* metric calculates pixel-level overlap to judge how well the predicted mask shape fits the true defect boundary. The *FPS* score counts the frames processed per second to show the running speed on the factory line.

3.3. Comparison with other methods

We compared YOLO11-SPMB with several top networks. Table 1 lists the baseline models. We chose two-stage models (Mask R-CNN, Cascade Mask R-CNN, and PointRend). We chose standard segmentation models (U-Net and DeepLabV3+). We also chose single-stage models (YOLOv9-CBAM, YOLOv10-DCNv3, and YOLOv11-seg).

Table 1 shows the test results. Two-stage models like Mask R-CNN are too slow. They only run at 9 to 15 FPS because their structure is too large. Factories cannot use them on fast lines. Models like YOLOv10-DCNv3 add attention or deformable convolutions. This improves accuracy but drops the speed to 42 FPS due to the extra math. See **Table 1** and **Figure 4**.

Table 1. Performance comparison table of mainstream instance segmentation networks

<i>Model</i>	<i>mAP@0.5</i>	<i>mAP@0.5:0.95</i>	<i>Recall</i>	<i>mIoU</i>	<i>FPS</i>
Mask R-CNN	0.902	0.612	0.861	0.728	12
Cascade Mask R-CNN	0.915	0.635	0.874	0.735	9
PointRend	0.912	0.648	0.870	0.758	15
U-Net	0.901	0.648	0.861	0.747	58
DeepLabV3+	0.922	0.632	0.884	0.715	35
YOLOv9-CBAM	0.928	0.655	0.894	0.753	47
YOLOv10-DCNv3	0.933	0.665	0.902	0.760	42
YOLOv11-seg	0.926	0.658	0.895	0.751	60
YOLO11-SPMB	0.954	0.702	0.928	0.782	60

Fig. 5. Visual Comparison of Tire Defect Instance Segmentation across Different Methods

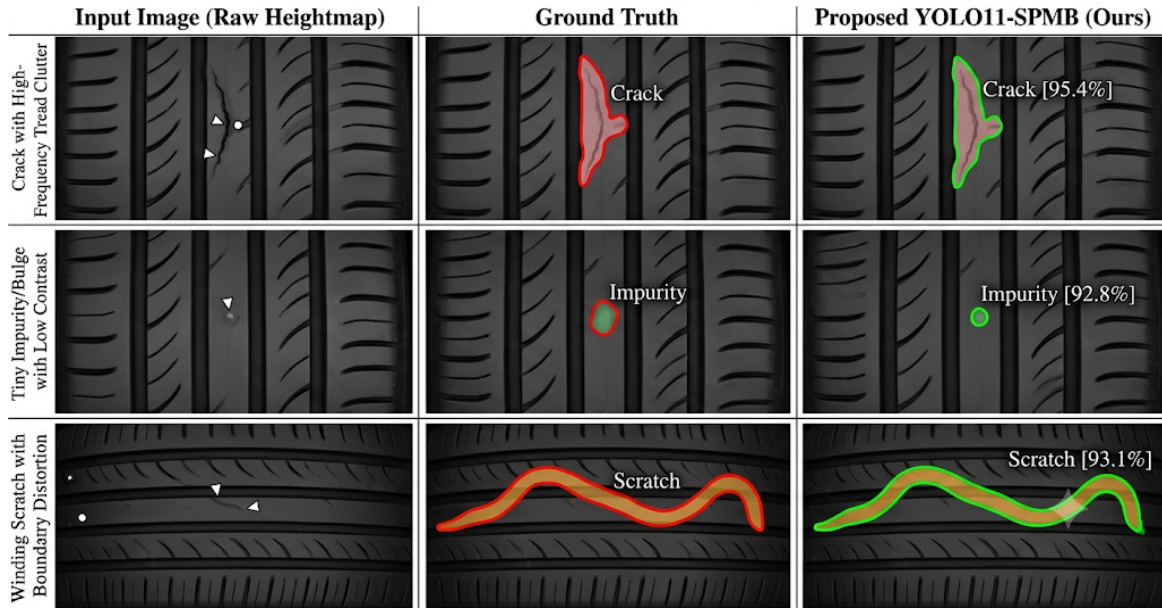


Figure 4. Tire defect images.

YOLO11-SPMB retains both speed and precision. The model reaches an $mAP@0.5:0.95$ of 0.702. The Recall score increases to 0.928, which means fewer small cracks are missed. The mIoU is 0.782. This score shows that the predicted masks fit the real defect shapes well. The MS-AFF and BE-IC modules add very little computing time. YOLO11-SPMB runs at 60 FPS, matching the baseline YOLOv11-seg exactly. The model balances mask quality and speed. It works well for automated tire checks in factories.

3.4. Ablation study

We ran an ablation study on the MS-AFF, BE, and IC modules. The baseline is the standard YOLOv11-seg network. Table 2 shows the step-by-step results. Adding the MS-AFF module increases the Recall score from 0.895 to 0.912. The module filters tread noise and finds small defects. Adding the BE mechanism uses Dice loss to improve boundary shapes. This increases the mIoU score to 0.770. The IC strategy handles the data imbalance from rare defects. The final $mAP@0.5:0.95$ score reaches 0.702. See **Table 2**.

Table 2. Ablation study results of the proposed modules

<i>MS-AFF</i>	<i>BE</i>	<i>IC</i>	<i>mAP@0.5:0.95</i>	<i>Recall</i>	<i>mIoU</i>
			0.658	0.895	0.751
✓			0.681	0.912	0.762
✓	✓		0.685	0.915	0.770
✓	✓	✓	0.702	0.928	0.782

4. Conclusion

This paper presents YOLO11-SPMB, a single-stage instance segmentation network designed to inspect tire surface defects. It handles common factory issues like busy tread patterns, very small flaws, and faint defect edges. We updated the feature fusion neck and the prediction head of the standard YOLOv11 model to get

better pixel-level results. Inside the network, the Multi-Scale Adaptive Feature Fusion (MS-AFF) module blocks out the repeating noise from normal tire treads and makes tiny defects easier to see. At the same time, the Boundary Enhancement and Imbalance Calibration (BE-IC) strategy uses a boundary Dice loss to fix messy mask edges and changes class weights to handle uneven data. Tests on our tire dataset show that YOLO11-SPMB balances speed and accuracy well. It gets a $\text{mAP}@0.5:0.95$ of 0.702, a recall of 0.928, and a mIoU of 0.782, while running at 60 FPS. We also tested it on the NEU-Seg dataset, and the results show the model works well on other industrial surfaces too.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zuo Y, Wang J, Song J, 2021, Application of YOLO Object Detection Network in Weld Surface Defect Detection. 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 704–710.
- [2] Xikun X, Changjiang L, Meng X, 2021, Application of Attention YOLOV4 Algorithm in Metal Defect Detection. 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), 465–468.
- [3] Mišković M, et al., 2021, Automated Defect Detection From Ultrasonic Images Using Deep Learning. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 68(9): 3126–3134.
- [4] Huang W, Wei P, Zhang M, 2021, PCB Defect Detection Using Deep Learning Methods. 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 1–6.
- [5] Sun S, Liu W, Cui R, 2022, YOLO Based Bridge Surface Defect Detection Using Decoupled Prediction. 2022 7th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), 117–122.
- [6] Zhang Y, Zhang Z, Fu K, et al., 2022, Adaptive Defect Detection for 3-D Printed Lattice Structures Based on Improved Faster R-CNN. IEEE Transactions on Instrumentation and Measurement, 71: 1–9.
- [7] Li J, et al., 2022, Surface Defect Detection Based on Improved YOLOv3-Tiny Algorithm. 2022 7th International Conference on Image, Vision and Computing (ICIVC), 1–6.
- [8] Ren X, He X, 2023, An Enhanced Instance Segmentation Network for Surface Defect Detection. 2023 9th International Conference on Computer and Communications (ICCC), 2098–2103.
- [9] Chen D, Xie S, Xie Y, 2023, A Contour-Based Instance Segmentation Method for Copper Tube. 2023 China Automation Congress (CAC), 7864–7869.
- [10] Shen M, Cai F, Xie J, 2023, A Multi-Scale Defect Detection for Steel Surface Based on Improved Faster R-CNN. 2023 China Automation Congress (CAC), 7448–7453.
- [11] Malik P, Lakshakar H, Hazra J, 2023, Aircraft Defect Localization and Detection Using Object Detection Techniques. 2023 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC), 1–6.
- [12] Haobo Y, 2024, A Survey of Industrial Surface Defect Detection Based on Deep Learning. 2024 International Conference on Cyber-Physical Social Intelligence (ICCSI), 1–6.
- [13] Raj A, Ahlawat P, 2024, Refined YOLOv8 for Enhanced Small Object Detection in PCB Defect. 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT),

1–6.

- [14] Yu L, Yan P, 2024, YOLOv8-LSD: A Steel Surface Defect Detection Algorithm Based on YOLOv8. 2024 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), 1394–1399.
- [15] Wang C, Yeh I, Liao H, 2024, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 6222–6232.
- [16] Sun H, Yao G, Zhu S, 2025, SOD-YOLOv10: Small Object Detection in Remote Sensing Images Based on YOLOv10. IEEE Geoscience and Remote Sensing Letters, 22: 1–5.
- [17] Mishra A, Narendra M, Sinha A, et al., 2025, Dynamic Backbone Optimization of YOLOv10 for Real-Time Object Detection in UAV-Based Search and Rescue Missions. IEEE Access, 13: 195975–195986.
- [18] Dang Z, Wang X, 2025, FD-YOLO11: A Feature-Enhanced Deep Learning Model for Steel Surface Defect Detection. IEEE Access, 13: 63981–63993.
- [19] Shen P, Mei K, Cao H, 2025, LDDFSF-YOLO11: A Lightweight Insulator Defect Detection Method Focusing on Small-Sized Features. IEEE Access, 13: 90273–90292.
- [20] Selvam P, Rajasekar R, Gunasundari C, 2025, YOLO-DefXpert: An Advanced Defect Detection on PCB Surfaces Using Improved YOLOv11 Algorithm. IEEE Access, 13: 143085–143101.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Structure- and Frequency-aware Domain Adaptation for SAR Raft Aquaculture Extraction

Yijie Wu, Shan Jiang*

School of Information Science and Engineering, Dalian Polytechnic University, Dalian 116034, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Marine aquaculture is an important part of the marine economy in China. Synthetic aperture radar (SAR) images can work at all times, which makes them suitable for large-scale monitoring of aquaculture areas. However, SAR images from different regions often have obvious differences in background scattering, aquaculture structures, and imaging conditions. These differences cause serious domain shift problems and reduce the segmentation performance of existing methods in cross-region tasks. To solve this problem, this paper proposes a structure- and frequency-aware domain adaptation method (SFDA) for raft aquaculture extraction from SAR images. The method includes a directional structure alignment (DSA) module and a frequency domain alignment (FDA) module. The DSA module aligns the directional structure features between the source and target domains, while the FDA module aligns texture and scattering features in the frequency domain. Experiments on the Gaofen-3 SAR dataset from Jiangsu province show that the proposed method is effective for cross-region raft aquaculture extraction.

Keywords: Transfer learning; Unsupervised domain adaptation; SAR; Marine aquaculture

Online publication: Jul 23, 2026

1. Introduction

In recent years, the offshore raft aquaculture industry has become an important component of the coastal economy and has developed rapidly in China ^[1]. Accurate extraction of raft aquaculture areas is of great practical significance for marine resource management, marine environmental monitoring, and sustainable aquaculture planning. Synthetic aperture radar (SAR) imagery is less affected by weather and illumination conditions, enabling all-time observation of offshore aquaculture areas. Therefore, SAR remote sensing technology has been widely applied in the extraction of raft aquaculture areas ^[2].

Traditional raft aquaculture extraction methods mainly rely on manually designed features and threshold-based segmentation strategies. Although these methods can achieve satisfactory results in simple scenes, they generally suffer from poor robustness in complex marine environments. With the rapid development of deep learning techniques, convolutional neural network-based semantic segmentation algorithms have

significantly improved the extraction accuracy of marine aquaculture targets and demonstrated excellent performance^[3]. Wang et al. proposed the MDOAU-net model, which integrates multi-scale feature fusion, atrous convolution, and deformable convolution into the attention U-Net architecture for SAR image aquaculture area segmentation^[4]. Fan et al. proposed a method based on complex-valued convolutional neural networks for SAR aquaculture recognition. The method uses phase information in polarimetric SAR data and improves detection results on Gaofen-3 (GF-3) images^[5]. Wang et al. proposed the IDUDL model for marine aquaculture extraction in SAR images^[6]. The model updates pseudo labels step by step to support semantic extraction and incremental learning. However, SAR images from different sea areas often have large distribution differences. These methods work well in one region, but their performance usually decreases when applied to other regions. This makes cross-region generalization difficult.

Unsupervised domain adaptation (UDA) is widely used to reduce differences between domains. It transfers knowledge from a labeled source domain to an unlabeled target domain. Common methods include feature alignment, style transfer, and pseudo-label generation^[7–9]. UDA achieved remarkable performance in urban scene segmentation and is widely applied in SAR tasks, such as vehicle classification, ship detection, land cover classification, and oil spill segmentation^[10–16]. These studies show that UDA can improve cross-domain performance. However, raft aquaculture extraction in SAR images is still difficult.

SAR images from different coastal regions often show large differences because of sea conditions, imaging geometry, incidence angles, and speckle noise. These differences lead to large domain gaps between the source domain and the target domain. Raft aquaculture areas also have regular strip-like shapes and obvious directional features. Most existing domain adaptation methods focus on semantic features and ignore structural information. Because of this, the extracted features from different domains often do not keep good structural consistency. At the same time, SAR images differ not only in texture information but also in spectral distributions, which further increases the difficulty of cross-region adaptation.

To reduce the domain gap in cross-region SAR raft aquaculture extraction, this paper develops a structure- and frequency-aware domain adaptation framework based on ADVENT^[17]. The framework contains two adaptation modules. One is the directional structure alignment (DSA) module. It aligns directional features from the source and target domains, helping the model preserve similar structural patterns across domains. The other is the frequency domain alignment (FDA) module. This module reduces spectral differences in SAR images by aligning low-frequency components in the Fourier domain.

2. Proposed method

2.1. Overview

The overall framework of SFDA is shown in **Figure 1**. The framework uses ADVENT as the basic domain adaptation model and further applies alignment in both spatial and frequency domains to reduce differences between SAR scenes.

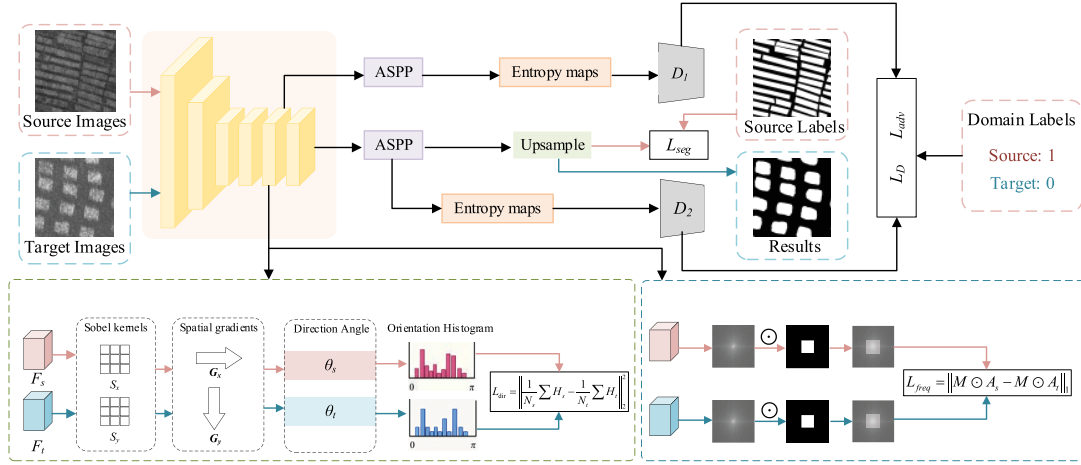


Figure 1. Architecture of the proposed structure- and frequency-aware domain adaptation framework.

2.2. ADVENT-based domain adaptation framework

Let the labeled source domain and the unlabeled target domain be denoted as $\mathcal{D}_s = \{(x_s, y_s)\}$ and $\mathcal{D}_t = \{x_t\}$, respectively. The segmentation network G predicts the probability map $P = G(x)$. For source-domain samples, standard supervised segmentation loss is employed:

$$L_{\text{seg}} = -\sum_{h,w} \sum_{c=1}^C y_s^{(h,w,c)} \log P_s^{(h,w,c)} \quad (1)$$

where C denotes the number of classes, and $P_s^{(h,w,c)}$ represents the predicted probability that the source-domain pixel belongs to class C . Following ADVENT, normalized entropy maps are constructed as

$$E(x) = -\frac{1}{\log C} \sum_{c=1}^C P^{(c)} \log P^{(c)} \quad (2)$$

The entropy map reflects the prediction uncertainty, where source-domain predictions usually exhibit lower entropy values, while target-domain predictions tend to have higher entropy values. Subsequently, the entropy maps of the source and target domains are fed into the domain discriminator D for adversarial learning. The domain discriminator, which aims to distinguish source-domain predictions from target-domain predictions, while the segmentation network attempts to confuse the discriminator. The discriminator loss is

$$L_D = -\mathbb{E}_{x_s} [\log D(E_s)] - \mathbb{E}_{x_t} [\log (1 - D(E_t))] \quad (3)$$

The adversarial objective for the segmentation network is

$$L_{\text{adv}} = -\mathbb{E}_{x_t} [\log D(E_t)] \quad (4)$$

Through entropy adversarial learning, the prediction distribution of the target domain is gradually aligned with that of the source domain.

2.3. Direction structure alignment

Raft aquaculture regions in SAR imagery usually exhibit highly regular directional structures due to an artificial parallel arrangement. Although the appearance characteristics vary significantly across regions, the directional continuity and structural consistency remain relatively stable. Conventional domain adaptation

methods mainly focus on semantic feature alignment while ignoring such structural priors. Consequently, the extracted features may still suffer from structural inconsistency under large domain shifts. To address this issue, a direction structure alignment module is designed to align the directional distribution of source and target features explicitly.

Let $F_s \in \mathbb{R}^{C \times H \times W}$ and $F_t \in \mathbb{R}^{C \times H \times W}$ denote the intermediate features extracted by the encoder from the source and target domains, respectively. Spatial gradients are first computed using Sobel operators.

$$\begin{aligned} G_x &= F * S_x \\ G_y &= F * S_y \end{aligned} \quad (5)$$

where S_x and S_y denote horizontal and vertical Sobel kernels, respectively, and $*$ denotes the convolution operation. The local directional angle is then calculated as

$$\theta = \arctan\left(\frac{G_y}{G_x + \epsilon}\right) \quad (6)$$

where, ϵ epsilon is a small constant to avoid numerical instability. To characterize the global directional distribution, orientation histograms are constructed over the feature maps:

$$H(\theta) = \text{Histogram}(\theta) \quad (7)$$

The histogram bins represent different orientation intervals within $[0, \pi]$.

To reduce directional discrepancy across domains, the orientation distributions of source and target domains are aligned using Maximum Mean Discrepancy (MMD) ^[18]:

$$L_{\text{dir}} = \left\| \frac{1}{N_s} \sum H_s - \frac{1}{N_t} \sum H_t \right\|_2^2 \quad (8)$$

where, H_s and H_t denote orientation histograms of source and target domains, N_s and N_t represent the number of samples.

Through direction structure alignment, the network can learn more stable cross-domain structural features, thereby improving the generalization capability under complex marine conditions.

2.4. Frequency domain alignment

In SAR imagery, domain discrepancy is not only reflected in spatial appearance but also in spectral characteristics caused by imaging conditions, sea states, and speckle noise. Low-frequency components mainly encode global imaging style and illumination statistics, while high-frequency components preserve structural details of raft targets. Therefore, aligning low-frequency spectral information can effectively reduce style discrepancy between domains while maintaining target structures.

For a feature map F , the 2D Fourier transform is computed as:

$$\mathcal{F}(u, v) = \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} F(x, y) e^{-j2\pi(ux/H + vy/W)} \quad (9)$$

The Fourier representation can be decomposed into amplitude and phase components:

$$\mathcal{F}(u, v) = A(u, v) e^{j\phi(u, v)} \quad (10)$$

where, $A(u, v)$ denotes the amplitude spectrum, $\phi(u, v)$ denotes the phase spectrum. Since low-frequency components dominate domain-specific imaging styles, only the central low-frequency region is aligned. A low-frequency mask M is defined as

$$M(u, v) = \begin{cases} 1, & (u, v) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where Ω denotes the low-frequency region. The frequency alignment loss is formulated as

$$L_{freq} = \|M \odot A_s - M \odot A_t\|_1 \quad (12)$$

where, A_s and A_t denote the amplitude spectra of source and target features, $\odot$ represents element-wise multiplication.

By aligning low-frequency spectral distributions, the proposed module effectively mitigates SAR imaging-style discrepancy across domains.

2.5. Overall optimization

The final training objective is

$$L_{total} = L_{seg} + \lambda_{adv} L_{adv} + \lambda_{dir} L_{dir} + \lambda_{freq} L_{freq} \quad (13)$$

where, L_{seg} denotes supervised segmentation loss on the source domain, L_{adv} denotes the adversarial entropy alignment loss in ADVENT, L_{dir} denotes the proposed direction structure alignment loss, L_{freq} denotes the proposed frequency-domain alignment loss.

During training, the optimization process alternately updates the segmentation network to minimize the segmentation and alignment losses, while updating the discriminator to enhance its capability of distinguishing source and target domains.

Through joint spatial-directional and frequency-domain alignment, the proposed framework learns more robust domain-invariant representations for cross-region SAR raft aquaculture extraction.

3. Experiments

3.1. Datasets

Both the source and target domain datasets are constructed using 3 m spatial resolution GF-3 ultra-fine strip map (UFS) mode SAR images. The source domain is collected from Changhai County, Liaoning province, while the target domain is collected from Jiangsu province. All images are cropped into image patches of 512×512 pixels. A total of 1000 image patches are generated for the source domain, whereas 96 image patches are selected for the target domain and divided into training and testing sets at a ratio of 84:12.

3.2. Evaluation metrics

To comprehensively evaluate the segmentation performance of the proposed method on the target domain, five commonly used evaluation metrics were used are adopted in this paper: mean Intersection over Union (*mIoU*), Overall Accuracy (*OA*), Precision (*P*), Recall (*R*), and F1-score (*F1*). Among them, the positive class corresponds to aquaculture areas, while the negative class corresponds to the surrounding seawater regions^[19].

3.3. Implementation details

All experiments are conducted in PyTorch 1.9.0 using an Intel Xeon W5-3435X with a clock speed of 3.1GHz and an NVIDIA GeForce RTX 4090. The hyperparameters λ_{adv} , λ_{dir} , and λ_{freq} are empirically set to 0.001, 0.01, and 0.01, respectively.

3.4. Comparison results

To validate the effectiveness of the proposed SFDA method, three representative methods are selected for comparison: K-Means and ADVENT, and RoadDA^[17,20]. All comparative experiments are conducted under the same settings. The visualization results of the predictions are shown in Figure 2, while the quantitative metrics are presented in **Table 1**. The proposed method achieves an mIoU of 0.7677 on the target domain, which is 0.0487 higher than that of RoadDA.

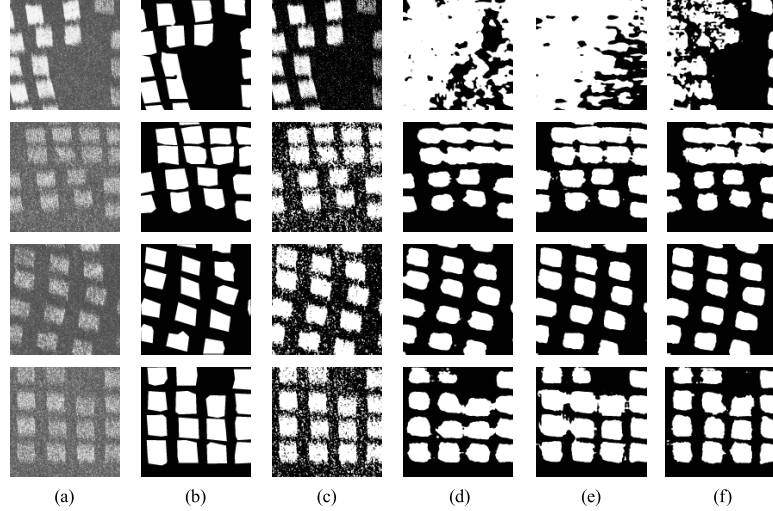


Figure 2. Comparative experimental results of floating raft aquaculture segmentation in Jiangsu province. (a) Original images. (b) Ground truth. (c) K-Means (d) ADVENT. (e) RoadDA. (f) SFDA.

Table 1. Evaluation results on the shared Jiangsu dataset

Methods	<i>mIoU</i>	<i>OA</i> (%)	<i>P</i> (%)	<i>R</i> (%)	<i>F1</i>
K-Means	0.6719	81.04	68.96	85.30	0.7627
ADVENT ^[17]	0.7092	83.79	73.86	84.52	0.7883
RoadDA ^[20]	0.7190	84.30	73.26	88.21	0.8004
SFDA(Ours)	0.7677	87.92	84.56	80.95	0.8272

3.5. Ablation experiment

To further validate the effectiveness of each module, ablation experiments are conducted, and the results are summarized in Tab. II. First, the performance of models equipped with only the DSA module or only the FDA module is evaluated. The corresponding mean mIoU values reach 0.7427 and 0.7513, respectively. When both DSA and FDA are jointly incorporated, the mIoU is further improved to 0.7677. These results demonstrate the effectiveness and complementary characteristics of the proposed modules within the overall framework. See **Table 2**.

Table 2. Results of ablation experiments

DSA	FDA	<i>mIoU</i> (%)	<i>OA</i> (%)	<i>F1</i>
		0.7092	83.79	0.7883
✓		0.7427	86.19	0.8105
	✓	0.7513	86.92	0.8136
✓	✓	0.7677	87.92	0.8272

4. Conclusion

This paper presents a structure- and frequency-aware domain adaptation framework, SFDA, for cross-region raft aquaculture extraction in SAR images. The framework includes the DSA module and the FDA module to align features between the source and target domains from both the structure and frequency views. This helps reduce the domain gap caused by different regions in SAR data and improves segmentation accuracy and generalization.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Pörtner H O, Scholes R, Arneth A, et al., 2023, Overcoming the Coupled Climate and Biodiversity Crises and Their Societal Impacts. *Science*, 380(6642): eabl4881.
- [2] Fan J, Zhao J, Song D, et al., 2018, Marine Floating Raft Aquaculture Dynamic Monitoring Based on Multi-Source GF Imagery. 2018 7th International Conference on Agro-Geoinformatics, 1–4.
- [3] Chen W, Li X, 2024, Deep-Learning-Based Marine Aquaculture Zone Extractions from Dual-Polarimetric SAR Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17: 8043–8057.
- [4] Wang J, Fan J, Wang J, 2022, MDOAU-Net: A Lightweight and Robust Deep Learning Model for SAR Image Segmentation in Aquaculture Raft Monitoring. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.
- [5] Fan J, Zhao J, An W, et al., 2019, Marine Floating Raft Aquaculture Detection of GF-3 PolSAR Images Based on Collective Multikernel Fuzzy Clustering. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8): 2741–2754.
- [6] Wang X, Zhou J, Fan J, 2022, IDUDL: Incremental Double Unsupervised Deep Learning Model for Marine Aquaculture SAR Images Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–12.
- [7] Zhao J, Li Y, Zhou Y, et al., 2025, DDCL: Unsupervised Domain Adaptation for Remote Sensing Images Based on Diffusion Causal Distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–12.
- [8] Zhu J, Guo Y, Sun G, et al., 2024, Causal Prototype Inspired Contrast Adaptation for Unsupervised Domain Adaptive Semantic Segmentation of High-Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–17.
- [9] Zhu C, Liu K, Tang W, et al., 2025, Hard-Aware Instance Adaptive Self-Training for Unsupervised Cross-Domain Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 5655–5671.
- [10] Shi Y, Du L, Li C, et al., 2024, Unsupervised Domain Adaptation for SAR Target Classification Based on Domain- and Class-Level Alignment: From Simulated to Real Data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 207: 1–13.
- [11] Yang Y, Chen J, Sun L, et al., 2024, Unsupervised Domain-Adaptive SAR Ship Detection Based on Cross-Domain Feature Interaction and Data Contribution Balance. *Remote Sensing*, 16(2): 420.
- [12] Liu S, Li D, Song H, et al., 2025, SAR Ship Detection Across Different Spaceborne Platforms with Confusion-Corrected Self-Training and Region-Aware Alignment Framework. *ISPRS Journal of Photogrammetry and Remote Sensing*, 228: 305–322.
- [13] Yang Y, Yang X, Yang D, 2025, Unsupervised Domain Adaptation for SAR Ship Detection Based on Multitask

- Decoupling. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18: 12684–12696.
- [14] Tu H, Wang W, Guo Y, et al., 2025, Mamba-UDA: Mamba Unsupervised Domain Adaptation for SAR Ship Detection. *IEEE Geoscience and Remote Sensing Letters*, 22: 1–5.
- [15] Ren Z, Du Z, Zhang Y, et al., 2024, Multi-Step Unsupervised Domain Adaptation in Image and Feature Space for Synthetic Aperture Radar Image Terrain Classification. *Remote Sensing*, 16(11): 1901.
- [16] Cui G, Fan J, Zou Y, 2025, Enhanced Unsupervised Domain Adaptation with Iterative Pseudo-Label Refinement for Inter-Event Oil Spill Segmentation in SAR Images. *International Journal of Applied Earth Observation and Geoinformation*, 139: 104479.
- [17] Vu T, Jain H, Bucher M, et al., 2019, ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2512–2521.
- [18] Arthur G, Karsten M B, Malte J R, et al., 2012, A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13(25): 723–773.
- [19] Fan J, Li M, Wang X, 2025, Unsupervised Transformer with Generative Label Optimization for Marine Aquaculture Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–14.
- [20] Zhang L, Lan M, Zhang J, et al., 2022, Stagewise Unsupervised Domain Adaptation with Adversarial Self-Training for Road Segmentation of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–13.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on Development and Performance of Intelligent Protective Mask Based on Flexible Heating and Piezoresistive Sensing Technology

Ziyi Dang¹, Miao Wang¹, Xiangyi Hou¹, Mingze Sun¹, Yin He^{1,2*}

¹School of Textile Science and Engineering, Tiangong University, Tianjin 300387, China

²Shaoxing Keqiao Research Institute, Tiangong University, Shaoxing 312030, China

*Corresponding author: Yin He, heyin@tiangong.edu.cn

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Aiming at the problems of worsening air pollution and the single function of traditional masks, this paper develops an intelligent mask integrating heating disinfection and respiratory monitoring. A copper wire heating layer is fabricated by hot pressing on a non-woven lining and combined with a sensing film to realize the integration of temperature regulation and respiratory monitoring. The results show that the mask can rise to 60 °C within 50 s and reach 72 °C in 200 s, with stable multi-level temperature control at 30 °C, 40 °C, and 70 °C. It can effectively identify different breathing patterns: the resistance change rate of normal breathing is 5–10%, deep breathing is 20–30%, and shallow breathing is 2–5%. The peak value exceeds 30% after exercise and returns to the baseline within 3–5 s, showing excellent respiratory recognition ability. This research provides a new idea for the development of intelligent protective equipment integrating environmental adaptability and health monitoring.

Keywords: Flexible heating; Flexible piezoresistive sensor; Intelligent mask; Respiratory monitoring

Online publication: Jul 27, 2026

1. Introduction

Traditional protective masks mainly block airborne particulate matter and droplets through filtration and physical barrier functions, playing an important role in preventing the spread of respiratory infectious diseases. However, traditional masks have limitations such as limited protection capacity, high breathing resistance, poor wearing comfort for long-term use, and non-reusability^[1–4]. With the development of intelligent technology, intelligent masks have become a research hotspot^[5–8]. Domestic institutions and Xiamen University, including Xiaomi, Shanghai Yuantian Electronic Technology Co., Ltd., and Xiamen University, have taken the lead in developing intelligent masks with respiratory monitoring and physiological data analysis functions. Nevertheless, existing intelligent masks still suffer from a single function and low

reusability.

The core technology of flexible heating clothing elements lies in hot-press bonding. Copper wires are arranged in a specific circuit on non-woven fabrics and bonded by hot pressing to form flexible heating liners. Such materials possess excellent electrothermal, thermal, and mechanical properties as well as temperature control performance, which can flexibly adjust heating output according to actual demands and maintain structural integrity and functionality during long-term service^[9–11]. Flexible piezoresistive sensors are prepared by the solution-blending method. Modified graphite/polyurethane (MGT/PU) composite films exhibit good electrical conductivity, mechanical performance, and sensitivity, suitable for wearable health monitoring systems to collect real-time physiological signals such as pulse and blood pressure^[12–17].

In view of the increasingly severe air pollution and the single function of conventional masks, this study develops an intelligent protective mask with self-cleaning, health monitoring, and comfort improvement functions by combining flexible heating and piezoresistive sensing technologies^[18–20]. The mask not only meets the high-efficiency protection needs in special environments, but also provides personalized temperature regulation and health status monitoring for users, greatly improving wearing comfort and service safety.

2. Experimental materials and methods

2.1. Experimental materials and instruments

2.1.1. Main materials

0.1 mm red copper wire (Quanzhou Jiadai New Material Co., Ltd.); 20 g/m² non-woven fabric (Guangzhou Nanqixing Non-woven Fabric Co., Ltd.); adhesive interlining (Erqi District Yongzhuo Garment Accessories Store, Zhengzhou); CNT/PU composite film.

2.1.2. Experimental equipment

Self-developed fabric thermal performance testing system, consisting of infrared camera (TP-8, Wuhan Guide Infrared Co., Ltd.), Agilent multimeter, DC regulated power supply and infrared image acquisition software; the infrared camera has a resolution of 384 × 288 pixels with a shooting distance of 30 cm from the sample platform; flexible sensor performance detection system; digital dual-display multimeter (U3402A, Agilent, USA).

2.2. Preparation of flexible heating non-woven functional layer

Copper wires are arranged along marked routes and carefully placed on non-woven fabric. A small amount of glue is dipped with a toothpick and applied to the fixing positions to paste copper wires onto the mask. Longitudinal copper wire circuits are arranged on both sides to reserve space for the nasal sensing film. Conducting wires are led out from appropriate positions of the copper circuit for power and sensor connection, with copper foil serving as electrode terminals.

2.3. Integrated fabrication of an intelligent protective mask

A flexible pressure sensor layer is constructed based on the pressure difference generated during breathing to explore the response performance for respiratory signal acquisition. The sensing film and copper foil electrodes are arranged on a non-woven lining to fabricate multifunctional piezoresistive sensors.

The functional layer is integrated with a non-woven fabric and a melt-blown layer to assemble the intelligent mask, and its heating circuit and sensor response characteristics are investigated. The heating circuit and strain sensor are arranged at designated positions of medical protective masks to fabricate the intelligent mask (**Figure 1**).

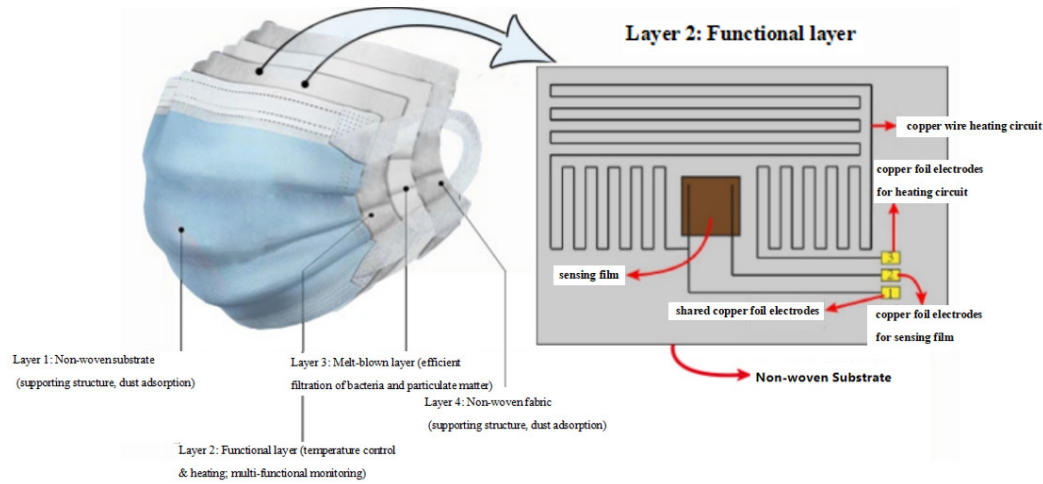


Figure 1. Structure of the intelligent protective mask.

2.4. Performance test of the intelligent protective mask

After integration, the mask is worn simulating the actual wearing state for the heating performance test by the fabric thermal performance tester. Volunteers wear the mask to simulate four breathing states: normal breathing, deep breathing, shallow breathing, and post-exercise heavy breathing. The flexible sensor detection system and digital multimeter are adopted to monitor resistance changes.

3. Results and discussion

3.1. Analysis of heating performance

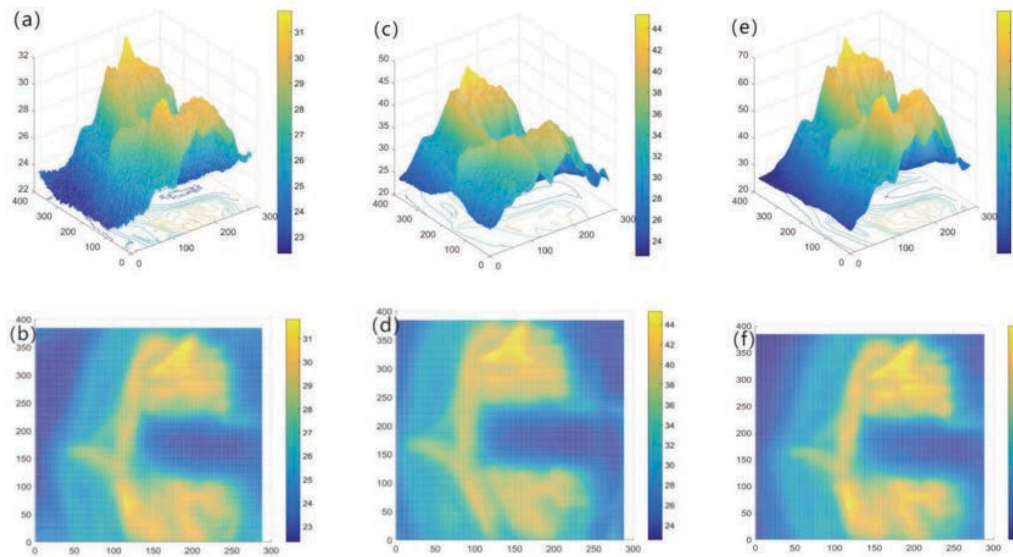


Figure 2. (a-b) 3D infrared thermography and thermal images at first gear; (c-d) 3D infrared thermography and thermal

images at second gear; (e-f) 3D infrared thermography and thermal images at third gear.

When switched to the first gear (green indicator), the maximum temperature reaches about 30 °C with an average temperature around 26 °C, providing a comfortable wearing experience for rhinitis patients. At the second gear (blue indicator), the maximum temperature rises to about 40 °C, supplying warm inhaled air in cold weather. At the third gear (red indicator), the maximum temperature approaches 70 °C; continuous heating for 30 min achieves a sterilization effect.

3.2. Respiratory monitoring performance

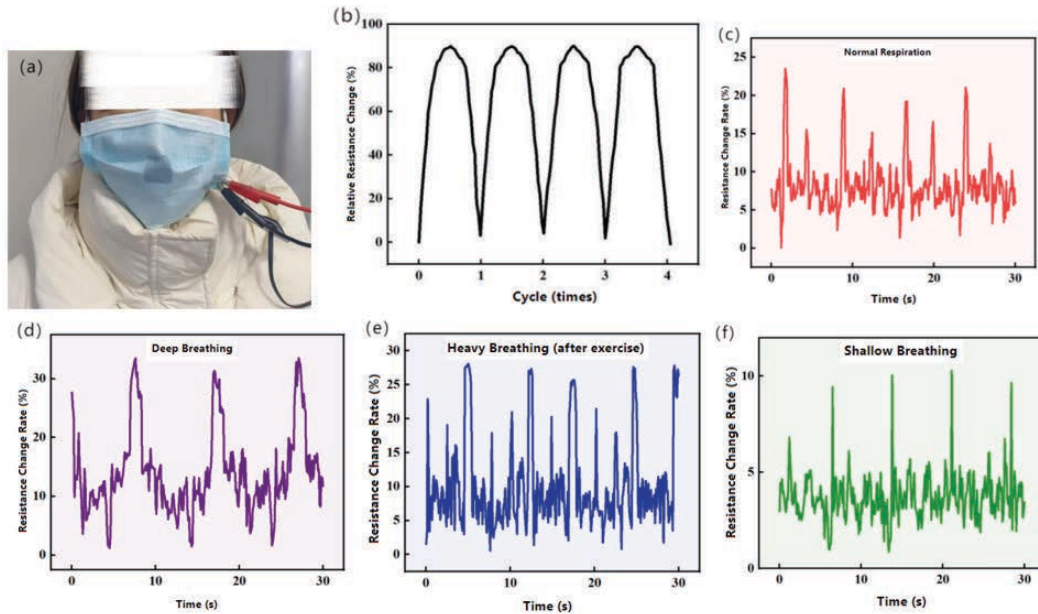


Figure 3. (a) Physical wearing photo; (b) Relative resistance change rate of multiple cycles; (c-f) Resistance change rate under different breathing patterns.

The PU-modified graphite sensor presents typical periodic response characteristics. During normal breathing, resistance rises regularly in inhalation and drops in exhalation, highly consistent with the respiratory rhythm. The excellent conductivity of modified graphite enables sensitive capture of slight thoracic undulation, while the good flexibility of the PU substrate ensures stable sensor performance under regular deformation. The resistance change rate maintains 5–10% under normal breathing with good baseline stability and repeatability.

Under deep breathing, the thoracic expansion increases sensor stretching, extending the conductive path and raising the resistance change rate to 20–30%. The sensor shows fast response (< 0.5 s) and high signal fidelity, effectively distinguishing deep breathing from normal breathing. After strenuous exercise, the respiratory frequency and depth increase; the peak resistance change rate exceeds 30% with a baseline recovery time of 3–5 s. Under shallow breathing, the resistance change rate drops to 2–5% with slight signal fluctuation due to irregular breathing rhythm. The sensor maintains stable signal consistency and repeatability during long-term use, providing reliable technical support for clinical applications of intelligent masks.

4. Conclusion

This study develops an intelligent protective mask integrating flexible heating and piezoresistive sensing, realizing integrated heating disinfection and respiratory monitoring. With a copper wire spacing of 0.5 cm and an applied voltage of 8 V, the fabric temperature rises to 60 °C within 50 s and reaches 72 °C in 200 s. Steady temperature control at 30 °C, 40 °C, and sterilization.70 °C is achieved via three gears, meeting the demands of warmth, comfort, and sterilization. The flexible piezoresistive sensor exhibits high sensitivity and recognition capability: resistance change rate is 5–10% for normal breathing, 20–30% for deep breathing, and 2–5% for shallow breathing; the peak value exceeds 30% after exercise with baseline recovery within 3–5 s, realizing effective identification of various breathing patterns. The intelligent mask shows stable heating response and respiratory recognition performance, possessing broad application prospects.

Funding

Tianjin College Students Innovation Training Program (Project No.: 202410058077);the Zhejiang Province Applied Basic Research Program (Project No.: 2026C02A1070)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhang B, Zheng Y, Qiu Y, et al., 2020, Overview and New Technology Research of Intelligent Masks. *Tianjin Textile Science & Technology*, 2020(3): 1–6.
- [2] Liu R, Xing D, Zhen X, et al., 2023, Research Progress of Functional Masks Against COVID-19. *Medical Equipment*, 44(1): 97–106.
- [3] Zhang C, Zhu G, Zhang Y, et al., 2025, Research on Polylactic Acid Nanofiber Membrane for Respiratory Monitoring and PM0.3 Filtration. *Membrane Science and Technology*, 45(1): 48–57 + 69.
- [4] Langman L E, 2025, Commentary on Female Mask (Unknown). *Academic Medicine*, 100(3): 289.
- [5] Huang Q, Jiang Y, Duan Z, et al., 2023, Electrochemical Self-Powered Strain Sensor for Static and Dynamic Strain Detections. *Nano Energy*, 118: 108997.
- [6] Yixiao C, Hui S, Quhong S, et al., 2024, A Review of Utility of Wearable Sensor Technologies for Older Person Frailty Assessment. *Experimental Gerontology*, 200: 112668.
- [7] Zhao L, Ferraro P, Shorten R, 2024, A Smart Mask to Enforce Social Contracts Based on IOTA Tangle. *PLoS One*, 19(3): 0292850.
- [8] Ullah A, Zulfiqar H M, Khan A M, et al., 2025, Fast-Response Humidity Sensor-Based Smart Face Mask for Multifunctional Applications. *Sensing and Imaging*, 25(1): 43.
- [9] Bai Y, 2018, Preparation and Performance Research on Flexible Heating Clothing Elements, thesis, Tiangong University.
- [10] Zhao L, 2020, Preparation and Application of Flexible Piezoresistive Sensors Based on Carbon Nanotubes, thesis, Tiangong University.
- [11] Boukhamla A, Azizi N, Belhaouari B S, 2025, Intelligent Mask Image Reconstruction for Cardiac Image Segmentation. *Applied Intelligence*, 55(4): 257.
- [12] He Y, Qiao Z, Liu H, et al., 2024, Preparation and Performance of Wearable Pulse Monitoring Sensors. *Cotton*

Textile Technology, 52(11): 1–8.

- [13] Wang Z, 2018, Research on Respiratory Measurement Technology of Intelligent Mask Based on Air Pressure Sensor. *Mechanical & Electrical Information*, 2018(21): 131–133.
- [14] He S, Zhang Y, Zhang G, et al., 2019, Experimental Study on Human Respiratory Parameter Changes Under Different Labor Intensities. *China Safety Science and Technology*, 15(5): 167–172.
- [15] Ma L, Wu R, Patil A, et al., 2019, Full-Textile Wireless Flexible Humidity Sensor for Human Physiological Monitoring. *Advanced Functional Materials*, 29(43): 1904549.
- [16] Bosry R S, Arpita D, Tanfeez F M, 2024, Fluid-Structure Interaction of a Thin Flexible Heater Inside a Lid-Driven Cavity. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 238(1): 75–93.
- [17] Obara S, Hirata N, Hagihira S, 2024, What Are Standard Monitoring Devices for Anesthesia in Future? *Journal of Anesthesia*, 38(4): 537–541.
- [18] Exley T, Johnson D, Jafari A, 2024, Comparative Analysis of Peltier Devices and Flexible Heater Strips for Thermo-Active Soft Actuators. *Journal of Medical Robotics Research*, 9(3–4): 2450002.
- [19] Liu J, Xiang H, Wang W, 2024, Highly Stable, Bending-Tolerant, and Sustainable Flexible Heater Through a Scalable Papermaking Procedure. *Materials*, 17(14): 3507.
- [20] Min L, Tao W, Xiang-Le L, 2023, Highly Stable Phosphotungstic Acid/Au Dual Doped Carbon Nanotube Transparent Conductive Films for Flexible Heaters. *Carbon*, 207: 219–229.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Research on Real-Time Intelligent Rendering Method for Immersive Interactive Virtual Reality Scenes

Wenzhen Wang

School of Culture And Communication, Zibo Polytechnic University, Zibo 255300, Shandong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Virtual reality scenes oriented towards immersive interaction have strict requirements for real-time rendering, such as high frame rate, low latency, and visual consistency. Traditional rendering methods are difficult to ensure both real-time performance and immersive experience under limited computing resources. This article systematically explores the key technologies and implementation paths of intelligent rendering methods: firstly, it analyzes the rendering pressure bottleneck, multi view continuity, and visual dual constraints caused by immersive interaction; Furthermore, an intelligent rendering mechanism based on gaze point prediction, scene semantic modeling, and closed-loop resource allocation will be constructed; On this basis, adaptive strategies such as behavior driven detail level adjustment, dynamic and static differentiation processing, spatiotemporal consistency, and load coordination are proposed; Finally, the evaluation indicators, stability analysis, and scalability verification of real-time performance and immersion are discussed. This article provides a system framework for the theoretical construction and engineering implementation of intelligent rendering methods.

Keywords: Immersive interaction; Real-time intelligent rendering; Fixation point prediction; Adaptive level of detail

Online publication: Jul 23, 2026

1. Introduction

With the rapid development of virtual reality technology, immersive interaction puts extremely high requirements on real-time performance and visual authenticity of scene rendering, while the traditional graphics rendering pipeline is limited by fixed logic and limited computing capacity, making it difficult to simultaneously guarantee immersive experience and computational efficiency in high frame rate, low latency dynamic interactive environments^[1]. Therefore, for virtual reality scenes oriented to immersive interaction, real-time intelligent rendering methods have become a research hotspot, the core of which lies in introducing intelligent perception, dynamic resource allocation, and adaptive strategies to break through existing rendering bottlenecks. This paper aims to systematically explore technical challenges in this field,

construct an intelligent rendering mechanism oriented to immersive features, and also analyze scene adaptive strategies, thus exploring the path of balance between real-time performance and immersiveness, in order to provide a theoretical reference for subsequent research and engineering practice.

2. Immersive interaction poses technical challenges for real-time rendering

2.1. Rendering pressure bottleneck under high frame rate and low latency

Immersive interaction requires rendering frame rate to reach 90 FPS or higher, and end-to-end latency also needs to be below 20 milliseconds. Users are prone to dizziness and perceptual disorders, which places direct pressure on the traditional rendering pipeline, with available computation time per frame being compressed to approximately 11 milliseconds, and in high-resolution headsets, pixel filling and geometry processing volume grow exponentially. General rendering algorithms such as ray tracing or complex shading are difficult to complete full computation within this time window, thus forcing system to make trade-off between image quality and response speed, more importantly, unpredictability of interactive behavior further amplifies above pressure, when users suddenly turn their head or move, rendering system must regenerate complete view within an extremely short time, with insufficient precomputation buffer available ^[2]. And breaking through hard constraints of frame rate and latency has become a fundamental bottleneck that intelligent rendering methods must prioritize to solve.

2.2. Continuity requirements for multi-perspective and dynamic scenes

Immersive interaction not only requires high-quality rendering from a single viewpoint, but also needs to maintain spatiotemporal consistency among multiple viewpoints during user's continuous head movement, with viewpoint moving at speed of several meters per second, scene projection relationship between adjacent frames will undergo drastic changes, and any rendering inconsistency, such as shadow jumping, texture flickering and geometric misalignment, will be immediately captured by visual system, thereby destroying sense of immersion ^[3]. Moving objects in dynamic scenes further aggravate this challenge, as their positions, postures, and occlusion relationships change rapidly, which requires the rendering system to complete verification and update of multi-viewpoint consistency within an extremely short period of time ^[4]. Traditional methods based on inter-frame difference or simple interpolation are difficult to simultaneously meet the requirements of speed and continuity, thus producing motion ghosting or afterimages, with building a rendering continuity mechanism that supports smooth transition among multiple viewpoints being a prerequisite for achieving immersion.

3. Construction of an intelligent rendering mechanism for immersive features

3.1. Intelligent region refinement rendering based on gaze point prediction

Gaze point prediction technology, by tracking movement of user's eyeballs, estimates in real time center of attention in current field of view, and guides rendering resources to lean toward that region, with the human visual system having physiological characteristics, which determines that fovea is only highly sensitive to details within about 2 degrees of center of visual field, while peripheral regions show significantly increased tolerance for resolution, shadow precision, and texture details ^[5]. Intelligent rendering methods can cleverly utilize this characteristic, performing full-precision rendering in gaze point region, that is, high-resolution

textures and precise lighting calculations, and gradually reducing sampling rate and shading complexity in peripheral regions, with the key lying in foresight of prediction algorithm, since rendering pipeline has inherent latency, thus requiring the system to predict landing point of gaze point several frames in advance, to avoid obvious resolution jumps when user turns their head ^[6]. Combined with positional encoding and attention mechanisms, deep learning plays an important role here, and the prediction accuracy of the gaze point can reach the sub-pixel level, thereby providing reliable support for region-refined rendering.

3.2. Perception sensitivity modeling based on scene semantics

Elements of different scenes contribute significantly differently to the immersive experience, and an intelligent rendering mechanism should dynamically adjust the rendering priority of objects according to the semantic category of objects. Semantic objects, such as dynamic characters, interactive objects, and text information, are often the focus of user attention, and their rendering defects (such as jagged edges, update delays, and missing shadows) are easily perceived. However, subtle degradation of distant backgrounds, static decorations, and repeated textures is not easily noticeable, while perceptual sensitivity modeling is a quantitative description of this difference, which decomposes 3D scene into a semantic map through pre-trained scene understanding network, and assigns sensitivity weights to each type of object, with this weight being not only determined by category of object, but also influenced by its spatial distance, motion state, and position relative to gaze point ^[7].

3.3. Dynamic trade-off logic between intelligent clipping and global features

Rendering resources are limited, and at this time, the intelligent clipping mechanism plays a role, which is responsible for identifying and removing visual information that contributes less to the sense of immersion, while retaining global features that are sufficient to maintain the understandability of the scene. Unlike traditional frustum clipping or occlusion clipping, intelligent clipping not only considers geometric visibility but also incorporates semantic importance and scene coherence constraints ^[8]. There is a small fast-moving object in distance, which can be temporarily clipped without attracting attention, but if object is only reference point indicating direction of movement, such as a flight guidance arrow, minimum visual presentation of it must be retained, with the logic of dynamic balance playing a core role in this process, thereby adjusting clipping threshold in real time according to currently available computing budget, complexity of scene, and user interaction state.

3.4. Closed-loop adaptive allocation between rendering tasks and computing resources

The resource allocation problem of real-time rendering is essentially an online optimization problem under dynamic constraints, which requires the construction of a closed-loop feedback system between rendering tasks and computing resources. The input of the system includes the rendering task queue for the current frame (each task comes with an estimated cost and priority), available GPU computing power, video memory bandwidth, and power consumption limit; The output end is task scheduling and resource allocation instructions ^[9]. The difficulty of closed-loop control lies in the strong coupling between task cost and resource state—high-priority gaze point rendering may occupy other tasks, leading to global frame time exceeding the limit. The intelligent allocation method adopts a cyclic architecture of prediction execution evaluation adjustment: based on historical frame data, it predicts the resource requirements of the next frame, evaluates

whether the frame rate target is met after allocation, and if it deviates, the allocation strategy is corrected through a proportional integral derivative controller or a learning scheduler.

4. Scene adaptive rendering strategy under intelligent driving

4.1. Detail level strategy for real-time adjustment based on user interaction behavior

The core of the level of detail strategy is to use user interaction behavior as a direct trigger signal for rendering resource scheduling, rather than relying solely on scene geometric distance. Traditional LOD based on line-of-sight static mapping cannot reflect interactive intent—the object that the user is staring at and preparing to grab should maintain high accuracy even if it is far away, while distant non-interactive backgrounds can be greatly simplified^[10]. Intelligent methods dynamically identify potential interaction targets and improve their LOD level in advance by analyzing interaction trajectories such as handle ray pointing, hand bone movement, or head approach speed. At the same time, an interaction urgency measure is introduced to enable expensive effects such as subsurface scattering and high-precision normal mapping for objects that are about to be touched, and conservative LOD is performed on areas with no signs of interaction in the near future.

4.2. Differentiated processing mechanism for dynamic objects and static environments

In the scene, there is a fundamental difference in rendering consistency between dynamic objects and static environments. Two have different needs for differentiated processing pipelines, static environments such as architecture and terrain have pre-computable properties, with their lighting information, along with shadow maps and global illumination caches being baked in advance and reused across multiple frames. In contrast, dynamic objects such as moving characters and interactive props need to recalculate their positions, poses, and real-time occlusion relationships with the environment every frame^[11]. Intelligent rendering mechanism makes full use of this difference, allocating low-frequency update channels for static environments, thereby adopting lower refinement frequencies and coarser spatial sampling rates, allocating high-frequency update channels for dynamic objects to ensure smooth motion trajectories and no visual breaks at seams with static scenes. The core challenge lies in interactive lighting between dynamic objects and static environments. Dynamic objects cast moving shadows and also receive environmental reflections, which requires lightweight real-time global illumination approximation solutions (such as radiometric probes or screen space reflections) to maintain visual fusion between the two and avoid a ‘floating feeling’ that breaks immersive experience.

4.3. Intelligent collaboration of spatiotemporal consistency and instantaneous computing load

Spatiotemporal consistency requires that rendered results maintain smooth transitions between continuous time frames and adjacent spatial viewpoints, but this has an inherent contradiction with dynamic fluctuation of instantaneous computation load, when system detects that load is too high, it usually takes degradation measures, such as reducing resolution and turning off special effects, however if these operations suddenly change within single frame, they may cause visual flickering or texture jumping, thereby directly destroying spatiotemporal consistency^[12]. Intelligent collaborative method solves this contradiction by introducing ‘smooth degradation corridor’, which sets multiple continuous visual quality levels, and difference between each level is progressive step that is not easily perceived by human eye, when load exceeds limit, system

does not directly jump levels, but slides frame by frame along corridor to target level, with temporal anti-aliasing and motion vector compensation being combined to conceal traces of change, and vice versa, upgrading frame by frame as load decreases. In addition, the system also needs to estimate the load trend of future frames, thus avoiding frequent changes in quality levels caused by short-term peaks.

4.4. Rendering strategy and minimization model of immersive experience loss

The essence of any intelligent rendering strategy is to minimize the immersive experience loss perceived by users under computational resource constraints, requiring the establishment of formal loss functions and optimization frameworks. Define immersion loss as a weighted sum of multidimensional indicators, including visual fidelity loss (such as texture blur, shadow loss), temporal discontinuity loss (such as inter-frame flicker, motion drag), and interactive response loss (such as operation feedback delay). Each strategy action—adjusting LOD, enabling clipping, changing sampling rate—will result in different loss distributions in the aforementioned dimensions. The core of the minimization model is to find the optimal strategy combination for each frame, so as to minimize the total immersion loss under a given computational budget. Due to the large strategy space and non-convex loss function, traditional exhaustive search is not feasible.

5. Balance and verification path between real-time performance and immersion

5.1. Construction of evaluation indicators for real-time rendering performance and immersive perception intensity

The evaluation indicators should cover both the objective performance of the system and the subjective immersion experience of users. Objective indicators include frame rate, frame time variance, end-to-end latency, and rendering load utilization, which can be directly measured. The subjective immersion intensity needs to be calibrated through psychophysical experiments, such as using a dual task paradigm to measure the user's perception threshold of scene changes, or inferring the level of attention engagement through head movement trajectories and gaze point distributions. Suggest building a composite indicator called 'immersion maintenance rate', defined as the ratio of the user's continuous immersion time to the total usage time under given hardware constraints. At the same time, the ratio of perceived quality score to system overhead is introduced as an efficiency indicator to avoid unthinkingly pursuing high image quality and wasting computing power. The standardization of the indicator system provides a quantitative benchmark for subsequent algorithm comparisons.

5.2. Stability analysis of intelligent methods under different scene complexities

Intelligent rendering methods need to maintain stable performance in environments with significant differences in complexity, such as open terrain, dense urban scenes, and enclosed indoor spaces. The dimensions of stability analysis include frame rate fluctuation coefficient, rendering quality variance, and strategy switching frequency. Using the method of controlling variables: a fixed hardware platform is used to sequentially run the same intelligent algorithm in large-scale vegetation scenes, multi-dynamic character scenes, and high reflection material scenes, and record the degree of performance degradation in each scene. The analysis should focus on two failure modes: one is the strategy oscillation that occurs when the scene suddenly changes, manifested as frequent LOD level jumps; The second is the crashing degradation of the system after the complexity exceeds the threshold, that is, the frame rate drops sharply below the dizziness

threshold. By extracting the correlation between feature parameters of different scenarios and algorithm performance, the applicable boundaries and vulnerable links of intelligent methods can be identified.

5.3. Threshold discussion on the impact of delay and visual feedback deviation on immersion

The disruption of immersion caused by delay and visual deviation has nonlinear characteristics and clear perceptual thresholds. Experimental studies have shown that users are almost imperceptible when the end-to-end latency is less than 15 milliseconds; 15 to 25 milliseconds can only be perceived in high dynamic motion; If it exceeds 30 milliseconds, it generally causes obvious discomfort and disconnection from operation. Visual feedback bias includes perspective drift, texture mapping misalignment, and motion ghosting. Its tolerance threshold is influenced by the type of bias and scene semantics: static bias can be ignored within a 0.5-degree perspective, while dynamic bias exceeding 0.1 degrees may be captured per frame. The discussion should further distinguish the different sources of delay—rendering delay, transmission delay, and display delay have different weights for disrupting immersion, with the former having the highest sensitivity. Determining these thresholds helps to set quantifiable engineering constraint goals for intelligent rendering algorithms.

5.4. Verification of scalability and cross-scene adaptability of intelligent rendering strategy

Scalability requires intelligent rendering strategies to maintain effective degradation rather than complete failure on heterogeneous computing platforms, from mobile VR to workstation level PC. The verification method adopts cross-platform comparative experiments: under the same algorithm framework, the maximum allowable frame time constraint is relaxed from 6 milliseconds to 25 milliseconds in sequence, and the upward curve of the immersion loss index is observed. The adaptability is verified through scenario transfer testing: the strategy trained in Class A scenarios is directly deployed to Class B scenarios, and the degree of performance degradation is measured. The ideal intelligent strategy should have zero sample generalization ability, which means it can maintain acceptable immersion quality in scenes of different geometric complexities, material features, and dynamic densities without retraining. If the attenuation is significant, meta-learning or domain adaptation modules need to be introduced to improve cross-scene robustness.

6. Conclusion

In summary, the core of the real-time intelligent rendering method for immersive interactive virtual reality scenes lies in breaking through the hard constraints of traditional rendering pipelines in frame rate, latency, and resource allocation, and establishing a dynamic optimization mechanism centered on user perception. Starting from the technical challenges of immersive interaction, this article analyzes the requirements of high frame rate, multi view, and visual style dual constraints on rendering systems; Furthermore, an intelligent rendering mechanism based on gaze point prediction, scene semantic modeling, and closed-loop resource allocation was constructed, and adaptive strategies such as behavior driven LOD adjustment, dynamic and static differentiation processing, spatiotemporal consistency, and load coordination were proposed. On this basis, an evaluation index system that includes objective performance and subjective immersion intensity is proposed, and the issues of delay threshold and cross-scene adaptability are discussed. Future research can

further integrate brain-computer interfaces with generative AI to achieve more predictive and personalized immersive rendering.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Liu Y, Yang S, Yang Y, et al., 2026, Research Progress on Gaze-Based Rendering Techniques for Visual Comfort Optimization in Virtual Reality (Invited). *Acta Optica Sinica*, 46(9): 85–103.
- [2] Wu R, 2026, Application and Challenges of Digital Media Technology in Constructing Virtual Reality Scene Systems. *Software*, 47(3): 123–125.
- [3] Wang X, 2026, Real-Time Rendering Optimization Algorithms for Virtual Reality Scenes Based on Deep Learning. *Journal of Xinyang Agriculture and Forestry University*, 36(1): 89–96.
- [4] Jiang D, Yan X, 2025, Research on Real-Time Image Rendering Algorithms in Virtual Reality Environments. *Information and Computers*, 37(22): 10–12.
- [5] Zhang F, 2025, Application of Virtual Reality Technology in Film and Television Production. *Television Technology*, 49(10): 157–159.
- [6] Wang Z, 2025, GPU-Accelerated Methods for Real-Time Rendering Optimization in Virtual Reality Scenes. *China New Technology & New Products*, 2025(8): 34–36.
- [7] Wang M, 2025, Analysis of the Integration of Virtual Reality and Electronic Information Processing Technology. *Electronics Technology*, 54(2): 306–307.
- [8] Wang S, Ren D, 2024, Maze, Miniature Worlds, and Experimental Fields: Investigating Landscape Experiences in Real-Time Interactive Films. *Contemporary Cinema*, 2024(11): 111–118.
- [9] Liang Z, Hao X, 2024, Virtual Simulation Modeling and Rendering Technology Based on Virtual Reality. *Science & Technology Innovation and Application*, 14(14): 37–40.
- [10] Zhang R, 2024, Integration and Challenges of Virtual Reality and 3D Rendering Technology in Game Design. *Toy World*, 2024(4): 118–120.
- [11] Gao Q, 2023, Virtual Space Interaction Systems Based on Unity3D. *Information Recording Materials*, 24(9): 176–178.
- [12] Zeng X, Huang Z, Dong J, 2023, Improvement and Implementation of Scene Rendering Methods in Virtual Reality Power Training Systems. *Electric Technology*, 2023(1): 68–71.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Exploration of Digital Twin-Driven Heterogeneous Terminal Linkage and Intelligent Control Technology in the Internet of Things

Tao Zhang

School of Integrated Circuits and Electronic Information, Zibo Polytechnic University, Shandong Zibo, 255300, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Digital twin technology provides a new theoretical path for solving the difficulties of heterogeneous terminal interconnection and low control efficiency in the Internet of Things. Starting from the endogenous logic of empowering IoT terminals with digital twins, this article analyzes the mirror mapping and closed-loop interaction mechanism between physical entities and virtual models, achieving a paradigm leap from data synchronization to semantic collaboration. On this basis, a digital twin collaboration architecture for heterogeneous terminal linkage was constructed, and core issues such as unified twin model construction, virtual real dynamic association maintenance, twin inter collaboration triggering, and cross protocol communication decoupling were discussed. Then focus on the key technologies of intelligent management and control based on digital twins, including real-time data stream processing and edge computing integration, terminal state autonomous awareness and exception diagnosis, policy adaptive generation under the centralized-decentralized hybrid architecture, and the management and control evolution path from rule-driven to target-driven. The research provides a systematic theoretical framework for intelligent linkage and efficient control of large-scale heterogeneous IoT terminals.

Keywords: Digital twin; Internet of things; Heterogeneous terminals; Intelligent control

Online publication: Jul 23, 2026

1. Introduction

With the rapid expansion of the Internet of Things, the significant differences in communication protocols, data models, and computing capabilities among massive heterogeneous terminals have led to difficulties in terminal interconnection and complex unified control. Traditional rule-based or centralized gateway-based methods are increasingly exposed to limitations in terms of dynamism, scalability, and semantic interoperability. Digital twin technology, with its ability to establish precise mapping, real-time interaction, and closed-loop optimization between physical entities and virtual spaces, provides a new theoretical perspective for solving the linkage and control problems of heterogeneous terminals in the Internet of Things

^[1]. This article aims to systematically explore the technological path of digital twin-driven heterogeneous terminal linkage and intelligent control in the Internet of Things. Firstly, analyze the interconnection limitations and control difficulties of heterogeneous terminals in the Internet of Things, and reveal the endogenous logic of digital twin empowerment; Furthermore, a digital twin collaboration architecture for heterogeneous terminal linkage is proposed, exploring the construction of a unified twin model, virtual real dynamic mapping, and the mechanism of collaboration between twin entities; Subsequently, focus on key technologies such as real-time oriented data stream processing, autonomous state perception, policy adaptive generation under hybrid architecture, and evolution from rule driven to goal driven; Finally, summarize the core values and future directions of digital twins in this field.

2. Digital twins empower the endogenous logic of IoT terminals

2.1. Internet of things heterogeneous terminal interconnection limitations and control challenges

In the Internet of Things, terminal devices are highly heterogeneous in terms of communication protocols, formats, computing power, and membership domains, which limits their interconnectivity and unified control. Some vendors use private interfaces and closed data models to sense, execute, and edge nodes, and these terminals share no standard semantic interoperability between terminals and high levels of information silos. While traditional gateway or middleware solutions can partly solve protocol conversion problems, they have problems like poor scalability, insufficient real-time response, and rigid control logic when the terminal size grows, and cross-domain collaboration requests become large ^[2]. At the same time, the uncertainty of terminal status, unpredictability of connections, and the low computing power of limited nodes also complicate the conflict between centralized control and distributed deployment. The contradiction is exemplified by the centralized control node's inability to understand the dynamic change of millions of terminals in real time, and distributed terminals lack enough computing and communication resources to achieve independent collaboration. In addition, the terminals belonging to different domains have limited security and data sovereignty requirements, and cannot share the state information directly with each other. Linking them to other domains often requires manual handover or pre-set interfaces and much less response efficiency. Additionally, existing control systems often rely on static configuration or fixed rule chains. When tasks or environmental conditions change, the linkage logic needs to be updated without any ability to adapt to unknown scenarios. These limitations not only limit the growth of IoT systems from data collection to collaboration linkage, but also limit the feasibility of implementing intelligent scheduling and closed-loop control for complex tasks. There is a need for a technology architecture that can uniformly describe heterogeneous terminals at the semantic level and dynamically coordinate virtual and real behaviors.

2.2. Mirror mapping and closed-loop interaction mechanism of digital twins

Digital twins create virtual images of physical terminals with high-quality semantic models for each type or group of IoT terminals, including their characteristics, behaviors, interfaces, and state space, which ensures that underlying physics and protocols are heterogeneous. With this, dynamic matching between virtual and real is preserved via bidirectional data flows: the physical terminal uploads observation information to drive the twin to run, while the twin may add control instructions to reverse the physical object. The key link in the closed-loop interaction mechanism is a feedback loop: the twin simulates and predicts the end behavior

and writes deviation or optimization instructions back to the physical end and performs continuous execution of perception decisions ^[3]. This does not rely on a common communication standard, but instead maps the cross-terminal by the intermediate semantic layer of the twin model, building a basis for linkage control in heterogeneous environments.

2.3. Driving paradigm leap from data synchronization to semantic collaboration

Traditional IoT control is based on data synchronization: terminal status reporting and issuing instructions. There is no common understanding of information meaning between terminals, and collaborative behavior often requires fixed rules. Digital twins have led this control to a new paradigm that moves the definition of control from data synchronization to semantic collaboration. In such a paradigm, twins are not only containers of data but also nodes of semantic interaction: each terminal twin describes its capabilities, needs, and constraints within a common ontology model, and different twins negotiate and decompose the tasks via the same semantic space, thus automatically forming linkage logic ^[4]. Semantic collaboration enables the system to infer cross-terminal meanings and intent without pre-defined point-to-point links (e.g., automatically detecting causal relationships between temperature anomalies and ventilation equipment by twins and inducing linkage). This leads to the reduction of heterogeneous integration, and makes the IoT systems able to adapt and restructure themselves in a basic fashion—from “transmitting” data to “understanding semantics”, to “collaborating action”.

3. Digital twin collaborative architecture in heterogeneous terminal linkage

3.1. Method for constructing a unified twin model for multi-source heterogeneous terminals

A unified twin model is the final step to eliminate heterogeneity among IoT terminals. Instead of iterating each protocol separately, this method builds an extensible meta model in which terminals are divided into four basic entities: attributes, services, events, and constraints. Attributes describe static features and state changes in the terminal, services describe their actions in an executable fashion, events describe the trigger conditions for state changes, and limitations reveal operating limits and limitations ^[5]. A twin can be compiled with consistent expression form using meta model instantiation and semantic annotation for each manufacturer and type of terminals. The construction of the model can be applied for both upstream data-driven automatic learning and explicit downstream manual protocols for known and unknown terminals. A unified model may also be used for cross-terminal relationships modeling, such as spatial inclusion, time collaboration, functional complementarity, etc. This model can explicitly model the relationship with a structured knowledge support for linkage execution.

3.2. Dynamic correlation and behavior consistency maintenance under virtual real mapping

The binding of physical terminals to twins is not static, but is dynamic over time and operation. In order to be behaviorally consistent, the twin model must be able to observe real-time state change, parameter shift, mode-switching, and correct representation bias of the virtual terminal. The dynamic association mechanism is implemented with light heartbeat, differential synchronization, and event-triggered updates (high-frequency updates for fast state changes, refreshing for stable attributes to balance accuracy and communication cost ^[6]).

If the difference between reality and virtuality is greater than a set value, the system automatically triggers alignment, such as state adjustment, model reparameterization, or inference compensation. The correlation of several twins will also be dynamically updated with connection topology, task roles, or environment context changes of the physical terminal. A dynamic correlation and behavioral consistency will not be satisfied by a dynamic linkage decision that deviates from physical reality.

3.3. Operating mechanism of the terminal linkage triggered by twin collaboration

Twin bodies map physical terminals separately and coordinate actual terminals by jointly computing them. Their functioning can be summarized as follows: perception, negotiation, arbitration, and execution;

- (1) Each twin produces linkage requirements according to local states and external events, e.g., temperature twin threshold violations;
- (2) The twin performs semantic negotiation in the collaborative space, decoupling the linkage logic with task objectives, terminal capabilities, and load, and selecting what terminals should perform which operations;
- (3) By using a distributed consensus or a lightweight auction for obtaining lower bottlenecks at the center ^[7].
- (4) A set of linked instructions is formed in the arbitration process, and decision-making conflicts and resource conflicts are resolved; and
- (5) The instruction is sent to the physical terminal for execution by the downstream channel.

Throughout the process, Twin acts as a decoupled layer between the decision maker and the physical object, allowing a physical terminal to complete complex collaborative tasks without communicating directly.

3.4. Twin intermediaries and communication decoupling paths in cross-protocol interactions

Heterogeneous terminals in the Internet of Things rely on different protocols, and if the terminals are directly linked, then they will be highly complex and difficult to maintain. Twin intermediaries provide an efficient communication decoupling pathway: all terminals do not need to communicate with their twin bodies, and twin bodies communicate in a unified digital space. The physical terminal reports its native protocol to the twin, which translates the data into standard semantic messages and publishes them to the co-buses bus. The target twin that needs to be linked subscribes to relevant messages, makes a decision, and provides instructions in a format that the target terminal is able to understand ^[8]. This architecture makes the cross-protocol conversion from the physical network layer to the twin service layer highly efficient and reduces its reliance on gateways. Twin mediators provide dynamic registration and hot plugging of protocol adapters, allowing terminals of new protocol types to be integrated in the linkage system without resorting to the modification of existing systems, to achieve highly decoupled and scalable communication paths.

4. Key technologies for intelligent control based on digital twins

4.1. Real-time oriented twin data stream processing and edge computing integration

The twin data delivered by the IoT terminals are high-frequency, multi-source, heterogeneous, and centralized, which cannot be processed in real-time. Connect twin data stream processing and edge computing along a vital line. Edge nodes are deployed close to the end and perform light inference tasks such as data cleaning, feature extraction, and local twin models. Only upload high-value or cross-domain

collaborative data to the cloud. By lowering some of the twin power to the edge, the system can complete state updates and linkage decisions in milliseconds, and thus, at low end-to-end latency ^[9]. The fusion has data sharing and joint data sharing between edge nodes, and no single node overload occurs. Twin data stream processor of edge cloud collaboration not only achieves real-tactical response in emergency linkage applications, but also is able to utilize cloud computing power to achieve global optimization and long-term model evolution, which complement each other in time and space.

4.2. Twin model-driven terminal state autonomous perception and anomaly diagnosis

The traditional IoT terminal state perception relies on preset levels of control or simple rules, so complex or hidden abnormal behaviors can be detected. The twin model-driven autonomous perception detects anomalies by comparing the deviation of the physical terminal and the prediction of the twin. The two models not only maintain static parameters of the terminal, but also learn its dynamic laws during normal operation, such as timing dependence, operating condition dependence, and device coupling ^[10]. When a significant residual between the real-time stream and the predicted values for the twin model is found, the system automatically identifies whether the state is not broken or the performance is not. Based on the distribution of residuals across other dimensions, the presence of anomalies can be identified—sensor drift, actuator failure, or communication interference. This diagnostic is not required by labeling fault samples, or belongs to the unsupervised or self-supervised model. It can adapt to different fault modes and very few samples of the Internet of Things, and significantly helps to improve terminal control in challenging environments.

4.3. Adaptive generation of control strategies under a centralized-decentralized hybrid architecture

IoT management faces a high single-point risk in centralized architecture and low global coordination for decentralized architecture. The centralized-decentralized hybrid architecture provided by digital twins provides an ideal balance: The global twin provides a global view, centered around the cloud or central node, and handles cross-domain strategy formulation, global resource optimization, and long-term goal decomposition; the local twin automatically implements local control at the edge and can respond quickly to changes in state and linkage requirements in the domain ^[11]. The control strategy is not fixed, but can be generated by adapting to network conditions, task priority, and end-to-end availability. Dynamic changes are made when the local decision is being challenged, and the global goal is being negotiated via twins. The hybrid architecture also supports elastic scaling of control granularity, preferred in centralized coordination where the communication is good, and automatically switches to decentralized autonomous mode during network shutdown or emergencies, and thus, the system can maintain a stable control status with respect to large-scale faults and dynamic environments.

4.4. Evolution of terminal collaborative control from rule-driven to goal-driven

Traditionally, IoT control systems adopt a rule-driven model, which automatically pre-enforces the “if then” logic to trigger the final actions. This mode cannot cope with unexpected situations, and a massive number of rule combinations leads to a very high maintenance cost ^[12]. Digital twin technology is leading control models towards goal-driven approaches. In a goal-driven case, we no longer focus on the particular operation steps, but only on the high-level goals of users or higher-level tasks (e.g., “reduce energy consumption and comfort”). Twin clusters automatically learn the linkage solutions to reach the goals by semantic knowledge

of terminal capabilities and environment perception. The evolution has been guided by the twin model that can support target decomposition, scheme search, and effect estimation, and can be optimized by lightweight programming or reinforcement learning for decision optimization. Since shifting from rules to goals, the system has become flexible, with fewer humans, and also implements a flexible control model for large-scale, long-term, and dynamic IoT applications.

5. Conclusion

Digital twins can overcome heterogeneous terminals and control issues in the Internet of Things by building semantic mirrors and collaborative interactions. From designing a twin model, maintaining virtual and real dynamic associations, the link and communication distance between twins, real-time data stream processing, autonomous anomaly detection, and target-driven control evolution, digital twins advance the intelligent control of the Internet of Things from synchronization to collaboration to a feasible theory for adaptive link and efficient control of large-scale heterogeneous terminals.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Ma Z, Xu T, Liu H, 2026, Intelligent Response and Data Management Model for Power Internet of Things Edge Devices Based on CNN-LSTM. *Ningxia Electric Power*, 2026(1): 53–61.
- [2] Lu T, He Z, Wang C, et al., 2026, Research and Practice of a Unified Edge-side AI Architecture Based on Heterogeneous Edge Terminals. *Communications World*, 2026(1): 44–47.
- [3] Fu L, Chen P, Wang T, et al., 2026, Research on Secure Access Applications for Edge Collaboration Oriented Toward Heterogeneous Terminals. *Information and Computers*, 38(1): 154–156.
- [4] Dai H, Zhao Z, Chang Z, et al., 2026, Data Frame Repair Mechanism for Heterogeneous Network Switching of Integrated Terminals in Complex Power Distribution Environments. *Sichuan Electric Power Technology*, 49(2): 38–44.
- [5] Tian J, Fang G, Zhang C, et al., 2025, Research on Management and Security Technologies for Heterogeneous Terminals in Local Area Networks. *Computer Era*, 2025(10): 61–65.
- [6] Zhao Y, Guan Y, Zhao X, et al., 2024, A Vertical Handover Algorithm for Trusted WLAN and 5G Heterogeneous Networks. *Shandong Electric Power Technology*, 51(9): 57–64.
- [7] Wang W, Cai Z, Chen Y, 2024, Heterogeneous Wireless Network Terminal Access Strategies for the Power Consumption Internet of Things. *Ship Electronic Engineering*, 44(3): 101–105.
- [8] Wu X, Gong Y, Zhang N, 2022, Design of a Secure Access System for Heterogeneous Terminals in the Energy Internet. *Microcomputer Applications*, 38(11): 193–196.
- [9] Zhang H, Chen K, 2022, Research on Data Extraction Technology for Tunnel Group Equipment Based on Multi-source Heterogeneous Information Fusion. *Tunnel Construction (Chinese and English)*, 42(1): 41–47.
- [10] Zhao P, Pu T, Wang X, et al., 2022, Key Technologies and Prospects of the Power Internet of Things for Energy Internet Digital Twins. *Proceedings of the Chinese Society for Electrical Engineering*, 42(2): 447–458.

- [11] Zhou S, Wang Y, Yu W, et al., 2020, Research on Edge Computing Based on Internet of Things Technology. Value Engineering, 39(36): 195–196.
- [12] Lin F, Zhang Q, Yang F, 2020, Design and Development of an Integrated Management Platform for Heterogeneous Internet of Things Systems. Internet of Things Technologies, 10(7): 72–75

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Vessel Segmentation Method Based on Prior-Guided False Positive Suppression Gating

Jintao Zhang

College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, Henan, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Vessel segmentation is a fundamental task in medical image analysis and plays an important role in computer-aided diagnosis, lesion localization, vascular morphology analysis, and subsequent three-dimensional reconstruction. However, blood vessels usually exhibit elongated shapes, complex branching patterns, significant scale variations, and locally low contrast. Under challenging conditions such as complex backgrounds, noise interference, and blurred boundaries, tiny vessels are prone to missed detection, while background textures and spurious edges are easily misclassified as vessels, resulting in increased false positives. To address these issues, this paper proposes a prior-guided vessel segmentation method with false positive suppression. The proposed method adopts an encoder-decoder architecture as the backbone and introduces a Vessel False Positive Suppression Gate (VFPSGate) into the skip connections during the decoding stage. By integrating vessel region priors and edge priors, the shallow features are recalibrated in a suppression-oriented manner, thereby reducing the interference of background noise and non-vascular high responses on segmentation results. In addition, a Vessel False Positive Suppression Loss (VFPSLoss) is designed to impose extra constraints on abnormally high responses in background regions that are not supported by the priors, thus enhancing the model's targeted suppression ability against false positives at the optimization level. Experimental results on the DCA dataset demonstrate that the proposed method achieves competitive performance, with IoU, DSC, ACC, and SEN reaching 66.49%, 79.72%, 97.88%, and 85.16%, respectively. Overall, the proposed method can more effectively distinguish real vessels from pseudo-vessels, providing a feasible solution for vessel segmentation under complex background conditions.

Keywords: Vessel segmentation; Medical image analysis; False positive suppression; Prior guidance

Online publication: Jul 23, 2026

1. Introduction

Cardiovascular diseases are among the major threats to human health worldwide, and medical image analysis plays an essential role in early disease screening, accurate diagnosis, and treatment evaluation^[1-3]. As one of the fundamental tasks in medical image analysis, vessel segmentation has important clinical applications

in various scenarios, including retinal imaging, cerebral angiography, pathological sections, and X-ray angiography^[4]. On the one hand, accurate vessel extraction can provide clinicians with clear anatomical structural information and support quantitative analysis of morphological indicators such as vessel diameter, curvature, and bifurcation angle, thereby offering reliable evidence for computer-aided diagnosis and lesion localization^[5,6]. On the other hand, high-quality two-dimensional vessel segmentation is also a critical prerequisite for subsequent three-dimensional vascular network reconstruction, hemodynamic analysis, and related computer-assisted surgical planning^[7,8].

However, vessel structures usually exhibit characteristics such as elongated morphology, complex branching patterns, significant scale variations, and locally low contrast, which make vessel segmentation a challenging task^[9,10]. In particular, under conditions of complex backgrounds, imaging noise, and blurred boundaries, tiny vessels are prone to discontinuity or omission, while background textures, noisy edges, and other non-vascular high-response regions are often incorrectly identified as vessels, leading to increased false positives^[11,12]. Such false detections not only reduce segmentation accuracy but may also disrupt vascular topological continuity and interfere with subsequent morphological measurement and clinical quantitative analysis^[13]. Therefore, how to effectively suppress false positives while preserving real vessel details as much as possible has become a key issue in current vessel segmentation research.

Early vessel segmentation methods mainly relied on handcrafted features, such as region growing, threshold-based segmentation, filtering enhancement, and morphological rules. Although these methods can achieve certain effects in specific scenarios, their generalization ability is usually limited due to their strong dependence on imaging conditions and expert-designed rules^[14,15]. With the rapid development of deep learning, especially the wide application of convolutional neural networks based on encoder-decoder architectures in medical image segmentation, the performance of vessel segmentation has been significantly improved^[16–18]. In recent years, the introduction of attention mechanisms, Transformers, and related global modeling techniques has further enhanced the representation capability of models for multi-scale features and long-range dependencies, providing new technical pathways for the extraction of complex vascular structures^[19,20].

Despite the substantial progress achieved by existing methods in vascular feature representation and structural recovery, explicit modeling of false-positive regions remains relatively insufficient. Specifically, although skip connections in the decoding stage can supplement rich spatial detail information, they may also transmit background textures, noisy edges, and non-vascular high responses into the subsequent segmentation process, thereby increasing the risk of false detections^[21]. Meanwhile, most existing methods still rely primarily on data-driven feature learning, and the joint utilization of vessel region priors and boundary priors remains inadequate. As a result, explicit structural constraints for distinguishing real vessel responses from background pseudo-responses are still lacking. In terms of loss function design, commonly used losses such as cross-entropy loss and Dice loss focus more on overall pixel classification accuracy or regional overlap, but lack targeted optimization mechanisms for the specific error type of false positives^[22]. In summary, existing methods have paid more attention to enhancing vessel representation, while relatively less effort has been devoted to systematically suppressing false positives from both the perspectives of structural constraints and objective function design.

To address the above issues, this paper proposes a prior-guided false-positive suppression framework for vessel segmentation. In the decoding stage, vessel region priors and edge priors are introduced to construct a

Vessel False Positive Suppression Gate (VFPSGate), which performs suppression-oriented recalibration on shallow features in the skip connections. By leveraging prior information, the proposed module identifies and filters potential false-positive regions, effectively suppressing interference caused by background textures and pseudo-edges while preserving responses that are more likely to correspond to real tiny vessels. Furthermore, a Vessel False Positive Suppression Loss (VFPSLoss) is designed to impose additional constraints on abnormally high responses in background regions that are not supported by the priors, thereby enhancing the model's false-positive suppression capability at the optimization level. Through the collaboration between architectural design and loss design, the proposed method can more effectively distinguish real vessels from pseudo-vessels under complex background conditions, thereby improving the accuracy, completeness, and robustness of vessel segmentation results.

2. Method

This section describes in detail the proposed prior-guided false-positive suppression method for vessel segmentation. Overall, the proposed framework suppresses false positives in vessel segmentation through the joint design of network architecture and optimization objective. First, auxiliary information, including a vessel region prior and an edge prior, is constructed to provide explicit region-level and boundary-level structural constraints. Second, VFPSGate is introduced into the skip connections of the encoder-decoder framework to perform suppression-oriented recalibration of shallow features. Finally, VFPSLoss is further designed to enhance the model's targeted suppression ability against background false detections from the perspective of the optimization objective.

2.1. Overall framework

As shown in **Figure 1**, the proposed method adopts a classical encoder-decoder architecture as the backbone and further incorporates a prior-guided false-positive suppression mechanism. The encoder progressively extracts multi-scale contextual semantic features through multiple convolutions and downsampling operations, while the decoder gradually restores spatial resolution through layer-by-layer upsampling and feature fusion to generate the final vessel segmentation result. Unlike the conventional U-Net, which directly transmits shallow encoder features to the decoder through skip connections, the proposed method embeds VFPSGate into the skip connections at multiple decoding stages to selectively filter and recalibrate shallow features, thereby reducing the interference of background textures, noisy edges, and non-vascular high-response regions during subsequent decoding.

Specifically, given an input image, the encoder first extracts feature representations at different scales. During decoding, before each decoder feature is fused with its corresponding encoder feature, the encoder feature is first modulated by VFPSGate. At the current scale, this module jointly exploits encoder features, decoder features, and the vessel region prior and edge prior constructed from the input image to perform false-positive suppression-oriented recalibration of shallow features. The gated features are then fused with decoder features for subsequent segmentation recovery. Meanwhile, during training, a dedicated false-positive suppression loss is introduced to cooperate with the structural gating mechanism, so as to further reduce false detections under complex backgrounds while preserving real tiny vessel structures.

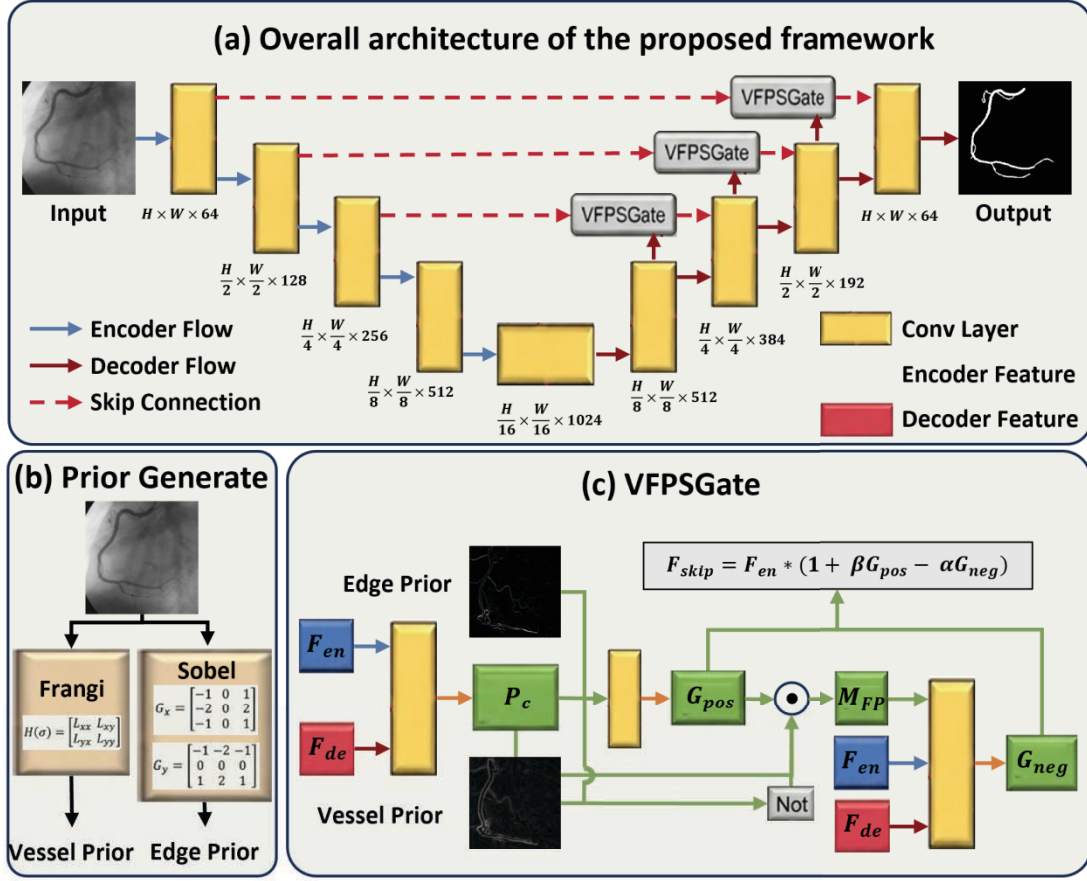


Figure 1. Overall framework of the proposed method.

2.2. Prior construction

To provide explicit region-level and boundary-level structural constraints for the network, a vessel region prior and an edge prior are constructed based on the input image, as shown in **Figure 1**. Different from purely data-driven feature learning, explicit priors provide more direct structural guidance before the gating process, thereby helping the model better distinguish real vessel responses from background pseudo-responses. Considering that vessels typically exhibit elongated, continuous, and tubular structures, while their boundaries are easily affected by noise and pseudo-edges under complex backgrounds, prior information is constructed from both region enhancement and boundary supplementation perspectives to improve the ability of the subsequent gating module to identify real vascular structures.

For the vessel region prior, since vessels in medical images usually appear as elongated and continuous tubular structures, the Frangi filter is employed to enhance the input image and highlight vessel-like responses^[23]. The Frangi filter selectively enhances tubular structures based on the eigenvalues of the Hessian matrix and is able to emphasize regions consistent with vessel morphology under multi-scale settings. Compared with general edge detection or intensity enhancement methods, the Frangi filter focuses more on the geometric characteristics of the target rather than solely on local intensity variations, thereby providing relatively stable candidate information for vessel regions. After normalization, the resulting map is regarded as the vessel region prior, P_v , which indicates the prior likelihood of each location belonging to a vessel region.

For the edge prior, to further supplement local variation information related to vessel boundaries, contours, and fine structures, the Sobel operator is used to compute the gradient magnitude of the input image^[24]. After normalization, the resulting map is regarded as the edge prior P_e . The Sobel operator can effectively reflect the intensity of local grayscale changes and is sensitive to fine boundaries and contour transitions, making it suitable as boundary-level structural supplementary information. Different from the vessel region prior, which mainly provides regional support, the edge prior pays more attention to local boundary variation and structural continuity, and can provide additional contour discrimination cues in regions with low contrast, blurred boundaries, or complex background textures.

Overall, the vessel region prior and the edge prior are highly complementary. The former emphasizes regional consistency and the distribution of tubular structures, whereas the latter emphasizes boundary rationality and local contour continuity. By jointly introducing these two types of priors, the model can not only determine whether a location is likely to belong to a vessel region, but also further assess whether it exhibits a plausible boundary pattern. These two priors are finally fed into the subsequent VFPSGate to provide explicit guidance for positive gating construction and false-positive suppression from both region and boundary perspectives, thereby enhancing the model's ability to distinguish real vessels from pseudo-vessels under complex background conditions.

2.3. Vessel false positive suppression gate

In a standard encoder-decoder architecture, skip connections provide rich, shallow detail information for the decoder, which is crucial for recovering tiny vessels. However, while preserving textures and edge information, shallow features may also contain background noise, pseudo-edges, and other non-vascular high-response regions. If such features are directly passed to the decoder, they may amplify false positives during subsequent segmentation. To address this issue, the proposed VFPSGate performs prior-guided recalibration of shallow encoder features in the skip connections. By combining decoder features, the vessel region prior, and the edge prior, the module identifies and suppresses potential false-positive regions while retaining responses that are more likely to correspond to real vessels. The overall computational process is illustrated in **Figure 1**.

First, at the current scale, the encoder feature F_{en} and the decoder feature F_{de} are fused, and a coarse vessel prediction is generated through a lightweight prediction branch P_c :

$$P_c = \sigma\left(f_c\left([F_{en}, F_{de}]\right)\right) \quad (1)$$

where $[\cdot]$ denotes channel-wise concatenation, $f_c(\cdot)$ denotes the coarse vessel prediction branch, and $\sigma(\cdot)$ denotes the Sigmoid activation function. P_c is not the final segmentation output, but rather an initial estimation of vessel regions at the current scale, which serves as the basis for subsequent false-positive candidate localization.

After obtaining the coarse vessel probability map, the vessel region prior P_v and the edge prior P_e are further incorporated to construct a positive gating branch, which is used to select responses that are more likely to correspond to real vessels from the perspectives of both regional consistency and boundary structural rationality. The positive gate G_{pos} is generated jointly from the coarse vessel prediction and the two priors:

$$G_{pos} = \sigma\left(f_p\left([P_c, P_v, P_e]\right)\right) \quad (2)$$

where $f_p(\cdot)$ denotes the positive gate generation branch. Different from relying solely on the coarse prediction, the positive gate simultaneously integrates region-level and boundary-level prior constraints, thereby more effectively highlighting responses associated with real vessels and providing guidance for the construction of subsequent false-positive candidate regions.

Based on the principle that the model may produce high responses in regions unsupported by the priors, a false-positive candidate map M_{FP} is further constructed. If a location exhibits a high response in the coarse vessel prediction and still maintains strong vessel-related activation after positive gating, but shows a weak response in the vessel region prior, then this location is more likely to correspond to a false-positive background response rather than a real vessel. The false-positive candidate map M_{FP} is defined as

$$M_{FP} = G_{pos} \odot P_c \odot (1 - P_v) \quad (3)$$

where $\odot$ denotes element-wise multiplication. This design explicitly extracts suspicious regions that the model considers vessel-like but are not supported by the vessel prior, thereby providing direct evidence for the subsequent negative suppression gate.

Subsequently, a negative gating branch is constructed using F_{en} , F_{de} , and M_{FP} to suppress responses that are more likely to be background false detections. The negative gate is formulated as

$$G_{neg} = \sigma\left(f_n\left([F_{en}, F_{de}, M_{FP}]\right)\right) \quad (4)$$

where $f_n(\cdot)$ denotes the negative gate generation branch. Unlike the positive gate, which focuses on preserving real vessel responses, the negative gate is specifically designed to model potential false-positive regions, thereby enabling explicit suppression of background high responses and pseudo-edge interference.

Finally, to preserve real vessel details while suppressing false-positive background responses, an asymmetric recalibration strategy is adopted to update the shallow encoder features:

$$F_{skip} = F_{en} \odot (1 + \beta G_{pos} - \alpha G_{neg}) \quad (5)$$

where F_{skip} denotes the gated skip connection feature, α and β are learnable parameters. According to preliminary experiments, they are initialized as 2.0 and -1.0, respectively. Through this asymmetric dual-path gating mechanism, VFPSGate can preserve fine vessel and boundary details while effectively reducing the interference of background textures, noisy edges, and non-vascular high-response regions during subsequent decoding, thereby improving the accuracy and structural integrity of vessel segmentation under complex backgrounds.

2.4. Vessel false positive suppression loss

To further enhance the model's targeted suppression ability against false-positive regions, a Vessel False Positive Suppression Loss is designed and jointly optimized with VFPSGate. Different from traditional segmentation losses that rely only on final output supervision, the overall loss function in this work simultaneously considers the final segmentation output, intermediate coarse vessel predictions, and background false-detection penalties, thereby constraining the learning process from both output-level and intermediate-level perspectives.

First, the main segmentation loss is used to constrain the consistency between the final output and the ground-truth label. In this work, a combination of binary cross-entropy (BCE) loss and Dice loss is adopted as the basic supervision term to account for pixel-level classification accuracy and region overlap quality jointly^[25,26]

$$L_{main} = L_{BCE}(\hat{Y}, Y) + L_{Dice}(Y, Y) \quad (6)$$

where $\hat{Y}$ denotes the final prediction and Y denotes the ground-truth label.

Second, to improve the reliability of the intermediate coarse vessel predictions in each VFPSGate, auxiliary supervision is imposed on the coarse predictions at multiple decoding stages. Let $P_c^{(k)}$ denote the coarse vessel prediction at the k -th scale and $Y_c^{(k)}$ denote the corresponding ground-truth label at that scale. The auxiliary supervision loss is defined as

$$L_{aux} = \frac{1}{K} \sum_{k=1}^K L_{seg}(P_c^{(k)}, Y_c^{(k)}) \quad (7)$$

where K denotes the number of VFPSGate modules involved in auxiliary supervision. Consistent with the main segmentation loss, the auxiliary loss also adopts the combination of BCE loss and Dice loss. By applying segmentation supervision to these intermediate predictions, they are not merely internal variables of the gating module, but also carry explicit vessel semantic information. More reliable coarse predictions contribute to more stable positive gating construction and false-positive candidate generation, thereby enhancing the stability and interpretability of the entire gating mechanism.

On this basis, a false-positive penalty term is further designed to constrain regions that satisfy the following conditions explicitly: they belong to the background in the ground-truth label, are not supported by the vessel region prior, but still receive high responses in the coarse vessel prediction. For the k -th scale, if a location corresponds to the background in $Y^{(k)}$ It has a low response in $P_v^{(k)}$, but still exhibits a high value in $P_c^{(k)}$ Then the model tends to misclassify this background region as a vessel. Accordingly, the multi-scale false-positive penalty term is defined as

$$L_{fp} = \frac{1}{K} \sum_{k=1}^K \frac{\sum (1-Y^{(k)})(1-P_v^{(k)})P_c^{(k)}}{\sum (1-Y^{(k)}) + \epsilon} \quad (8)$$

where K is the number of scales involved in auxiliary supervision, $Y^{(k)}$, $P_v^{(k)}$, $P_c^{(k)}$ and denote the ground-truth label, vessel region prior, and coarse vessel prediction at the k -th scale, respectively, and ϵ is a smoothing constant. This loss term imposes an additional penalty on locations in background regions that violate prior constraints, thereby enabling targeted suppression of false positives. Unlike traditional segmentation losses, which mainly focus on overall region overlap, this term directly addresses the specific error type of “background being misclassified as vessel,” and is therefore more consistent with the design objective of this work.

Finally, the main segmentation loss, auxiliary supervision loss, and false-positive penalty term are combined in a weighted manner to form the overall optimization objective:

$$L = L_{main} + \lambda_{aux} L_{aux} + \lambda_{fp} L_{fp} \quad (9)$$

where λ_{aux} and λ_{fp} denote the weighting coefficients for the auxiliary supervision term and the false-positive penalty term, respectively. Through the joint optimization of these loss terms, the network is able not only to learn more accurate vessel segmentation representations but also to explicitly reduce abnormally high responses in prior-unsupported regions during training, thereby improving the accuracy, completeness, and robustness of vessel segmentation results under complex background conditions.

3. Results

3.1. Experimental settings and evaluation metrics

Experiments were conducted on the DCA dataset to validate the effectiveness of the proposed method^[27]. This dataset contains coronary angiography images and their corresponding vessel segmentation annotations, and can well reflect practical segmentation challenges such as complex backgrounds, low contrast, and abundant fine vessel structures. The dataset was divided into training, validation, and test sets with a ratio of 7:1:2. To improve the generalization ability of the model, data augmentation strategies including random 90°

rotation, random horizontal flipping, and random vertical flipping were adopted during training. All input images were resized to 256×256 .

To comprehensively evaluate the effectiveness of the proposed method, several representative medical image segmentation models were selected for comparison, including the classical convolutional architectures UNet, ResUNet, UNet++, the attention-based Attention UNet, and models with stronger feature modeling capability, including TransUNet, Swin UMamba, and UKAN^[28–34]. For fair comparison, all methods were trained and tested under the same dataset and unified experimental settings.

To assess segmentation performance from different perspectives, Intersection over Union (IoU), Dice Similarity Coefficient (DSC), Accuracy (ACC), and Sensitivity (SEN) were adopted as evaluation metrics^[35–38]. Among them, IoU and DSC mainly reflect the overlap between the prediction and the ground truth, ACC measures overall pixel-wise classification accuracy, and SEN emphasizes the detection ability for true vessel regions, thereby better reflecting the model’s capability to preserve fine vessels and low-contrast vessels.

All experiments were implemented in PyTorch and conducted on an NVIDIA RTX 4090 GPU. During training, the batch size was set to 8, and the initial learning rate was set to 1×10^{-4} . The Adam optimizer was adopted, and the total number of training epochs was 400. For loss function configuration, VFPSLoss was used as the overall optimization objective, where the weight of the auxiliary supervision term $\lambda_{aux}=0.2$, and the weight of the false-positive penalty term $\lambda_{fp}=0.05$.

3.2. Comparison experiments

Table 1 presents the quantitative comparison results between the proposed method and several mainstream segmentation models on the DCA dataset. As can be seen, the proposed method achieved the best performance across all evaluation metrics. Specifically, the proposed method reached 66.49%, 79.72%, 97.88%, and 85.16% in terms of IoU, DSC, ACC, and SEN, respectively. These results indicate that the proposed method not only improves the overall overlap quality of segmentation but also enhances the detection ability for vessel regions.

Table 1. Quantitative comparison of different methods for vessel segmentation on the DCA dataset

	IOU	DSC	ACC	SEN
UNet	65.43	78.95	97.85	82.53
ResUNet	65.59	79.08	97.84	83.48
UNet++	65.99	79.33	97.85	84.26
Attention UNet	66.22	79.54	97.85	84.89
TransUNet	66.39	79.67	97.90	84.19
Swin UMamba	63.90	77.83	97.68	83.84
UKAN	65.34	77.58	97.68	81.76
Our	66.49	79.72	97.88	85.16

From the performance trends, the improvements achieved by the proposed method are particularly noticeable in terms of DSC and SEN. The improvement in DSC indicates a higher overall overlap between the prediction and the ground-truth vessel region, while the clear gain in SEN suggests that the proposed method is more effective in preserving real vessels, especially fine vessels, low-contrast vessels, and branching structures. Considering that there is usually a trade-off between fine-detail recovery and false-

positive suppression in vessel segmentation, the proposed method still maintains relatively high ACC and IoU while improving SEN. This indicates that the gain in sensitivity is not achieved by simply enlarging the predicted vessel regions, but rather by effectively suppressing false-positive responses in background regions while preserving real vessels.

These results are consistent with the design objective of the proposed method. First, the vessel region prior provides region-level structural support, helping the model emphasize candidate regions that are more consistent with tubular morphology. Second, the edge prior further supplements local contour and boundary information, enabling the model to distinguish real vessel boundaries from background pseudo-edges better. Finally, VFPSLoss imposes additional penalties on regions that belong to the background and are unsupported by the prior but still receive high model responses, thereby further reducing false detections at the optimization level. As a result, the structural gating mechanism and the targeted loss constraint complement each other, endowing the model with a stronger capability to identify real vessels and suppress false positives under complex background conditions.

To provide a more intuitive comparison of segmentation performance, **Figure 2** shows the visualization results of different methods on two representative samples. As can be observed, although the compared methods can extract the main vessel trunks to a certain extent, they still show deficiencies in recovering fine branches, curved boundaries, and low-contrast regions, and may also produce false detections in local background areas. In contrast, the proposed method achieves better performance in terms of main trunk continuity, fine branch completeness, and background suppression.

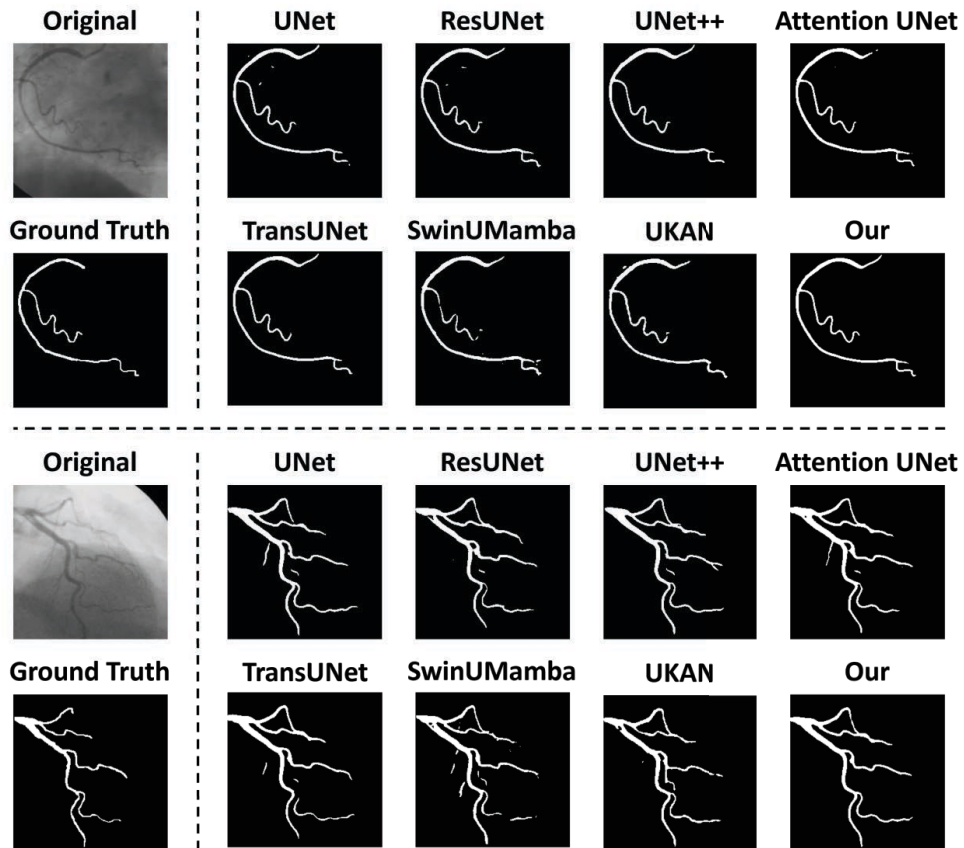


Figure 2. Visual comparison of vessel segmentation results produced by different methods.

Specifically, in the first sample, the proposed method preserves the fine branches within the curved main trunk more completely, produces smoother and more continuous segmentation contours, and yields fewer scattered false responses in the background. In the second sample, when faced with regions containing more bifurcations and more complex structures, the proposed method better maintains the connectivity among branches while reducing erroneous responses in non-vascular regions. Overall, the segmentation results generated by the proposed method are cleaner, more continuous, and more consistent with the true morphological characteristics of vessels.

3.3. Ablation studies

To further verify the individual contributions of each component and their collaborative effect, ablation studies were conducted on the DCA dataset. Specifically, based on the baseline model, the vessel region prior, the edge prior, the false-positive suppression loss, and the complete model were introduced step by step. The experimental results are shown in **Table 2**.

Table 2. Ablation study of different components of the proposed method on the DCA dataset

Vessel	Edge	VFPSLoss	IOU	DSC	ACC	SEN
-	-	-	65.43	78.95	97.85	82.53
√			66.31	79.57	97.92	82.63
√	√		66.03	79.38	97.90	82.71
√		√	66.09	79.43	97.84	85.06
√	√	√	66.49	79.72	97.88	85.16

Each component improves the model performance to a certain extent. When only the vessel region prior is introduced, IoU, DSC, and ACC all improve, indicating that the region prior can provide effective structural guidance for the model. After introducing the false-positive suppression loss, SEN increases from 82.53% to 85.06%, demonstrating that this loss plays an important role in both real vessel detection and background false-positive suppression. In comparison, the improvement brought by the edge prior alone is relatively limited, but it still provides useful supplementary information for local contour and boundary modeling.

When the vessel region, edge prior, and false-positive suppression loss are used together, the complete model achieves the best performance. Compared with the baseline, IoU, DSC, ACC, and SEN are improved by 1.05%, 0.77%, 0.03%, and 2.63%, respectively. These results indicate that the proposed components are complementary to each other and can collaboratively suppress false positives from both the perspectives of structural constraint and optimization objective, thereby improving the overall vessel segmentation performance.

4. Discussion

The experimental results on the DCA dataset demonstrate that the proposed prior-guided false-positive suppression strategy can effectively improve vessel segmentation performance. By combining quantitative results with visual analysis, it can be observed that the proposed method not only improves the overlap between the predictions and the ground truth but also enhances the detection of fine vessels, low-contrast

vessels, and complex branching structures. Meanwhile, the visualization results further show that, while maintaining the continuity of the main vessel trunks and the integrity of fine branches, the proposed method can more effectively reduce scattered false responses and erroneous connections in background regions. These findings indicate that the performance gain of the proposed method does not stem from simply enlarging the predicted vessel area, but rather from the simultaneous enhancement of real vessel preservation and background false-positive suppression.

From the perspective of the method design, this performance improvement mainly benefits from the collaboration between the architectural design and the loss design. On the one hand, the vessel region provides region-level structural support, enabling the network to focus more on candidate regions that are consistent with tubular morphology. The edge prior further supplements local contour and boundary information, thereby improving the discrimination between real vessel boundaries and background pseudo-edges. On the other hand, VFPSGate performs suppression-oriented recalibration of shallow features in the skip connections, which helps preserve fine details while weakening the interference caused by noisy textures and non-vascular high responses. In addition, VFPSLoss imposes additional constraints on high responses in background regions that are not supported by the priors, thereby enhancing the model's targeted suppression ability against false-positive regions at the optimization level. The fact that the complete model achieves the best performance in the ablation experiments also indirectly verifies the good complementarity among the proposed components.

Another advantage of the proposed method lies in its practicality and scalability. For tasks such as coronary angiography, where the background is complex, the boundaries are ambiguous, and fine vessels are abundant, reducing false positives is beneficial not only for improving segmentation accuracy but also for enhancing the reliability of subsequent morphological analyses such as vessel diameter, curvature, and bifurcation measurement. Moreover, the prior-guided false-positive suppression idea proposed in this work does not rely on a specific backbone network. In principle, it can be conveniently transferred to other encoder-decoder-based segmentation frameworks, indicating a certain degree of general applicability.

Nevertheless, the proposed method still has some limitations. First, the model performance depends to some extent on the quality of the vessel region prior to and the edge prior to. When the input image contains stronger noise or even lower contrast, the priors themselves may be degraded, which may further weaken the gating effect. Second, in regions containing extremely weak vessels, severe structural overlap, or complex lesion interference, the difference between real vessel responses and background pseudo-responses may still be subtle, leaving room for further improvement in discrimination capability. Finally, the experiments in this work were mainly conducted on the DCA dataset, and the cross-dataset generalization ability of the proposed method has not yet been systematically evaluated on more diverse vessel segmentation tasks.

Future work can be carried out in three directions. First, more adaptive prior construction strategies can be explored, extending the current priors generated by traditional image processing methods to learnable or dynamically updated prior representations. Second, based on false-positive suppression, topological continuity or structural consistency constraints can be further introduced to better balance fine vessel preservation and overall connectivity. Third, the proposed framework can be extended to retinal vessel, cerebrovascular, and other multimodal vessel segmentation tasks to validate its generality and clinical application potential further.

5. Conclusion

This paper proposes a prior-guided false-positive suppression framework to address the problems of missed detection of fine vessels and increased false positives under complex backgrounds in vessel segmentation. The proposed method introduces VFPSGate into the skip connections of the encoder-decoder architecture and combines vessel region priors and edge priors to recalibrate shallow features. In addition, VFPSLoss is designed to impose extra constraints on abnormally high responses in background regions that are not supported by the priors, thereby effectively reducing background false detections while preserving real vessel details.

Experimental results on the DCA dataset demonstrate that the proposed method achieves competitive performance across multiple evaluation metrics. Quantitative comparisons, visual analyses, and ablation studies further verify the effectiveness of each component and its collaborative effect. Overall, the proposed method can more effectively distinguish real vessels from pseudo-vessels, thereby improving the accuracy and robustness of vessel segmentation under complex background conditions.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Naser M, Majeed A, Alsabah M, et al., 2024, A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges. *Algorithms*, 17(2): 78.
- [2] Restrepo Tique M, Araque O, Sanchez-Echeverri L, 2024, Technological Advances in the Diagnosis of Cardiovascular Disease: A Public Health Strategy. *International Journal of Environmental Research and Public Health*, 21(8): 1083.
- [3] Sanyaolu S, 2025, Integration of Machine Learning in Imaging Analysis for Clinical Diagnosis of Cardiovascular Diseases. *Premier Journal of Cardiology*, 2: 100006.
- [4] Xia W, et al., 2025, vesselFM: A Foundation Model for Universal 3D Blood Vessel Segmentation. *arXiv preprint arXiv:2411.17386*.
- [5] Oliveira A, Pereira S, 2021, Retinal Vessel Segmentation for Morphological Analysis and Computer-Aided Diagnosis of Cardiovascular Diseases. *Medical Image Analysis*, 71: 102066.
- [6] Khan M, Ahmad B, Soomro S, 2022, Coronary Artery Segmentation and Morphological Quantification for Atherosclerotic Lesion Detection. *IEEE Journal of Biomedical and Health Informatics*, 26(7): 3104–3114.
- [7] Kryvoshei O, Kamencay P, Polak L, 2026, Neural Vessel Segmentation and Gaussian Splatting for 3D Reconstruction of Cerebral Angiography. *AI*, 7: 22.
- [8] Narra S, et al., 2025, A Multi-Stage Deep Learning Pipeline for 3D Coronary Artery Reconstruction from X-Ray Angiograms. *IEEE Xplore (Conference Proceedings)*.
- [9] Yuan Y, Zhang Y, Zhu L, et al., 2024, Exploiting Cross-Scale Attention Transformer and Progressive Edge Refinement for Retinal Vessel Segmentation. *Mathematics*, 12: 264.
- [10] Xu B, Yang J, Hong P, et al., 2024, Coronary Artery Segmentation in CCTA Images Based on Multi-Scale Feature Learning. *Journal of X-Ray Science and Technology*, 32(4): 973–991.
- [11] Song R, Liu L, Zhang Y, et al., 2026, VFSGS-Net: Frequency-Guided State-Space Learning for Topology-

- Preserving Retinal Vessel Segmentation. arXiv preprint arXiv:2602.10978.
- [12] Xu H, Meng P, Wang M, et al., 2025, DB-KAUNet: An Adaptive Dual Branch Kolmogorov-Arnold UNet for Retinal Vessel Segmentation. arXiv preprint arXiv:2512.01657.
 - [13] Wang W, Xia Q, Yan Z, et al., 2024, AVDNet: Joint Coronary Artery and Vein Segmentation with Topological Consistency. *Medical Image Analysis*, 91: 102999.
 - [14] Paulauskaite Taraseviciene A, Siaulys J, Jankauskas A, et al., 2025, A Robust Blood Vessel Segmentation Technique for Angiographic Images Employing Multi Scale Filtering Approach. *Journal of Clinical Medicine*, 14(2): 354.
 - [15] Zhong Y, Chen T, Zhong D, et al., 2024, Vessel Segmentation in Fundus Images with Multi-Scale Feature Extraction and Disentangled Representation. *Applied Sciences*, 14(12): 5039.
 - [16] Kande G, et al., 2024, MSR U-Net: An Improved U-Net Model for Retinal Blood Vessel Segmentation. *IEEE Access*, 12: 534–551.
 - [17] Ramos-Cortez J, Alvarado-Carrillo D, Ovalle-Magallanes E, et al., 2025, Lightweight U-Net for Blood Vessels Segmentation in X-Ray Coronary Angiography. *Journal of Imaging*, 11: 106.
 - [18] Hernandez-Gutierrez F, Avina-Bravo E, Ibarra-Manzano M, et al., 2025, Retinal Vessel Segmentation Based on a Lightweight U-Net and Reverse Attention. *Mathematics*, 13: 2203.
 - [19] Soni T, Gupta S, Bharany S, et al., 2025, Retinal Vessel Segmentation Using Multi-Scale Feature Attention with MobileNetV2 Encoder. *Scientific Reports*, 15(1): 43369.
 - [20] Zhang Y, Chung A, 2024, Retinal Vessel Segmentation by a Transformer-U-Net Hybrid Model with Dual-Path Decoder. *IEEE Journal of Biomedical and Health Informatics*, 28(9): 5347–5359.
 - [21] Chen S, Fan J, Ding Y, et al., 2024, PEA-Net: A Progressive Edge Information Aggregation Network for Vessel Segmentation. *Computers in Biology and Medicine*, 169: 107766.
 - [22] Li A, Sun M, Wang Z, 2024, TD Swin-UNet: Texture-Driven Swin-UNet with Enhanced Boundary-Wise Perception for Retinal Vessel Segmentation. *Bioengineering*, 11: 488.
 - [23] Shi G, Lu H, Hui H, et al., 2025, Benefit from Public Unlabeled Data: A Frangi Filter-Based Pretraining Network for 3D Cerebrovascular Segmentation. *Medical Image Analysis*, 101: 103442.
 - [24] Bai H, Ma Z, Gao C, et al., 2025, SVSNet: Scleral Vessel Segmentation with a CNN-Transformer Hybrid Network. *Journal of Innovative Optical Health Sciences*, 18(6): 2550017.
 - [25] Yeung M, Sala E, Schönlieb C, et al., 2022, Unified Focal Loss: Generalizing Dice and Cross Entropy-Based Losses to Handle Class Imbalanced Medical Image Segmentation. *Computerized Medical Imaging and Graphics*, 95: 102026.
 - [26] Yeung M, Rundo L, Nan Y, et al., 2023, Calibrating the Dice Loss to Handle Neural Network Overconfidence for Biomedical Image Segmentation. *Journal of Digital Imaging*, 36(2): 739–752.
 - [27] Cervantes-Sanchez F, Cruz-Aceves I, Hernandez-Aguirre A, et al., 2019, Automatic Segmentation of Coronary Arteries in X-Ray Angiograms Using Multiscale Analysis and Artificial Neural Networks. *Applied Sciences*, 9: 5507.
 - [28] Shariaty F, Mohebi M, Barzegar-Golmoghani E, et al., 2025, Deep Vessel Segmentation with U-Net and Texture Representation of Image (TRI) Features Provides a Foundation for Improved Objective and Automated Analysis of Coronary Artery Disease from Angiography. *Computer Methods and Programs in Biomedicine*, 272: 109072.
 - [29] Amine J, Mourad M, 2025, Toward Accurate Alzheimer's Detection: Transfer Learning with ResNet50 for MRI-Based Diagnosis. *Frontiers in Neuroscience*, 19: 1664418–1664418.

- [30] Zhou Z, Siddiquee M, Tajbakhsh N, et al., 2020, UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Transactions on Medical Imaging*, 39(6): 1856–1867.
- [31] Cao J, Chen J, Gu Y, et al., 2023, MFA-UNet: A Vessel Segmentation Method Based on Multi-Scale Feature Fusion and Attention Module. *Frontiers in Neuroscience*, 17: 1249331.
- [32] Zhang Y, Chung A, 2024, Retinal Vessel Segmentation by a Transformer-U-Net Hybrid Model with Dual-Path Decoder. *IEEE Journal of Biomedical and Health Informatics*, 28(9): 5347–5359.
- [33] Liu J, et al., 2025, Swin-UMamba: Adapting Mamba-Based Vision Foundation Models for Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 44(10): 3898–3908.
- [34] Li C, Liu X, Li W, et al., 2025, U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(5): 4652–4660.
- [35] Prokop K, Polap D, 2024, Image Segmentation Enhanced by Heuristic Assistance for Retinal Vessels Case. *Proceedings of IEEE Congress on Evolutionary Computation*, Yokohama, Japan: 1–6.
- [36] Seker M, Kartal M, Moniri A, et al., 2025, Hepatic Vessel Segmentation and Classification in CTA Images Using nnU-Net with Centerline Regression. *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing*, Istanbul, Türkiye: 1–4.
- [37] Attah M, Elloumi Y, Kachouri R, 2024, High Performance and Low Complexity Retinal Vessel Segmentation Method Based on Extended DCNN. *Proceedings of IEEE International Conference on Image Processing Theory, Tools and Applications*, Rabat, Morocco: 01–06.
- [38] Xu J, Liu Q, Shen J, et al., 2025, Image-Text Guided Fundus Vessel Segmentation via Attention Mechanism and Gated Residual Learning. *Frontiers in Cell and Developmental Biology*, 13: 1710343.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Design and Simulation of Wireless Charging Circuit for Electric Vehicles

Zijian Lu

College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130012, Jilin, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: In the 21st century, wireless charging technology has developed rapidly, and realizing wireless charging for electric vehicles is a key focus of future charging work. Building upon previous research, this paper studies and designs a wireless charging circuit based on the principle of electromagnetic induction, which includes a transmitter and receiver, a bridge rectifier, and an RLC series resonant circuit. A set of feasible key parameters was analyzed, and some optimization schemes were proposed. The feasibility and safety of the circuit were verified through simulation software testing. The circuit was found to be relatively stable and applicable to charging most practical electric vehicles. This could provide a new solution for wireless charging of electric vehicles. It also has the potential to contribute to the efficient and convenient development of electric vehicle charging and is worthy of further research.

Keywords: Electric vehicles (EVs); Wireless charging; Electromagnetic induction; High efficiency and convenience

Online publication: Jul 23, 2026

1. Introduction

1.1. Research background

Technological advancements have brought cars into every household. While cars help people travel, environmental protection has become a focal point of discussion. Electric vehicles can significantly reduce vehicle exhaust emissions, which is very effective in energy conservation and emission reduction. Therefore, electric vehicles have been widely used in recent years. How to charge electric vehicles efficiently and conveniently is one of the urgent problems that the automotive industry needs to solve. Compared to traditional wired charging, wireless charging technology is renowned for its superior safety, efficiency, and stability. It has a broader prospect in the field of charging.

1.2. Previous research and limitations

In recent years, numerous advances and achievements have been made in the development and research of wireless charging circuits for automobiles. Oak Ridge National Laboratory in the United States has a long history of research in the field of high-performance Technology and has been able to increase the performance of wireless charging methods to over 120 kW after years of testing. The transmission efficiency

is over 97 %, thus forming the basis for fast charging ^[1]. The Korea Advanced Institute of Science and Technology in South Korea developed an electric vehicle that uses electromagnetic induction and wireless charging and was put into operation in Seoul in 2011 ^[2]. In 2023, in Detroit, Michigan, North America's first dynamic Wireless charging track opened. The 400-meter-long track reached a charging power of 19 kW during a demonstration drive. The Israeli company Electron Wireless installed a dynamic Wireless charging system on a main road southwest of Paris ^[3]. Thanks to this research, Wireless charging technology has made great progress and is now gradually entering practical application. Energy transmission efficiency and range are generally sufficient to meet the basic charging needs of electric vehicles today. However, there is still room for improvement in terms of performance, heat generation, and stability, and the transmission efficiency and range are not optimal and need further optimization.

1.3. Research content and significance

This study utilizes Faraday's principle of electromagnetic induction. After carefully studying the working principle of wireless charging, we designed a dual-port charging circuit equipped with a transformer. By utilizing the resonance of inductors and capacitors, it can significantly reduce energy loss during the charging process. A bridge rectifier circuit was added to the output of the circuit. It can convert AC power to DC power. This provides the load with a relatively stable charging voltage. We analyzed whether this circuit is safe and feasible, and proposed methods for improvement. The goal of this research is to make wireless charging both safe and practical. It also enables fast, convenient, and efficient charging. This also offers new ideas and approaches for the wireless charging of electric vehicles.

2. Scheme principle and design

2.1. General design principle and logic

Electromagnetic induction wireless charging occurs when the coils of the transmitting end and the receiving end are close; the transmitting end generates an alternating magnetic field through the alternating current, which then induces a current in the receiving coil to realize the transmission of energy. This method is simple and direct, and can complete energy transmission at a short distance. It has relatively mature technology, and the cost is relatively controllable ^[4]. When the parameters ω_0 , L, C in a circuit satisfy $\omega_0 L = \frac{1}{\omega_0 C}$, which implies $f_0 = \frac{1}{2\pi\sqrt{LC}}$, the circuit state in which the port voltage and current of a one-port network happen to be in phase is called a resonant state. An RLC series circuit has the minimum impedance and the maximum current at resonance. It can achieve relatively high energy transmission efficiency. The transmitter design in this scheme includes a resonant network comprising an alternating power supply, resistors, capacitors, and inductors. Alternating current is transmitted to the receiving end through a transformer. At the receiving end, an RLC series resonant circuit is still used to play a crucial "frequency selection filtering" role. In the wireless charging system, by locking the resonant point, the influence of input voltage fluctuations can be offset, and the load voltage can be more stable. A bridge rectifier circuit is added between the resonant circuit and the load, and the alternating current becomes a one-way pulsating direct current, which is more in line with the needs of wireless charging.

The design of wireless charging circuits cannot be separated from the design of inductive devices. Modern inductive devices can be divided into two categories: air-core inductors (including solenoid inductors,

ring inductors, flat inductors, etc.) and magnetic core inductors (including ferrite magnetic ring filter inductors, electrically controlled adjustable inductors, etc.). A magnetic core inductor is an inductor with a magnetic core, which uses magnetic cores (usually iron, ferrite, or other magnetic materials) to enhance the inductance effect. The magnetic core can improve the self-inductance value of the inductor, enhance the response of the inductor to the external magnetic field, and reduce the magnetic coupling noise. Compared with air-core inductors, magnetic core inductors are more widely used in high-power, medium-low-frequency, or some high-frequency occasions that require energy storage and filtering. The magnetic core can store more magnetic energy and generate a higher induced voltage. In the field of wireless charging of electric vehicles, it is suitable for inductors with magnetic cores. It can significantly improve transmission efficiency and reduce energy waste. It realizes large inductance and large current in a smaller volume to meet the needs of electric vehicles better. How to place the ferrite magnetic core will also affect the focusing effect of the magnetic field. Different core shapes such as U-shaped, E-shaped, and zig-shaped have their own uses ^[5].

Based on the principles discussed above, the key parameters for the circuit are established as shown in **Table 1**.

Table 1. Key parameter settings

Operating frequency	50Hz
Resonant capacitance	100 μ F
Resonant inductance	100mH
Transformer turns ratio	1:1.2
Target RMS output voltage	263V
Load resistance	50 Ω
Rated output power	1.38338kW

Verification of resonance conditions: $f_0 = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{2\pi\sqrt{0.1 \times 10^{-4}}} \approx 50.3\text{Hz}$, which is close to the operating frequency and satisfies the resonance condition. The transformer turns ratio. $n = \frac{N_2}{N_1} = 1.2$, therefore the effective value of the transformer secondary no-load voltage is: $U_1 = n \times U_0 = 1.2 \times 220 = 264\text{V}$ (where U_0 is the total input voltage of the power supply), and the forward voltage drop of the 1N4007 diode is $U_D \approx 0.7\text{V}$. Therefore, the effective value of the voltage across the load R_2 is $U_2 = U_1 - \frac{2U_D}{\sqrt{2}} = 264 - \frac{1.4}{1.414} \approx 263\text{V}$. The load output power $P = \frac{U_2^2}{R_2} = \frac{263^2}{50} = 1.38338\text{kW}$. The quality factor of the transmitter is $Q = \frac{2\pi f_0 L}{R_1} = \frac{2 \times 3.1416 \times 50 \times 0.1}{1} \approx 31.42$, indicating that the loss of the transmitter circuit is relatively small.

2.2. Safety and feasibility analysis of the scheme

This solution does not use contact charging, which effectively avoids the risk of leakage and electric shock. Furthermore, the magnetic field is relatively concentrated, the external radiation is low, and it has certain overvoltage, overcurrent, and overtemperature protection. This set of schemes is simple in structure, universal in devices, and low in cost. It is suitable for static charging in the car and in the parking lot, and basically has the value of engineering promotion. *RLC* resonant charging technology allows the transmitting coil and receiving coil of the system to be at the same operating frequency. At a specific working frequency, the transmitting coil and the receiving coil will resonate. In this way, electricity can be transmitted efficiently

^[6]. In practical applications, you can also add controllers, integrated sensors, communication modules, and so on. The wireless charging system can be remotely monitored, fault-diagnosed, and adaptively adjusted, and the user experience and intelligence level can also be improved ^[7].

2.3. Circuit structure design

This scheme uses a dual-port network between the transmitter and the receiver, plus the *RLC* oscillation circuit. Within the range allowed by the error, a relatively stable DC pulsation voltage can be obtained at both ends of the load. The circuit topology based on the principles mentioned above is shown in **Figure 1**.

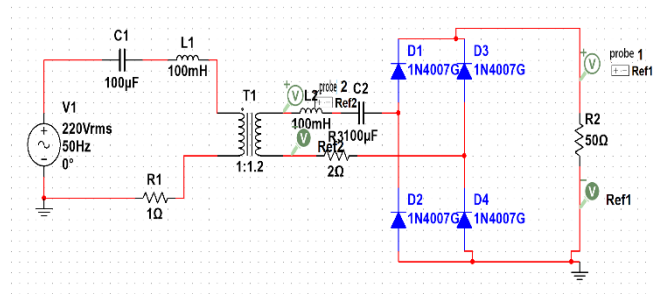


Figure 1. Circuit topology diagram.

3. Test results

Use a voltmeter probe to measure the total voltage at the receiving end and the output voltage across the load. The circuit structure was built using NI Multisim 14.3 simulation software, and the waveforms of the total output voltage and the voltage across the load were plotted, as shown in **Figure 2**.

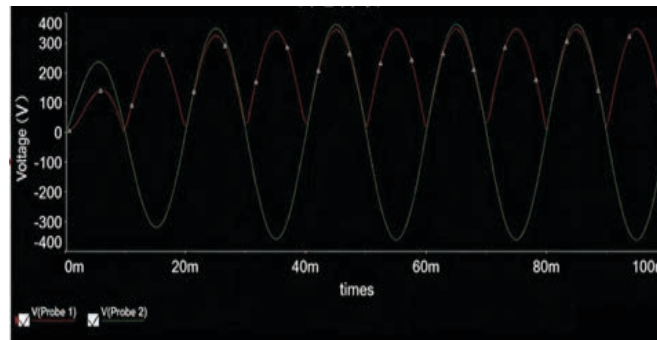


Figure 2. Waveforms of the total output voltage and the voltage across the load.

Analysis of the waveform in **Figure 2** shows that: Within the allowable error range, the total output voltage exhibits a basically stable sine wave, indicating that the *RLC* resonant system has achieved basic stable operation. Energy transfer via magnetic coupling is normal. The voltage across the load is a rippled DC voltage. The peak voltage is approximately 380–400V. This is the expected output waveform after rectification by the wireless charging system. It can directly charge the load of electric vehicles. It meets the basic requirements for wireless charging. The system detects indicators such as the state of resonance, operational stability, safety, and feasibility. The results show that the waveform is relatively smooth, and no

drastic voltage changes have been observed. The entire charging system operates in a coordinated and stable manner, and its overall performance is good. Able to provide a basically stable pulsed DC power to the load. Its kilowatt-level power output gives it a long battery life. The temperature, voltage, and other properties of each component all meet safety standards.

4. Optimization research

4.1. Research limitations

The main shortcomings of this solution are as follows: The transmission distance is relatively short, and the efficiency will decrease significantly as the distance increases; When the system is working, the coil that is supplied with a high-frequency signal may generate significant heat, which accelerates the aging of the coil; If the coil deviates, the power will fluctuate, and the overall stability still needs to be improved. These shortcomings of wireless charging systems can be optimized in the following aspects to meet the charging needs of electric vehicles better.

4.2. Transmission distance optimization

The transmission distance will directly determine the upper limit and application areas of wireless charging. Optimizing transmission distance is a top priority for improving various wireless charging systems. Adjusting the magnetic field focusing effect is the most important part of increasing the transmission distance. Optimizing the distance for power transmission requires consideration of energy loss, transmission efficiency, and cost. The reference proposes a coil optimization design method for a high-efficiency mid-range magnetic resonance coupled wireless charging system for electric vehicles that adapts to demand-side management^[8]. The structure of the charging coupling coil was designed, and the efficiency changes caused by the relative offset of the coil were optimized and analyzed. The optimal operating frequency of the system was determined, which improved the transmission distance of the system.

4.3. Coil heating optimization

According to Joule's law, current flowing through various circuit components will definitely produce thermal effects. Long-term heating will reduce the efficiency of the equipment and accelerate the aging of the components. If it is serious, it may also bring danger. The temperature rise of the wireless charging coil is also a key factor affecting the reliability and efficiency of the system. If the temperature is too high, it will accelerate the aging of the insulation and also increase the copper loss and magnetic core loss. Using the electromagnetic-thermal coupling model, coupled with the multi-objective Pareto optimization method. By optimizing the winding structure, the layout of ferrite, and the heat dissipation structure, it can effectively reduce the steady-state temperature of the coil. It can effectively reduce the steady-state temperature of the coil^[9].

4.4. System efficiency optimization

Improving the energy conversion efficiency of the system can enhance the overall efficiency of the charging system, making charging tasks more convenient and faster. Reference suggests that high-frequency copper loss can be effectively reduced by setting the coil structure parameters reasonably (such as optimizing the wire diameter and number of strands, using Litz wire for winding, and optimizing the number of turns

distribution)^[10]. Choose magnetic core materials with high magnetic conductivity to reduce the loss of the magnetic core itself. Then make adaptive adjustments to the size and geometric shape of the magnetic core; this can reduce iron loss and enhance magnetic field coupling. Improving the design of the magnetic shielding structure and the spatial layout of the coil can suppress the phenomenon of magnetic leakage and improve the utilization rate of magnetic flux.

4.5. System stability optimization

The primary prerequisite for designing wireless charging circuits is a stable charging process. The stability of the system is mainly reflected through the dynamic balance of output power. To solve the power fluctuation problem caused by coil misalignment, researchers have introduced some feedback mechanisms. For example, closed-loop power control, power factor correction, and automatic power control are used. They are used to reduce or eliminate the impact of uncertainties. In this way, the output power and various voltage and current parameters can basically be stable. The overall stability of the system in the dynamic working state has also been improved.

5. Conclusion

This paper designs a wireless charging circuit based on the principles of electromagnetic induction and resonance. It is divided into two networks: the transmitting end and the receiving end. The transmitter uses AC power as the excitation source. The receiving end, in conjunction with a bridge rectifier circuit, provides pulsed DC voltage to the load. The circuit structure of this solution is relatively simple and has strong feasibility. It facilitates the detection of the source of the fault when a circuit malfunctions. The overall cost is low. Non-contact magnetic coupling charging is used, and the components are functioning well. Safety is guaranteed. The voltage across the load terminals exhibits a basically stable pulsed DC voltage. The power output reached the kilowatt level, and both the *RLC* series resonant and bridge rectifier systems operated normally. Good energy transfer effect. However, there is still room for improvement in transmission distance, heat generation, system efficiency, and stability. This research can now be applied to charging in fields such as wireless charging for household electric vehicles, automatic parking lot recharging, and on-board low-voltage auxiliary power supplies. In the future, this technology will continue to develop towards long-distance, high-efficiency charging. Further optimize heat generation and operational stability. To better adapt to the high power and dynamic wireless charging of electric vehicles, giving it a broader application prospect.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Chen P, Niu H, Ji S, 2025, Research on Magnetic Coupling Resonant Wireless Charging System for Electric Vehicles. *Special Purpose Vehicle*, 2025(06): 21–23.
- [2] Yang F, 2023, Research on Electromagnetic Shielding Design and Safety of Wireless Charging Coupling Mechanism for Electric Vehicles, thesis, Chang'an University.

- [3] Li X, 2024, Research on Power Stability Control Optimization of Dynamic Wireless Charging System for Electric Vehicles, thesis, Southeast University.
- [4] Xing R, 2025, Efficiency Improvement and Electromagnetic Compatibility Analysis of Wireless Charging Technology for Electric Vehicles. Popular Science & Technology (Volkswagen Qiche), 2025(07): 43–45.
- [5] Shi W, 2026, Research on Wireless Charging Technology for Power Batteries of New Energy Vehicles on Expressways. Internal Combustion Engine & Parts, 2026(06): 104–106.
- [6] Zhang Q, 2021, Research on Static Resonant Wireless Charging System for Electric Vehicles, thesis, Anhui University of Science and Technology.
- [7] Qin S, 2025, Design and Optimization of Wireless Charging System for Electric Vehicles. Auto Time, 2025(02): 124–126.
- [8] Zheng G, 2017, Optimal Design of High-Efficiency Medium-Distance Magnetic Resonance Wireless Charging Coil for Electric Vehicles Adapting to Demand-Side Management. Transactions of China Electrotechnical Society, 32(S1): 209–216.
- [9] Ma J, Li Z, Liu Y, et al., 2023, Thermal Analysis and Optimization of the Magnetic Coupler for Wireless Charging System. IEEE Transactions on Power Electronics, 38(12): 16269–16280.
- [10] Zhong Y, 2025, Optimization Research of Magnetic Coupling Mechanism for Electric Vehicle Wireless Charging System. Popular Science & Technology (Volkswagen Qiche), 2025(11): 70–72.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Strategic Analysis of Dynamic Detection and Tracking Methods for Autonomous Vehicles

Chenhe Kang

School of Mechanical Engineering, Shandong University, Jinan 250061, Shandong, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To solve the problem of perception failure due to vehicle occlusion in urban areas, this paper will review and analyze many dynamic detection and tracking methods proposed in recent years in detail. The main idea of the first representative study is to use laser beams and geometric figures for rigid matching. They are very effective in open areas, but because there is no information on the surface, the algorithm cannot work when there is occlusion. Earlier, researchers put forward a soft-assignment method by building a “four-rectangle measurement model”, and it was found that probabilistic modeling is better than binary matching; now, a laser ray does not need to hit an object, and the matching probability of local sparse point clouds can be calculated. A good way to improve the efficiency of tracking is the scaled sequence algorithm. The first problem of this algorithm is that it involves a high-dimensional state search. The detection rate of severe occlusion in the KITTI dataset for the new model is higher than that of traditional methods. As shown in the above case, although this method is relatively stable in the face of missing values, it has some defects; that is, it is too sensitive to the initial prior and may be trapped in a local optimum under difficult conditions. Due to the shortcomings of the previous models, such as fixed dimensions and an inability to address all kinds of traffic flexibly, it is hoped that future research will be based on multi-scale adaptive geometric models to achieve all-weather sensing and all-conditions operation. Vehicle occlusion, Likelihood Field Model, Scaling sequence algorithm, self-motion compensation.

Keywords: Likelihood field model; Four-rectangle measurement model; Scaling sequence algorithm; Adaptive multiscale modeling

Online publication: Jul 23, 2026

1. Introduction

With the development of artificial intelligence and sensor technology, autonomous vehicles have gradually gone from closed-track tests to public road operation in many cities. The first need for safety in a crowded area of the city is to detect and comply with the operation of nearby moving vehicles. Due to the frequent occlusion of urban traffic, the traditional perception algorithm has serious problems; that is to say, when local features are absent, the bounding box regression of the object detector is inaccurate, and there is considerable

drift. To solve the problems of vehicle detection and tracking, many studies have been carried out. First, only the grid method was studied, so grid cells were used to represent the environment. These methods have a large number of calculations and are restricted by the grid resolution; thus, they cannot meet the demand for real-time high-speed driving. Then, models based on geometric features began to be employed gradually. Among them, the “beam model” is the most typical one. The two-dimensional ray path of LiDAR is used to determine the geometric shape of the vehicle; it is very fast in an open area. The beam model has a serious defect; it determines whether a ray “hits” the surface of an object in the calculation of probabilities. That is to say, if the target vehicle is covered by other objects, the laser rays will be blocked from reaching the surface of the vehicle directly; thus, the model cannot acquire valid measurement data, and the detection and tracking function will fail. Due to the problem of perception failure in a complex urban environment, the previous methods for dynamic vehicle detection and tracking will be introduced and analyzed systematically in this paper. The purpose of the likelihood field model is to determine how probable it is that a measurement point is close to the shape of the target, and it is no longer necessary to consider only the ray-surface intersection point. Researchers have made some new applications of this idea in vehicle detection and constructed a four-rectangular vehicle measurement model to cover all visible parts of the vehicle, as well as its interior and exterior. This model can make full use of sparse 3D point cloud information; even when the target vehicle is largely occluded, it is still possible to obtain a good estimate of the vehicle’s pose as long as a small number of laser points fall in the high-probability region of the likelihood field. To solve the problem of real-time computation in high-dimensional state spaces, an efficient Bayesian state estimation method called the scaling sequence algorithm has also been proposed and optimized; it is based on an “annealing” scheme and can increase the speed of state estimation. This algorithm is a multi-resolution search method that first takes a large step, then a small step, and so on; that is to say, it quickly finds an approximate location of the target and then refines it step by step to save computation. In this paper, we will present the likelihood-field-based vehicle measurement model and its optimized calculation method, introduce the whole system workflow that includes improved virtual scanning, hypothesis generation and tracker update, show experimental results on the KITTI dataset, focus on analyzing the detection performance in occlusion scenarios, and finally summarize this paper and put forward future research directions^[1].

The basic task of the entire dynamic detection and tracking system for autonomous driving in complex cities is to find the position of a car accurately. Due to the instability of the old way in occlusion environments, Tongtong Chen and others have carried out a large number of studies on the likelihood field-based vehicle measurement model and optimization strategies to promote the development of this technology^[2].

2. Construction of vehicle likelihood field models and algorithm performance optimization for complex occlusion scenarios

2.1. Reconstruction of the likelihood field vehicle measurement model

Traditionally, the beam model is used for vehicle detection; Petrovskaya and Thrun first mentioned in their early work that one way to find an obstacle is to determine the length of a two-dimensional laser beam, and the beam model has the advantage of very high computation speed^[3]. The problem with this type of detection is that if the target vehicle is blocked by other cars or trees on the road, the laser beam cannot reach the ground for reflection, so the model does not work and cannot detect it; although a particle filter was later added to the tracking framework to estimate velocity, it also performed poorly in a complex and heavily

occluded environment ^[4].

Therefore, scholars have been studying the basic theorems of probabilistic robotics. At the core of the above theoretical adjustment is the systematic model philosophy of probabilistic robotics. To address the uncertainty of the urban environment in the model of sensor perception more reasonably, this system will represent the uncertainty as a probability distribution and provide strong mathematical support for the quadrature measurement model ^[5]. Based on the work by Thrun et al. on robust Monte Carlo localisation, a new way of calculating probabilities has been introduced for the “likelihood field”, and according to Bayesian filtering theory, the likelihood field measurement model in this paper no longer needs to scan the entire outline of the vehicle; instead, it is proposed that “four rectangular boxes” (that is to say, the front, body, interior and exterior regions) be used for geometric modelling of the target vehicle ^[6]. The advantage of the above method is that the sparse 3D point cloud can be transformed into a four-layer rectangular model, and then matching probabilities are calculated; in the traditional direct line-of-sight method, a likelihood field-based soft assignment mechanism can be used to infer the whole position and orientation of the vehicle by maximising the probability, even when only a small number of point clouds from the partial vehicle region are available. With the recent cutting-edge research in dynamic object detection for LiDAR, a likelihood field model can be obtained from local observations to obtain the global pose, and it has shown good results in heavy urban traffic ^[7].

2.2. Algorithm acceleration and state optimization strategies

The model of the likelihood field is theoretically free from occlusion, but it is too computationally intensive. To improve the speed of computation for the autonomous driving system, which must respond within milliseconds, two optimizations have been made to the algorithm.

Probability calculations in the likelihood field have a large number of complicated exponential and distance operations. At the beginning of the system, frequently used mathematical formulas can be pre-calculated and stored as high-precision offline data tables; then, to get the probability values in real time, only a simple memory lookup is needed. Engineering optimization has increased the overall speed of computation by about 4.5 times compared with real-time calculation and reduced the computational load on the onboard computer significantly; therefore, this way of trading computational efficiency for data preprocessing has also been widely used in recent research on point cloud perception. For example, a small encoder can be used to reduce the amount of data that needs to be transmitted from the car’s sensors in real time, so all the original point cloud data do not have to be transferred ^[8].

A full-space global optimum search would be too far-reaching; it is not easy to find a needle in a haystack. Therefore, when the algorithm first searches for a potential vehicle, it has a loose matching criterion (that is to say, it is a rough search with a high threshold); after the system finds a target in a certain area with a high probability, it will gradually lower the criterion (reduce the threshold) and perform a local, high-precision pose alignment. In other words, it is a “coarse-to-fine” method that can be both fast in terms of search speed and very accurate for final localization, thus supporting continuous tracking.

3. Dynamic vehicle detection and tracking system

Tongtong Chen and others have built a very stable single-frame pose estimation model, and based on this model, a whole dynamic vehicle detection and tracking system has been constructed according to the

temporal tracking framework.

3.1. Improved virtual scan and hypothesis generation

Data preprocessing and the extraction of the Region of Interest (ROI) need to be carried out for the tracking system. In the construction of an occupancy grid in traditional virtual scanning technology, it is often the case that occluded unknown regions are either ignored or marked as unobstructed. It is suitable for an open area, but due to the large number of vehicles at this busy intersection, continuous observation of the obscured vehicle is not feasible.

To solve the above problem, the improved virtual scanning method has changed the logic of global uniform processing and carries out separate clustering and analysis on each object in the scan image. A spatiotemporal difference mechanism is used in the system to accurately align the radar images of the previous frame with those in the current frame based on the own GPS and inertial navigation data of the autonomous vehicle, remove static background elements, and if a point cloud cluster shows a large spatial displacement, it will be promptly marked as a “suspected vehicle” and trigger the likelihood field model for pose verification. According to the latest research results on dynamic vehicle detection in road scenes, this method is based on prior pose estimation and dynamic differentiation to filter out interference from leaves and pedestrians and reduce the false positive rate of the system to a certain extent ^[9].

3.2. Tracker updates and self-motion compensation

The main problem in the target-tracking stage is how to track a moving object over a long period stably. The first type of algorithm that the researchers have been studying is the very efficient Bayesian estimator called “Scaling Series” ^[10]. It was first designed for overseas adaptation in terms of touch feedback for robotic arms, and it already has good support for all kinds of data distributions.

It does not change much if it is different. After a large number of alterations, it has been successfully employed in the area of vehicle tracking for autonomous driving. The first addition was the self-motion compensation function. Autonomous driving is not only about sensing objects in the external environment but also about moving the car at a high speed. Self-motion compensation is to subtract the vehicle’s own motion data, such as vibrations due to acceleration and deceleration, and angular velocity produced during turns at intersections, from the relative motion; thus, the system does not have a change in its own coordinate system and only contains the absolute motion trajectory of the target vehicle in the global coordinate system. Chen Fu and others have found that to guarantee the accuracy of tracking in a highly dynamic environment, strict spatio-temporal synchronization and self-motion compensation need to be carried out; at the start of a traffic light change or during a complex turn at an intersection, it can be observed that ^[11].

4. Technical challenges and development

Although dynamic vehicle detection and tracking algorithms based on likelihood field models have achieved some success in solving the problem of vehicle occlusion and reached a detection rate of up to 65% in experiments, autonomous driving perception systems still have many technical problems that need to be solved before they can be deployed on a large scale in open urban roads.

First of all, there is a problem of computation time and real-time demands. Although the previous research has presented lookup table methods and improved scaling sequence algorithms to accelerate high-

frequency computation and search in high-dimensional state spaces, keeping the response time in the millisecond range on the limited onboard computing platform is still a serious engineering problem due to the increasing congestion in urban traffic. Second, the main bottleneck of the current system is that it has fixed geometric dimensions for the models; Secondly, the core problem of the current system is the fixed size of the models, and as a result, the evaluation framework for 3D multi-object tracking is also relatively unchanged. Although the basic Bayesian framework can handle occlusion, it cannot adaptively adjust according to different vehicle sizes and performs poorly in mixed-traffic conditions. Therefore, in order to achieve all-scenario perception, a 3D tracking benchmark with good generalization ability needs to be constructed^[12]. Although the existing quadrilateral measurement model can use sparse point clouds for local pose estimation, fixed-size models are not very suitable for the complex reality of mixed traffic flow in urban areas; thus, the generalization and perception capabilities of the system are limited by all kinds and various sizes of vehicles. In addition, due to the changes in traffic conditions and high speeds of vehicles, keeping up with and constantly tracking them both in space and time has technical difficulties.

In light of the problems above, the next step in technological progress will be to make perception models and multi-dimensional scene understanding more adaptable. In the future, more studies will be conducted on the development of adaptive multiscale geometric modeling techniques. Free from the constraints of a fixed size, the system can now flexibly match and reassemble the geometric features of all kinds of objects to meet the full-scene perception requirements of complex mixed traffic, and one of the necessary developments to improve the foresight prediction ability of the perception system is automatic identification of object categories at long distances and construction of stable spatio-temporal correlations among multiple objects. Finally, by continuous adjustment of algorithms and improvements in the mechanisms for self-motion compensation and autonomous driving perception technology, the problems of occlusion and tracking in complex environments will be fully solved, and safe application in all scenarios and various weather conditions will be realized.

5. Conclusion

With the start of autonomous driving on the road, vehicle occlusion has been an obstacle to traditional perception technology for a long time. To solve the problem of a perception bottleneck, a four-rectangle measurement model based on likelihood fields was constructed, and then it was employed in conjunction with an annealing search strategy. Through the combination of spatio-temporal differentiation and self-motion compensation, a stable dynamic tracking system has been established, and in cooperation with the previous beam-based perception model, it has been found that our new system can effectively solve the problem of detection omission and tracking failure caused by direct line-of-sight obstruction in heavy traffic and vehicle occlusion. It can significantly improve both the success rate and stability of dynamic target detection, provide strong technical support for dynamic targets in complex and rapidly changing road conditions, but current models still have some issues in actual application; that is to say, most current models have fixed-dimension geometric models and limitations in generalization and reconstruction when faced with all kinds of mixed traffic flow in urban areas.

Research has also been carried out on new ways of likelihood field theory and probabilistic soft assignment for vehicle measurement. The research carried out by the above scholars has provided some references for engineers in the design of high-level perception systems and researchers in the field of urban

traffic safety; currently, these methods can be applied to solve the problem of continuous target tracking under severe occlusion and offer a reliable and practical technical path. Looking ahead, it will be necessary to add the functionality of multi-scale adaptive geometric model construction and precise identification of distant targets on the horizon to the perception system to address problems in complex environments and truly advance the commercial application of all-weather, all-scenario safety.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Geiger A, Lenz P, Stiller C, et al., 2013, Vision Meets Robotics: The KITTI Dataset. *The International Journal of Robotics Research*, 32(11): 1231–1237.
- [2] Chen T, Wang R, Dai B, et al., 2016, Likelihood-Field-Model-Based Dynamic Vehicle Detection and Tracking for Self-Driving. *IEEE Transactions on Intelligent Transportation Systems*, 17(11): 3142–3158.
- [3] Petrovskaya A, Thrun S, 2008, Efficient Techniques for Dynamic Vehicle Detection. *Proceedings of the 11th International Symposium on Experimental Robotics*: 261–271.
- [4] Petrovskaya A, Thrun S, 2009, Model-Based Vehicle Detection and Tracking for Autonomous Urban Driving. *Autonomous Robots*, 26(2–3): 123–139.
- [5] Thrun S, Burgard W, Fox D, 2005, *Probabilistic Robotics*. MIT Press.
- [6] Thrun S, Fox D, Burgard W, 2001, Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence*, 128(1–2): 99–141.
- [7] Cellina M, Corno M, Savaresi S, 2025, LiDAR-Based Vehicle Detection and Tracking for Autonomous Racing. *arXiv*.
- [8] Lang A, Vora S, Caesar H, et al., 2019, PointPillars: Fast Encoders for Object Detection from Point Clouds. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*: 1–8.
- [9] Motro M, Ghosh J, 2018, Measurement-Wise Occlusion in Multi-Object Tracking. *arXiv*.
- [10] Petrovskaya A, Khatib O, 2011, Global Localization of Objects via Touch. *IEEE Transactions on Robotics*, 27(3): 569–585.
- [11] Fu C, Dong C, Zhang X, et al., 2019, Low-Cost LIDAR-Based Vehicle Pose Estimation and Tracking. *arXiv*.
- [12] Weng X, Wang J, Held D, et al., 2020, AB3DMOT: A Baseline for 3D Multi-Object Tracking and New Evaluation Metrics. *arXiv*.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Distributed Flexible Job Shop Scheduling Based on an Improved Genetic Algorithm

Wenwen Chen*, Bowen Du

School of Electrical and Automation Engineering, Liaoning Institute of Science and Technology, Benxi, Liaoning, China

*Corresponding author: Wenwen Chen, 1375400124@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Aiming at the multi-dimensional and highly complex decision-making characteristics of distributed flexible job shop scheduling, and the local convergence and slow convergence speed of the traditional genetic algorithm, an improved genetic algorithm- the dynamic hierarchical genetic algorithm is proposed. The algorithm enhances the global search ability and local optimization ability of the traditional genetic algorithm through a dynamic hierarchical structure, a multi-neighborhood local search strategy, and adaptive adjustment of the crossover and mutation rates. The simulation results show that after 10 independent runs on several improved Kacem (MK) series test cases, the algorithm has a lower minimum average maximum completion time than the traditional genetic algorithm. Especially in the complex cases of MK03 and MK09, not only is the completion time reduced by 24%, but also the number of iterations is reduced. This fully verifies the ability of the algorithm to improve scheduling efficiency and stability in distributed flexible manufacturing workshops. The completion time of MK09 is reduced by about 24%, and the number of iterations is also reduced. This fully verifies the ability of the algorithm to improve the scheduling efficiency and stability in the distributed flexible manufacturing workshop scheduling problem.

Keywords: Distributed flexible job shop scheduling; Dynamic population classification; Multi-neighborhood local search

Online publication: Jul 23, 2026

1. Introduction

At present, most mechanical manufacturing industries adopt a flexible job shop scheduling mode, which can flexibly adjust processing procedures and allocate multiple machines according to resource load ^[1]. With the popularization of the distributed production mode, the distributed flexible job shop scheduling problem has become the core optimization problem in the field of intelligent manufacturing ^[2]. DFJSP is an NP-hard problem, which needs to complete three key decisions at the same time: the selection of workpiece processing plants, the allocation of processing machines, and the determination of processing sequence.

The scheduling plan directly affects equipment utilization, production cycle, and energy consumption

level ^[3]. Reasonable scheduling can significantly improve production efficiency, while unreasonable scheduling will lead to idle equipment and production delay ^[4].

For the solution of DFJSP, numerous studies have been conducted by scholars both at home and abroad. SANG Y W constructed a multi-objective DFJSP model to solve the production scheduling problem in distributed intelligent factories and designed a memetic algorithm for its solution ^[5]. XU W X considered the scenario of outsourcing of workpiece operations and combined the tabu search algorithm with the genetic algorithm to solve the multi-objective DFJSP model ^[6]. WANG S Y, for the DFJSP model with assembly machines, completed the model solution in a vector manner ^[7]. LI J proposed an improved artificial bee colony algorithm for DFJSP and improved the performance by optimizing the initial population and local search ^[8]. Li Y W employed an improved heuristic algorithm to solve the DAFJSP model for multi-objective optimization ^[9]. Although the existing algorithms have made certain progress, the traditional genetic algorithm (GA) still has the defects of slow convergence speed and an easy tendency to fall into a local optimum when solving DFJSP, and is difficult to handle large-scale and complex scheduling scenarios efficiently ^[10].

To address the aforementioned issues, this paper proposes the Dynamic Hierarchical Genetic Algorithm (DH-GA). By implementing a dynamic population hierarchical strategy to balance the retention of high-quality solutions and the global search capability, it designs multi-neighborhood local search operators to deeply explore the improved space of the solutions. Combined with adaptive genetic operations to enhance the algorithm's stability, it ultimately provides a more efficient and feasible solution for the engineering application of DFJSP.

2. Problem description

The mathematical problem of DFJSP can be specifically described as:

- (1) The set of workpieces $J = \{J_1, J_2, \dots, J_i, \dots, J_n\}$, where n represents the total number of workpieces, and J_i indicates the i -th workpiece.
- (2) The factory set $U = \{U^1, U^2, \dots, U^i, \dots, U^q\}$, where q represents the total number of factories, and each factory U^i is a flexible production workshop.
- (3) Each factory U^i has a set of processing units (machines) forming $M = \{M^{i1}, M^{i2}, \dots, M^{ik}, \dots, M^{im}\}$, where m represents the number of processing units in this factory, and M^{ik} indicates the k th processing unit of factory U^i .
- (4) Process set $O_i = \{O_{i1}, O_{i2}, \dots, O_{ij}, \dots, O_{in_i}\}$, where n_i represents the total number of processes for workpiece i , and O_{ij} indicates the j th process of workpiece i .

The core optimization objective of DFJSP is to minimize the maximum completion time for all workpieces (see **Figure 1**).

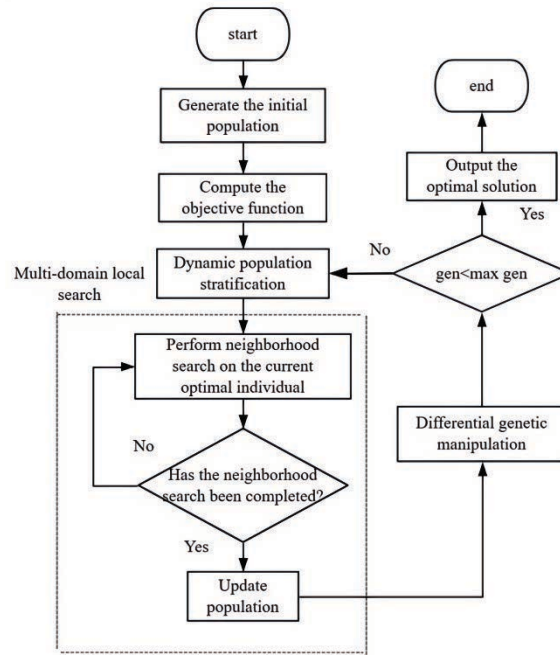


Figure 1. DH-GA Algorithm flowchart.

$$C_{\max} = \max_{i,j,l,k} \{C_{i,j,l,k}\} \quad (1)$$

3. DH-GA algorithm

The flow of the DH-GA algorithm is shown in Figure 1, starting with parameter initialization, and finally completing population update and iteration termination through coding population initialization, decoding and fitness calculation, dynamic population stratification, differential genetic operation, and three-layer multi-neighborhood local search. Through the synergy of dynamic population stratification and multi-neighborhood search, this framework realizes the dynamic balance between global exploration and local development, and provides a feasible technical way to solve the DFJSP problem efficiently.

Dynamic population classification adjusts the retention rate of excellent individuals in the iterative process. For example, in the early iteration stage, the retention rate of high-quality individuals is set to 0.5. Quickly screen elite features at the beginning of iteration to prevent the loss of high-quality genes at the beginning of iteration. In the middle of the iteration, the retention rate dropped linearly to 0.3, which expanded the population of ordinary individuals and enhanced the global search ability. In the later stage, the retention rate increased to 0.4, and the population diversity was increased to avoid the algorithm falling into a local optimum.

Multi-neighborhood local search adopts a three-layer multi-neighborhood search based on DFJSP decision-making feature design, and carries out local optimization search for high-quality individuals produced in the population iteration process. Firstly, the adjacent process of the optimal individual is exchanged; Secondly, some processes are randomly selected. Finally, the processes in the selected factory are redistributed to other idle processing units.

Differential genetic operation fully combines the linear decreasing crossover rate with the adaptive mutation rate. The crossover rate decreased linearly from 0.9 to 0.7, which promoted population diversity in the early iterations and remained stable in the later period to protect elite individuals. In order to ensure the diversity of the population, when the population coincidence rate exceeds the threshold, the mutation rate increases to 0.02; otherwise, it is kept at 0.01 to prevent the algorithm from falling into a local optimum.

4. Analysis of experimental results

In this paper, MATLAB R2023b is used for programming, and the DH-GA algorithm and the traditional GA algorithm are compared and analyzed, and the minimum and maximum completion times of the target are optimized according to the experimental results.

In order to verify the effectiveness of the algorithm, this paper selects a subset of MK test cases, and the corresponding test data set is shown in **Table 1**.

Algorithm parameters: iteration number 150, population size 200, initial crossover probability 0.9, initial mutation probability 0.01. In order to ensure the fairness of the experiment, the parameters of the genetic algorithm are consistent.

Table 1. Example parameters

Example	Workpiece count	Machine count	Factory count	Process count
MK01	10	6	2	55
MK02	10	6	2	58
MK03	15	8	3	150
MK04	15	8	3	90
MK05	15	4	2	106
MK06	10	10	3	150
MK07	20	5	2	100
MK09	20	10	3	240

The case of MK03 is tested by the DH-GA and GA algorithms, respectively. In **Figure 2**, at the beginning of the iteration, the DH-GA algorithm converges faster than the GA algorithm. In the middle of the iteration, the search ability of the GA algorithm drops and falls into a local optimum, while the DH-GA algorithm continues to optimize and finally converges in 155 minutes. This shortens the completion time by about 24% compared with the GA algorithm. This shows that the DH-GA algorithm provides superior optimization performance and greater stability.

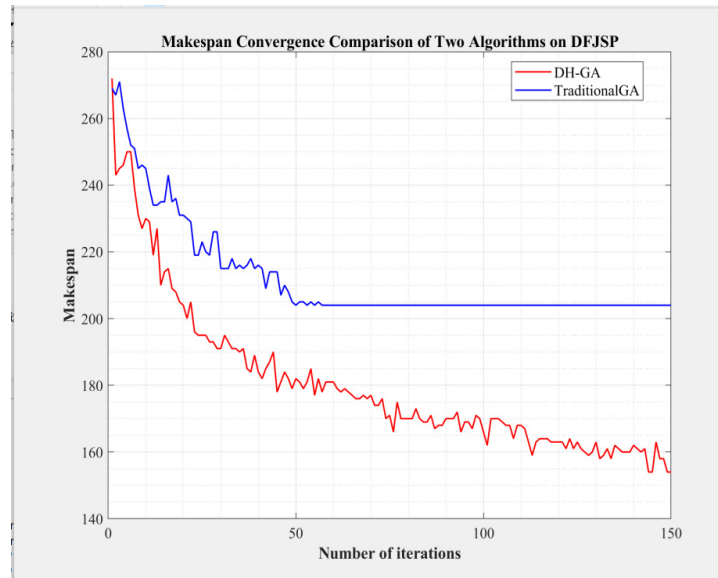


Figure 2. Iteration comparison of DH-GA and GA algorithms under case study MK03.

Based on the MK case, DH-GA and GA have been tested 10 times respectively, and the average results of 10 runs are shown in Table 2. The results show that compared with the traditional GA algorithm, the proposed DH-GA algorithm achieves better completion time in the MK case. It is worth noting that in the complex MK09 case, its ability to search the global optimum is significantly improved, which fully verifies that the DH-GA algorithm improves the scheduling efficiency (see **Table 2**).

Table 2. Average value of the calculation example

Example	DH-GA/min	GA/min
MK01	42	42
MK02	29	30
MK03	199	204
MK04	64	66
MK05	176	179
MK06	73	75
MK07	147	149
MK09	308	321

5. Conclusion

This study solves the limitations of local convergence and slow convergence of the traditional genetic algorithm in solving NP-hard DFJSP. In order to overcome these challenges, the DH-GA algorithm is proposed. Through dynamic population stratification, multi-neighborhood local search, and adaptive genetic operation, the algorithm effectively solves the multi-dimensional optimization problem of DFJSP. Compared with the traditional genetic algorithm, it has a significant improvement in convergence speed, solution accuracy, and load balance, and the maximum completion time is reduced by 24% for complex MK test cases. This research provides a feasible technical solution for distributed flexible workshop scheduling in

intelligent manufacturing. At present, the research is limited to single-objective optimization. Future work will extend to multi-objective optimization, adaptive parameter adjustment based on machine learning, and practical application in dynamic scheduling scenarios.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhang H, Chen Y, Xu G, et al., 2025, Distributed Assembly Flexible Job Shop Scheduling with Dual-Resource Constraints Via a Deep Q-Network Based Memetic Algorithm. *Swarm and Evolutionary Computation*, 98: 102086.
- [2] Cui X, Wan L, Zhao H, et al., 2023, A Deep Reinforcement Learning-Based Scheduling Method for Flexible Manufacturing Workshops. *Manufacturing Technology and Machine Tools*, 2023(12): 165–170.
- [3] Fan Y, 2023, Research on Dynamic Scheduling of Multi-Objective Flexible Manufacturing Workshops Based on an Improved NSGA-II Algorithm, thesis, Hefei University of Technology.
- [4] Kaikai Z, Guiliang G, Ningtao P, et al., 2023, Dynamic Distributed Flexible Job-Shop Scheduling Problem Considering Operation Inspection. *Expert Systems with Applications*, 224: 119840.
- [5] Yanwei S, Jianping T, 2022, Intelligent Factory Many-Objective Distributed Flexible Job Shop Collaborative Scheduling Method. *Computers & Industrial Engineering*, 164: 107884.
- [6] Wenxiang X, Yongwen H, Wei L, et al., 2021, A Multi-Objective Scheduling Method for Distributed and Flexible Job Shop Based on Hybrid Genetic Algorithm and Tabu Search Considering Operation Outsourcing and Carbon Emission. *Computers & Industrial Engineering*, 157: 107318.
- [7] Wang S, Wang L, 2015, An Estimation of Distribution Algorithm-Based Memetic Algorithm for the Distributed Assembly Permutation Flow-Shop Scheduling Problem. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(1): 139–149.
- [8] Li J, Bai S, Duan P, et al., 2019, An Improved Artificial Bee Colony Algorithm for Addressing Distributed Flow Shop with Distance Coefficient in a Prefabricated System. *International Journal of Production Research*, 57: 1–15.
- [9] Li Y, Song L, 2025, Research on Scheduling for Distributed Flexible Manufacturing Workshops Considering Energy Consumption. *Manufacturing Technology and Machine Tools*, 2025(03): 157–165.
- [10] Meng G, Huang J, Wei Y, 2024, Solving Flexible Job Shop Scheduling Problems Using a Hybrid Beluga Whale Optimization Algorithm. *Computer Engineering and Applications*, 60(12): 325.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Virtual Simulation Modeling and Validation of Traffic Measurement Equipment Based on Multi-Physical Field Coupling

Lin Zhou, Jing Wang, Nan Wang, Chi Zhang, Zhongtian Jin, Lucheng Wang

School of Automotive Engineering, Beijing Polytechnic University, Beijing 100176, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: As the smart transportation system continues to evolve, the precise and stable operation of traffic measurement equipment directly determines the overall effectiveness of traffic data monitoring and system management. Traditional field tests are limited by specific operational scenarios, narrow coverage of driving conditions, high equipment wear and maintenance costs, and fail to meet the rigorous performance verification requirements under complex environments. This study integrates theories from mechanics, thermodynamics, and electromagnetism to establish a virtual simulation framework for traffic measurement equipment, enabling accurate replication of real-world operating conditions, conducting performance simulations, and facilitating continuous model refinement. This approach overcomes the limitations inherent in single-physical-field simulations. The outcomes provide robust digital support for equipment performance testing, structural optimization, and condition-specific calibration, thereby advancing the development and modernization of measurement systems in smart transportation applications.

Keywords: Multi-physics coupling; Traffic measurement equipment; Virtual simulation; Modeling techniques; Performance verification

Online publication: Jul 23, 2026

1. Introduction

The development pace of smart transportation infrastructure continues to accelerate. As core hardware for traffic data acquisition and operational status monitoring, traffic measurement equipment covers various application scenarios including vehicle speed detection, traffic flow analysis, and load measurement. The operational quality of these devices directly impacts the accuracy and stability of the entire transportation system. Operating in complex outdoor environments, traffic measurement equipment faces interactions from external factors such as mechanical vibrations, temperature fluctuations, electromagnetic interference, and airflow variations, which can lead to measurement accuracy deviations and reduced operational stability. Traditional physical testing methods are costly, time-consuming, and unable to replicate diverse complex

operating conditions, resulting in significant limitations in comprehensive performance evaluation. This study employs multi-physical-field coupled virtual simulation technology to establish a standardized modeling and validation framework that accurately replicates actual equipment operation conditions, providing reliable technical support for optimizing performance and adapting to complex operational environments of traffic measurement devices.

2. Fundamental theories of transportation measurement equipment and multi-physical field coupling simulation

2.1. Core structure and working principle of traffic measurement equipment

Traffic measurement equipment comprises various components including flow statistics sensors, vehicle load measurement devices, speed detection systems, and road condition sensing modules. Each component serves specific functional purposes, with the overall architecture consisting of sensing units, data processing units, fixed support structures, and signal transmission components. The equipment utilizes built-in sensing elements to capture diverse physical signals during traffic operations; these signals are subsequently converted, filtered for noise, and processed through computational algorithms within the system to generate standardized measurement data. Continuous collection of external dynamic physical parameters, such as structural deformation, temperature variations, and electromagnetic interference, can alter the sensitivity of sensing components, leading to measurement deviations. A thorough understanding of the equipment's fundamental structure and signal acquisition mechanisms enables effective virtual simulation modeling, precise identification of core performance fluctuation causes, and provides robust structural and theoretical foundations for developing multi-physical field coupling models.

2.2. Core concepts and simulation characteristics of multi-physical field coupling

Multi-physics field coupling simulation transcends the analytical limitations of individual physical fields by simultaneously accounting for interactions and intrinsic relationships among mechanics, thermodynamics, electromagnetism, fluid dynamics, and other physical domains, thereby accurately reproducing the actual operational behavior of systems in complex environments. Magnetic retention relays feature intricate structures with poor component consistency and suboptimal parameter coordination, leading to increased springing cycles between the armature and core as well as between moving and stationary contacts. This exacerbates arc erosion and contact wear, and in severe cases may result in contact fusion welding, causing significant changes in static and dynamic parameters such as engagement time, engagement voltage, and contact pressure^[1]. The multi-physics coupling approach enables bidirectional parameter correlation and dynamic iteration across all physical fields, realistically simulating the combined effects of temperature deformation, vibration interference, electromagnetic influences, and other factors during equipment operation, ensuring high alignment with real-world application scenarios.

2.3. Advantages of applying virtual simulation modeling technology

To meet the high reliability requirements of magnetic holding relays in railway, aerospace, and other applications, this study establishes methods for theoretical analysis of static and dynamic characteristics, develops corresponding theoretical and simulation models, and examines the patterns governing these characteristics. The digital modeling approach enables precise control of various parameters while capturing

subtle deformations, signal fluctuations, and precision deviations during equipment operation, thereby providing more comprehensive performance analysis. Virtual simulation testing eliminates the need for physical equipment and specialized test materials, significantly reducing R&D costs and accelerating product iteration cycles. All data generated during simulations can be fully archived and retrospectively analyzed, offering a robust data foundation for structural improvements, algorithm optimization, and performance enhancement.

2.4. Development of a virtual simulation modeling system for multi-physical field coupling in transportation measurement equipment

2.4.1. Overall design of the simulation modeling framework

By leveraging the operational characteristics of transportation measurement equipment and the principles of multiphysics coupling simulation, a hierarchical, modular virtual simulation framework can be established. This framework encompasses multiple core components: device geometric modeling, physical field parameter configuration, coupling relationship definition, simulation computation, and data output analysis. The simulation modeling of magnetic holding relays' static and dynamic behaviors involves comprehensive analysis of electromagnetic-mechanical motion interactions. Three distinct models are developed: a 3D electromagnetic field simulation model, a rigid-flexible coupled dynamics model, and a multi-physics coupled simulation model integrating coil voltage equilibrium equations^[2]. The geometric modeling layer performs digital reconstruction based on actual structural parameters to ensure high accuracy between digital models and physical devices. The physical field parameter configuration layer incorporates real-world operational conditions from transportation environments, including mechanical, thermal, and electromagnetic parameters. The coupling relationship definition layer clarifies interaction mechanisms among physical fields, while simulation computations replicate complete operational states and generate comprehensive data outputs, ultimately forming a fully integrated closed-loop digital modeling system.

2.4.2. Refined construction of the equipment's geometric model

Geometric model construction serves as the foundational step in virtual simulation modeling, where the quality of model formulation directly impacts the reliability of multiphysics simulation results. The modeling process adheres to core principles of refinement, equivalence, and operational condition adaptation, involving digital reconstruction based on the actual structural dimensions of transportation measurement equipment. Minor redundant components that have negligible effect on overall performance are eliminated, while critical functional elements such as sensors, load-bearing structures, and signal transmission pathways are retained to complete the three-dimensional digital model. Given that core equipment components exhibit heightened sensitivity to various physical fields, localized mesh refinement is employed to enhance simulation accuracy in critical areas, whereas auxiliary components with stable structures and uniform force distribution utilize conventional meshing methods, achieving a balance between simulation precision and computational efficiency. Model dimensional parameters and assembly relationships undergo repeated calibration against physical prototypes to prevent simulation artifacts caused by geometric deviations, thereby establishing a robust structural foundation for subsequent multiphysics coupled simulations.

2.4.3. Parameter setting for multi-physical field coupling and implementation of coupling mechanisms

By leveraging the outdoor operational environment of transportation measurement equipment, we identify structural mechanical fields, temperature fields, electromagnetic fields, and fluid fields as the core physical parameters for simulation, with precise boundary conditions tailored to actual operating conditions. The mechanical field replicates vibrations and impacts generated during vehicle passage as well as structural deformations caused by equipment weight; the temperature field captures temperature variations resulting from day-night cycles and seasonal changes, along with associated thermal expansion/contraction effects; complex outdoor electromagnetic environments are simulated through electromagnetic field modeling; dynamic forces from wind, rain, and airflow are modeled using fluid dynamics to represent environmental disturbances on equipment casings and sensing units accurately. A bidirectional coupling mechanism among multiple physical fields is established to define interrelationships between parameters: temperature-induced structural deformation alters mechanical states, electromagnetic fluctuations interfere with signal acquisition, and airflow-generated vibrations exacerbate structural deformation. This multidimensional integrated simulation approach faithfully reproduces real-world composite operational conditions of equipment.

3. Simulation model validation and application optimization analysis

3.1. Design of the validation scheme for the validity of the simulation model

To ensure the multi-physics coupled simulation model aligns with practical engineering applications and demonstrates reliable implementation value, this study employs a hierarchical validation approach. By comparing it with experimental data from physical equipment, the simulation model is validated across three core dimensions: structural deformation, operational accuracy, and signal stability. The simulation conditions closely replicate those of actual experiments, with all environmental parameters and operating parameters fully synchronized throughout the process. Operational data from both digital models and physical devices are collected simultaneously for cross-comparison; by analyzing variation trends, fluctuation patterns, and error ranges between the two datasets, the model's correspondence with real-world equipment performance is assessed ^[3]. The validation covers various extreme scenarios including high/low temperatures, intense vibrations, and high-current interference. Through repeated testing, systematic errors are mitigated, subtle deviations in model parameters and coupling mechanisms are corrected, and discrepancies between simulation and experimental data are continuously reduced, thereby significantly enhancing the model's reliability in practical engineering applications.

3.2. Analysis of simulation results and tracing of performance deviations

Multi-physics coupled simulation enables comprehensive acquisition of structural deformation, measurement accuracy, and signal transmission stability data for transportation measurement devices under complex operating conditions, allowing precise analysis of how various physical fields and their interactions influence device performance. Simulation results demonstrate that when mechanical vibrations and temperature fluctuations act simultaneously on the device, stress distributions in structural components do not follow the simple additive pattern expected under independent effects, exhibiting significant nonlinear amplification effects, details that are difficult to identify through traditional single-physics simulations. In single-physics environments, performance deviations exhibit fixed patterns and limited variability; however, multi-

physics coupling further exacerbates structural deformation and measurement inaccuracies, with overall error magnitudes typically exceeding the sum of errors caused by individual physical effects. Simulation data enables precise identification of root causes of performance issues, quantifies the contributory ratios of environmental factors such as temperature spikes, high-frequency vibrations, and electromagnetic interference, and pinpoints critical weaknesses including structural vulnerabilities, inadequate sensor module compatibility, and insufficient signal immunity. Furthermore, simulation analyses allow differentiation of dominant physical field impacts across operational scenarios, providing targeted recommendations for structural and performance optimization while addressing the limitations of traditional experimental methods in accurately pinpointing root causes of performance deviations under complex conditions.

3.3. Iterative optimization of the simulation model and engineering adaptation upgrades

Based on model validation results and performance deviation traceability analysis, the multiphysics coupled simulation model enables iterative optimization. By adjusting physical field coupling parameters, mesh refinement accuracy, and boundary condition settings, the model's multi-field interaction logic is refined to enhance simulation accuracy further. The study employs adaptive mesh refinement technology to improve mesh resolution in critical sensing regions, utilizes bidirectional coupling convergence criteria to strengthen data exchange stability between temperature and mechanical fields, and introduces specialized filtering boundary conditions for high-frequency electromagnetic interference components ^[4]. Accounting for variations across different traffic scenarios, the model undergoes modular adaptation upgrades by adjusting corresponding physical field parameter combinations for urban roads, highways, tunnels, and other operating environments, achieving cross-scenario simulation compatibility. The optimized model accurately replicates actual operational conditions of traffic measurement devices under diverse environments, serving not only for factory performance verification but also providing technical support for structural improvements, sensor algorithm optimization, and anti-interference design enhancements.

4. Epilogue

This study addresses the limitations of traditional physical testing and single simulation approaches for traffic measurement equipment by establishing a comprehensive virtual simulation modeling and validation framework based on multi-physical field coupling theory. It systematically conducts research encompassing device geometric modeling, multi-field parameter configuration, coupling relationship establishment, and simulation verification optimization. The findings overcome issues inherent in single simulations, such as insufficient operational condition coverage and low practical accuracy, enabling accurate representation of equipment performance characteristics and trend changes under complex conditions. This provides a robust digital methodology for performance analysis and structural improvement of traffic measurement devices. Future work should focus on refining coupling algorithms and model precision, enhancing simulation scenario adaptability, and promoting widespread application of virtual simulation technology in R&D and performance evaluation processes for traffic measurement systems.

Funding

Development and Research of Specialization Metrology Equipment Based on Intelligent Virtual Simulation

(Project No. 2025X004-KXD)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Zhao J, Li N, Zhang X, et al., 2024, Establishment and Experimental Analysis of a Multi-Physical Field Rigid-Flexible Coupling Simulation Model for Magnetic Retention Relays with Multi-Leaf Structures. *Journal of Electrical Engineering Technology*, 39(10): 3192–3205.
- [2] Zhang J, Liu Q, Wang S, et al., 2024, Research on the Pulsed Electromagnetic EMAT–Eddy Current Detection Composite System. *Sensors and Microsystems*, 43(3): 43–46 + 51.
- [3] Wang J, Liu X, 2023, Analysis and Discussion on Improving the Accuracy of Highway Automatic Scales. *China Metrology*, 2023(10): 62–64.
- [4] Wang K, Ye P, Guo J, et al., 2024, Research on Simulation Testing Methods for Multi-Objective Radar Speedometers. *China Inspection and Testing*, 32(2): 29–31.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Structural Design and Optimization of Tesla Valve Based on Bezier Curves

Mingtao Gong*

High School Affiliated to Shanghai Jiao Tong University, Shanghai, China

*Corresponding author: Mingtao Gong, 770515287@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: To reduce forward vortex losses caused by discontinuous curvature in conventional polyline Tesla valves, this study introduces Bezier curves into Tesla valve design to improve unidirectional flow control through curvature continuity. Eight differentiated Bezier-curve configurations (ta-th) are designed, covering concave, convex, and baseline combinations of Curve 1 (inlet to bifurcation midsection) and Curve 2 (bifurcation midsection to outlet). The channel section is set to 3 mm × 2 mm with a bifurcation angle of 20°. A CFD model is established using the standard k- ω turbulence model, with a velocity inlet of 5 m/s and a pressure outlet of 0 Pa gauge to analyze forward and reverse pressure drops, diodicity, and the distributions (pressure, velocity, and turbulent kinetic energy). The results show that the tc structure (concave Curve 1 and convex Curve 2) achieves the best unidirectional flow performance, with a diodicity of 1.700183, a reverse pressure drop of 16035.53 Pa, and a forward pressure drop within 9400–9700 Pa. A concave Curve 1 improves forward flow uniformity and reduces energy loss, whereas a convex Curve 2 strengthens the contraction-expansion effect in reverse flow and enhances blocking capacity. This study confirms that Bezier curves address the structural deficiencies of conventional Tesla valves, establishing links among quantitative geometry, flow, and performance and parametric design rules. This fills a research gap in continuous-curvature structures and provides high-performance passive fluid control for EV thermal management, flame safety, wastewater treatment, as well as theoretical guidance for other passive fluidic devices.

Keywords: Bezier curve; Tesla valve; CFD; Diodicity; Flow-field optimization

Online publication: Jul 23, 2026

1 Introduction

The Tesla valve is a passive unidirectional fluid control device with no moving parts. It was first patented by Tesla, and its diodicity was later experimentally confirmed by Forster et al. ^[1,2]. The device has a flow diode, with low resistance in the forward direction and high resistance in the reverse direction ^[3,4]. Due to its simple structure, high reliability, and low maintenance cost, it shows application potential in thermal management, safety protection, ecological engineering, and environmental treatment, and is gradually expanding to new

scenarios such as high-pressure differential regulation and system efficiency improvement ^[5–10]. With the increasing demands on fluid control precision and energy efficiency in modern industry, the traditional polyline Tesla valve is prone to generating local vortices and additional pressure drops due to discontinuous curvature in forward flow, and has problems such as insufficient stability of separation bubbles in reverse flow, limiting its further application ^[3,11].

Bezier curves have the advantages of continuous curvature, controllable shape, and simple parametric expression, and can achieve smooth flow channel transition through control point adjustment, which helps to reduce forward flow loss and enhance the contraction-expansion effect in reverse flow. Although existing studies have been conducted on bifurcation angles, series, channel widths, and adaptability to multiple scenarios, systematic analysis of continuous curvature structures is still insufficient ^[3,12–16]. In particular, a unified discussion on the relationship between curve combination morphologies and the pressure, velocity, and turbulent kinetic energy fields is lacking. The unidirectional conduction characteristics of the Tesla valve result from the differential regulation of forward and reverse flow by the flow channel structure ^[3]. In forward flow, the fluid moves along the main channel attached to the wall, and if the turn is smooth, there will be fewer vortices and lower resistance; in reverse flow, the bifurcation section induces vortices and separation bubbles. Vortices dissipate kinetic energy and separation bubbles occupy the effective channel, thereby significantly increasing flow resistance ^[2,3,11]. Compared with the traditional polyline structure, the continuous curvature of the Bezier curve mitigates the impact of the forward sharp-angle abrupt change and improves the generation position and stability of the reverse separation bubble.

This study thus designs eight Bezier-curve Tesla valve structures, uses CFD to analyze their pressure drops, diodicity, and flow fields, summarizes optimization rules for curve combinations, and offers a parametric design reference for passive fluidic devices.

2. Characteristics of Tesla valves and theoretical basis of CFD simulation

2.1. Operating principle and key metrics of the Tesla valve

The key performance metrics of a Tesla valve are forward pressure drop ΔP_f , reverse pressure drop ΔP_r , and diodicity R_i . Here, ΔP_f reflects the energy cost of forward transport, ΔP_r indicates the flow blocking capability, and a higher diodicity means stronger unidirectional control.

2.2. Technical approach

This study uses Creo for 3D modeling, Ansys Meshing for mesh generation, Ansys Fluent for flow field simulation, and Tecplot for post-processing and visualization.

3. Experimental method and CFD model construction

3.1. Structural design and parameter setting

This paper designs eight different types of Bezier-curve structures (ta-th), including concave, convex, and baseline combinations for Curve 1 (inlet- middle bifurcation) and Curve 2 (middle bifurcation - outlet): ta is Curve 1 concave + Curve 2 baseline, tb is Curve 1 concave + Curve 2 concave, tc is Curve 1 concave + Curve 2 convex, td is Curve 1 convex + Curve 2 baseline, te is Curve 1 convex + Curve 2 concave, tf is Curve 1 convex + Curve 2 convex, tg is Curve 1 baseline + Curve 2 concave, th is Curve 1 baseline + Curve 2

convex. The model uses a $3\text{ mm} \times 2\text{ mm}$ channel cross-section with a bifurcation Angle of 20° and a single-stage length of approximately 35 mm; continuous control of the channel curvature is achieved by adjusting the coordinates of Bezier control points. Table 1 and 2 present the basic parameters and control point coordinates, respectively.

3.2. Simulation scenarios, boundary conditions, and solver settings

The simulation domain is a single-stage flow channel with equal inlet and outlet lengths. The boundary conditions are set at a velocity inlet of 5 m/s and a pressure outlet of 0 Pa gauge, with a non-slip boundary on the wall. The turbulence intensity is 5%. The solver uses the steady-state pressure-based algorithm, with pressure-velocity coupling handled by the SIMPLEC method. The residual convergence criterion is 1×10^{-5} , and convergence is judged in combination with outlet flow rate and pressure drop variation.

3.3. CFD model construction procedure

Geometric modeling begins with creating a two-dimensional sketch in Creo and drawing the inlet channel. Based on the quadratic Bezier curve formula, the bifurcation start point, end point, and control points are defined. Afterward, the outlet channel and 3D solid model are completed. Then, local refinement is applied to the curved segments and boundary layer regions in Meshing, followed by forward and reverse flow calculations in Fluent (see **Figure 1**).

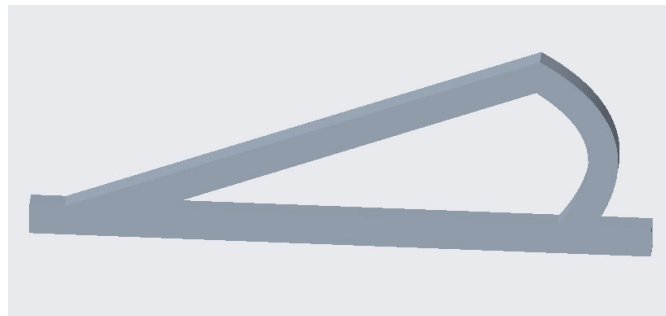


Figure 1. 3D modeling of the Tesla valve.

4. Results and discussions

4.1. Bezier curve analysis of Tesla valve pressure drop and diodicity

To compare the effects of different Bezier-curve structures on the forward and reverse flow performance of Tesla valves, Ansys Fluent is used to simulate the forward pressure drop ΔP_f , reverse pressure drop ΔP_r , and diodicity R_i of eight structures (ta-th) at flow rates of 0.5–2.0 m/s. The results are shown in **Figure 2**.

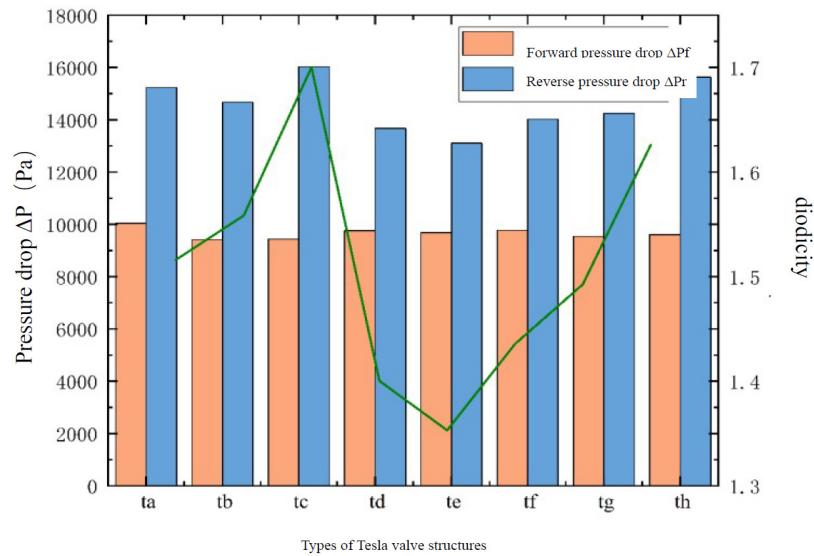


Figure 2. Comparison of Tesla valve performance.

The forward pressure drops of the structures mainly fall between 9400 and 10050 Pa, with little overall difference. Among them, ta has the highest forward pressure drop (10050.32 Pa), while tc, td, and tf are relatively low (about 9400–9700 Pa), indicating that the energy loss in forward flow is similar for most schemes. In scenarios such as thermal management and fluid delivery where forward resistance is emphasized, tc, td, and tf can be prioritized to reduce system energy consumption. The difference in reverse pressure drop is more pronounced, ranging from 13105 to 16035 Pa. tc is the highest (16035.53 Pa), followed by th (15626.48 Pa) and ta (15236.79 Pa), and te is the lowest (13105 Pa). The greater the reverse pressure drop, the stronger the blocking effect of the structure on the reverse fluid. Therefore, tc, th, and ta have an advantage in scenarios that require strong reverse flow blocking, such as fire suppression and wastewater treatment; te, with its lower reverse pressure drop, is more suitable for microfluidic scenarios where the requirements for flow blocking are not high.

The diodicity directly reflects the unidirectional conduction ability of the Tesla valve. The greater the R_i , the better the unidirectionality. tc has the highest diodicity (1.700183), followed by th (1.625938) and tb (1.558484), and te has the lowest (1.353051). This indicates that tc does not significantly increase forward loss while significantly enhancing reverse blocking, and thus achieves the best overall unidirectional conduction performance.

4.2. Bezier curve characteristics of pressure distribution in tesla valve flow field

Based on the Ansys Fluent simulation results, the static pressure nephograms (**Figure 3**) of the ta and td structures under the forward (tal/tdl) and reverse (tar/tdr) conditions are extracted, and the influence of the concave and convex of Curve 1 on the pressure distribution is analyzed.

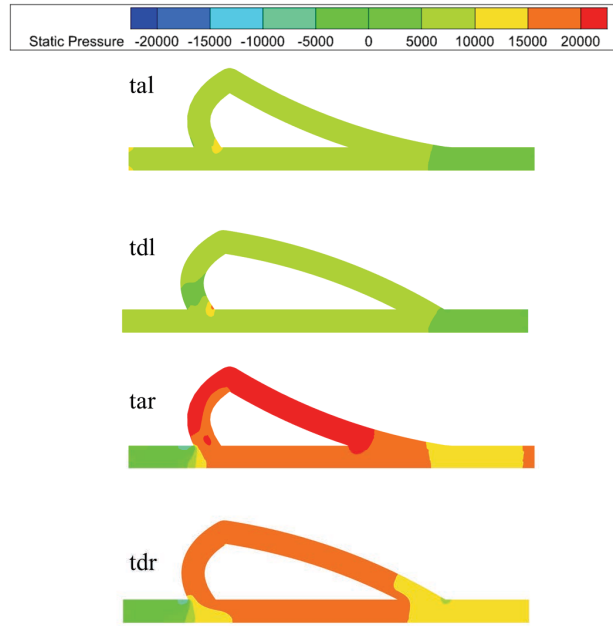


Figure 3. Diagram of pressure variation affected by Curve 1.

Both structures present the basic feature of high pressure at the inlet and low pressure at the outlet, but the curve shape affects the uniformity of the pressure gradient and local losses. In forward flow, the pressure distribution of *ta* is relatively uniform, mainly concentrated in the 0–10000 Pa range, with very small local low-pressure areas, indicating that the forward pressure is reduced and the flow is relatively smooth; *td*, on the other hand, has a local low-pressure area of –5000 to 0 Pa, with slightly poorer pressure uniformity and slightly higher forward pressure drop. The concave curve is more conducive to smooth flow diversion, while the convex curve is more likely to cause local pressure loss. In reverse flow, *ta* exhibits large high-pressure (15000–20000 Pa) and strong low-pressure (–20000 to –15000 Pa) regions, indicating strong reverse blocking capability and thus a higher diodicity. In contrast, *td* has a smaller high-pressure region and weaker low-pressure intensity, so its reverse blocking effect is less pronounced than that of *ta*. Thus, when Curve 1 adopts a concave morphology, it tends to generate a stronger pressure gradient, which favors low forward resistance and high reverse resistance.

As can be seen from **Figure 4**, *tg* and *th* have similar pressure distributions in forward flow, both mainly 0–10000 Pa, with little difference in forward flow performance; however, in reverse flow, the range and intensity of the high-pressure zone of *th* are stronger than those of *tg*, and the flow-blocking effect is more prominent. Therefore, both *tg* and *th* are suitable for the forward low-resistance scenario, while *th* is more advantageous when a stronger reverse blocking is required.

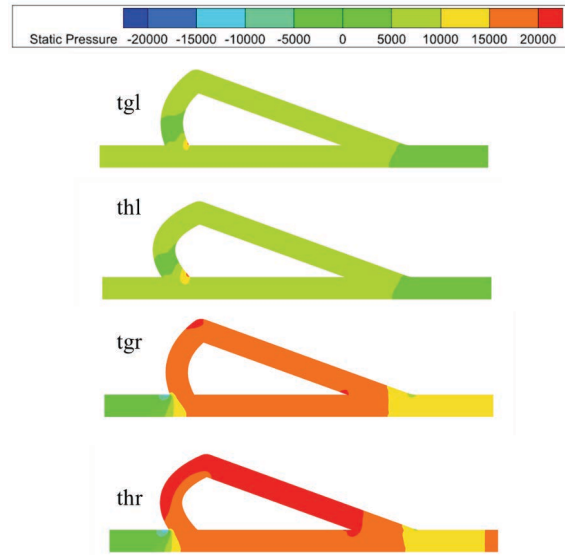


Figure 4. Diagram of pressure variation affected by Curve 2.

As shown in **Figure 5**, a comparison of tbl, tcl, tel, tfl and their reverse counterparts reveals that in forward flow, tb has the most uniform pressure distribution, and tc, though showing some local low-pressure areas, still performs well overall. Although the pressure gradients of te and tf are relatively gentle, their reverse flow blocking capacity is insufficient. In reverse flow, tc has the largest high-pressure region and the strongest low-pressure region, which corresponds to the highest diodicity, indicating its strongest reverse blocking performance; tb ranks second, but its overall performance is still inferior to that of tc and th; te and tf have weaker high-pressure regions and have poorer blocking performance. Overall, tc performs best in terms of pressure distribution and diodicity, and is more suitable for scenarios that require strong reverse flow blocking, such as fire blocking and wastewater treatment.

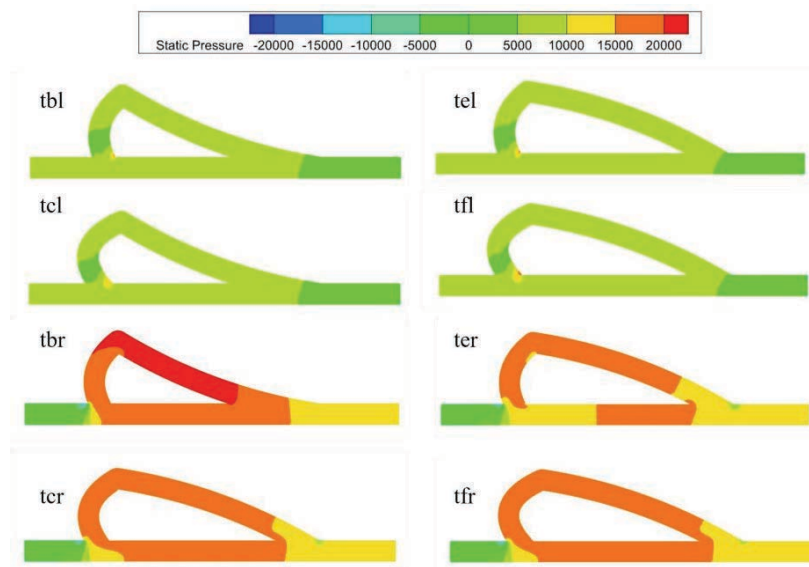


Figure 5. Diagram of the combined effect of Curves 1 and 2 on pressure variation.

4.3. Bezier curve characteristics of velocity distribution in the flow field of Tesla valve

Using the velocity nephograms (**Figure 6**) from Ansys Fluent, the velocity distribution differences between ta and td in forward and reverse flow are analyzed:

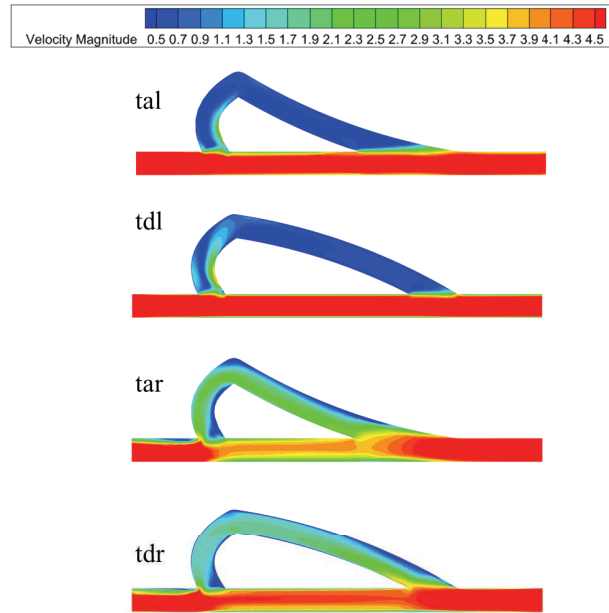


Figure 6. Velocity variation affected by Curve 1.

In forward flow, the mainstream velocity of ta (2.5–3.5 m/s) is uniformly distributed along the flow channel, with a small velocity gradient (≤ 1.0 m/s), and the continuous curvature of the concave curve makes the fluid movement trajectory smooth without obvious velocity change zones, meeting the requirement of uniform flow velocity in thermal management systems; td has a lower mainstream velocity (2.0–3.0 m/s), a velocity change zone (≥ 1.5 m/s) appears at the junction of curve 1 and the bifurcation section, and the convex curve expansion–contraction pattern causes a sudden change in the local channel cross-section, resulting in velocity fluctuations and increased energy dissipation. In reverse flow, the ta bifurcation forms a large high-speed area (3.5–4.5 m/s) with a large velocity gradient (≥ 2.0 m/s). The concave curve guides the reverse fluid to form a stable large-scale vortex where the velocity is low at the center and high at the periphery, thereby enhancing the flow blocking effect. For td, the high-velocity region is smaller (3.0–4.0 m/s), and the velocity gradient is lower than that of ta (≤ 1.5 m/s). The convex curve disperses the reverse fluid, leading to small-scale vortices with poor stability. Thus, its reverse blocking effect is weaker than that of the concave curve.

With Curve 2 unchanged in its baseline morphology, the concave Curve 1 optimizes forward velocity uniformity and enhances reverse velocity gradient, balancing low forward resistance with high reverse resistance; the convex form leads to forward velocity fluctuations and a weaker reverse gradient, with both forward efficiency and reverse resistance being weaker than the concave form. This rule corroborates the pressure distribution analysis. The concave-convex shape of Curve 1 is the core variable for regulating the forward and reverse performance of the Tesla valve. The preference for the concave shape of Curve 1 provides a clear direction for geometric optimization.

Based on the velocity nephogram (**Figure 7**), the velocity distributions of tbl, tcl, tel, tfl, and their reverse structures tbr, tcr, ter, tfr are analyzed.

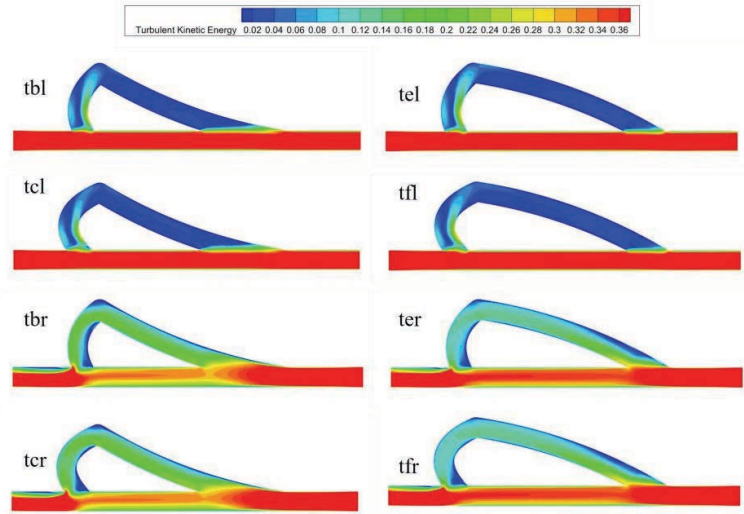


Figure 7. Combined effect of Curve 1 and Curve 2 on velocity variation.

In forward flow, the mainstream velocity of tb is relatively uniform (2.5–3.5 m/s), the velocity gradient is small, and the forward flow efficiency is relatively high, which is basically consistent with its higher diodicity. tc has a slightly higher velocity gradient, but its diodicity indicates that the reverse blocking performance is more prominent, so this part of the forward disadvantage is still acceptable. The velocity distribution of the te and tf structures is relatively stable, but the diodicity is relatively low, indicating insufficient reverse optimization. Overall, the more the channel geometry conforms to the fluid flow path, the more uniform the velocity distribution tends to be. In reverse flow, tc forms a large high-speed region (3.5–4.5 m/s) in the bifurcation section, with the maximum velocity gradient corresponding to its maximum diodicity, indicating that its vortex energy dissipation effect is the strongest. te has a slower velocity gradient and a weaker reverse flow blocking capacity, which is consistent with its lower diodicity. Overall, the reverse velocity characteristics are closely related to the blocking and guiding effects of the structure. The easier it is to form a stable large-scale vortex in the flow channel, the more obvious the velocity gradient and the stronger the reverse blocking ability. The velocity distribution results are largely consistent with the diodicity ranking. tc is the structure with the best unidirectional performance, followed by th and tb.

4.4. Bezier curve characteristics of turbulent kinetic energy distribution of Tesla valve

Based on the turbulent kinetic energy nephogram (**Figure 8**), the turbulent kinetic energy distribution of ta and td in the forward and reverse flow is analyzed under the condition that Curve 2 maintained the baseline morphology to investigate the influence of Curve 1 shape on the turbulent intensity.

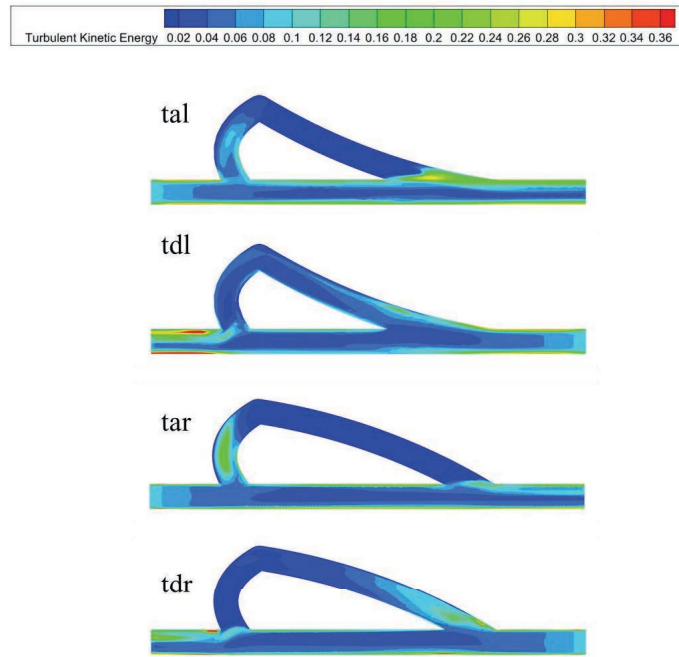


Figure 8. Turbulent kinetic energy affected by Curve 1.

In forward flow, the turbulent kinetic energy of ta is generally low (≤ 0.2). Only at the junction between the bifurcation section and the main channel does a local high-value region (0.2–0.25) appear. This indicates that the concave curve forms a smoother flow-guiding channel. The fluid motion is more stable, and turbulent disturbance is weaker. These observations are consistent with its relatively uniform velocity field and mild pressure gradient. Thus, the concave curve effectively suppresses turbulent dissipation in forward flow. In contrast, td has a larger range of high values of turbulent kinetic energy (0.25–0.3) and a local peak (≥ 0.3) at the root of curve 1, indicating that the abrupt change of cross-section and local vortices caused by the convex curve can enhance the forward turbulent disturbance and increase the energy loss. In reverse flow, ta shows a concentrated and stable high-TKE region (0.25–0.3) in the bifurcation section, indicating that the concave curve more easily induces large-scale stable vortices and sustained turbulent energy concentration. This aligns with the distinct high-velocity gradient and strong low-pressure zone, demonstrating a strong reverse turbulent blocking effect. The high-value area of td is more dispersed and has a lower intensity (0.2–0.25). This indicates that the small-scale vortices induced by the convex curve have poor stability. They can hardly sustain a continuous turbulent energy dissipation process, so the reverse blocking capability of td is weaker than that of ta. Overall, with Curve 2 unchanged, the concave form of Curve 1 can balance forward low dissipation and reverse high blocking, while the convex form can result in greater forward energy loss and weaker reverse blocking effect. This result corroborates the previous analyses of velocity and pressure fields. It indicates that the concave or convex form of Curve 1 is a key factor affecting the turbulent characteristics of the Tesla valve; it is more reasonable to give priority to the concave shape of Curve 1 in scenarios such as thermal management and flame blocking where both forward efficiency and reverse flow blocking are concerned.

Based on the turbulent kinetic energy nephogram (**Figure 9**), the turbulent kinetic energy distributions of tbl, tcl, tel, tfl, and their reverse structures tbr, tcr, ter, tfr are analyzed.

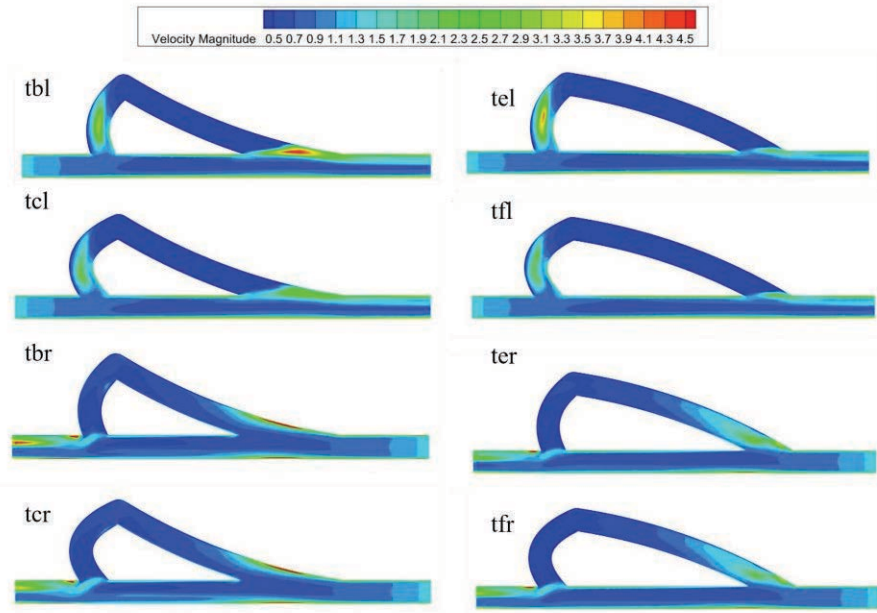


Figure 9. Combined effect of Curves 1 and 2 on turbulent kinetic energy variation.

In forward flow, the turbulent kinetic energy of the tb structure is lower (≤ 0.2), and the local high value area of the bifurcation section is smaller, indicating less forward energy dissipation, which is basically consistent with its higher diodicity. The turbulent kinetic energy of tc is slightly higher, but the conduction ratio indicates that its reverse advantage is sufficient to compensate for the forward loss. The turbulent gradients of the te and tf structures are relatively gentle, but the diodicity is relatively low, and the overall performance is still weak. This indicates that the more the channel geometry fits the fluid motion, the less turbulent dissipation is usually. In reverse flow, the high-value region of turbulent kinetic energy of tc is most concentrated (0.25–0.3), and the distribution is stable, corresponding to the highest diodicity, indicating the strongest turbulent blocking effect; tb has the second-best turbulent kinetic energy, but its overall performance is still inferior to tc; th has a similar turbulent intensity to tb but a higher diodicity; te has weaker turbulent kinetic energy and poorer reverse flow blocking capacity. Overall, the distribution of turbulent kinetic energy is largely consistent with the change in diodicity. tc shows the best combined performance through stronger reverse turbulent energy dissipation, and th and tb can be considered as alternatives.

5. Conclusion

This paper presents a parametric design of the Tesla valve flow channel using Bezier curves and systematically analyzes the flow performance of different structural configurations via CFD. The results show that among the eight Bezier-curve structures, tc (concave Curve 1 and convex Curve 2) has the best overall performance, with a diodicity of 1.700183, a reverse pressure drop of 16035.53 Pa, and a stable forward pressure drop in the range of 9400–9700 Pa. This structure effectively balances low forward resistance with high reverse resistance. Further analysis reveals that the concave form of Curve 1 is conducive to improving the uniformity of the forward flow field and reducing local vortices and energy dissipation; Curve 2, with its convex shape, can enhance the contraction-expansion effect in reverse flow and strengthen the separation and blocking capacity. The research demonstrates that the continuous curvature flow channel design can

effectively improve the flow characteristics of the traditional Tesla valve. It provides a theoretical basis for subsequent structural parameter optimization and engineering applications, and offers a reference for improving the design level of passive fluid control devices.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Tesla N, 1920, Valvular Conduit. US1329559A.
- [2] Forster F, Bardell L, Afromowitz M, et al., 1995, Design, Fabrication and Testing of Fixed-Valve Micro-Pumps. 39–44.
- [3] Hu P, 2023, Research on the Diodictiy Characteristics and Its Optimal Design Method of Tesla Valve, thesis, Zhejiang University.
- [4] Zhang Z, Chen Z, Chen T, 2020, The Structural Design of Tesla Valves Based on Finite Element Analytical Simulation and 3D Printing Technology. *College Physics*, 40(4): 73–77.
- [5] Zhang W, Guo Y, 2024, Numerical Study on Heat Transfer Characteristics of Tesla Valve-Type Pulsating Heat Pipe. *Chemical Engineering*, 52(8): 37–42.
- [6] Liu X, et al., 2025, Study on Flame Propagation and Quenching Characteristics in Tesla Valve Structure Flame-Retardant Units. *Journal of Safety and Environment*, 25(5): 1700–1709.
- [7] Du P, 2023, Analysis of the Influence of Channel Form and Layout on the Hydraulic Performance of the Emitter, thesis, Taiyuan University of Technology.
- [8] Li P, 2024, Numerical Simulation of Hydraulic Cavitation Based on Tesla Valve, thesis, North China Electric Power University (Beijing).
- [9] Zhong Z, 2023, Application of the Tesla Valve in Pressure Regulation Under Large-Gradient and High-Pressure-Difference Conditions. *Proceedings of the 5th Gas Safety Exchange Seminar*, 2: 114–116.
- [10] Ren P, Long W, 2021, Numerical Analysis of Flow Field Characteristics and Hydrodynamic Cavitation Mechanism of Tesla Valves. *Journal of Southwest University (Natural Science Edition)*, 43(8): 161–166.
- [11] Gamboa A, Morris C, Forster F, 2005, Improvements in Fixed-Valve Micropump Performance Through Shape Optimization of Valves. *Journal of Fluids Engineering*, 127(2): 339–346.
- [12] Su C, et al., 2025, Research on the Influence of Tesla Valve Structural Parameters on One-Direction Flow Characteristic and Modification. *Journal of Chemical Engineering of Chinese Universities*, 39(6): 993–1001.
- [13] Zhou R, et al., 2020, Simulating the Performance of Tesla Valve. *Physical Experiment*, 40(9): 44–50.
- [14] Thompson S, Walters D, Paudel B, et al., 2013, A Numerical Investigation of Multi-Staged Tesla Valves, ASME 2013 Fluids Engineering Division Summer Meeting.
- [15] Zhang Y, Tong J, Zhu Z, 2023, Numerical Calculation of Forward and Reverse Flow in Tesla Valves with Different Longitudinal Width-to-Narrow Ratios. *Scientific Reports*, 13(1): 12496.
- [16] Xi C, 2023, Design and Characterization of Tesla Runner With Flow Diodictiy Property and Its Metamaterial Structure, thesis, Yanshan University.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CVaR-Based Optimal Scheduling of an Integrated Energy System with Battery Storage, Hydrogen Storage, and Demand Response

Delong Yan

Hubei Minzu University College of Intelligent Systems Science and Engineering, Enshi 445000, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: High renewable penetration improves the low-carbon performance of integrated energy systems, but it also increases scheduling uncertainty and renewable curtailment risk. This paper proposes a CVaR-based optimal scheduling model for an electric-heat-hydrogen integrated energy system with battery energy storage, hydrogen storage, and demand response. The proposed model minimizes a weighted objective that combines expected operating cost and tail-risk cost, while considering electricity purchase and sale, gas consumption, carbon emissions, battery degradation, hydrogen conversion, demand response compensation, and renewable curtailment penalty. Wind power, photovoltaic generation, and electric load uncertainty are represented by multiple scenarios, and the same scenario set is used for all comparative cases to ensure fairness. Five operation schemes are studied, including no storage, battery energy storage only, hydrogen storage only, battery-hydrogen storage, and the proposed battery-hydrogen-demand response scheme. The numerical results show that the proposed scheme achieves the lowest weighted objective, expected cost, and CVaR cost. Compared with the no-storage case, the proposed scheme reduces the weighted objective from 12337.22 to 9677.39, increases renewable utilization from 92.3% to 99.8%, and reduces expected carbon emissions from 4771.88 to 2994.13. These results indicate that coordinated scheduling of battery storage, hydrogen storage, and demand response can improve economic performance, reduce operational risk, and enhance renewable energy accommodation in high-renewable integrated energy systems.

Keywords: Integrated energy system; Optimal scheduling; CVaR; Battery energy storage; Hydrogen storage; Demand response; Renewable curtailment

Online publication: Jul 23, 2026

1. Introduction

The rapid growth of renewable energy has changed the operation pattern of modern energy systems. Wind power and photovoltaic generation can reduce fossil fuel consumption and carbon emissions, but their output is strongly affected by weather conditions. When renewable generation fluctuates, the system operator must

balance electricity, heat, and storage resources under greater uncertainty. Integrated energy systems provide an effective framework for this problem because they coordinate multiple energy carriers, such as electricity, heat, gas, and hydrogen. Through energy conversion devices and flexible resources, an integrated energy system can improve renewable energy utilization and reduce dependence on conventional fossil-based supply.

Battery energy storage, hydrogen storage, and demand response have different operational characteristics so that they can provide complementary flexibility. Battery energy storage has high conversion efficiency and fast response speed, which makes it suitable for short-term power balancing and peak shaving. Hydrogen storage has lower round-trip efficiency, but it can absorb surplus renewable electricity through electrolysis and release energy later through fuel cells. It is therefore more suitable for longer-duration renewable energy accommodation. Demand response adjusts flexible load on the demand side and can reduce peak-hour pressure without installing additional generation capacity. If these resources are scheduled independently, their flexibility may not be fully used. A coordinated scheduling strategy is needed to determine when each resource should operate and how it should contribute to cost reduction, renewable utilization, and emission mitigation.

The uncertainty of renewable generation and load demand makes deterministic scheduling insufficient for high-renewable integrated energy systems. A dispatch scheme that performs well under forecasted conditions may lead to high operating costs when wind power, photovoltaic generation, or load demand deviates from the expected profile. Scenario-based stochastic optimization can describe different operating conditions, but the expected-cost criterion may underestimate low-probability adverse scenarios. Conditional value-at-risk provides a practical way to measure the average cost in the high-cost tail of the scenario distribution. By introducing CVaR into the scheduling model, the operator can reduce the exposure to extreme operating costs while still considering the average economic performance of the system.

Existing studies have investigated long-term planning and key technologies for high-renewable power systems, green-electricity direct connection and source-grid-load-storage coordinated scheduling, electricity-hydrogen-carbon ecosystems and power-hydrogen-ammonia coupling systems, and hydrogen-electric-thermal or electric-hydrogen coordinated dispatch strategies^[1–12]. In addition, energy storage configuration and flexible load integration have been studied for regional integrated energy systems and distribution networks, while low-carbon scheduling considering flexible-load uncertainty has also been discussed^[13–15]. However, relatively limited attention has been paid to the joint coordination of battery storage, hydrogen storage, and demand response under a CVaR-based tail-risk framework for high-renewable electric-heat-hydrogen integrated energy systems.

This paper develops a CVaR-based optimal scheduling model for an electric-heat-hydrogen integrated energy system. The main contributions are threefold. First, battery energy storage, hydrogen storage, and demand response are jointly coordinated in a unified scheduling framework, which allows short-term and long-duration flexibility to be used in a complementary manner. Second, renewable curtailment is explicitly modeled and penalized in the objective function, so the model can evaluate renewable accommodation more directly than approaches that only infer curtailment from power export. Third, a weighted objective combining expected operating cost and CVaR cost is adopted to balance average economic performance and tail-risk control under renewable generation and load uncertainty.

2. System description

The studied system is an electric-heat-hydrogen integrated energy system with renewable generation, conventional energy supply, energy conversion devices, storage units, grid interaction, and flexible demand. The electric side includes wind power, photovoltaic generation, electric load, grid electricity purchase and sale, gas turbine generation, battery energy storage, electrolyzer consumption, fuel cell generation, heat pump consumption, and demand response. The heat side includes heat load, gas turbine heat recovery, gas boiler heat production, heat pump heat output, and fuel cell waste heat recovery. The hydrogen side consists of an electrolyzer, a hydrogen storage tank, and a fuel cell. The electrolyzer converts surplus electricity into hydrogen, while the fuel cell converts stored hydrogen back into electricity and also provides usable heat.

Battery energy storage and hydrogen storage are assigned different operational roles in the system. Battery energy storage has high charging and discharging efficiency, so it is mainly used to smooth short-term renewable fluctuations and reduce peak electricity purchase. Hydrogen storage has a lower electric round-trip efficiency, but it can store a larger amount of surplus renewable energy over a longer period. This makes it useful when photovoltaic generation is high, and the battery capacity is insufficient to absorb all renewable output. Demand response is modeled as an interruptible electric load. It reduces the net electric demand during selected periods and provides additional flexibility when storage resources alone cannot fully relieve peak-hour pressure.

3. Mathematical model

A two-stage scenario-based scheduling model is formulated for the integrated energy system. All scenarios share the first-stage decisions and represent the scheduled outputs of the main conversion devices before the realization of renewable generation and load uncertainty. The second-stage decisions are scenario-dependent and describe corrective operation after the uncertain variables are observed. These decisions include grid exchange, battery charging and discharging, hydrogen production and utilization, demand response, and renewable curtailment. The scheduling horizon is divided into 24 hourly periods. The objective is to minimize a weighted sum of the expected operating cost and the CVaR cost, while satisfying electricity balance, heat balance, storage dynamics, device limits, and risk-related constraints.

3.1. Objective function

The objective function consists of the expected operating cost and the CVaR cost. The expected operating cost represents the average economic performance of the system over all scenarios. The CVaR cost measures the expected loss in the high-cost tail of the scenario distribution. By combining these two terms, the proposed model can reduce the average operating cost while controlling the risk caused by renewable generation and load uncertainty. Let C_s denote the total operating cost in scenario s , and let S denote the number of scenarios. The weighted objective function is formulated as follows:

$$\min F = (1 - \beta) \frac{1}{S} \sum_{s=1}^S C_s + \beta \left[\eta + \frac{1}{(1 - \alpha)S} \sum_{s=1}^S z_s \right]$$

where β is the risk weight, α is the confidence level of CVaR, η is the value-at-risk auxiliary variable, and z_s is the non-negative excess cost variable in scenario s . The CVaR auxiliary constraints are expressed as:

$$z_s \geq C_s - \eta, \quad \forall s$$

$$z_s \geq 0, \quad \forall s$$

The scenario operating cost C_s includes electricity purchasing cost, electricity selling revenue, gas consumption cost, operation and maintenance cost, carbon emission cost, battery degradation cost, hydrogen conversion cost, demand response compensation, and renewable curtailment penalty. It can be written as:

$$C_s = C_s^{\text{grid}} + C_s^{\text{gas}} + C_s^{\text{om}} + C_s^{\text{carbon}} + C_s^{\text{bat}} + C_s^{\text{h2}} + C_s^{\text{dr}} + C_s^{\text{curt}}$$

In this formulation, the expected-cost term encourages economical operation under normal scenarios, while the CVaR term limits the average cost of adverse scenarios. Therefore, the model is not only cost-oriented but also risk-aware. This is important for high-renewable integrated energy systems, because renewable output and load demand may deviate from their forecasted values and cause expensive corrective operation.

3.2. Energy balance constraints

The integrated energy system must satisfy electricity and heat balance in each time period and under each scenario. The electricity balance links renewable generation, grid interaction, gas turbine output, battery operation, hydrogen conversion, heat pump consumption, and demand response. For time period t and scenario s , the electric power balance is expressed as:

$$P_{t,s}^{\text{wt}} + P_{t,s}^{\text{pv}} - P_{t,s}^{\text{curt,wt}} - P_{t,s}^{\text{curt,pv}} + P_{t,s}^{\text{gt}} + P_{t,s}^{\text{fc}} + P_{t,s}^{\text{bat,dis}} + P_{t,s}^{\text{grid,buy}} = L_{t,s}^e - P_{t,s}^{\text{dr}} + P_{t,s}^{\text{hp}} + P_{t,s}^{\text{el}} + P_{t,s}^{\text{bat,ch}} + P_{t,s}^{\text{grid,sell}}$$

where $P_{t,s}^{\text{wt}}$ and $P_{t,s}^{\text{pv}}$ are the available wind and photovoltaic power, respectively. $P_{t,s}^{\text{curt,wt}}$ and $P_{t,s}^{\text{curt,pv}}$ represent curtailed wind and photovoltaic power. $P_{t,s}^{\text{gt}}$, $P_{t,s}^{\text{fc}}$ and $P_{t,s}^{\text{bat,dis}}$ are the electric outputs of the gas turbine, fuel cell, and battery. $P_{t,s}^{\text{grid,buy}}$ and $P_{t,s}^{\text{grid,sell}}$ are the purchased and sold electricity from the main grid. On the demand side, $L_{t,s}^e$ is the electric load, $P_{t,s}^{\text{dr}}$ is the interrupted demand response load, $P_{t,s}^{\text{hp}}$ is the electric consumption of the heat pump, $P_{t,s}^{\text{el}}$ is the electric consumption of the electrolyzer, and $P_{t,s}^{\text{bat,ch}}$ is the battery charging power.

The heat balance ensures that the heat load is supplied by heat recovery and heat-producing devices. It is given by:

$$H_{t,s}^{\text{gt}} + H_{t,s}^{\text{gb}} + H_{t,s}^{\text{hp}} + H_{t,s}^{\text{fc}} = L_{t,s}^h$$

where $H_{t,s}^{\text{gt}}$ is the recovered heat from the gas turbine, $H_{t,s}^{\text{gb}}$ is the heat output of the gas boiler, $H_{t,s}^{\text{hp}}$ is the heat output of the heat pump, $H_{t,s}^{\text{fc}}$ is the recovered heat from the fuel cell, and $L_{t,s}^h$ is the heat load. The heat pump output is related to its electric consumption by:

$$H_{t,s}^{\text{hp}} = \text{COP}^{\text{hp}} P_{t,s}^{\text{hp}}$$

The renewable curtailment variables are constrained by the available renewable generation:

$$0 \leq P_{t,s}^{\text{curt,wt}} \leq P_{t,s}^{\text{wt}}, \quad \forall t, s$$

$$0 \leq P_{t,s}^{\text{curt,pv}} \leq P_{t,s}^{\text{pv}}, \quad \forall t, s$$

These balance constraints ensure that electricity and heat demands are met at all times. They also allow

surplus renewable energy to be absorbed by battery charging, hydrogen production, grid export, or demand-side adjustment. If these flexibility resources are insufficient, renewable curtailment occurs and is penalized in the objective function.

3.3. Storage and hydrogen constraints

Battery energy storage is used to transfer electric energy across time periods. Its state of charge changes according to the charging and discharging power. For time period t and scenario s , the battery energy dynamics are formulated as:

$$E_{t,s}^{\text{bat}} = E_{t-1,s}^{\text{bat}} + \eta^{\text{ch}} P_{t,s}^{\text{bat,ch}} \Delta t - \frac{P_{t,s}^{\text{bat,dis}} \Delta t}{\eta^{\text{dis}}}$$

where $E_{t,s}^{\text{bat}}$ is the stored energy of the battery, $P_{t,s}^{\text{bat,ch}}$ and $P_{t,s}^{\text{bat,dis}}$ are the charging and discharging power, η^{ch} and η^{dis} are the charging and discharging efficiencies, and Δt is the scheduling time interval. The battery power and energy limits are given by:

$$0 \leq P_{t,s}^{\text{bat,ch}} \leq u_{t,s}^{\text{bat,ch}} P_{\text{max}}^{\text{bat}}$$

$$0 \leq P_{t,s}^{\text{bat,dis}} \leq u_{t,s}^{\text{bat,dis}} P_{\text{max}}^{\text{bat}}$$

$$E_{\text{min}}^{\text{bat}} \leq E_{t,s}^{\text{bat}} \leq E_{\text{max}}^{\text{bat}}$$

To avoid simultaneous charging and discharging, the following binary constraint is imposed:

$$u_{t,s}^{\text{bat,ch}} + u_{t,s}^{\text{bat,dis}} \leq 1$$

The initial and terminal battery energy levels are constrained to be equal:

$$E_{0,s}^{\text{bat}} = E_{\text{ini}}^{\text{bat}}$$

$$E_{T,s}^{\text{bat}} = E_{\text{ini}}^{\text{bat}}$$

Hydrogen storage is coupled with the electrolyzer and the fuel cell. The electrolyzer consumes electricity to produce hydrogen, while the fuel cell consumes hydrogen to generate electricity and heat. The hydrogen storage dynamics are described as:

$$H_{t,s}^{\text{h2}} = H_{t-1,s}^{\text{h2}} + \eta^{\text{el}} P_{t,s}^{\text{el}} \Delta t - \frac{P_{t,s}^{\text{fc}} \Delta t}{\eta^{\text{fc}}}$$

where $H_{t,s}^{\text{h2}}$ is the hydrogen storage level, $P_{t,s}^{\text{el}}$ is the electrolyzer power, $P_{t,s}^{\text{fc}}$ is the electric output of the fuel cell, η^{el} is the hydrogen production efficiency of the electrolyzer, and η^{fc} is the electric conversion efficiency of the fuel cell. The operation limits of the electrolyzer, fuel cell, and hydrogen storage tank are expressed as:

$$0 \leq P_{t,s}^{\text{el}} \leq u_{t,s}^{\text{el}} P_{\text{max}}^{\text{el}}$$

$$0 \leq P_{t,s}^{\text{fc}} \leq u_{t,s}^{\text{fc}} P_{\text{max}}^{\text{fc}}$$

$$H_{\text{min}}^{\text{h2}} \leq H_{t,s}^{\text{h2}} \leq H_{\text{max}}^{\text{h2}}$$

The electrolyzer and fuel cell are not allowed to operate at the same time, which is enforced by:

$$u_{t,s}^{\text{el}} + u_{t,s}^{\text{fc}} \leq 1$$

The initial and terminal hydrogen storage levels are also required to be equal:

$$H_{0,s}^{h2} = H_{ini}^{h2}$$

$$H_{T,s}^{h2} = H_{ini}^{h2}$$

These constraints describe the intertemporal behavior of the two storage resources. Battery energy storage mainly provides short-term electric flexibility with high efficiency, while hydrogen storage provides longer-duration renewable energy absorption. The terminal constraints prevent the model from obtaining an artificial cost reduction by simply depleting stored energy at the end of the scheduling horizon.

3.4. Demand response and operational constraints

Demand response is modeled as an interruptible electric load. It provides load-side flexibility during high-cost or high-risk periods. For each time period and scenario, the interrupted load is limited by a fixed proportion of the electric demand:

$$0 \leq P_{t,s}^{dr} \leq \gamma^{dr} L_{t,s}^e, \quad \forall t, s$$

where $P_{t,s}^{dr}$ is the interrupted demand response load, γ^{dr} is the maximum interruption ratio, and $L_{t,s}^e$ is the electric load. To avoid unrealistic load reduction over the whole scheduling horizon, the total interrupted energy is further constrained by:

$$\sum_{t=1}^T P_{t,s}^{dr} \Delta t \leq \Gamma^{dr} \sum_{t=1}^T L_{t,s}^e \Delta t, \quad \forall s$$

where Γ^{dr} is the maximum daily interruption ratio. This constraint reflects the limited availability of flexible demand. In this study, demand response should be interpreted as interruptible load rather than shiftable load, because the reduced demand is not transferred to other periods.

The gas turbine and the heat pump are coupled with the heat balance through their conversion relationships. The recovered heat from the gas turbine is proportional to its electric output:

$$H_{t,s}^{gt} = a^{gt} P_{t,s}^{gt}, \quad \forall t, s$$

where a^{gt} is the heat-to-power ratio of the gas turbine. The heat pump output is determined by its coefficient of performance:

$$H_{t,s}^{hp} = COP^{hp} P_{t,s}^{hp}, \quad \forall t, s$$

The main grid interaction is also constrained. The system can purchase electricity from the grid or sell surplus electricity to the grid, but a binary variable prevents simultaneous buying and selling:

$$0 \leq P_{t,s}^{grid,buy} \leq u_{t,s}^{grid} P_{max}^{grid,buy}$$

$$0 \leq P_{t,s}^{grid,sell} \leq (1 - u_{t,s}^{grid}) P_{max}^{grid,sell}$$

where $u_{t,s}^{grid}$ is a binary grid operation variable. This treatment avoids artificial arbitrage between buying and selling electricity in the same period. The export capacity is set lower than the purchase capacity, which represents a practical limitation on renewable power delivery to the external grid.

The risk constraints are embedded in the CVaR formulation. For each scenario, the excess cost variable z_s measures the part of the scenario cost that exceeds the auxiliary threshold η . This structure penalizes high-cost scenarios and encourages the dispatch strategy to remain economical under renewable and load uncertainty. In the optimization results, the CVaR value is therefore used as a risk-oriented performance index in addition to the expected operating cost.

3.5. Solution procedure and case design

The proposed model is solved over a 24-hour scheduling horizon with an hourly time resolution. Wind power, photovoltaic power, and electric load are represented by a set of uncertainty scenarios. The same scenario set is used for all comparative cases, so the performance differences are caused by the configuration of flexible resources rather than by random scenario variation. Five cases are designed for comparison: a system without storage, a system with battery energy storage only, a system with hydrogen storage only, a system with both battery energy storage and hydrogen storage, and the proposed system with battery energy storage, hydrogen storage, and demand response. **Figure 1** presents the weighted objective, expected operating cost, CVaR cost, and computational time of the five cases.

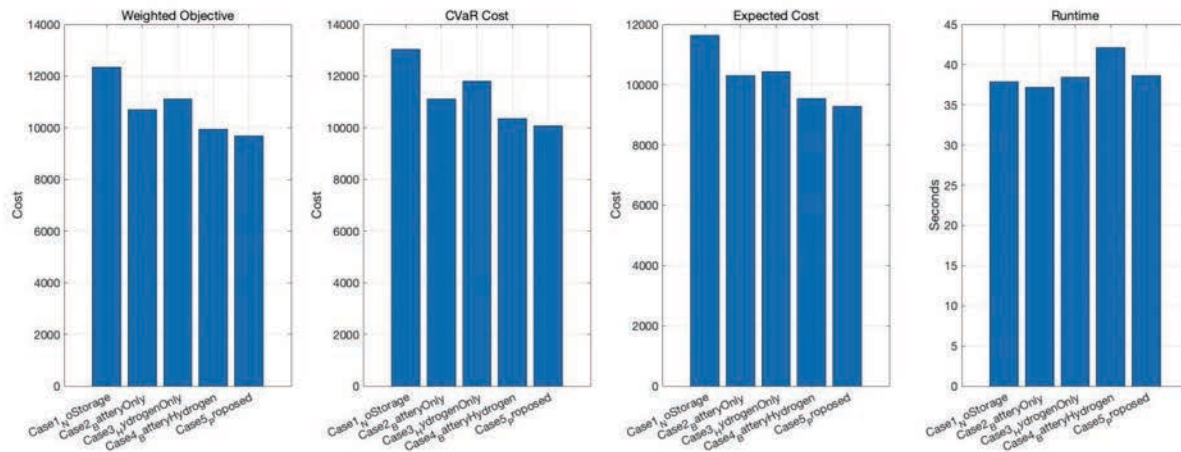


Figure 1. Cost and runtime.

4. Case study

A 24-hour case study is conducted to evaluate the proposed risk-aware optimal dispatch model for the electric-heat-hydrogen integrated energy system. The scheduling interval is one hour. Wind power, photovoltaic power, and electric load are treated as uncertain variables and represented by multiple scenarios. To ensure a fair comparison, the same scenario set is used for all cases. Five configurations are compared, including no storage, battery storage only, hydrogen storage only, battery-hydrogen storage, and the proposed battery-hydrogen-demand response scheme.

4.1. Economic and risk performance analysis

As shown in **Figure 1**, the proposed battery-hydrogen-demand response scheme achieves the lowest weighted objective, expected operating cost, and CVaR cost among the five cases, while maintaining an acceptable computational time. This indicates that the coordinated use of battery storage, hydrogen storage, and demand response improves both economic performance and tail-risk control under renewable generation and load uncertainty.

The results show that the proposed case achieves the best economic and risk performance. Compared with the no-storage case, the weighted objective decreases from 12337.22 to 9677.39, which corresponds to a reduction of 21.56%. The expected operating cost decreases by 20.15%, and the CVaR cost decreases by

22.82%. This indicates that the proposed scheduling strategy can reduce both average operating cost and tail-risk cost under renewable generation and load uncertainty. The battery-only case also performs well, since battery storage has high round-trip efficiency and can respond quickly to short-term renewable fluctuations. However, the hydrogen-only case has a higher weighted objective than the battery-only case, even though it improves renewable utilization. This result is reasonable because hydrogen conversion involves larger energy losses than battery storage. The battery-hydrogen case further reduces both expected cost and CVaR cost, showing that short-term battery flexibility and longer-duration hydrogen storage are complementary. After demand response is added, the proposed case obtains the lowest weighted objective. This improvement suggests that load-side flexibility can reduce high-cost peak-period operation and improve the risk-aware dispatch performance.

4.2. Renewable utilization and curtailment analysis

As shown in **Figure 2**, renewable utilization increases and curtailment decreases as flexible resources are gradually introduced. The proposed scheme achieves the highest renewable utilization and the lowest curtailment rate, and it also reduces expected carbon emissions compared with the no-storage case. These results demonstrate that the coordination of battery storage, hydrogen storage, and demand response can enhance renewable energy accommodation and improve the low-carbon performance of the integrated energy system.

The no-storage case exhibits the highest renewable curtailment because excess wind and photovoltaic generation cannot be shifted to other periods. Introducing battery storage significantly improves renewable utilization by absorbing short-term surplus power and releasing it during high-demand periods. The hydrogen-only case also reduces curtailment by converting excess renewable electricity into hydrogen, demonstrating its advantage in long-duration energy storage. When battery and hydrogen storage are jointly deployed, their complementary operating characteristics enable the system to accommodate a larger proportion of renewable generation. The proposed case achieves the highest renewable utilization and the lowest curtailment after demand response is incorporated. This result indicates that coordinating short-term battery storage, long-duration hydrogen storage, and flexible demand can effectively mitigate renewable power fluctuations and enhance the system's overall renewable energy accommodation capability.

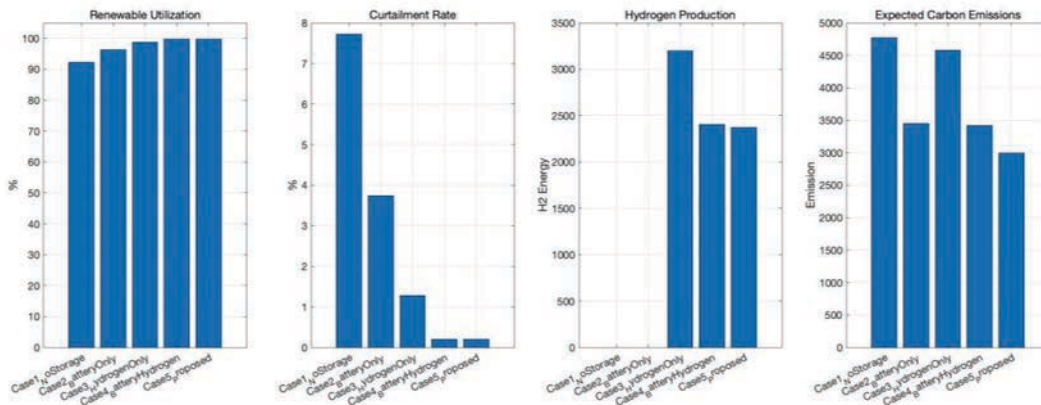


Figure 2. Energy and emission metrics.

4.3. Energy storage and demand response operation analysis

As shown in **Figure 3**, the best-case dispatch profile illustrates the complementary operation of battery storage, hydrogen storage, and demand response. The battery charges during renewable-rich or low-price periods and discharges during peak-load periods, while hydrogen production absorbs sustained renewable surplus and the fuel cell supports the system when a power shortage occurs. Demand response further reduces peak demand, thereby lowering grid dependence and improving operational flexibility.

The battery mainly charges during periods of surplus renewable generation or low electricity prices and discharges during peak-load or high-price periods. This operating pattern confirms its role in managing short-term power fluctuations and peak–valley differences. In contrast, the hydrogen storage system absorbs sustained renewable surplus through the electrolyzer and supplies energy through the fuel cell when power shortages occur. Demand response further adjusts the load profile by transferring flexible demand from peak periods to periods with abundant renewable generation. The coordinated operation of these resources reduces peak grid dependence, improves renewable energy consumption, and lowers exposure to high-cost scenarios. Therefore, battery storage, hydrogen storage, and demand response perform complementary functions in the proposed risk-aware scheduling framework. (See **Figure 3**)

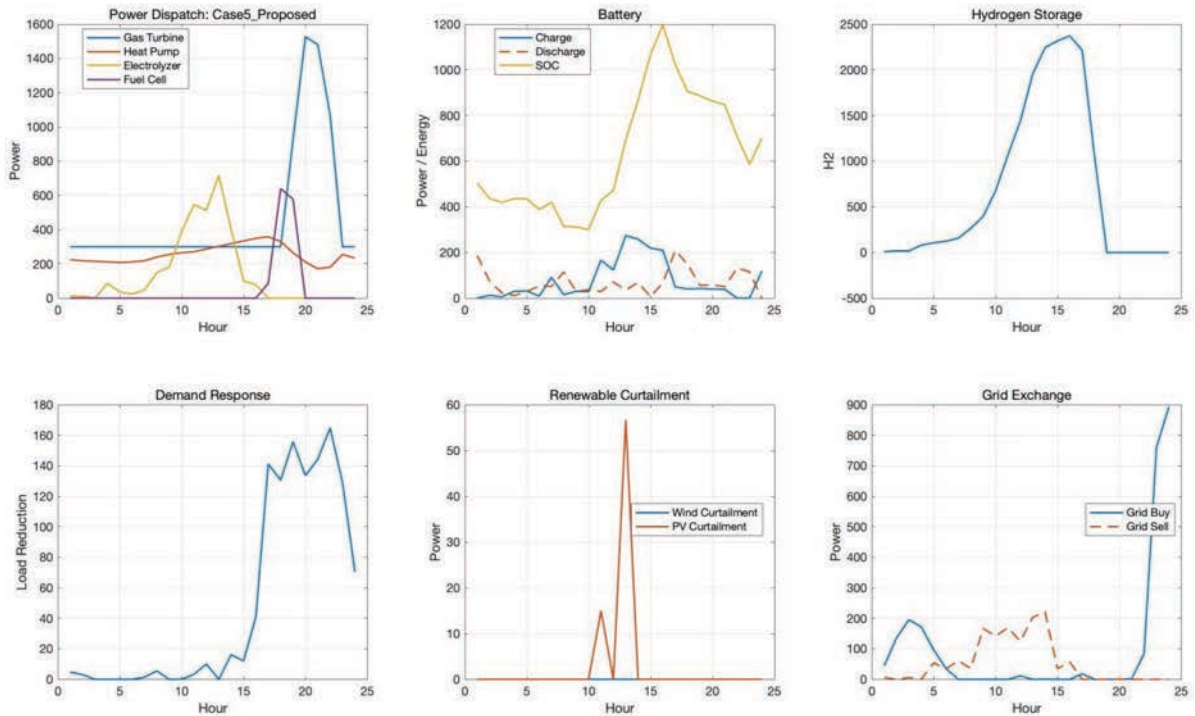


Figure 3. Best case dispatch.

4.4. Chapter summary

This chapter evaluates the proposed two-stage CVaR-based scheduling strategy under five system configurations. The results demonstrate that coordinating battery storage, hydrogen storage, and demand response achieves the lowest weighted objective, expected operating cost, and CVaR cost. The proposed case

also improves renewable energy utilization and reduces curtailment through the complementary operation of short-term battery storage and long-duration hydrogen storage. Demand response further enhances scheduling flexibility by shifting loads away from peak and high-cost periods. Overall, the proposed strategy improves the economic efficiency, renewable energy accommodation capability, and operational robustness of the integrated energy system under uncertainty.

5. Conclusion

This study proposes a two-stage CVaR-based scheduling model for an integrated energy system incorporating renewable generation, battery storage, hydrogen storage, and demand response. The model considers renewable generation and load uncertainties while balancing expected operating cost and tail-risk cost.

The results show that the coordinated configuration achieves the best overall performance among the five cases. Compared with the no-storage case, the proposed case reduces the weighted objective by 21.56%, the expected operating cost by 20.15%, and the CVaR cost by 22.82%. Battery storage efficiently manages short-term power fluctuations, whereas hydrogen storage supports longer-duration energy shifting. Demand response further improves system flexibility by transferring loads away from peak and high-cost periods. Their coordinated operation reduces renewable energy curtailment, grid dependence, and exposure to high-cost scenarios.

Overall, the proposed strategy provides an effective approach for improving the economy, renewable energy accommodation capability, and operational robustness of integrated energy systems under uncertainty. Future research may consider equipment degradation, multi-day scheduling, forecast correlations, and real-time operation to further enhance the practical applicability of the model.

Disclosure statement

The author declares no conflict of interest.

Reference

- [1] Li B, Chen M, Zhong H, et al., 2023, A Review of Long-Term Planning of New Power Systems with Large Share of Renewable Energy. *Proceedings of the CSEE*, 43(2): 555–581.
- [2] Zhuo Z, Zhang N, Xie X, et al., 2021, Key Technologies and Development Challenges of Power Systems with High Proportion of Renewable Energy. *Automation of Electric Power Systems*, 45(9): 171–191.
- [3] Liu Z, 2026, Research on Future Development Path of Renewable Energy: Analysis Based on Renewable Energy Consumption Responsibility Weight and Green Power Direct Connection Policy. *China Energy Observation*, 2026(2): 92–96.
- [4] Yang Z, Cao J, Shi Y, et al., 2026, Capacity Allocation Optimization of Wind-Solar-Load-Storage for Grid-Connected Green Power Direct Supply Projects Considering Load-Side Response. *Northwest Hydropower*, 2026(2): 124–130.
- [5] Zhao D, An Y, Sun Z, et al., 2026, Research on Optimization and Scheduling Technology of Green Power Direct Connection System for AIDC Computing Power Base. *Journal of Electric Power*: 1–11.
- [6] Ding H, Shang H, Tao Y, et al., 2026, A Review on Scheduling Optimization of Source-Grid-Load-Storage

- Integrated System. *Journal of Chinese Society of Power Engineering*, 46(5): 50–63.
- [7] Lin W, Jiang R, Ren X, 2026, Definition and Key Collaborative Technologies for Electricity-Hydrogen-Carbon Ecosystem. *Power System Technology*: 1–19.
 - [8] Zhang B, Cai Y, Liu D, et al., 2025, An Intelligent Regulation Solution for Integrated Wind-Solar-Storage and Power-Hydrogen-Ammonia Coupling Systems. *Automation Panorama*, 42(10): 54–57.
 - [9] Huang Z, Wen Z, Wang C, 2025, Optimal Scheduling of Integrated Energy System Considering Multiple Utilization of Ammonia Energy. *Modern Electronics Technique*: 1–6.
 - [10] Wang S, Hu S, Cheng D, et al., 2026, Two-Stage Distributionally Robust Scheduling of Off-Grid Wind-Solar-Hydrogen-Ammonia Systems Adopting Multi-State Flexible Synthesis. *Journal of Xi'an Jiaotong University*: 1–10.
 - [11] Cheng J, Zhang B, Bao G, 2025, Two-Stage Optimal Scheduling Strategy for Hydrogen-Electric-Thermal Multi-Energy Coupled Hybrid Hydrogen Production Integrated Energy System. *Southern Power System Technology*: 1–12.
 - [12] Wang R, Liu J, Ju W, et al., 2025, Coordinated Optimal Scheduling of Electric-Hydrogen System Considering Hybrid Energy Storage in the Day-Ahead and Intra-Day Stages. *Journal of Shandong University (Engineering Science)*: 1–9.
 - [13] Xu Z, Sun Y, Xie D, et al., 2020, Optimal Energy Storage Configuration of Regional Integrated Energy System Considering Electric/Thermal Flexible Loads. *Automation of Electric Power Systems*, 44(2): 53–59.
 - [14] Qi N, Cheng L, Tian L, et al., 2020, Review and Prospect of Distribution Network Planning Considering Flexible Load Integration. *Automation of Electric Power Systems*, 44(10): 193–207.
 - [15] Zhang J, Lyu H, Hu S, et al., 2026, Low-Carbon Optimization Scheduling of Electric-Thermal Integrated Energy System Considering Uncertainty of Flexible Loads. *Machine Building & Automation*, 55(1): 213–217.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

A Brief Analysis of the Development of the Wire and Cable Industry and Quality Test Technology in Shaanxi Province in the Age of Artificial Intelligence

Beini An

Shaanxi Inst Prod Qual Supervis & Inspection, Xi'an 710048, Shaanxi, China

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: As the “industrial blood vessel” of the national economy, wire and cable products underpin the safety of infrastructure and power systems. The rise of artificial intelligence (AI) is driving the transformation of cable quality testing from manual operation to intelligent mode. As a major energy base in China, Shaanxi Province has formed a distinctive cable industrial cluster, yet faces prominent problems such as low intelligent quality test level and insufficient high-end supply. This paper sorts out the development status of Shaanxi’s wire and cable industry, analyzes the bottlenecks of traditional quality test technology, and elaborates the enabling value of AI from four dimensions: defect identification, non-destructive testing, whole-process quality control, and intelligent supervision. On this basis, it summarizes the practical challenges of AI-industry integration, and proposes targeted development paths including building a provincial AI quality test service center, promoting industry-university-research collaboration, upgrading industrial clusters, formulating local standards and constructing an intelligent supervision system, so as to provide references for the high-quality development of Shaanxi’s cable industry.

Keywords: Artificial intelligence; Wire and cable; Quality test technology; Shaanxi province

Online publication: Jul 23, 2026

1. Introduction

Wire and cable are widely applied in energy, power, transportation, and aerospace sectors, serving as critical infrastructure for national economic operation. With the advancement of the digital economy, AI technology has become a core driving force for industrial upgrading, bringing transformative opportunities for cable quality testing.

As an important comprehensive energy base and UHV power transmission hub in China, Shaanxi Province boasts a large-scale cable industrial cluster closely linked to local pillar industries. According to Shaanxi’s 15th Five-Year Plan, AI will be empowered to penetrate key industrial scenarios including R&D design, pilot verification, manufacturing and operation management ^[1]. For provincial quality supervision

institutions, promoting the integration of AI and cable quality testing is not only an inevitable requirement to serve local industrial development, but also a concrete practice to implement the national strategy of quality and technological power. This paper takes Shaanxi as a research sample to discuss the integration path of AI technology and the cable industry.

2. Status of wire and cable industry in Shaanxi province

2.1. Growing demand driven by characteristic industries

Shaanxi's huge demand for cables comes from three major fields: first, the large-scale development of new energy and UHV transmission projects in northern Shaanxi, which drives the demand for high-voltage and extra-high-voltage cables; second, the strong demand for special cables from aerospace, rail transit and high-end equipment manufacturing industries; third, the Xianyang power transmission and transformation equipment industrial cluster, which has formed a collaborative industrial pattern of "power transmission equipment & wire and cable" led by leading enterprises.

2.2. Urgent demand for intelligent upgrading of industrial clusters

Shaanxi's cable industry has formed a large industrial cluster dominated by medium and low voltage power cables, mining cables and general rubber-sheathed cables, with some enterprises extending to UHV and new energy special cables. However, the industry is generally large but not strong, with a high proportion of small and medium-sized enterprises, insufficient high-end production capacity, and low intelligence level in production and quality test links. With the construction of new power systems and the development of Qinchuangyuan Innovation-driven Platform, local enterprises have an increasingly urgent demand for digital and intelligent transformation.

3. Pain points of traditional cable quality test technology

Wire and cable are core infrastructure of power systems, and their service life directly affects the stability and safety of the whole system ^[2]. Cable quality test covers appearance, size, electrical performance, mechanical performance, and other links, which is the key to ensuring product quality. At present, cable quality test mainly relies on manual visual inspection and mechanical testing, with prominent limitations in efficiency and precision.

3.1. High dependence on manual work with limited efficiency and accuracy

Traditional appearance inspection and partial discharge quality test rely heavily on manual operation, which is susceptible to subjective factors and personnel status. For micro defects such as millimeter-scale water marks and insulation scratches, manual quality tests have a high miss rate, leading to defective products flowing into the market and potential safety hazards.

3.2. Disconnection between quality test and production without real-time control

The traditional mode is dominated by offline sampling inspection, which cannot realize online synchronous quality testing on production lines. Quality problems in production cannot be found in time, often resulting in batch unqualified products. Most quality test data are recorded on paper and scattered in different links, failing to support process optimization through data analysis.

3.3. Insufficient high-end quality test supply for special scenarios

For special cables used in aerospace and UHV fields, high-precision non-destructive testing is required for internal defects and weather resistance, which cannot be well met by traditional testing equipment. Meanwhile, traditional technology lacks support for full life cycle management, quality traceability, and fault early warning.

3.4. Scattered quality test resources with low supervision efficiency

Cable manufacturers are widely distributed across the province. Traditional supervision relies on offline spot checks, and quality test data from different regions and institutions cannot be shared, making it difficult for regulatory authorities to dynamically grasp the overall product quality situation.

4. Innovative Application of AI in cable quality test

4.1. High-precision appearance defect quality test based on machine vision

Intelligent cable quality inspection equipment equipped with machine vision and deep learning algorithms can convert visual information in the physical world into computable data through digital image processing and pattern recognition, so as to realize the quality test, identification, and tracking of objects or processes^[3]. It can fully meet the real-time quality test requirements of high-speed production lines, control appearance quality from the production source, and reduce the outflow of defective products.

4.2. Internal defect quality test combining AI and non-destructive testing

AI technology is integrated with rapid testing technologies including infrared thermal imaging, X-ray fluorescence spectroscopy, ultrasonic testing, and partial discharge quality test, to automatically identify and quantitatively analyze internal defects such as conductor wire breakage, insulation layer bubbles, and shielding layer damage, providing reliable technical support for refined quality control of special cables^[4].

4.3. Whole-process quality control based on multi-source big data

With the architecture of “terminal real-time collection + cloud intelligent analysis”, quality test data can be synchronized to the cloud platform in real time. AI algorithms can conduct quality trend analysis and fault cause tracing, promoting the transformation of quality management from post-inspection to pre-prevention.

4.4. Provincial intelligent quality supervision platform

A provincial cable quality big data platform integrating enterprise production data and supervision sampling data can realize dynamic monitoring, accurate traceability, and risk early warning of product quality across the province, providing intelligent decision support for market supervision departments.

5. Challenges of AI-industry integration in Shaanxi

(1) Low willingness of small and medium-sized enterprises

Most small and medium-sized cable enterprises still attach more importance to production than quality testing. The high upfront cost of intelligent testing equipment leads to insufficient investment in intelligent transformation.

(2) Lack of compound talents

The application of AI in local testing institutions is still in the initial stage. Professionals who master both cable quality test knowledge and AI algorithm technology are in short supply, resulting in a mismatch between technology and practical demand.

(3) Unsmooth achievement transformation and lack of standards

The transformation efficiency of AI technological achievements through platforms such as Qinchuangyuan needs to be improved. There is still a lack of unified national standards for quality test indicators, algorithm accuracy, and data format.

(4) Imperfect collaborative mechanism

AI enterprises, cable manufacturers, testing institutions, and research institutes have not established a normalized cooperation mechanism. A complete integrated development ecology of “R&D-testing-application-standard” has not yet been formed.

6. Targeted development paths

6.1. Build a provincial AI cable quality test technology service center

Provincial testing institutions should give full play to their own testing capabilities and technical advantages, and take the lead in promoting the R&D and application of AI testing technology. First, upgrade intelligent testing equipment and introduce advanced devices such as AI cable quality inspection machines and X-ray intelligent non-destructive testing systems, to build an authoritative AI testing technology verification platform in the province. Second, build a provincial wire and cable quality big data platform, integrating enterprise testing data, supervision sampling data, and risk monitoring data, to realize quality trend analysis, fault tracing, and risk warning through AI algorithms, and provide data services for enterprises and market supervision departments ^[5]. Third, carry out AI testing technology promotion services to reduce the transformation threshold for small and medium-sized enterprises.

6.2. Construct a government-industry-university-research collaborative innovation system

Led by testing institutions and supported by the Qinchuangyuan platform, an innovation alliance for AI cable quality test technology should be established together with universities, leading enterprises, and AI companies. The alliance will carry out joint technical research for high-end special cables, promote the transformation of scientific and technological achievements, and co-build talent training bases to cultivate compound professionals.

6.3. Promote intelligent upgrading of industrial clusters

Taking the Xianyang power transmission and transformation equipment cluster as a pilot, leading enterprises such as XD Group and Shaanxi Chint Intelligent Electric will be guided to take the lead in intelligent transformation, realize full-process intelligent quality testing on production lines, and form a demonstration effect ^[6]. Supported by AI quality test technology, local enterprises will be encouraged to extend to high-end fields such as aerospace, UHV and new energy, and build a whole-industry-chain quality control system.

6.4. Accelerate standard formulation for standardized development

Testing institutions should take the lead in formulating local standards for AI cable quality test technology,

clarifying requirements for quality test indicators, algorithm accuracy and data format, and promoting the connection between local standards and industry/national standards. Standard publicity and implementation supervision should be carried out to guide the standardized development of the industry.

6.5. Build a provincial AI quality supervision system

A quality supervision system featuring “intelligent quality test & big data analysis & precise supervision” should be constructed. By integrating production and testing data across the province, quality portraits of enterprises will be formed. AI algorithms will be used for risk monitoring and targeted spot checks, so as to improve the accuracy and efficiency of quality governance.

7. Conclusion

AI technology brings historic opportunities for the upgrading of the wire and cable industry and quality test technology. Shaanxi has unique industrial scenarios and a demand foundation for AI application. Provincial quality testing institutions should give full play to their technical advantages, take intelligent quality test upgrading as the starting point, build a collaborative innovation system and a big data platform, and promote the in-depth integration of AI and the local cable industry, so as to accelerate the industry’s transformation toward high-end, intelligent and characteristic development.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Shaanxi Provincial People’s Government, 2026, Outline of the 15th Five-Year Plan for National Economic and Social Development of Shaanxi Province. Shaanxi Daily, 2026-04-16(003).
- [2] Wang W, 2025, Research on Application of Artificial Intelligence Technology in Cable Life Assessment. Electrical Technology & Economy, 2025(6): 75–78.
- [3] Qu H, Lyu X, Xiao X, 2024, Research on Application of Machine Vision Technology in Wire and Cable Test. China Quality Supervision, 2024(8): 86–87.
- [4] Zhang F, 2025, Application of Rapid Test Technology in Wire and Cable Inspection. Engineering and Technical Research, 10(14): 110–112.
- [5] Meng F, Yang Q, Gao Y, et al., 2024, Research and Enlightenment on Construction Mode of Provincial Government Big Data Platform. Big Data, 10(3): 27–39.
- [6] Guo Z, 2026, Current Situation, Problems and Paths of Digital Transformation of Manufacturing Industry in Shaanxi Province. New West, 2026(4): 188–192.

Publisher’s note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Integrated Services Platform of International Scientific Cooperation

Innoscience Research (Malaysia), which is global market oriented, was founded in 2016. Innoscience Research focuses on services based on scientific research. By cooperating with universities and scientific institutes all over the world, it performs medical researches to benefit human beings and promotes the interdisciplinary and international exchanges among researchers.

Innoscience Research covers biology, chemistry, physics and many other disciplines. It mainly focuses on the improvement of human health. It aims to promote the cooperation, exploration and exchange among researchers from different countries. By establishing platforms, Innoscience integrates the demands from different fields to realize the combination of clinical research and basic research and to accelerate and deepen the international scientific cooperation.

Cooperation Mode



Clinical Workers



In-service Doctors



Foreign Researchers



Hospital



University



Scientific institutions

OUR JOURNALS



The *Journal of Architectural Research and Development* is an international peer-reviewed and open access journal which is devoted to establish a bridge between theory and practice in the fields of architectural and design research, urban planning and built environment research.

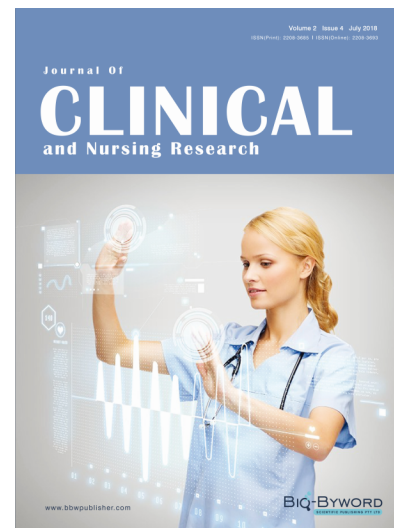
Topics covered but not limited to:

- Architectural design
- Architectural technology, including new technologies and energy saving technologies
- Architectural practice
- Urban planning
- Impacts of architecture on environment

Journal of Clinical and Nursing Research (JCNR) is an international, peer reviewed and open access journal that seeks to promote the development and exchange of knowledge which is directly relevant to all clinical and nursing research and practice. Articles which explore the meaning, prevention, treatment, outcome and impact of a high standard clinical and nursing practice and discipline are encouraged to be submitted as original article, review, case report, short communication and letters.

Topics covered by not limited to:

- Development of clinical and nursing research, evaluation, evidence-based practice and scientific enquiry
- Patients and family experiences of health care
- Clinical and nursing research to enhance patient safety and reduce harm to patients
- Ethics
- Clinical and Nursing history
- Medicine



Journal of Electronic Research and Application is an international, peer-reviewed and open access journal which publishes original articles, reviews, short communications, case studies and letters in the field of electronic research and application.

Topics covered but not limited to:

- Automation
- Circuit Analysis and Application
- Electric and Electronic Measurement Systems
- Electrical Engineering
- Electronic Materials
- Electronics and Communications Engineering
- Power Systems and Power Electronics
- Signal Processing
- Telecommunications Engineering
- Wireless and Mobile Communication

