ISSN 3137-0829

JOURNAL OF DIGITAL INTELLIGENCE FOR INDUSTRY

Vol.1, No.1, October 2026



HONG KONG OCEANWISH PRESS LIMITED

jdii.org

CONTENTS

1 Research on the Application of Embodied Intelligent Welding Robots in Construction Steel Structure Scenarios
LI Yong

12 The Application of Digital Twin Technology Integrated with AI in Intelligent Oilfield Production
HOU Lei

24 Research on Strategies and Methods for Improving the Efficiency of Exoskeleton Servo Drives
WANG Bin

35 Research on Environmental Perception Technology in Unmanned Delivery Vehicles
LI Yanze

49 Application of Smart Valves in Wastewater Treatment Systems
TAN Tu

Research on the Application of Embodied Intelligent Welding Robots in Construction Steel Structure Scenarios

LI Yong

Abstract:Through a closed-loop architecture of "perception-decision-execution", embodied intelligent welding robots deeply integrate visual perception, process large models and robot motion control, realizing a paradigm shift from "fixed program execution" to "autonomous intelligent welding". This paper systematically sorts out the technical connotation and core architecture of embodied intelligent welding robots, and analyzes the applicable characteristics of three types of robots in construction steel structure scenarios, including seven-axis ground-track inverted installation robots, intelligent collaborative welding robots and mobile (tracked) robots. Combined with practical application cases from enterprises such as Jiangsu Beiren, CSCEC Science and Industry Group, China Railway Baoji Bridge Group and Zhongke Optoelectronics, this paper demonstrates the technical advantages of embodied intelligence in addressing core pain points such as complex weld seam identification, adaptive adjustment of process parameters and compatibility with multi-variety workpieces. The research results indicate that embodied intelligent welding robots can cut the welding process development cycle by 60%, reduce comprehensive production costs by more than 40%, achieve a first-pass welding qualification rate exceeding 99%, and multiply the operation efficiency of single-person multi-machine deployment.

Key words:Embodied Intelligence; Welding Robot; Construction Steel Structure; Visual Perception; AI Large Model; Human-Machine Collaboration

1 INTRODUCTION

1.1 Research Background

As a key structural form in modern construction, building steel structures have gained increasing application in high-rise buildings, large-span venues, industrial plants, and bridge projects due to their comprehensive advantages such as high strength, light self-weight, short construction cycles, and excellent seismic performance. According to statistics from the China Steel Structure Association, during the 14th Five-Year Plan period, the domestic steel structure industry has made significant progress in digitalization and intelligent transformation. By the end of 2024, among the 571 surveyed steel structure enterprises, 371 had adopted digital processing equipment, reaching a penetration rate of 65%.

However, welding—the core process in steel structure manufacturing—remains significantly behind in automation compared to discrete manufacturing industries such as automotive and shipbuilding. Although the number of companies using welding robots has increased to 243, accounting for 42% of surveyed firms, the overall welding automation rate still remains below 30%, with many critical processes still relying heavily on manual operation. This reality has led to multiple challenges, including poor quality consistency, low production efficiency, high labor costs, and a shortage of skilled workers, increasingly becoming a bottleneck that constrains the industry's high-quality development.

1.2 The "capability boundary" of traditional welding robots

Traditional industrial welding robots operate in a "teach-and-learn" mode, where engineers manually guide the robot's end effector along a predetermined welding path, recording all trajectory points. The robot then repeats this program. This approach works well in mass production scenarios with few product variations, such as automobile manufacturing—where a single vehicle model may require thousands of weld points. Once the program is debugged, it can run stably for months or even years.

However, directly applying this technology to the construction steel structure industry leads to fundamental incompatibility. Steel structure fabrication in construction is inherently project-based rather than product-based—each building has unique structural designs, requiring custom-tailored components, specifications, and weld types. Lin Tao, Vice President of Jiangsu Beiren, vividly described this challenge: "Previously, when welding these components, engineers had to repeatedly adjust robot parameters based on drawings; switching to a new batch with different components meant spending half a day re-teaching the robot." In industries such as shipbuilding, elevators, and heavy steel, small-batch, multi-variety

production is the norm. The traditional industrial robot's one-task-one-function limitation drastically reduces line changeover efficiency and requires significant manual effort for parameter tuning.

1.3 The Breakthrough Value of Embodied Intelligence

The emergence of embodied intelligence technology offers a novel solution to this challenge. Embodied intelligence emphasizes that intelligent agents perceive the world, make decisions, and execute actions through physical interactions between their bodies and the environment, forming a closed-loop cognitive framework of "perception-decision-execution." When applied to welding robots, this means the robot is no longer merely a mechanical device passively executing pre-programmed instructions, but rather an intelligent welding partner capable of "seeing" workpiece geometry, "understanding" welding requirements, and autonomously planning process paths.[1,2]

In early 2026, Boqing Technology launched the first industrial-grade embodied AI agent in the smart welding industry. Its chairman, Feng Xiaobing, used a vivid analogy to explain its essence: "If traditional welding robots are 'limbs,' then the embodied AI agent is equipping them with a 'brain' and 'nervous system.'" In May of the same year, Xiaoyu Intelligent Manufacturing signed a strategic cooperation agreement with China Construction Science & Technology Group's Intelligent Technology Division, planning to jointly promote the large-scale application of 1,000 embodied intelligent welding robots within one year—a milestone marking the official entry of this technology into the industrialization and commercialization phase.

1.4 Research Objectives and Framework

This paper focuses on the application of embodied intelligent welding robots in construction steel structure scenarios, aiming to systematically elaborate on their technical connotations, core architecture, product forms, practical cases, and development trends. Section 2 discusses the paradigm shift from "teach-and-learn" to "embodied intelligence" from a technological philosophy perspective, and provides a detailed analysis of key technical modules. Section 3 delves into the core challenges in welding scenarios for construction steel structures. Section 4 categorically examines the technical characteristics and applicable scenarios of three types of embodied intelligent welding robots. Section 5 presents implementation cases and performance data from multiple enterprises. Section 6 discusses key technical challenges and future development trends. Section 7 summarizes the entire paper and offers forward-looking insights.

2 Technical Connotation and Architecture of Embodied Intelligent Welding Robots

2.1 From "Teaching and Reproduction" to the Paradigm Shift of "Embodied Intelligence"

To understand the innovative value of embodied intelligent welding robots, it is essential to first grasp their fundamental differences from traditional industrial robots. This difference represents not merely a technological upgrade, but a paradigm shift in manufacturing philosophy.

Traditional Paradigm: Programming-Execution Conventional industrial robots operate on a linear "programming-execution" logic. Engineers must manually guide the robot's end effector along the welding path for each workpiece and every weld seam using a teach pendant, recording trajectory points and setting process parameters. This method is highly efficient in assembly-line production, as the same program can be reused tens of thousands of times. However, at its core, this approach reduces the robot to a mere "program repeater"—it does not understand what it is doing, simply executing pre-defined actions mechanically. When the workpiece position deviates beyond acceptable limits or the joint gap changes, the robot still faithfully follows the original program, resulting in compromised weld quality.

Embodied Intelligent Paradigm: Perception-Decision-Execution Closed Loop. In contrast, embodied intelligent welding robots establish an entirely new technical framework based on a "perception-decision-execution" closed-loop system. At its heart lies a multi-modal perception system (including vision, laser, and tactile sensors) that continuously captures 3D information about the workpiece and environmental conditions. An AI large model acts as the "brain," analyzing sensory data to understand welding requirements and generating optimal welding parameters and motion trajectories. A high-precision robotic arm serves as the "body," carrying out these decisions. Meanwhile, real-time feedback during execution—such as arc signals, molten pool images, and positional deviations—is fed back into the perception and decision-making layers, enabling dynamic optimization within a closed loop.

The essence of this paradigm shift lies in transforming welding from "program repetition" to "task solving." The robot no longer cares about "what number

this weld has in the program," but instead interprets it as "this is a T-joint fillet weld between the flange and web of an H-beam, with a plate thickness of 16 mm and a joint gap of 2 mm, requiring a multi-layer, multi-pass welding process." This is precisely the "welding intelligence" that embodiment and artificial intelligence confer upon the machine.[3-5]

2.2 Core Technology Module

2.2.1 Multimodal Perception System: The Robot's "Eyes"

Perception is the starting point of embodied intelligence. Welding environments are harsh, characterized by high temperatures, intense light, smoke and dust, and electromagnetic interference, making it difficult for a single sensor to operate reliably under all conditions. Embodied intelligent welding robots typically integrate multiple sensing technologies to achieve "redundant complementarity."

3D structured-light/laser vision sensors are responsible for capturing the overall three-dimensional profile of the workpiece. By projecting structured light patterns and analyzing their deformation on the workpiece surface, these sensors can rapidly reconstruct 3D point cloud models of the workpiece, identifying weld positions, types, and geometric dimensions. [6] The seven-axis floor-mounted inverted robot developed by Xi'an Zhongke Optoelectronics employs a "long-range and high-precision scanning sensor," enabling accurate acquisition of workpiece pose and accommodating workpieces with plate thicknesses ranging from 3 to 25 mm and alignment errors not exceeding 20 mm.[7]

Laser profile sensors play a crucial role in seam tracking. By projecting a laser line and capturing its reflection from the weld cross-section, they precisely calculate geometric parameters—such as gap width, bevel angle, and misalignment—using triangulation principles, providing quantitative data for adjusting process parameters.

Arc sensors operate continuously during welding. By detecting periodic variations in welding current and voltage, they can determine in real time the positional deviation of the torch relative to the weld centerline and make immediate corrections. The frequency of arc signals is typically correlated with welding speed, enabling millisecond-level position feedback that compensates for potential failures of vision sensors in dusty or smoky environments.

Pool monitoring cameras represent a newly emerging sensing technology. China Wuye's self-developed high-altitude welding robot integrates a proprietary molten pool camera and real-time monitoring system, capable of accurately detecting subtle deviations such as weld offset and current fluctuations that are imperceptible to the human eye, thereby enabling both construction and inspection to be completed in a single pass.[8]

2.2.2 Welding Process Big Model: The Robot's "Brain"

If perception enables robots to "see" workpieces, then process decision-making allows them to "understand" how to weld. Traditional robots rely on engineers inputting welding parameters based on experience, whereas embodied intelligent welding robots are equipped with AI large models that integrate vast amounts of process knowledge.

The Jiangsu Beiren BR-iWeld welding process expert database is a representative achievement in this field. By integrating multidimensional knowledge from materials science, welding metallurgy, and process parameter optimization, the system can automatically generate matching welding plans—covering parameters such as current, voltage, speed, oscillation amplitude, and interpass temperature control—after identifying the workpiece type, and further adjust these parameters in real time based on actual welding results.

Boqing Technology's "Jiluan" industrial-grade welding large model takes this even further. According to information released in February 2026, the model has constructed a welding knowledge graph encompassing millions of data points and a training metadata system comprising tens of millions of records, enabling intelligent generation and real-time optimization of welding processes. Verified in industrial settings, it reduces welding process development cycles by 60% and lowers overall production costs by over 40%.

From a technical standpoint, training a welding process large model requires extensive datasets linking "operating conditions-parameters-quality." Traditionally, this knowledge is scattered across experienced technicians' expertise, process manuals, and experimental records, making systematic organization difficult. The key contribution of the large model lies in encoding this fragmented knowledge into a computable and inferable parameter space, allowing it to not merely "retrieve" the most similar historical case when facing new conditions, but instead "reason" out the optimal parameter combination.

2.2.3 Intelligent Trajectory Planning System: The "Bridge" Connecting Brain and Body

Trajectory planning is the critical link connecting the

"brain" to the "body." Traditional robots rely on manually programmed sequences of fixed waypoints, whereas embodied intelligent welding robots autonomously plan optimal welding paths based on 3D models of workpieces and weld seam recognition results.

Trajectory planning must consider multiple constraints: geometric constraints—ensuring the robotic arm reaches each weld point in an optimal posture within its reachable workspace while avoiding collisions with the workpiece or fixtures; process constraints—arc initiation position, welding direction, oscillation motion, arc termination, and other parameters must comply with welding specifications; and efficiency constraints—the path should be as short as possible to minimize idle travel time, subject to satisfying the first two conditions.

According to product documentation from Xi'an Zhongke Optoelectronics, their embodied intelligent welding robot features a complete set of functions including autonomous path planning, arc tracking, pulse welding, circumferential welding, oscillatory welding, intermittent welding, corner wrapping, and multi-layer multi-pass welding. This means the robot can independently plan and execute welding operations—from simple fillet welds to complex multi-layer, multi-pass groove welds—without requiring manual teaching for each individual pass.

2.2.4 Human-Machine Collaborative Interaction Interface: From "Replacement" to "Collaboration"

Embodied intelligence emphasizes collaboration between humans and machines rather than replacement. Currently, mainstream collaborative welding robots typically employ teach-by-demonstration technology—workers can manually guide the robot's end effector, "pointing where to weld" just like using a handheld welding torch, while the robot automatically records the path and optimizes process parameters without requiring any coding.

Qiao Zhongliang, founder of Xiaoyu Intelligent Manufacturing, pointed out when discussing technology implementation: "When we apply general embodied intelligence to complex industrial environments, we deeply feel a significant gap—a natural barrier between sophisticated algorithm demonstrations in the lab and the demands of factories for round-the-clock, highly reliable, and low-cost operations." This insight reveals the deeper logic behind human-machine collaboration: embodied intelligence is not about replacing humans in the lab, but about assisting and empowering them in real industrial settings—liberating welders from physically demanding tasks and transforming them into equipment managers and process optimizers.[9,10]

2.3 System Integration Architecture of Embodied Intelligent Welding Robots

Integrating the above modules, the system architecture of embodied intelligent welding robots can be abstracted into three layers: Perception Layer: Composed of 3D structured light cameras, laser profile sensors, arc sensors, and molten pool monitoring cameras, this layer is responsible for real-time acquisition of workpiece geometry, weld seam position and cross-sectional parameters, and welding process status data.

Decision Layer: Consists of a large-scale welding process model, motion planning algorithms, and a digital twin system. Preprocessed perception data are fed into the large model, which outputs welding process parameters and trajectory planning schemes. Meanwhile, the digital twin system performs virtual simulation and verification to anticipate potential process issues and equipment failures.

Execution Layer: Includes multi-axis robotic arms, welding power sources, wire feeding mechanisms, and external axes (floor tracks/positioning devices). This layer receives instructions from the decision layer to carry out welding tasks, while simultaneously transmitting real-time feedback—such as positional deviations, current and voltage fluctuations, and molten pool images—back to the perception and decision layers, enabling closed-loop optimization.

3 Core points in steel structure welding scenarios for construction

3.1 Small-Batch Multi-Variety and Flexible Manufacturing Requirements

The most prominent characteristic of building steel structure fabrication is that "every project is unique". A single super high-rise building may require thousands of different types of steel components, including H-shaped steel columns, box girders, bracing trusses, connection joints, embedded parts and more. Each component varies in cross-sectional dimensions, plate thickness, weld configuration and fillet weld leg size. Even within the same project, component specifications for different floors may be adjusted to accommodate changes in structural load-bearing forces.

This multi-variety, small-batch production model stands in stark contrast to mass production in automotive manufacturing, where tens of thousands of identical standard parts are produced in a single run. For traditional teach-pendant programming robots,

switching to a new type of component entails hours of downtime for commissioning: re-teaching motion paths, resetting process parameters, and revalidating programs. Factories are forced to choose between frequent production halts caused by robotic deployment and full reliance on manual welding. In practice, most steel structure manufacturers opt for the latter. While manual welding delivers lower efficiency and inconsistent quality, it eliminates the need to develop dedicated programs for every new component.

A typical case encountered by Jiangsu Beiren in 2023 fully illustrates this dilemma: a well-known domestic special steel structure enterprise faced a situation where "the assembly state, material thickness and weld configuration of components varied from batch to batch". Conventional robots could not adapt to such variable conditions at all, forcing the enterprise to rely entirely on manual welding, which brought low production efficiency and uncontrollable welding quality.[11]

3.2 Large Span of Component Dimensions and Reachability Constraints

Building steel components cover an extremely wide size range. Small connecting fittings are less than one meter long, while giant steel columns for super high-rise buildings can reach tens of meters with a single unit weighing over one hundred tons. Such a broad spectrum of sizes imposes stringent spatial coverage requirements on welding equipment.

Meanwhile, steel components feature highly complex and diverse weld positions, including open flat fillet welding zones as well as confined spaces such as the inner cavities of narrow box partition plates and stiffener areas at column bases. The technical team of MCC Group No.5 Engineering Co., Ltd. pointed out: "High-altitude steel structure operations have long been a persistent headache for the construction industry. Workers carry heavy loads while working hundreds of meters above ground. Their protective suits become stifling in sweltering summer heat, and operational accuracy drops amid cold winter winds, with potential safety hazards lingering at all times." Restricted by fixed bases and limited arm spans, traditional industrial robots often fail to reach target positions or operate in awkward postures in these complex working scenarios.

Furthermore, environmental conditions at construction sites are far less controllable than those inside workshops. Combined factors including fluctuating lighting, wind and rain interference, cramped spaces and high-altitude work create greater challenges for the deployment and operation of automated equipment.

3.3 Consistency of Weld Quality and Defect Control

Welding quality of steel structures is directly tied to building safety. Any defect in a weld, whether porosity, slag inclusion, incomplete fusion or cracks, can become the initiation point of structural failure under high-stress conditions. National standards impose strict requirements on the quality grades of steel structure welds, and Class I welds must pass 100% non-destructive testing. Nevertheless, the quality of manual welding heavily relies on welders' proficiency, mental state and physical stamina. Even top-tier welders cannot maintain perfectly consistent welding quality throughout a full workday. Particularly under harsh working conditions such as high temperatures and high altitudes, human fatigue and carelessness are far more likely to trigger quality fluctuations, constituting a systematic quality risk plaguing the steel structure industry.

While conventional automated welding delivers more consistent quality than manual welding, it is premised on consistent workpieces and high assembly precision. Fabrication assembly of building steel components usually carries substantial cumulative errors, which cannot be accommodated by the rigid, scripted execution mode of traditional robots. Embodied intelligent welding robots make up for this shortcoming to a considerable extent through real-time perception and adaptive adjustment. Instead of welding strictly following pre-set programs, they weld by visually sensing workpieces and adjusting parameters proportionally according to detected deviations.[12]

3.4 Structural Labor Shortage and Cost Pressure

The labor crisis in the building steel structure industry is intensifying. On the one hand, welding work features harsh working conditions, heavy labor intensity and high occupational health risks, making the industry unattractive to young people in general. On the other hand, cultivating skilled welders takes several years, and veteran experienced welders are retiring at an accelerating pace. Industry surveys show that the average age of welders in most steel structure manufacturers exceeds 45 years old, creating a severe generational gap.

Meanwhile, labor costs keep rising steadily. In economically developed regions, the monthly salary of a proficient welder has exceeded 10,000 RMB and is still on the rise. Enterprises are caught in a dilemma between failing to recruit enough workers and being unable to afford labor expenses. The emergence of embodied intelligent welding robots offers a technical solution to

resolve this structural contradiction. Rather than simply replacing human workers with machines, the technology transforms welders from manual laborers holding welding torches into technical operators managing multiple robots, achieving a dramatic efficiency boost through the "one operator, multiple machines" model.

4 Technical Characteristics and Application Adaptability of Three Types of Embodied Intelligent Welding Robots

In response to the diverse demands of building steel structure scenarios, three product forms of embodied intelligent welding robots have been developed in the industry, each with distinct strengths and complementary to one another.

4.1 Seven-Axis Overhung Intelligent Welding Robot with Ground Rail

4.1.1 Technical Feature

The seven-axis overhung robot with ground rail consists of a traveling ground rail (the 7th axis), an overhung manipulator arm and an intelligent control system. Compared with traditional six-axis industrial robots, the 7th axis (ground travel axis) enables the robot to move freely along the length of workpieces, greatly expanding its working envelope. The overhead suspended layout allows the robot to perform welding from above the workpiece, eliminating collisions between ground-mounted equipment and workpiece fixtures, delivering more flexible motion postures and superior reachability.

Xi'an Zhongke Optoelectronics Precision Engineering Co., Ltd. (its core technologies originate from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences) is a representative enterprise in this field. Its TR-iW series seven-axis overhung robots with ground rail adopt self-developed embodied intelligent architecture, equipped with manipulator arms from overseas brands (Yaskawa / Kawasaki / FANUC) or domestic brands (Inovance). The core functions include intelligent identification, automatic path planning, autonomous deviation correction and adaptive process parameter adjustment.

The technical parameters of this product series reflect the typical performance limits of seven-axis overhung ground rail robots: its working range spans from 6m×3m to 24m×3m, compatible with slender components such as H-beams and I-beams of varying lengths. It accommodates plate thicknesses ranging from 3 mm to 25 mm with assembly alignment errors within 20 mm. It supports various weld types including flat fillet welds, vertical fillet welds, lap fillet welds and groove butt welds. The robot integrates a full set of functions: autonomous path planning, arc seam tracking, pulse welding, circumferential welding, weaving welding, intermittent welding, corner wrapping welding and multi-layer multi-pass welding.[13,14]

4.1.2 Applicable Scenarios, Advantages and Limitations

Seven-axis overhung robots with ground rails are mainly suitable for batch welding of slender workpieces such as H-beams, I-beams and plate units, and applied to standardized production line scenarios inside workshops. A single-machine workstation can cover all welds of one type or a category of components; dual-machine collaborative models enable simultaneous welding on both sides, doubling production efficiency.

Its strengths lie in high precision, excellent stability and suitability for continuous mass production. As the manipulator arm is mounted on a fixed ground rail, the repeat positioning accuracy reaches sub-millimeter level and is immune to ground vibration interference. For bulk orders of components with identical specifications, its advantages in efficiency and consistency are extremely prominent.

Its limitations include large equipment footprint and fixed installation, resulting in insufficient flexibility. When the production line switches to completely different component types, although re-teaching is unnecessary (the embodied intelligent system can automatically identify and adapt to workpieces), physical constraints such as ground rail travel range and workpiece clamping methods may still require a certain amount of adjustment time.[15]

4.2 Intelligent Collaborative Welding Robot

4.2.1 Technical Features

Collaborative robots represent the optimal carrier for the embodied intelligence concept. Instead of simply replacing human workers, such robots prioritize human-robot collaboration, with three core design philosophies as follows: First, no safety perimeter fences required. Equipped with force perception and speed limitation functions, collaborative robots automatically stop or decelerate upon physical contact with operators. Unlike traditional industrial robots that demand physical isolation fences, they enable safe co-working with humans within a shared workspace.

Second, lead-through drag teaching with zero

programming. Workers can grasp the robot end-effector and drag it to target positions "just like operating an intelligent welding torch". The robot automatically records path waypoints and optimizes trajectory parameters without any manual coding throughout the process. This highly intuitive interaction drastically lowers operational barriers; ordinary welders can master the equipment after short-term training.

Third, lightweight design and rapid deployment. Collaborative robots generally weigh only dozens of kilograms and can be transported and installed by a single person. The flexible welding robot independently developed by CSCEC Steel Tech Intelligent Technology weighs merely 9.8 kg and can be fully deployed within 3 minutes.

4.2.2 Applicable Scenarios, Advantages and Limitations

Collaborative robots integrate flexibility and intelligence, boasting an extremely wide range of applicable scenarios: welding of small and medium-sized components in workshops, confined-space operations on box girders, box columns and round tubes at construction sites, flexible welding for multi-variety small-batch workpieces, as well as temporary repair welding and rework.

The most strategically significant application mode lies in clustered operation with the "one operator, multiple robots" model. For Tower C of the Shenzhen Bay Super Headquarters Base undertaken by CSCEC Steel Tech, the intelligent steel structure construction center was equipped with 10 flexible welding robots, which undertook on-site welding of approximately 7,000 tons of steel components, accounting for 90% of all column welding work on the project. The construction center, sized like a standard shipping container, expands into a movable intelligent capsule factory with a working radius of 35 meters. It can be hoisted to upper floors and towed to different positions, greatly facilitating vertical transportation and horizontal relocation on construction sites.

Its advantages include flexible deployment, easy operation and safe human-robot collaboration. However, it has limitations: the absolute productivity of a single unit cannot match large dedicated automated equipment. Operators need to master relevant operation and management skills, and its stability is inferior to dedicated machinery for continuous welding tasks involving extra-thick plates and ultra-long weld seams.[16]

4.3 Mobile (Track-Type) Intelligent Welding Robot

4.3.1 Technical Features

Tracked mobile welding robots mount welding manipulators on tracked chassis, enabling an operation mode where workpieces remain stationary while the robot actively travels to target welding positions. Equipped with lead-through drag teaching and programming-free intelligent systems, they can access work areas unreachable by fixed equipment such as gantry cranes and ground rail robots. Certain models are additionally fitted with magnetic wall-climbing functions, allowing them to crawl vertically on 90° walls and rotate 360°, delivering dead-angle-free welding across all vertical surfaces.

The high-altitude welding robot independently developed by MCC No.5 Engineering serves as a typical representative of this product category. Its magnetic wall-climbing technology allows easy access to high-difficulty and high-risk work sites including high-altitude zones and confined enclosed spaces. The integrated six-axis extra-long manipulator (1200 mm) achieves a positioning accuracy of 0.02 mm.

Patented technology from Jiangsu Bafang Steel Structure Group demonstrates another innovative development direction for this robot type: integrated ultrasonic weld inspection. An electric push rod delivers coupling agent, and the ultrasonic probe maintains stable contact with workpieces via spring preloading, realizing integrated "welding-inspection" operation and eliminating the need for separate manual inspection procedures.

4.3.2 Applicable Scenarios, Advantages and Limitations

Tracked robots are mainly applied to on-site welding for equipment maintenance, field assembly of heavy steel structures, on-site fabrication of non-standard workpieces, and repair welding at scattered flexible workshop stations. They boast irreplaceable advantages in scenarios without fixed tooling, such as bridge construction sites and large-scale equipment installation sites.

For the steel tower fabrication of Zhangjinggao Yangtze River Bridge, China Railway Baoji Bridge (Yangzhou) developed mobile workstations with tracked collaborative robots and electrically driven lifting welding platforms to address complex welds including curved seams on diaphragm units and corner welds of steel shells, realizing fully automatic welding. During the steel box girder fabrication of Xiamen-Jinmen Bridge, aiming at the massive quantity and highly variable positions of U-rib notch welds, the company deployed a lightweight electrically driven sliding rail welding system integrated

with AI vision weld recognition, which enables continuous automatic welding of multiple notches. Relevant research projects have been appraised and identified as internationally advanced by China Highway Construction Industry Association.

Its strengths lie in all-terrain mobility, powerful on-site adaptability and multi-purpose functionality. Its limitation is that the load capacity of the chassis restricts the size of the manipulator, resulting in relatively limited continuous operation capacity for extra-thick plates and ultra-long weld seams.[17]

5 Practical Cases and Efficiency Analysis

5.1 Jiangsu Beiren: "One Intelligent Brain, Multiple Hardware Forms" Solution for Steel Structure Welding

In 2023, a well-known domestic special steel structure enterprise raised a challenging demand to Jiangsu Beiren: the special steel structural components it produces differ in assembly state, material thickness and weld type from batch to batch. Traditional robots failed to adapt to such varied workpieces, and manual welding suffered from low efficiency and unstable welding quality.

Centered on the closed-loop "perception-decision-execution" architecture, the R&D team of Jiangsu Beiren developed embodied intelligent welding robots equipped with the BR-iWeld welding process expert database. After half a year of commissioning and iterative optimization, the first unit was officially put into operation. When faced with diverse welding requirements, the robot requires no manual teaching; it only generates exclusive welding schemes via vision and laser scanning.

Of greater strategic significance is the technical concept of "One Intelligent Brain, Multiple Hardware Forms" proposed by Jiangsu Beiren. A single set of intelligent systems is compatible with multiple robot forms including tracked mobile, gantry, cantilever and ground rail types. This upgrade shifts equipment from dedicated to general-purpose scenarios and breaks the limitation of "one robot for one single task" inherent to conventional robots.

The welding scenario of elevator car platforms has verified this technical concept. The dual-robot dual-station setup runs continuously for 24 hours, compatible with car platforms of various specifications and doubling production efficiency. Lin Tao, Vice President of Jiangsu Beiren, stated: "This technical architecture not only breaks the traditional restriction of 'one robot for a single dedicated task', but also upgrades robots from simple production tools to intelligent collaborative partners."

5.2 CSCEC Science and Industry: From In-house R&D Institute to Pioneer of Intelligent Welding Industry

CSCEC Steel Tech Intelligent Technology originated from the Intelligent Manufacturing Research Institute of CSCEC Steel Structure and was registered as an independent legal entity in 2025. This transformation marks a landmark event in the intelligent upgrading of the building steel structure industry. By commercializing its internal R&D capabilities, the large construction enterprise signals that intelligent welding technology has shifted from research exploration to industrial promotion.

In May 2026, Xiaoyu Intelligent Manufacturing signed a strategic cooperation agreement with CSCEC Steel Tech Intelligent Technology. The two parties plan to jointly advance the large-scale application of 1,000 sets of embodied intelligent welding robots in the building steel structure sector within one year. The core technology of Xiaoyu Intelligent Manufacturing adopts the "One Brain, Multiple Forms" architecture: a highly generalized embodied intelligent brain can adapt to and control robot bodies of various forms for diverse scenarios. Its founding team is led by Qiao Zhongliang, former head of MIUI R&D at Xiaomi, and Wang Wenlin, former general manager of Xiaomi's Software System Platform Department, with core team members recruited from leading tech enterprises including Xiaomi, ByteDance and Huawei.

This collaboration forms a closed industrial ecosystem integrating technology supply and on-site implementation. Embodied intelligent algorithms rely on data support and verification from real industrial scenarios, while building steel structure enterprises are in urgent need of intelligent technologies to resolve labor shortages and inconsistent quality issues. The two sides complement each other and achieve win-win results.

5.3 China Railway Baoji Bridge: Reverse-Driven Technological Innovation Oriented to Bridge Scenarios

Welding quality standards for bridge steel structures are even stricter than those for building steel structures. Bridges bear dynamic and fatigue loads, making the fatigue performance of welds critically important. Targeting the characteristics of bridge steel structures, China Railway Baoji Bridge (Yangzhou) has developed a variety of derivative equipment based on embodied intelligent welding technology.

During the fabrication of steel towers for

Zhangjinggao Yangtze River Bridge and steel shells for Guanyinsi Yangtze River Bridge, the enterprise addressed the pain points of stringent surface quality requirements for welds and low automation rates on steel tower and steel shell products. It developed mobile workstations equipped with tracked collaborative robots and electrically driven lifting welding platforms, which realized automatic welding of curved welds on diaphragm units and corner welds of steel shells. Two invention patents have been granted: An Automatic Welding Method for Cable Tower Steel Shell Blocks with Double-Layer Dense Grid Stiffened Wall Panels and An On-site Automatic Welding Method for Super-Large Section Steel Bridge Towers.

For the fabrication of steel box girders of Xiamen-Jinmen Bridge, in response to the large quantity, densely distributed and highly variable welding positions of U-rib notch welds plus rigorous quality requirements, the company developed a lightweight electrically driven sliding rail welding system integrated with AI vision weld recognition for steel box girder U-rib notches, enabling continuous automatic welding of multiple notches. This achievement won the Third Prize of Jiangsu Provincial Science and Technology Innovation Achievement Transformation Award. The relevant research project Automatic Welding Construction Technology of Portable Robots for Steel Box Girders has been appraised and certified as internationally advanced by the China Highway Construction Industry Association.

5.4 MCC No.5: High-Altitude Magnetic Adsorption Robots and Intelligent Construction Ecosystem Cluster

The high-altitude welding robot independently developed by MCC Group No.5 stands out with its proprietary magnetic wall-climbing technology. Capable of crawling vertically on 90° walls and rotating 360°, the robot can easily access high-difficulty and high-risk work sites such as high-altitude areas and confined enclosed spaces to achieve dead-angle-free welding on all vertical surfaces. Its built-in six-axis extra-long manipulator with a 1200 mm working stroke delivers a positioning accuracy of 0.02 mm.

In terms of efficiency improvement, the robot achieves remarkable breakthroughs. Conventional manual high-altitude welding requires operators to wear safety harnesses and conduct segmented welding, which comes with low productivity and high safety risks. In contrast, the robot supports continuous operation and completely eliminates the hazard of workers falling from heights. Multiple robots can work simultaneously on site under the control of a single operator. Its working efficiency can be up to four times higher than manual welding, cutting overall labor costs by 80%.

This high-altitude welding robot has passed practical verification on multiple key projects including Chengdu Gaotou Xinguang Intelligent Manufacturing Park and Tongliang Station of Chengdu-Chongqing Middle Line High-speed Railway, accelerating the transformation of construction welding operations toward intrinsic safety.[18]

5.5 Summary of Efficiency Data

Synthesizing the above practical cases and industrial statistics, the efficiency improvements brought by embodied intelligent welding robots in building steel structure scenarios can be quantified into the following indicators:

Table 1 Efficiency Data

Efficiency Indicators	Typical Data	Case
First Pass Welding Qualification Rate	≥99.5%	CSCEC Science and Industry Corporation LTD
Reduced Process Development Cycle	60%	BOTSING Technology
Comprehensive Production Cost Cut	above 40%	BOTSING Technology
Multi-operator Multi-robot Efficiency Improvement	One operator controls 3 to 10 robots	MCC No.5、CSCEC Science and Industry Corporation LTD
Sharp Reduction in Overall Labor Costs	80%	MCC No.5
High-Altitude Welding Positioning Accuracy	0.02mm	MCC No.5
Welding Automation Rate	80%	Qingdao East Steel Tower
Enterprise Production Efficiency Improvement	20%（Full-Process Digitalization）	Qingdao East Steel Tower

6 Key Technical Challenges and Development Trends

6.1 Sensor Fusion and Environmental Adaptability

The perception capability of current embodied intelligent welding robots performs well in standardized workshop environments. However, visual sensors suffer severe stability challenges under harsh construction site conditions such as complex illumination, welding fumes and electromagnetic interference. As summarized by Shi Yonghua et al. in a review published in Mechanical & Electrical Engineering Technology, improving sensing accuracy and realizing real-time multimodal data fusion is the key to transforming welding from the "teach-and-playback" mode to an adaptive, anthropomorphic and intelligent mode.[19]

The future direction lies in further deepening multimodal sensor fusion: information from vision,

LiDAR, arc sensing, tactile sensing and other sources will achieve redundant complementarity via AI algorithms to enhance the environmental robustness of the system. The technical team of China MCC5 has put forward a plan to develop acoustic models, which imitate veteran welders' ability to judge welding conditions by sound. This represents a valuable exploration for expanding multimodal perception from vision-dominated mode to audio-visual fusion.

6.2 Lightweight AI Models and Real-Time Inference

Welding process large models feature massive parameter sizes, leading to latency issues during cloud-based inference. Meanwhile, welding operations impose extremely high real-time requirements with millisecond-level response. Distilling large models into lightweight edge models that can be deployed on the robot's local controller while maintaining inference accuracy and speed constitutes a critical technical bottleneck for industrial application.

Academic research indicates that the core technologies of intelligent welding will focus on two major directions: deep multi-sensor fusion based on vision technology, and embodied intelligent welding robots driven by large AI models. The ultimate goal is to advance fully autonomous intelligent welding systems capable of autonomous perception, recognition, decision-making and execution.

6.3 Digital Twin and Predictive Maintenance

By constructing a full-element virtual mirror of the welding process, digital twin technology enables process simulation, parameter optimization and defect prediction in virtual space, drastically cutting the physical trial welding cost. The embodied intelligent agents of BOTSING Technology have been connected with digital twin systems and AI algorithms to predict equipment operating conditions and carry out predictive maintenance, eliminating the reliance of traditional welding on manual experience.

BOTSING Technology has also put forward the concept of "Robot-defined Super Factory". Based on embodied intelligence technology, integrated with industrial Internet and digital twin systems, a collaborative cluster of multi-type and multi-functional welding robots is constructed to realize three major upgrades: from single-point operation to collaborative decision-making, from single function to full-process coverage, and from passive maintenance to active optimization.

6.4 From Single-Robot Intelligence to Cluster Intelligence

With the advancement of CSCEC Steel Tech's large-scale application plan for up to 1,000 robots, multi-robot collaboration and cluster scheduling will become an inevitable trend. Each robot acts not only as an independent welding unit but also a node in the data network. Welding parameters, quality data and equipment status are uploaded to the cloud in real time. Continuous optimization of process models via big data analysis generates a swarm intelligence effect where the system grows smarter with operation.

China MCC5 has taken the lead in this field by releasing China's first self-developed intelligent construction robot ecological cluster equipped with a project-level intelligent scheduling and management system for robot clusters, realizing coordinated scheduling of multiple types of robots.[16]

7 CONCLUSION

Embodied intelligent welding robots are reshaping the production logic of construction steel structure manufacturing. Fundamentally, this technology transforms welding operations from programmed playback to task-oriented solving, enabling robots to truly understand welding demands, optimize processes independently and perform autonomous execution. In terms of industrial value, it directly addresses the core pain points of the steel structure industry, namely small-batch, multi-variety production and labor shortages, and delivers scalable solutions for widespread promotion.

Three product configurations, namely seven-axis overhead gantry robots, intelligent collaborative robots and crawler mobile robots, match differentiated scenarios including mass workshop production, flexible on-site deployment and positioning on complicated terrain, jointly forming a full product matrix for construction steel structure applications. Practical data demonstrates that embodied intelligent welding technology can cut the process development cycle by 60%, reduce comprehensive costs by 40%, achieve a first-pass yield of over 99%, and multiply efficiency under the single-operator multi-robot mode, fully verifying its technical and economic benefits.

Looking forward to the 15th Five-Year Plan period, with the continuous advancement of large AI model technology, declining costs of multi-sensor fusion, and gradual improvement of industrial standard systems, the penetration rate of embodied intelligent welding robots in the construction steel structure industry is expected

to rise from less than 20% at present to more than 50%. The practices of leading enterprises with 80% welding automation in the Jiaozhou steel structure industrial cluster prove that this target is well within reach.

This represents far more than an automation upgrade for welding procedures; it serves as a core pivot for the transformation of construction steel structure manufacturing from a labor-intensive industry to a data-driven industry. When robots can complete 90% of high-risk operations, workers will shift from manual laborers to technical professionals – transforming from welders holding torches into technical supervisors managing multiple robots. This delivers not only economic gains via industrial upgrading but also profound social value centered on putting people first.□

REFERENCES

[1] HE Y. Xiaoyu Intelligent Manufacturing Reaches Embodied Intelligence Strategic Cooperation with CSCEC Steel Tech Intelligent Technology[N]. Beijing Business Daily, 2026-05-09.

[2] Strategic Cooperation Agreement Signed Between Xiaoyu Intelligent Manufacturing and CSCEC Steel Tech Intelligent Technology to Promote the Application of Thousand-Unit Welding.

[3] HUANG S. Jiangsu North Human: "One Brain, Multiple Forms" Welding Robots Come into Reality[N]. People's Daily Online, 2025-10-21.

[4] Glimpsing the Future of Green Intelligent Manufacturing for Steel Structure Buildings from a Single Weld Seam[N]. China Metallurgical News, 2026-03-10.

[5] ZHANG Y. Intelligent Welding Technology Realizes Human-Robot Collaboration for Large Bridge Construction[N]. Science and Technology Daily, 2026-03-18.

[6] "Tech Brain" Vitalizes Traditional Industries, Platform Economy Restructures Iron and Steel Ecosystem[N]. Qingdao Bureau of Industry and Information Technology, 2026-05-26.

[7] SHI Y H, LING L J, ZHANG J Y, et al. Research and Application Progress of Intelligent Vision Sensing Technology for Robotic Welding[J]. Mechanical & Electrical Engineering Technology, 2025, 54(12): 1-23.

[8] Enterprises in Beijing Economic-Technological Development Area Released the First Industrial-Grade Embodied Intelligence Agent for Intelligent Welding Industry[N]. Beijing Youth Daily, 2026-02-06.

[9] Xi'an Zhongke Optoelectronic Precision Engineering Co., Ltd. Seven-Axis Overhead Gantry Welding Robot TR-iW110D-M-X(S) [EB/OL].

[10] Xi'an Zhongke Optoelectronic Precision Engineering Co., Ltd. Overhead Gantry (Dual-Robot) Welding Robot TR-iTR21D[EB/OL].

[11] China Steel Construction Society. Research Report on Digital Transformation of Steel Structure Industry[R]. 2025.

[12] Boqing Technology. Technical White Paper on Jiluan Industrial-Grade Welding Large Model[R]. 2026.

[13] Jiangsu North Human Intelligent Manufacturing Technology Co., Ltd. BR-iWeld Welding Process Expert Database Technical Manual[R]. 2024.

[14] China MCC5 Group. Technical Appraisal Report on High-Altitude Magnetic Wall-Climbing Intelligent Welding Robot[R]. 2025.

[15] China Railway Baoji Bridge (Yangzhou) Co., Ltd. Technical Appraisal Materials for Automatic Welding Construction Technology of Portable Robots for Steel Box Girders[R]. 2025.

[16] CSCEC Steel Tech Intelligent Technology Co., Ltd. Flexible Welding Robot Product Manual[R]. 2025.

[17] Qingdao East Steel Tower Co., Ltd. Construction Report of Digital Workshop for Intelligent Welding of Steel Structures[R]. 2025.

[18] FENG X B. Embodied Intelligence Agent: The "Brain" and "Nervous System" of Welding Robots[N]. China Industry News, 2026-02-10.

[19] QIAO Z L. From Laboratory to Factory: Barriers and Breakthroughs in the Industrial Implementation of Embodied Intelligence[J]. Robot Industry, 2026(3): 22-28.

Author Introduction: LI Yong, male, engineer, research direction: application of automation technology in construction engineering.

Copyright by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The Application of Digital Twin Technology Integrated with AI in Intelligent Oilfield Production

HOU Lei

Abstract:Based on actual engineering practices in oilfields, this paper systematically presents the architecture and operational mechanisms of integrating digital twin and AI technologies. Focusing on three core application scenarios—optimization of oil production schemes, predictive maintenance of oilfield equipment, and intelligent energy scheduling across the entire field—the paper deeply analyzes technical implementation pathways, methods, practical outcomes, and application advantages through real-world projects at Changqing, Shengli, and Daqing oilfields. Practical experience demonstrates that the integration of digital twin and AI technologies effectively addresses longstanding challenges in traditional oilfield operations, significantly enhancing precision management, reducing operational costs, minimizing equipment failure downtime, and optimizing energy consumption structures. This technology provides solid technical support for the digital transformation and intelligent upgrading of oilfields, improving quality and efficiency while enabling green and low-carbon production. It exhibits strong engineering practicality, implementability, and broad industry promotion value, offering reliable reference for intelligent retrofitting projects in similar oil and gas fields.
Key words:Digital twin; AI technology; smart oilfield; production optimization; equipment operation and maintenance; energy dispatching

1 INTRODUCTION

Under the dual strategic guidance of "Digital China" and "Smart Energy," traditional energy production models—characterized by high energy consumption, reliance on experience, and extensive operations—are undergoing fundamental transformation. As a core pillar of national energy security, the oil and gas industry has made digitalization and intelligent transformation an inevitable path toward high-quality development. With increasing years of field exploitation, most major domestic oilfields have gradually entered the mid-to-late stages of development, facing numerous operational challenges such as declining formation pressure, rising water cut, reduced single-well productivity, severe equipment aging, soaring maintenance costs, and persistently high energy consumption. The shortcomings and drawbacks of conventional management models are becoming increasingly evident.

Leveraging next-generation information technologies—including the Internet of Things (IoT), big data, 5G communication, digital twins, and artificial intelligence (AI)—to build smart oilfields is now the key technical pathway to address these challenges in mature fields. By establishing end-to-end data connectivity across geological exploration, drilling, oil production, gathering and transportation, equipment operation and maintenance, energy management, and safety control, a new production system driven by data, empowered by intelligence, and enabled by precise control can be established. In recent years, major domestic oil and gas fields have initiated intelligent upgrading and transformation programs. Digital twin and AI integration technology, serving as the core enabler for smart oilfield development, is progressively replacing traditional manual and experiential management approaches. Thanks to its advantages in virtual-physical integration, simulation-based forecasting, intelligent decision-making, and proactive prediction, this technology plays an irreplaceable role in optimizing oilfield operations and enabling fine-grained control.

For a long time, most domestic oilfield operations, equipment management, and energy dispatching have heavily relied on the on-site practical experience of frontline technicians and managers, resulting in a management model marked by strong dependence on experience, delayed responses, and coarse operations. During oil production, dynamic changes in reservoir parameters, real-time fluctuations in well conditions, and evolving block development scenarios make it difficult for human operators to accurately detect subtle changes. Adjustments to production parameters and optimization of production regimes often require multiple rounds of field trials and repeated tuning, leading to prolonged adjustment cycles, substantial consumption of labor, electricity, and materials, and frequent risks of unstable well performance, increased water content, and reservoir damage due to poorly adapted strategies—all of which negatively impact overall field development efficiency.

In equipment management, traditional fixed-interval maintenance schedules lack flexibility, while reactive emergency repairs are inherently delayed and unable to identify hidden faults or aging trends, making sudden equipment failures and production interruptions highly likely. In energy management, manual scheduling suffers from slow response times and rigid allocation, failing to adapt to dynamic production loads, thus causing persistent issues of energy waste and overconsumption. These problems severely hinder the efficient, refined, and sustainable development of modern oilfields, necessitating a comprehensive transformation of management practices through intelligent technologies.

Digital twin technology enables the precise creation of a virtual mirror of physical oilfield production environments, achieving real-time data synchronization and operational coordination between physical assets and their digital counterparts. When combined with AI-powered analytics, simulation, forecasting, and autonomous optimization capabilities, this integrated approach breaks through the technological bottlenecks of traditional management, enabling proactive prediction, real-time control, and post-operation review throughout the entire production lifecycle. Compared to conventional experience-based management, this fusion technology offers significant advantages including full-spectrum data integration, high simulation accuracy, intelligent decision-making, and strong implementation capability, effectively addressing various fine-grained management challenges in oilfield operations. Drawing on years of hands-on experience in oilfield engineering operations and technical management, this paper systematically outlines the core theoretical framework of digital twin and AI integration technologies. Focusing on practical implementation cases from major domestic oilfields, it conducts an in-depth analysis of the application pathways, outcomes, and optimization value of these technologies across three key scenarios: oil production, equipment operation and maintenance, and energy dispatching. The paper summarizes existing challenges and improvement directions in field applications, providing replicable and actionable engineering references for digital transformation, technology deployment, and enhanced operational efficiency in similar intelligent oilfields.[1,2]

2. Core Principles of Digital Twin and AI Integration

2.1 Core Architecture of Digital Twin Technology

Digital twin technology is a full lifecycle system that leverages core technologies such as IoT sensing, high-precision 3D modeling, 5G data transmission, virtual simulation, and edge computing to create a 1:1, high-fidelity digital replica of all physical assets in an oilfield—such as oil wells, pumping units, gathering and transportation equipment, pipeline networks, wastewater treatment stations, and production facilities. Its key technical advantage lies in establishing a four-dimensional closed-loop integration among physical entities, virtual models, real-time data, and business applications, effectively overcoming the traditional problems of isolated data, disconnected scenarios, and fragmented management in oilfield operations. By deploying numerous high-precision IoT devices—including pressure sensors, flow meters, temperature transmitters, vibration acquisition terminals, and video surveillance systems—on-site, the system continuously collects comprehensive operational data 24/7, including bottomhole pressure, wellhead flow rate, casing temperature, equipment rotational speed, vibration amplitude, pipeline pressure loss, and ambient temperature and humidity. Through 5G and industrial internet connectivity, this data is transmitted with low latency and synchronized in real time to cloud-based virtual models. Utilizing advanced 3D modeling techniques, the system reconstructs realistic scenes of terrain, equipment structures, pipeline layouts, and well pad distributions, enabling millisecond-level replication of actual field conditions. This allows for full-process visualization, dynamic simulation, historical state tracing, and scenario-driven inference, providing a precise, comprehensive, and real-time data foundation for subsequent AI-powered intelligent analysis and decision optimization.

2.2 The Enabling Role of AI Technology

A standalone digital twin model can only replicate physical conditions, visualize data, and trace production status—it represents a "passive perception" approach, lacking the capability for autonomous analysis of operational patterns, risk prediction, or production optimization. It cannot meet the demands of modern smart oilfields for autonomous decision-making, precise control, and proactive forecasting. Deep integration of AI algorithms fully addresses these limitations, enabling a complete upgrade across the entire workflow from "perception-analysis-prediction-decision-optimization." During implementation, machine learning, deep learning, big data clustering, and multi-objective

optimization algorithms are applied to analyze vast volumes of accumulated historical production data, real-time operational data, equipment maintenance records, and energy consumption metrics stored on the digital twin platform. These methods accurately identify oil production patterns under different reservoir blocks and operating conditions, uncover hidden features of equipment failures, detect correlations in energy consumption fluctuations, and track trends in subsurface parameters. Furthermore, through algorithmic iterative simulation, multi-scenario comparison, and intelligent parameter fitting, the system enables trend forecasting and global optimization for critical business scenarios such as production regimes, equipment operation states, energy distribution strategies, and maintenance scheduling. This fundamentally transforms traditional oilfield management practices—reliant on experience, reactive control, and trial-and-error approaches—into a proactive, intelligent decision-making framework. As such, it serves as the core technological pillar driving the digital transformation of smart oilfields.[3,4]

3. Digital Twin and AI Technology in Specific Engineering Applications for Smart Oilfields

3.1 Intelligent Optimization of Oil Production Schemes

Oil production is the core operation in oilfield management, and the scientific accuracy of production regimes, operational parameters, and water injection plans directly determines individual well productivity, reservoir recovery efficiency, and overall field development performance. Domestic oilfields feature complex and diverse geological conditions, with low-permeability fields, fault-block fields, mature fields, and high-water-cut fields coexisting. Significant variations exist across different blocks, formations, and wells in terms of formation permeability, porosity, reservoir pressure, and oil saturation, resulting in highly individualized characteristics. Traditionally, production planning and parameter optimization rely heavily on technicians' accumulated field experience combined with static geological data to establish fixed production schemes. These methods fail to adapt to dynamic changes in reservoir conditions, leading to poor generalizability, weak specificity, and low precision in manually designed production plans. To achieve optimal extraction conditions, multiple trial-and-error tests are often required at the site, repeatedly adjusting key parameters such as pumping unit stroke length, speed, production quotas, water injection intensity, and injection cycles—resulting in extremely high trial costs. Frequent on-site adjustments also disrupt stable production conditions, causing fluctuations in reservoir pressure, rising crude oil water content, and unstable production capacity. This not only wastes substantial amounts of electricity, water, labor, and resources but also shortens the period of stable production, severely impacting the economic returns of oilfield development.

By leveraging digital twin technology, a high-precision virtual model of the entire oilfield block can be established, enabling 1:1 replication of geological structures, wellpad layouts, equipment status, pipeline networks, and production conditions. This integrates comprehensive multidimensional data—including geological survey information, single-well formation parameters, historical production data spanning over five years, water flooding development records, operational logs, and maintenance data—into a complete digital foundation for oilfield operations. On this basis, an intelligent simulation and optimization system is built by embedding AI-based multi-objective genetic algorithms, operating condition fitting algorithms, and parameter iteration algorithms. Engineers no longer need to conduct on-site trials; instead, they can customize various combinations of production parameters within the virtual twin platform, simulating oilfield production under different production rhythms, water injection patterns, and allocation schemes. The system rapidly evaluates corresponding output metrics, including production rates, water cut trends, reservoir pressure changes, energy consumption levels, and extraction stability. The AI algorithm autonomously performs comparative analysis, iterative screening, and parameter fitting across thousands of operational scenarios, accurately identifying optimal parameter sets tailored to individual well characteristics. It automatically selects production plans that balance stable output, increased production, energy efficiency, and long-term sustainability, precisely aligning with reservoir oil production dynamics and energy evolution trends.

This intelligent simulation and optimization approach completely transforms the outdated practice of repetitive on-site trial-and-error and experience-based tuning. Without conducting blind physical experiments, multiple production schemes can be quickly iterated, compared, filtered, and optimized through virtual simulation, significantly reducing the debugging cycle. It effectively avoids waste of crude oil resources, power, labor, and

materials, stabilizes oilfield output, and substantially improves both single-well production efficiency and overall reservoir recovery rates. The technology is highly adaptable to various complex scenarios, including low-permeability wells, enhanced recovery in mature fields, high-water-cut well treatments, and development of complex fault-block reservoirs, demonstrating strong scenario compatibility and practical engineering value.[6]

Specific engineering application case: Optimization of oil production scheme in a low-permeability reservoir block at Changqing Oilfield

A major development block in Changqing Oilfield is located in the Ordos Basin and represents a typical "triple-low" reservoir characterized by low permeability, low porosity, and low abundance. The reservoir exhibits extreme heterogeneity, poor formation connectivity, rapid pressure decline, and high difficulty in both production and stable output, making it a representative example of complex and inefficient oilfields in China. Under traditional development models, production plans for this block were entirely based on engineers' experience, with core parameters such as pumping unit stroke rate, stroke length, allocated production quotas, and water injection intensity set uniformly without considering individual well differences. This lack of personalized optimization led to insufficient capacity utilization in many wells and a high proportion of low-yield, inefficient wells. To improve single-well productivity, field staff long relied on manual, repeated adjustments of production parameters, with each adjustment cycle lasting 7 to 15 days per well. During these periods, equipment was frequently started and stopped, and parameters changed often, consuming substantial labor and electricity resources while also easily causing operational instability, increased water cut, and imbalance in reservoir energy–further exacerbating inefficiencies. As a result, the overall recovery rate of the block has remained consistently below the industry average, severely limiting improvements in oilfield development efficiency.

To fully address the issues of inefficient development, high trial-and-error costs, and unstable production capacity in this reservoir block, a digital twin combined with an AI-based intelligent oil production optimization system was fully implemented in 2023. A high-precision digital twin virtual model covering all 236 producing wells was established, incorporating detailed geological parameters for each well–including formation permeability, porosity, original formation pressure, oil saturation, and reservoir thickness. The system integrated five years of production data, monthly water injection records, operational fluctuations, equipment operation logs, and maintenance history to build a dedicated large-scale database for reservoir production.

Leveraging this database, an AI multi-objective genetic optimization algorithm was deployed to create a simulation and forecasting model tailored for low-permeability reservoirs, which are characterized by rapid pressure decline, production sensitivity, and quick water cut increase. Engineers used the virtual platform to conduct multidimensional and multi-group simulation tests on 12 key production parameters–such as pumping unit stroke rate, stroke length, daily allocated production, water injection intensity, injection cycle, and shut-in time–completing over 1,200 sets of parameter combinations for comparative simulation. This enabled comprehensive analysis of how different parameter settings affect production, water cut, formation pressure, and energy consumption. Through deep learning of simulation data, the AI algorithm automatically iteratively optimized parameter combinations, identifying personalized optimal production regimes suited to the unique geological characteristics of each individual well, precisely aligning with reservoir oil production patterns and pressure trends.

After field implementation, the adjustment cycle for single-well production plans was reduced to just 1–2 days, eliminating blind on-site trial-and-error operations entirely. The block's average water cut decreased by 3.2%, average daily oil production per well increased by 8.6%, and overall recovery efficiency improved by 1.8 percentage points. Annual savings in electricity, labor, materials, and other production costs exceeded 1.2 million yuan. These significant improvements in quality and efficiency clearly demonstrate the substantial application value of this technology in complex oilfield development.[7-9]

3.2 Intelligent Predictive Maintenance for Oilfield Equipment

Core production equipment in oilfields–including oil well pumps, oil transfer pumps, compressors, gathering and transportation pipelines, and water injection systems–serves as the foundation for continuous and stable operations. The operational condition of this equipment directly determines the oilfield's production capacity and safety. These devices typically operate 24/7 under fluctuating loads, compounded by harsh outdoor environments involving wind, sand, humidity, temperature variations, and corrosion. As a result, frequent issues such

as wear, aging, fouling, and corrosion arise, leading to diverse failure types with complex root causes. For years, domestic oilfield equipment management has relied on a traditional model combining "scheduled maintenance" and "emergency repairs," which is coarse-grained and lacks precision.

Scheduled maintenance follows uniform cycles and standards without considering differences in equipment operating conditions, service life, or load intensity. This leads to unnecessary disassembly and servicing of otherwise healthy equipment, resulting in significant waste of labor, materials, and time. Meanwhile, latent faults such as bearing wear, motor degradation, minor pipeline blockages, and seal failures often go undetected until they cause sudden equipment shutdowns. Such incidents disrupt crude oil production and transportation, causing substantial output losses and posing serious safety risks including equipment burnout, pipeline leaks, and environmental compliance violations.

By leveraging digital twin technology, a 1:1 three-dimensional, high-fidelity model can be created for all core production equipment, establishing a virtual electronic ledger covering the entire equipment lifecycle. Comprehensive data—including factory specifications, installation and commissioning records, operational years, maintenance logs, fault history, and spare parts replacement details—are fully integrated, enabling digital, visual, and traceable equipment management. High-precision sensors—such as vibration, temperature, current-voltage acquisition terminals, and pressure modules—are installed at critical locations including pump bearings, motors, pump chambers, compressor inlets and outlets, and key pipeline nodes. These continuously collect dynamic operational data (e.g., vibration amplitude, housing temperature, working current, rotational speed, inlet/outlet pressure) around the clock, transmitting it via 5G networks to a cloud-based digital twin platform. This ensures millisecond-level synchronization between physical equipment status and its virtual model.

Combined with AI-based fault feature recognition algorithms and trend prediction models, real-time data is compared against historical normal operation data and known fault samples. This enables precise detection of subtle operational anomalies, effectively identifying hidden faults and aging trends that are difficult to spot during manual inspections. Potential risks such as bearing wear, motor overheating, pipeline clogging, seal deterioration, and abnormal current draw can thus be predicted in advance.

Based on AI analysis results, the system automatically generates optimized, tiered maintenance plans, accurately pinpointing fault locations, forecasting failure timing and severity, and distinguishing between routine maintenance, focused inspections, and emergency repairs. This marks a fundamental shift from reactive and scheduled maintenance to intelligent, predictive, and precise maintenance. On one hand, it proactively identifies latent faults, preventing unexpected outages such as pump failures, pipeline blockages, and transfer pump breakdowns, thereby avoiding production interruptions and ensuring continuous, stable operations. On the other hand, it eliminates ineffective, redundant, and excessive maintenance activities, significantly reducing material and labor costs, slowing equipment aging, extending the lifespan of core assets, and decreasing replacement frequency. Ultimately, it provides strong support for fine-grained, lifecycle-wide equipment management in oilfields.[10,11]

Specific engineering application case: Intelligent equipment operation and maintenance practice at a certain oil production plant in Shengli Oilfield

A coastal oil production plant in Shengli Oilfield is a key production area, currently operating over 320 producing wells and equipped with more than 460 sets of core equipment, including pumping units, oil transfer pumps, compressors, and gathering and transportation pipelines. Due to the long service life of these assets and their location in a coastal tidal flat region exposed to high humidity, high salinity, frequent sandstorms, and severe corrosion, problems such as equipment corrosion, wear, abnormal vibration, and seal failure occur frequently. The plant has long relied on traditional equipment management practices—primarily quarterly scheduled maintenance and reactive repairs after failures—which have significant drawbacks. On one hand, standardized, one-size-fits-all maintenance schedules ignore variations in equipment operating conditions and health status, leading to frequent disassembly and replacement of parts for otherwise stable and healthy equipment. This results in widespread ineffective and excessive maintenance, wasting substantial human resources, materials, and labor hours. On the other hand, latent faults cannot be predicted in advance; minor issues such as bearing wear, pipeline micro-blockages, and motor aging accumulate over time, eventually causing unexpected equipment shutdowns. According to operational records, prior to system upgrades, the plant

experienced an average of over 80 unplanned equipment outages annually, each lasting 4 to 8 hours. Each incident not only required extensive manpower and consumables for repair but also directly halted crude oil production, resulting in significant annual output loss and increased pressure on safety and operational control.

To fundamentally address the challenges of inefficient operations, high rates of unexpected failures, elevated maintenance costs, and substantial production losses, the plant fully deployed a digital twin-based AI-powered intelligent equipment operation and maintenance management system in 2022. It completed 1:1 three-dimensional digital modeling for all 460+ core equipment units and established an integrated digital platform for full lifecycle equipment management. The platform consolidates data from equipment factory documentation, installation and commissioning logs, historical maintenance records, fault registers, and spare part replacement history into standardized digital archives. Additionally, the plant upgraded its equipment sensing infrastructure by installing high-precision data acquisition terminals at critical points on key equipment such as pumping units, oil transfer pumps, and compressors. These terminals continuously collect vital operational parameters—including vibration, temperature, current, speed, and pressure—and leverage 5G industrial internet to ensure low-latency data transmission, enabling real-time synchronization between virtual models and physical equipment conditions.

The system incorporates a self-trained AI algorithm for fault feature recognition, which learns from nearly ten years of normal operation data, various fault samples, and degradation trends to build a predictive fault model tailored to the complex conditions of tidal-flat oilfields. This enables precise identification of subtle anomalies and hidden fault risks. The system can predict potential failures—such as bearing wear, motor overheating, pipeline blockage, and seal aging—7 to 15 days in advance, automatically generating accurate maintenance work orders and pushing them to maintenance personnel's devices, enabling "on-demand, precision maintenance." After two years of field application, the plant has reduced unplanned equipment failures by 65%, decreased unnecessary maintenance frequency by 58%, lowered average annual maintenance costs per unit by 30%, extended equipment lifespan by 2.5 years, and saved over 300 tons of crude oil production annually due to reduced downtime. While ensuring safe and stable production, the system has achieved significant cost reduction and efficiency improvements in equipment operations and maintenance.[12-15]

3.3 Intelligent Energy Dispatch Optimization for Oilfields

Oilfield production represents a typical high-energy-consumption, high-water-use, and high-resource-demand industrial scenario. The entire process—from oil lifting and crude oil gathering to water injection, wastewater treatment, equipment operation, and site heating—requires substantial support from electricity, industrial water, thermal energy, compressed air, and other resources. Energy consumption is widespread across numerous points, spans broad operational areas, features complex energy structures, and experiences significant dynamic load fluctuations. Traditional oilfield energy management relies primarily on manual data collection, fixed allocation, and post-event accounting, lacking capabilities for real-time monitoring, intelligent analysis, and proactive forecasting. As a result, energy dispatching remains highly inefficient.

In daily operations, energy quotas for each production unit remain fixed over long periods, unable to adapt dynamically to real-time production targets, equipment loads, or changing operating conditions. This often leads to insufficient energy supply during peak production periods, forcing critical oil extraction and water injection units to operate at reduced capacity and thereby compromising crude oil output. Conversely, during low-production periods or maintenance phases, auxiliary equipment runs idle, resulting in massive waste of electricity, water, and thermal energy. Additionally, manually collected energy data is often delayed and prone to errors, making it difficult to accurately identify high-consumption nodes and inefficient energy use. Consequently, energy control in oilfields has long remained reactive, with energy consumption per unit of oil produced remaining unacceptably high, failing to meet current policy requirements and production demands for green, low-carbon, energy-efficient, and high-quality development in the oil and gas industry.

By leveraging digital twin technology, a virtual model for comprehensive energy management across the entire oilfield can be established. This model integrates consumption, supply, and loss data for various energy types—including electricity, industrial water, thermal energy, and compressed air—and links them to real-time production loads, output targets, and operational statuses of individual production units, wells, and equipment. It builds a fully integrated, time-continuous,

and element-comprehensive energy data monitoring system. The model enables real-time visualization of energy flow paths, energy distribution patterns, and load fluctuation trends throughout the oilfield, achieving transparent and visible energy management.

Combined with AI-based dynamic scheduling algorithms, the system can analyze in real time the input-output ratios, equipment efficiency, and energy load characteristics of each production unit. Based on current operational conditions, it adjusts energy allocation quotas second by second, precisely matching power, water, and thermal energy to core production units that are under high load, high output, and high efficiency. Meanwhile, it automatically reduces or cuts off energy supply to idle equipment, inefficient units, and unused processes, minimizing unnecessary energy consumption and reducing transmission losses and wastage.

Furthermore, through continuous learning of historical energy consumption data, production load patterns, and seasonal variations, the AI algorithm can accurately predict peak energy demand and potential shortages under different production stages, seasons, and output requirements. This allows for the development of scientific, fine-grained energy dispatch plans in advance, enabling dynamic balancing of overall energy loads across the oilfield. By avoiding both peak overloads and low-load waste, the system effectively reduces total energy consumption, supporting smart oilfield initiatives toward energy conservation, emission reduction, and green, low-carbon development—fully aligning with national policies for sustainable transformation in the oil and gas sector.[16,17]

Specific engineering application case: Energy intelligent dispatching renovation project at a certain operation area of Daqing Oilfield

A key operational area of Daqing Oilfield is a comprehensive production zone integrating oil extraction, crude oil gathering and transportation, water injection development, wastewater treatment, and equipment operation and maintenance. The area hosts over a thousand types of production and auxiliary equipment, featuring long production chains, numerous energy consumption nodes, and high overall energy usage—making it a typical high-energy-consuming mature oilfield operation zone. For years, the area has relied on traditional manual energy dispatching methods, with rigid and fixed energy allocation schemes that fail to dynamically adjust energy supply based on monthly production plans, daily operational loads, or equipment operating conditions. During peak production periods, insufficient power and water supply to core units such as oil extraction, water injection, and gathering systems prevent equipment from running at full capacity, directly constraining crude oil output. Conversely, during low-production periods or equipment maintenance phases, numerous auxiliary devices, wastewater treatment systems, and supporting facilities operate under no load, resulting in significant energy waste. Additionally, manually collected energy data suffers from substantial delays, making it difficult to accurately identify high-consumption points and inefficiencies. As a result, the area's energy consumption per unit of oil produced has consistently exceeded the average for the entire oilfield, creating immense pressure to reduce energy use and meet green production targets set by the oilfield.

To address challenges related to inefficient energy management, high energy consumption, and resource waste, the operation zone completed the construction of a digital twin-based AI intelligent energy scheduling system in 2023. This system established a full-field digital twin simulation model, fully integrating energy consumption data from regional power, water supply, heating, and production equipment systems. It enables real-time collection, aggregation, analysis, and visualization of energy data across all energy types, production units, and time periods. The model precisely replicates the regional energy distribution networks, electrical layouts, and energy consumption node distributions, comprehensively linking production targets, equipment load levels, and operational conditions across all units to build a dynamic energy database. Based on this foundation, an AI multi-objective optimization scheduling algorithm model was developed, focusing on core objectives including ensuring production capacity, reducing energy consumption, improving efficiency, and stabilizing load. The system leverages deep learning techniques to analyze historical energy consumption data, seasonal patterns, and correlations between production capacity and energy use. It continuously monitors overall energy consumption trends and dynamically adjusts energy allocation quotas for each production unit within seconds, prioritizing energy supply to critical production units while shutting down idle or inefficient equipment to precisely minimize unnecessary energy consumption.

Meanwhile, by integrating AI algorithms with seasonal variations, monthly production plans, and equipment maintenance schedules, the system

proactively forecasts fluctuations in energy demand and formulates three-tier energy dispatch plans at the monthly, weekly, and daily levels, enabling proactive, intelligent, and fine-grained energy management. After implementation, the overall comprehensive energy consumption of the operation area decreased by 9.3%, industrial water utilization improved by 12.5%, and no-load energy losses from equipment were significantly reduced. The annual savings in various energy costs exceeded 1.8 million yuan, successfully meeting the green and energy-efficient production targets set by the oilfield and providing a proven technical model for the green and low-carbon transformation of mature oilfields.[18-20]

3.4 International Case Studies

3.4.1 BP Company – APEX Digital Twin Platform and AI-Driven Production Optimization

BP is a pioneer in digital transformation within the oil and gas industry, with its digital twin strategy centered on the APEX (Advanced Production Excellence) platform, built through over a decade of collaboration with Palantir Technologies. APEX is a real-time digital twin system that integrates data from more than 2 million sensors across BP's global oil and gas production facilities, creating a unified operational view. The platform enables real-time simulation of production system performance, allowing engineers to test adjustment scenarios in a virtual environment, predict operational outcomes, and proactively identify potential issues.

Building upon the APEX platform, BP developed an AI-enhanced tool called Optimization Genie in early 2024, which was fully deployed by September of the same year. Optimization Genie autonomously identifies and tests bottlenecks in production systems, simulates adjustment strategies, and recommends optimal configurations to engineers. The integration of AI has reduced capacity optimization analyses—previously requiring weeks of manual effort by engineers—to just a few hours.

Notable achievements include: at the Atlantis offshore facility in the U.S. Gulf of Mexico, Optimization Genie achieved approximately 2,000 additional barrels per day in production by optimizing well placement variables and prioritizing operations—all without incurring extra costs. In 2017, analysis from the APEX platform contributed an extra 30,000 barrels per day in output for BP. In exploration, BP leveraged AI to achieve nearly 97% reliability in upstream operations and a 98% accuracy rate in predicting drilling "kick" events, with plans to save $2 billion in costs by 2026.

Technical features: BP's approach reflects the evolution of digital twins from mere visualization to autonomous decision-making. Its technical architecture consists of three layers: the data layer (integrating over 2 million sensor data points), the analytics layer (using AI algorithms to detect bottlenecks and simulate adjustments), and the application layer (delivering actionable optimization recommendations). Human engineers retain final decision authority, establishing a human-AI collaborative model of "AI suggestion – human validation – closed-loop optimization."[21]

3.4.2 Shell – Deep Integration of Dynamic Digital Twins and Predictive Maintenance

Shell has adopted a technology path for digital twin applications that emphasizes dynamic simulation and predictive maintenance. At one of Norway's largest offshore gas fields, Shell has deployed a real-time dynamic digital twin system that integrates data, visualization, and dynamic modeling capabilities, enabling engineers to replicate field conditions in a virtual environment. Since its commissioning in 2007, the field has operated using two 120-kilometer multiphase gas pipelines and an underwater gas compression system. The introduction of the digital twin allows engineers to conduct "what-if" scenario analyses, benchmark models against real-time operational data, and anticipate risks before they impact production.

In equipment maintenance, Shell collaborated with Akselos to use digital twin technology for structural analysis of a 50-year-old platform in the southern North Sea, UK. This effort successfully identified critical upgrade needs, extending the platform's operational life by 20 years and avoiding substantial investment in building a new platform.

Key technological breakthrough: Shell also led the development of RTDIP (Real Time Data Ingestion Platform), an open-source platform for real-time data ingestion, which is made available to the entire industry through Linux Foundation Energy. Designed to address the "air gap" between OT (operational technology) and IT environments, RTDIP enables secure access to time-series and event data, providing a high-quality data foundation for AI analytics. Quantifiable results: The AI-driven digital twin solution developed jointly by Shell and Capgemini reduced analysis time for inefficient and optimization-targeted wells from seven hours to 1.5 hours, achieving production increases of 1–1.5% for onshore operations and 1–2% for offshore operations. Shell executives noted that combining real-time dynamic digital twins with predictive modeling enables engineers to detect

anomalies and optimize equipment parameters. Going forward, the company plans to further integrate AI-powered features to automatically flag key risks and recommend corrective actions.[22]

4. Technical Application Advantages and Engineering Value

Compared to traditional oilfield production management models dominated by manual experience and characterized by extensive, inefficient practices, the integration of digital twin and AI technologies has demonstrated remarkable technical advancement and engineering practicality through multi-scenario applications in smart oilfields. Its comprehensive advantages and engineering value are prominent and can be categorized into four core dimensions. First, it enables full-spectrum cost reduction and efficiency improvement, enhancing development profitability. By replacing conventional on-site trial-and-error operations with virtual simulation and scenario analysis, it significantly reduces waste of crude oil, electricity, labor, materials, and other resources. Precise equipment maintenance lowers operational costs, while intelligent energy scheduling minimizes unnecessary energy consumption. This holistic approach substantially cuts overall production and operation expenses, stabilizes and increases oilfield output capacity and recovery rates, and significantly improves overall economic returns from oilfield development. Second, it ensures precise and controllable production, breaking free from limitations imposed by subjective experience. By establishing a real-time linkage system between physical oilfield environments and virtual models, it enables visualization, data-driven, and intelligent management across geological conditions, production operations, equipment status, and energy consumption changes. Leveraging AI-powered big data analytics, it replaces traditional human-based, subjective judgment, greatly improving the accuracy and scientific rigor of production decisions, parameter optimization, fault handling, and energy dispatching. Third, it strengthens safety foundations and ensures stable operations. Through early prediction of equipment failures, real-time alerts for abnormal operating conditions, and virtual simulation of high-risk operations, it proactively mitigates various production safety hazards such as equipment breakdowns, operational imbalances, and excessive energy loads. This effectively reduces equipment downtime, production interruptions, and safety incidents, ensuring continuous, safe, and stable oilfield operations and elevating overall safety management standards. Fourth, it supports long-term iterative enhancement and drives industrial transformation. By continuously accumulating data throughout the entire oilfield production lifecycle—including operational, maintenance, and energy consumption data—on the digital twin platform, combined with ongoing AI algorithm optimization, it refines production models and decision-making systems. This creates a replicable, optimizable, and upgradable intelligent production management model, providing sustained technological support for the digitalization, intelligence, and green transformation of oilfields, thereby promoting high-quality development in the traditional oil and gas industry. As evidenced by three field implementation cases discussed above, this integrated technology is well-suited to meet the production modernization needs of diverse domestic oilfields, delivering significant results and offering exceptional engineering scalability and promotion value.[23-25]

5. Current Issues and Optimization

Considering the current status of intelligent transformation in major domestic oilfields, although digital twin and AI integration technologies have achieved notable success in smart oilfield applications, overall implementation remains in a transitional phase characterized by pilot projects and iterative refinement. Constrained by multiple factors—including oilfield construction age, hardware infrastructure, data systems, algorithm capabilities, and talent structure—large-scale deployment and high-precision applications still face significant shortcomings and bottlenecks, limiting the full realization of technological empowerment. First, older oilfields suffer from weak foundational data and incomplete sensing systems. Many key domestic oilfields were built decades ago with limited early-stage intelligent upgrades. The sensors and data acquisition terminals deployed on-site are outdated, low-precision, and unstable. In some remote well sites and edge production areas, sensor coverage is incomplete, creating blind spots in data collection. Moreover, frequent issues such as data transmission delays, packet loss, and data distortion from aging equipment result in untimely updates to digital twin models, causing slight discrepancies between virtual model conditions and actual field operations. This directly affects the accuracy of

AI-driven data analysis and intelligent decision-making. Second, data fragmentation across multiple systems creates prominent data silos. During oilfield operations, production systems, equipment management systems, energy consumption monitoring systems, safety surveillance systems, and geological databases were typically developed independently at different times and by different vendors. These systems lack standardized data formats, incompatible interfaces, and interoperability, leading to severe data fragmentation. As a result, a unified, comprehensive data ecosystem cannot be established, making it difficult to support high-precision, multi-dimensional simulation and intelligent optimization. Third, AI algorithms exhibit insufficient adaptability to complex operating conditions. Most intelligent algorithms currently used in oilfields are trained and optimized under conventional, stable production scenarios. There is inadequate training data for special or extreme situations such as sudden weather changes, abrupt formation parameter shifts, unexpected equipment failures, large-scale maintenance shutdowns, or abnormal operational fluctuations. Consequently, prediction accuracy, optimization performance, and adaptability degrade under complex conditions, indicating that algorithm robustness needs further improvement. Finally, there is a severe shortage of interdisciplinary technical talent. Digital twin and AI-based smart oilfield technologies span multiple disciplines, including petroleum engineering, reservoir development, IoT, 3D modeling, big data analytics, and artificial intelligence algorithms, demanding highly integrated expertise from professionals. Currently, frontline oilfield technicians primarily specialize in traditional oil extraction and equipment operation, lacking sufficient knowledge in digitalization, automation, and algorithmic modeling. Skilled digital professionals are scarce, and the development of cross-disciplinary teams lags behind, significantly hindering efficient system operation, deep application, and continuous iteration.[26,27]

To address the aforementioned challenges, and in line with the trends of oilfield digitalization and engineering application needs, future improvements to the technical system can be continuously advanced through four dimensions: equipment upgrades, data integration, algorithm optimization, and talent development, thereby promoting deeper implementation and widespread adoption of technologies. First, on-site sensing devices should be iteratively upgraded by fully replacing outdated and inefficient data acquisition terminals, addressing shortcomings in remote wellsite monitoring, and integrating 5G and edge computing technologies to enhance data accuracy, transmission speed, and stability, ensuring real-time and precise mapping of field conditions in digital twin models. Second, a unified oilfield data platform should be established to standardize data formats and interface specifications across all systems, breaking down data silos related to production, operations, energy consumption, geology, and equipment, enabling centralized storage, integrated sharing, and unified scheduling of data, thus providing a robust foundation for AI algorithm refinement and simulation analysis. Third, the intelligent algorithm framework should be optimized and iterated by expanding training datasets for complex geological conditions, extreme production scenarios, and sudden failures, refining model architectures to improve prediction accuracy, fault detection, and solution optimization in specialized situations, thereby broadening the range of applicable technical scenarios. [28,29,30]

6. CONCLUSIONS

The deep integration of digital twin and AI technologies represents a core technical pathway for the digital transformation, quality improvement, efficiency enhancement, and green development of smart oilfields in the new era. It has fundamentally broken the traditional, experience-based, extensive, and lagging production management models, effectively addressing key industry challenges such as high trial-and-error costs in oil recovery schemes, insufficient precision in production optimization, passive and delayed equipment maintenance, frequent unexpected failures, inefficient and unbalanced energy dispatching, and lack of scientific decision-making. This paper systematically reviews the core architecture, enabling mechanisms, and integration advantages of digital twin and AI technologies, and validates their application value and implementation effectiveness across three real-world projects: oil recovery optimization in low-permeability reservoirs at Changqing Oilfield, intelligent equipment maintenance at Shengli Oilfield, and smart energy scheduling at Daqing Oilfield. By constructing a high-precision virtual mirror that covers all elements, processes, and time sequences of an oilfield, and leveraging AI-powered big data analytics, simulation modeling, trend forecasting, and iterative optimization capabilities, this integrated technology enables intelligent iteration of oil production plans,

precise control of equipment operations, and scientific allocation of energy resources. It comprehensively enhances oilfield production efficiency, recovery rates, safety performance, and overall economic returns, providing solid technological support for maintaining and increasing output, reducing costs, and achieving green, low-carbon transformation in mature oilfields. The technology framework is adaptable to diverse production modernization needs across various types of domestic oilfields, featuring strong practicality, broad applicability, and significant efficiency gains, thus demonstrating exceptional engineering utility and wide-scale industrial promotion potential. Against the backdrop of digitalization, intelligence, and greening transformation in the energy sector, with continuous advancements in sensing devices, data platforms, and intelligent algorithms, the convergence of digital twin and AI will achieve full coverage and deep integration in smart oilfield development, gradually realizing unmanned operation, autonomous decision-making, and fine-grained control. This will comprehensively drive high-quality development in the oil and gas industry while offering replicable, implementable practical references and technical insights for frontline petroleum engineers and technicians engaged in intelligent oilfield construction and production management.□

REFERENCES

[1] Liu He, Yao Shanglin, Li Xin, et al. Reflections and Design Principles for Smart Oilfield Development [J]. Engineering Research–Engineering in an Interdisciplinary Perspective, 2025, 17(2): 243-258.

[2] Nie Xiaowei. Current Status and Development Trends of Key Technologies for Intelligent Oilfields [J]. Petroleum Geology and Recovery Efficiency, 2022, 29(3): 68-79.

[3] Liu Wenling, Han Dakuang. Digital Twin Reservoir: A New Direction for Smart Oil and Gas Field Development [J]. Acta Petrolei Sinica, 2022, 43(10): 1450-1461.

[4] Ren S S, Shen F, Zhang X Y, et al. Application and development of digital twinning in oil and gas exploitation and production: a focus on China and global perspectives[J]. Geoenergy Science and Engineering, 2025, 245: 213568.

[5] Yang Jianfeng, Du Jinhü, Yang Yong, et al. Research and Practice on Digital Transformation in the Oil and Gas Industry [J]. Petroleum Journal, 2021, 42(2): 248-258.

[6] Bigdeli A. The rise of digital twin: a new era for reservoir simulation and modeling[J]. Journal of Petroleum Exploration and Production Technology, 2026, 16(1): 45-62.

[7] Zhang Kai, Zhao Xinggang, Zhang Liming, et al. Research Status and Prospects of Big Data and Intelligent Optimization Theories and Methods in Smart Oilfield Development [J]. Journal of China University of Petroleum (Natural Science Edition), 2020, 44(4): 28-38.

[8] Liu He, Li Xin, Yao Shanglin, et al. Key Technologies for Intelligent Reservoir Injection-Production Control System [J]. Drilling & Production Technology, 2025, 48(1): 180–189.

[9] Lin Botao, Zhu Haitao, Jin Yan, et al. Construction Method and Application Cases of Digital Twin Models for Oil and Gas Drilling and Production [J]. Petroleum Science Bulletin, 2024, 9(2): 282–296.

[10] Du Jinhü, Shi Fugeng, Zhang Zhonghong, et al. Research and Practice on China National Petroleum Corporation's Exploration and Development Dream Cloud [J]. China Petroleum Exploration, 2020, 25(1): 58-66.

[11] Capgemini. Revolutionizing oilfield operations with AI-powered digital twin[R]. Paris: Capgemini Research Institute, 2025.

[12] Younger R. Agentic AI for oilfield optimization: bridging expert workflows and generative intelligence in multi-field facilities[R]. Abu Dhabi: ADNOC, 2026.

[13] Zeng Nannuo, Li Li, Li Jinsong, et al. Exploration and Practice of Digital Twin Construction for the "Shenhai No. 1" Smart Gas Field Production [J]. Petroleum Science Bulletin, 2025, 10(5): 945-958.

[14] Liang Wenfu. Achievements and Prospects in the Development of Intelligent Application Systems for Oilfield Development [J]. Daqing Petroleum Geology and Development, 2019, 38(5): 283-289.

[15] Lin Yang. Practices and Insights on Digital Transformation in Non-Digital-Native Enterprises: A Case Study of CNOOC's Smart Oilfield Demonstration Project [J]. Petroleum Drilling and Production Technology, 2021, 43(4): 552-558.

[16] Shell plc. Shell partners with NVIDIA to accelerate AI applications in energy sector[EB/OL]. (2024-03-18)[2026-06-20]. https://www.shell.com.

[17] Exxon Mobil Corporation. ExxonMobil and Microsoft accelerate digital transformation in oil and gas[EB/OL]. (2023-11-10)[2026-06-20]. https://corporate.exxonmobil.com.

[18] TotalEnergies SE. TotalEnergies leverages digital twin for integrated production optimization[R]. Paris: TotalEnergies, 2024.

[19] BP plc. BP advances digital transformation with AI-powered production optimization[R]. London: BP, 2024.

[20] Halliburton Company. Digital twin solutions for reservoir management[R]. Houston: Halliburton, 2023.

[21] Schlumberger Limited. DrillOps closed-loop drilling optimization platform[R]. Houston: Schlumberger, 2024.

[22] Equinor ASA. AI and digital twin applications in unmanned offshore platforms[R]. Stavanger: Equinor, 2023.

[23] Nie Xiaowei. Key Technical Research Directions and Prospects for Smart Oilfields [R]. Beijing: Research Institute of Petroleum Exploration and Development, Sinopec, 2023.

[24] Zhang Kai, Liu Wei, Zhao Xinggang. Research Progress on Dynamic Automatic Monitoring and Intelligent Control Technology in Production Sites [J]. Petroleum Drilling Techniques, 2021, 49(6): 1-9.

[25] Du Jinhü, Yang Jianfeng, Zhang Zhonghong. Intelligent Technology and Application of Next-Generation Industrial Big Data in Oilfields [J]. Petroleum Exploration and Development,

2022, 49(1): 178-188.
[26] Liu He, Li Xin, Yao Shanglin. Research Status and Prospects of Intelligent Optimization Technologies for Reservoir Development [J]. Petroleum Drilling and Production Technology, 2024, 46(2): 185-196.
[27] Zhang Kai, Zhao Xinggang, Zhou Yingcao. Architecture Design of a Cloud-Edge Collaborative Intelligent Injection and Production Control System [J]. Journal of China University of Petroleum (Natural Science Edition), 2025, 49(1): 112-121.
[28] Ren S S, Shen F, Zhang X Y. Challenges and countermeasures of digital twin data governance in China's oil and gas industry[J]. Energy, 2025, 312: 134256.
[29] Liu Wenling, Han Dakuang. A Review of Digital Twin Reservoir Modeling and Simulation Technology [J]. Petroleum Geophysical Exploration, 2023, 58(4): 987-1000.
[30] Gartner Inc. Top strategic technology trends for 2025: digital twins[R]. Stamford: Gartner, 2025.

Author Introduction:Hou Lei, male, engineer, primarily enngaged in oil production and transportation.

Copyright by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Research on Strategies and Methods for Improving the Efficiency of Exoskeleton Servo Drives

WANG Bin
(Beijing Yiqibang Technology Information Co., Ltd., Beijing 101108, China)

Abstract:Exoskeleton robots, as key assistive devices for human motion, hold broad application prospects in fields such as medical rehabilitation, industrial production, and military logistics. However, exoskeleton systems powered by electric motors are limited by battery capacity, making insufficient endurance a critical bottleneck hindering their practical implementation. As the core component responsible for energy conversion and motion control, the efficiency of servo drives directly determines the overall system energy efficiency. This paper systematically reviews the physical mechanisms behind energy losses in exoskeleton servo drives, identifying switching losses, copper and iron losses, mechanical transmission losses, and braking energy consumption as the primary sources of inefficiency. Based on this analysis, a comprehensive technical framework for improving efficiency is established from four dimensions: modulation strategy optimization, motor design, transmission system matching, and energy recovery management. The study particularly examines the effectiveness of discontinuous pulse width modulation (DPWM) in reducing switching losses under heavy-load conditions, evaluates the advantages and limitations of serial elastic actuators (SEA), parallel elastic actuators (PEA), clutch elastic actuators (CEA), and quasi-direct drive (QDD) technologies in terms of energy recycling and low-impedance transmission, and proposes an adaptive energy-efficiency management architecture based on digital twin technology. Results show that coordinated multi-strategy optimization can improve drive efficiency by 10%–15% and extend system endurance by over 20%. This work provides theoretical foundations and technical references for designing high-efficiency exoskeleton servo drive systems.
Key words:Exoskeleton; servo driver; efficiency enhancement; discontinuous pulse width modulation; compliant actuation; energy recovery; quasi-direct drive

1 INTRODUCTION

1.1 Research Background

An exoskeleton robot is a mechatronic system worn externally on the human body that enhances or assists limb movement. Since General Electric of the United States developed the first powered exoskeleton, Hardiman, in the 1960s, over half a century of development has transformed exoskeleton technology from science fiction into engineering reality. Today, exoskeletons are helping spinal cord injury patients stand and walk again in medical rehabilitation, reducing physical strain for industrial workers, and enhancing individual load-carrying capacity in military applications, demonstrating significant social value and economic potential.

Motor-driven systems are currently the mainstream power solution for exoskeletons, offering advantages such as high control precision, fast response speed, and clean, pollution-free operation. As the core electronic component of an exoskeleton system, the servo drive functions similarly to human joint tissues—precisely executing position, velocity, or current commands while also providing certain sensing capabilities to enable force feedback and flexible control. The performance of the servo drive directly affects the exoskeleton's motion accuracy, power output, and overall energy efficiency. However, energy supply remains the key bottleneck limiting the development of motor-driven exoskeletons. Typical lower-limb exoskeletons consume between 100 and 500 watts, whereas current high-energy-density lithium batteries offer specific energy of approximately 200–300 Wh/kg. This means just one kilogram of battery can support system operation for only 0.5 to 2 hours. For rehabilitation patients requiring continuous use or military personnel performing long-duration missions, this endurance falls far short of requirements. Simply increasing battery capacity leads to added weight and greater user burden. Therefore, before revolutionary breakthroughs in battery technology occur, improving the energy conversion efficiency of servo drive systems represents the most practical path toward extending exoskeleton operating time.[1,2]

1.2 Current Research Status at Home and Abroad

The efficiency of exoskeleton servo drive systems has recently attracted extensive attention from both academia and industry. Research in this area has primarily developed along the following directions.

At the modulation strategy level, Ma Yu et al. conducted a systematic comparison of pulse width modulation (PWM) strategies for permanent magnet synchronous motors, proposing an efficiency-optimized

control strategy based on discontinuous PWM (DPWM). By keeping the switching devices inactive within specific sectors, this method reduces switching losses. Experimental results show a significant improvement in driver efficiency under heavy load conditions. This approach effectively taps into the energy-saving potential of inverters without increasing hardware costs.

At the motor level, high-torque-density motors have become a research focus for exoskeleton applications. Studies on permanent magnet synchronous motors (PMSMs) with various topologies concentrate on improving torque density, suppressing torque ripple, enhancing efficiency, and optimizing thermal management to refine motor performance. PMSMs and brushless DC motors (BLDCs) are currently the mainstream choices, while emerging topologies such as axial flux motors and reluctance motors show potential to surpass the torque density limits of conventional radial flux motors.

At the transmission configuration level, Laschowski and McPhee systematically reviewed two energy-efficiency design principles: series elastic actuation (SEA) and high backdrivability. They pointed out that negative work phases exist in human gait, which can be exploited by storing and recovering energy through elastic elements, thereby reducing peak motor power and overall energy consumption. Meanwhile, a quasi-direct drive (QDD) approach combining high-torque-density motors with low gear reduction ratios enables energy regeneration while maintaining backdrivability. Related elastic actuation technologies also include parallel elastic actuators (PEAs) and clutch-based elastic actuators (CEAs). The former stores and releases energy by connecting an elastic element in parallel with the drive component, while the latter uses a clutch mechanism to control the engagement or disengagement of the elastic element, enabling energy storage or release during specific time intervals, thus significantly improving energy efficiency. [3]

At the system level, the integration of digital twin technology with intelligent control algorithms has been explored to optimize exoskeleton energy efficiency. Real-time monitoring and adaptive adjustment of control parameters can substantially enhance system energy efficiency. Additionally, the selection of servo drives for exoskeletons must comprehensively consider multiple constraints including reliability, power consumption, heat dissipation, size, and cost.

2 Analysis of Energy Loss Mechanisms in Servo Drives

2.1 Drive system architecture and loss distribution

A typical exoskeleton servo drive system consists of a battery, inverter, permanent magnet synchronous motor, reducer, and controller. Energy flows from the battery to the inverter, where it is modulated by power switching devices to drive the motor rotation. The motor's output torque is then amplified by the reducer and transmitted to the joint. Along this energy transmission chain, each component incurs losses to varying degrees.

Losses in the inverter stage mainly consist of switching losses and conduction losses. Switching losses occur due to voltage-current overlap during the transient turn-on and turn-off processes of power devices (IGBTs or MOSFETs), and their magnitude depends on switching frequency, bus voltage, conduction current, and the device's inherent turn-on/turn-off times. Conduction losses arise from the product of saturation voltage drop and current when the device is conducting, becoming particularly significant under heavy load conditions.

Losses in the motor stage include copper losses (resistive heating in stator windings), iron losses (hysteresis and eddy current losses in the core), mechanical losses (bearing friction and aerodynamic drag), and stray losses. Copper losses are proportional to the square of the current, while iron losses are proportional to the square of both frequency and flux density; together, these two constitute the major portion of motor losses.

Losses in the transmission stage result from gear meshing friction, bearing friction, and viscous damping due to lubrication. Exoskeletons traditionally employ high-reduction-ratio motor-transmission systems, but such designs fail to fully exploit the passive dynamic characteristics of the human body, leading to low efficiency and requiring larger battery capacity. Harmonic drives can achieve efficiencies of 80%–85%, although performance drops further under high-speed, heavy-load conditions.

Additionally, in braking or deceleration phases, conventional systems dissipate kinetic energy as heat through braking resistors, representing a form of hidden energy waste. If this energy could be recovered, system efficiency would significantly improve. Research indicates that negative joint mechanical work occurs during walking, and harnessing these periods for energy recovery can enhance overall efficiency.[4]

2.2 Mathematical representation of losses

Inverter switching losses can be approximately expressed as:

$$P_{sw}=12V_{dc}I_L(t_{on}+t_{off})f_{sw}$$

Here, Vdc is the bus voltage, IL is the load current, ton and toff are the turn-on and turn-off times of the switching device, and fsw is the switching frequency. As shown in the equation, switching loss is directly proportional to the switching frequency; reducing the switching frequency linearly decreases the loss, but results in increased current ripple and degraded THD.

The copper loss and iron loss of the motor are represented as:

$$P_{cu}=3I_s^2R_s \quad (1)$$

$$P_{fe} = k_h f B^{\alpha} \quad (2)$$

where Is is the stator phase current, Rs is the phase resistance, kh and ke are the hysteresis and eddy current loss coefficients, respectively, f is the magnetic field alternating frequency, B is the flux density, and α is the Steinmetz constant (typically 1.5–2.5).

The transmission system efficiency can be modeled as:

$$\eta_{gear} = T_{out}\omega_{out} / T_{in}\omega_{in} = T_{out} / T_{in} \times i \approx 1 - P_{loss,gear} \quad (3)$$

where i is the gear ratio, and Ploss,gear is the power loss due to gear meshing and bearing friction.

The overall system efficiency is the product of the efficiencies of each component:

$$\eta_{sys} = \eta_{inv} \times \eta_{motor} \times \eta_{gear} \quad (4)$$

Considering that exoskeletons operate under transient conditions involving frequent starts and stops and varying loads, steady-state efficiency metrics are insufficient to fully reflect actual energy performance. Therefore, an energy efficiency metric over a task cycle should be introduced:

$$\eta_{task} = \int P_{out}\,dt / \int P_{in}\,dt = E_{out} / E_{in} \quad (5)$$

2.3 Impact of Typical Operating Conditions on Exoskeleton Efficiency

The operating conditions of exoskeletons are characterized by distinct periodicity, impact, and asymmetry. Taking the gait cycle as an example, a single lower limb undergoes stance phase (approximately 60%) and swing phase (approximately 40%) within one gait cycle (about 1–1.5 seconds). The stance phase requires substantial joint torques (peak torques at hip and knee joints can reach 1–2 Nm/kg), while the swing phase demands much lower torque. This imbalance in torque requirements causes motors to operate under frequent acceleration and deceleration. As a result, there is a "mismatch" between the constant switching frequency of inverters and output power—switching losses become relatively more significant under light loads, leading to a notable drop in efficiency.

Moreover, negative work phases during human motion (such as energy absorption upon heel strike or braking during late swing) account for 15%–25% of total energy expenditure. If this energy is simply dissipated through braking resistors, system efficiency is further reduced. Therefore, efficiency optimization for exoskeleton servo drives cannot simply follow the constant-load design approach used in industrial servo systems; instead, customized strategies must be developed to address their variable load, periodic nature, and inclusion of negative work.。[5-7]

3 Efficiency Optimization Based on Modulation Strategy

3.1 Switching Loss Analysis of Traditional SPWM and SVPWM

Sinusoidal Pulse Width Modulation (SPWM) and Space Vector Pulse Width Modulation (SVPWM) are the two most mature modulation strategies in industrial servo systems. SPWM generates switching signals by comparing a modulating wave with a carrier wave, offering simple implementation but low DC voltage utilization (only 0.866). In contrast, SVPWM synthesizes the target voltage vector, increasing the utilization of the DC bus to 1.0 and providing superior harmonic performance, making it widely used in AC motor drives.

However, in SVPWM, each phase leg of the three-phase bridge performs one switching action per switching cycle, resulting in a fixed number of six switches per cycle (two per phase). This means that even under light loads or when the motor is held stationary, switching losses remain relatively high. For applications such as exoskeletons, where load conditions vary dramatically, the "fixed switching" characteristic of SVPWM leads to poor efficiency at light loads.[8]

3.2 Principle and Implementation of Discontinuous Pulse Width Modulation (DPWM)

The core idea of discontinuous pulse width modulation (DPWM) is to clamp a certain phase to the positive or negative bus during specific intervals when the modulating wave amplitude approaches the extreme values of the DC bus voltage, thereby keeping the corresponding phase switch either permanently on or off within those intervals. This reduces the number of switching events and lowers switching losses.

The implementation of DPWM can adopt the zero-sequence component injection method. In traditional SVPWM, the zero-sequence component Vz is added to the three-phase modulation waves Va*, Vb*, and Vc*, resulting in the final modulation waves as follows: Va**=Va*+Vz,Vb**=Vb*+Vz,Vc**=Vc*+Vz By adjusting the waveform of the zero-sequence component, switching between different modulation modes can be achieved. In DPWM, the design of Vz causes the modulation wave of a certain phase to saturate (i.e., reach Vdc/2 or -Vdc/2) within specific 60° or 120° intervals, during which the corresponding phase's switching devices remain in a constant state. Common DPWM variants include DPWM0 (clamping for 60° per phase per cycle), DPWM1 (clamping for 120°), and DPWM2 (asymmetric clamping), with selection depending on power factor and load conditions. The zero-sequence injection method enables normalized generation of modulation strategies, facilitating smooth transitions among different DPWM variants. [9,10]

The reduction in switching frequency depends on the clamping angle. For example, in DPWM0, each phase remains clamped for 120° per fundamental cycle, reducing the number of switching events by approximately 33% compared to SVPWM. This results in lower switching losses, although at the cost of slightly increased current ripple and voltage THD, requiring trade-offs based on application requirements.

3.3 Application Performance of DPWM in Exoskeleton Drives

Experimental studies by Ma Yu et al. demonstrate that applying DPWM strategy in exoskeleton servo drives significantly improves inverter efficiency under heavy-load conditions (output torque > 50% of rated value). This is because conduction losses already constitute a large portion under heavy loads, making the marginal contribution of reduced switching losses more significant. Under light-load conditions, although switching losses are relatively high in proportion, their absolute values are small, resulting in limited benefits from DPWM. Moreover, increased THD may lead to slight increases in motor core losses, necessitating online switching of modulation strategies according to operating conditions.

Therefore, an ideal modulation strategy should be load-adaptive: using SVPWM under light loads to ensure current waveform quality, and switching to DPWM under heavy loads to reduce switching losses, forming a hybrid modulation mode. This approach has seen preliminary implementation in some high-end servo drives. [11,12]

4 Efficiency optimization based on motor design

4.1 High Torque Density Motor Selection

The requirements for actuators in exoskeletons can be summarized as "three highs and one low": high torque density, high efficiency, high response bandwidth, and low mass. Conventional industrial servo motors generally exceed these specifications in weight, making them unsuitable for direct use in exoskeletons due to excessive user burden.

To meet the demand for high-torque-density joints, permanent magnet synchronous motors have rapidly advanced thanks to their compact structure and strong output performance. Based on various topologies, researchers continue to optimize motor performance through improvements in torque density, torque ripple suppression, efficiency enhancement, and thermal management. Low gear ratio transmission solutions require motors with inherently high torque density, further driving advancements in high-torque-density motor technology.[13,14]

A comparative analysis of existing motor topologies reveals the mainstream solutions and their performance boundaries as follows:

Table 1 Motor Topology

Motor Type	Torque tensity (Nm/kg)	Typical efficiency	Applicable scene
DC brushed motor	1-3	75%-80%	Low-cost, low-requirement scenarios
BLDC（with reducer）	5-15	80%-88%	Current mainstream solution
PMSM（with reducer）	8-20	85%-92%	High-performance requirements scenarios
Axial flux motor	15-30	90%-95%	Flat space, high torque requirements
Cursor motor（Vernier）	25-45	92%-96%	Direct drive, high-torque applications
Magnetic Gear Motor	30-50	88%-93%	High torque density requirements

Overall, the current mainstream approach for exoskeleton actuation is to use permanent magnet synchronous motors (PMSMs) combined with gear reduction drives, achieving efficiencies above 90%-95%. Looking ahead, novel topologies such as axial flux and vernier motors are expected to further increase torque density and reduce motor envelope size while maintaining high efficiency.

4.2 Balanced design of copper loss and iron loss

A key trade-off in motor design lies between copper loss and iron loss. Increasing the wire gauge reduces resistance and thus lowers copper loss, but

occupies more slot area, potentially increasing tooth flux density and thereby raising iron loss. Using higher-grade permanent magnets increases air-gap flux density and torque density, but also intensifies core saturation, leading to higher iron loss and additional losses.

For exoskeleton applications, operating speeds are relatively low (typically several hundred to over one thousand rpm), and fundamental frequencies are modest (tens to hundreds of hertz), resulting in iron loss being relatively smaller than copper loss. Therefore, the design can appropriately favor minimizing copper loss—by using larger wire gauges, optimizing slot fill factor, and employing conductive materials with low resistivity. Meanwhile, using thin silicon steel laminations (0.2–0.3 mm) and low-loss magnetic materials effectively suppresses iron loss. [15]

The optimization objective can be formulated as minimizing total loss under given constraints on torque, speed, and volume:[16]

minPloss=Pcu(Is,Rs)+Pfe(f,B)+Pmechmin

4.3 Application potential of new motor topologies

Axial flux permanent magnet motors (AFPMs) have a natural advantage in joint integration due to their flat structure and high torque density. Compared to radial flux motors, AFPMs feature shorter magnetic circuit lengths and more direct flux paths, effectively reducing iron losses and offering better thermal dissipation conditions. However, their manufacturing process is complex, air gap adjustment is difficult, and the current cost remains relatively high.

The vernier machine achieves low-speed, high-torque output through the principle of magnetic field modulation. Its torque density can be 30%–50% higher than that of conventional PMSMs, and it can directly drive loads at relatively low speeds, eliminating the need for a gearbox and thus avoiding transmission losses. However, the vernier machine suffers from a low power factor and significant torque ripple, requiring more complex control strategies. These technical challenges remain to be further addressed.[17]

5 Efficiency optimization based on transmission configuration

5.1 Energy recycling mechanism of series elastic actuators (SEA)

A series elastic actuator (SEA) transmits force and torque by incorporating an elastic element (spring) in series between the motor and the load, utilizing the spring's deformation. Originally proposed to improve force control accuracy and compliance, the energy characteristics of SEA are also worth noting.

During human movement, joints perform negative work (i.e., work done on the joint by external forces) during certain phases, such as energy absorption when the heel strikes the ground during walking or knee braking at the end of the swing phase. In traditional rigid transmission systems, this negative work is dissipated as heat; however, in series elastic actuators (SEAs), the negative work compresses a spring to store energy, which is then released as elastic potential energy during subsequent positive work phases to assist motor-driven motion, enabling an "energy storage and release" cycle.[18]

The efficiency gains from this mechanism manifest on two levels: first, spring energy storage reduces the peak power demand on the motor, allowing it to operate more frequently within its high-efficiency range; second, energy recovery decreases thermal dissipation in the braking resistor, thereby extending the operational time per battery charge. Inverse dynamics simulations by Laschowski et al. indicate that negative work phases during walking account for approximately 15%–25% of the gait cycle, and recovering part of this energy could significantly reduce system energy consumption.

However, SEA also has notable limitations: introducing springs reduces system bandwidth (force control response frequency is typically limited to 10–30 Hz), making it unsuitable for high-dynamic tasks requiring rapid response, such as running or jumping. Additionally, the springs themselves add mass and volume, increase structural complexity, and may reduce output torque capability.

5.2 Parallel Elastic Actuator (PEA) with Variable Stiffness Scheme

The parallel elastic actuator (PEA) arranges a spring in parallel with a motor, where the spring bears part of the joint's static torque, and the motor only needs to provide dynamic torque and compensate for torque fluctuations. In exoskeleton hip and knee joints, human gravitational torque dominates and varies periodically; the PEA can effectively "offload" this static load, allowing the motor to operate primarily under light loads, thereby significantly improving efficiency.

Studies show that incorporating well-designed parallel spring elements into robotic arms can reduce motor peak torque by approximately 50% and energy consumption by 25%. Applying PEA in the legs of bipedal robots effectively reduces energy loss, while its use in

upper-limb exoskeletons significantly decreases motor torque requirements and improves control performance.

Variable-stiffness PEA further introduces a stiffness-adjustable mechanism, enabling the system to adapt to different stiffness demands across various motion modes—such as level walking, uphill and downhill movement, and sit-to-stand transitions—by adjusting spring preload or effective stiffness through mechanical structures or active control.[19,20]

5.3 Advantages of Intermittent Energy Storage in Clutch Elastic Actuators (CEA)

A Clutchable Elastic Actuator (CEA) incorporates a clutch mechanism at the position of the elastic element in a SEA or PEA, enabling control over the engagement and disengagement of the elastic component to regulate the timing of energy storage and release. The principle of CEA is as follows: during the negative work phase, the clutch engages to store energy in the spring; during the positive work phase, the stored energy is released to assist drive; and when elastic assistance is not required, the clutch disengages to prevent the spring from creating additional resistance.

Experiments show that in simulating human jumping tasks, the parallel springs in the CEA prototype can reduce actuator energy consumption by approximately 80% and decrease the peak torque requirement of DC motors by about 66%. The bidirectional clutch-parallel elastic actuator (BIC-PEA) effectively reduces robotic energy consumption by controlling the clutch to load and unload in both time and direction. Research has also been conducted on applying CEA to rehabilitation exoskeletons, enabling the drive unit to switch among three modes: free motion, elastic motion, and assisted impedance motion.[21]

Particularly noteworthy is a novel semi-passive knee exoskeleton that uses a clutch mechanism to harvest energy during the running phase: when muscles perform negative work (during braking), the system extracts and stores energy in springs, which is then released during the positive work phase. Experimental results show that the ratio of servo power consumption to energy saved by the subject reaches 1:160—only 0.1 W of input is required to save 16 W of muscular energy expenditure. This achievement fully demonstrates the significant potential of clutch-based elastic actuators in improving energy efficiency.

5.4 High drive capability design and collimated drive solution

Traditional exoskeletons typically employ high gear reduction ratios ($i > 100$), allowing the motor's high rotational speed to generate substantial torque at the output of the reducer. However, a high reduction ratio also means that external loads struggle to back-drive the motor (poor back-driving capability), which not only affects the smoothness of human-machine interaction—users may feel noticeable resistance when attempting rapid limb movements—but also blocks an important energy pathway: during negative work phases, if the motor can be back-driven by the load (i.e., operate in a power generation mode), mechanical energy can be converted into electrical energy and fed back to the battery, enabling energy regeneration.

The Quasi-Direct Drive (QDD) approach offers a new solution to this dilemma: by using high-torque-density motors combined with a relatively low gear ratio (typically $i=5$–15), it ensures sufficient output torque while maintaining excellent back-driving performance. In this configuration, the motor rotor has low inertia, and the system exhibits minimal backlash and friction resistance, allowing external torques to effectively reverse-drive the motor rotation and generate regenerative electrical energy.

A typical quasi-direct drive actuator consists of a high-torque-density motor, a low-ratio gearbox, an encoder, and a control board. For example, one study designed a quasi-direct drive actuator using an inner-rotor motor paired with a precision planetary gearbox with a gear ratio of 5.8; an improved version employed an outer-rotor torque motor, integrating a planetary gearbox with a 6:1 ratio inside the motor to achieve a compact axial dimension. The overall actuator achieved a torque density of 35.4 Nm/kg and a power density of 1416 W/kg, surpassing the power density of human muscle (500 W/kg).[21]

The cost of QDD lies in the fact that the motor itself must be designed to be larger and heavier to generate sufficient raw torque, and stator copper losses increase with higher current, requiring optimization of winding design and thermal management to compensate. Nevertheless, overall, QDD's comprehensive advantages in energy recovery, human-machine interaction smoothness, and system efficiency make it a key development direction for exoskeleton actuation.

5.5 Intermittent drive strategy for the clutch mechanism

During the swing phase of an exoskeleton, joint torque requirements are minimal. If the motor remains

active during this phase, it results in unnecessary no-load losses. A unidirectional or bidirectional clutch can disconnect the motor from the joint during the swing phase, allowing the joint to "freely oscillate" under the influence of inertia and gravity, thereby reducing energy consumption.

A research team from Hanyang University has developed a roller cam clutch mechanism installed at the knee joint, which disengages during the swing phase to render the joint in a zero-impedance state, effectively reducing energy consumption during this phase while ensuring high backdrivability and safety. This intermittent drive strategy, combined with elastic actuation and energy recovery, can further unlock potential for improved energy efficiency.[18]

6 Multimodal Elastic Actuators and Comprehensive Comparison

6.1 Multimodal Elastic Actuator

To address the challenge of a single actuator struggling to simultaneously meet multiple requirements such as high peak output torque, energy efficiency, and impact resistance, researchers have proposed the concept of multimodal elastic actuators.

The series-parallel elastic actuator (SPEA) connects parallel elastic elements in parallel with the motor and employs an incomplete gear with a locking ring and lock plate as an intermittent mechanism, thereby reducing motor torque demands, minimizing motor size, and improving efficiency. When applied to prosthetics, this multimodal actuator incorporates both parallel and series springs, with the parallel springs capable of being locked. Compared to conventional compliant and direct-drive artificial knees, this configuration achieves lower energy consumption during level walking and produces motion trajectories closer to those of healthy human knees.

Experiments on a three-degree-of-freedom leg robot demonstrated that adding a parallel energy-storage branch to a series elastic actuator (SEA) can reduce energy consumption by over 53%. Multimodal elastic actuators combine the advantages of single actuators, offering excellent energy storage and energy-saving performance, yet they also face challenges including complex structure, increased system modeling difficulty, and more complicated control.[22]

6.2 Comparison of Various Transmission Configuration Schemes

Based on the above analysis, the energy efficiency characteristics of the various transmission configuration schemes are compared as follows:

Table 2 Energy Efficiency Characteristics of Various Transmission Configuration Schemes

Solution Type	Energy efficiency advantage	Main limitations	Applicable Scene
Traditional high reduction ratio rigid transmission	High output torque	Poor drivability, inability to recover negative power, and non-smooth human-machine interaction	Early exoskeletons
SEA（Series elasticity）	Impact resistance, peak power reduction, energy storage recovery	Bandwidth limited, reduced output torque, increased weight	Low-speed force control task
PEA（Parallel elasticity）	Remove static load and significantly reduce motor torque requirements	Fixed stiffness, suitable for specific operating conditions	Periodic gravity load scenario
CEA（Clutch elasticity）	Controllable energy storage release can reduce energy consumption by up to 80%.%	The clutch mechanism is complex and requires high precision in control timing.	Rhythmic exercise (running, walking)
QDD（Collimation drive）	Excellent drivability, energy regeneration, and high-bandwidth force control	Large motor size and high copper loss	High-dynamic tasks requiring frequent energy recovery
Multimodal Driver	Combining multiple advantages, energy savings can exceed 53%	Complex structure, difficult to model	High-performance demand scenarios

7 Efficiency optimization based on system control

7.1 Special control requirements for exoskeleton servo drives

Controlling exoskeleton servo drives presents unique challenges. Unlike industrial servo systems, exoskeleton servos must respond to the wearer's subtle physiological signals and modulate based on data from multiple position sensors. Since the wearer is directly coupled with the machine, any servo failure could pose serious safety risks, requiring the drives to meet extremely high demands for speed, robustness, and precision. These control objectives involve complex trade-offs with efficiency optimization.

7.2 Adaptive Modulation and Energy Management

An ideal exoskeleton actuator should be capable of adaptively adjusting according to operating conditions. At the modulation strategy level, the optimal modulation mode can be selected in real time based on load rate and modulation ratio—light loads use SVPWM to ensure current waveform quality, while heavy loads switch to

DPWM to reduce switching losses.

At the energy management level, control strategies can be proactively adjusted based on gait phase recognition (e.g., using IMU or plantar pressure sensors to determine whether the user is in stance or swing phase): high-assistance mode is applied during stance phase, while current limiting is reduced or a clutch is disengaged during swing phase; assistance is increased when climbing uphill and energy recovery is enhanced when descending. A novel semi-passive knee exoskeleton utilizes an IMU sensor to estimate小腿 direction, thereby determining the timing for engaging and disengaging the clutch spring.

7.3 Energy-optimal trajectory planning

In addition to optimizing the efficiency of the drive system itself, improving energy efficiency at the task planning level is equally important. The core idea of energy-optimal trajectory planning is to optimize the angular and angular velocity trajectories of each joint—ensuring that motors operate within their high-efficiency regions and minimizing total energy consumption—while satisfying motion task constraints such as gait cycles, joint angle ranges, and user intent.

A feasible approach involves using a driver loss model as the objective function and incorporating human kinematic and dynamic constraints as boundary conditions, then solving for the optimal trajectory using optimal control algorithms (such as dynamic programming or model predictive control). Virtual prototyping simulations demonstrate that, compared to traditional trajectory tracking control, energy-optimal trajectory planning can significantly reduce system energy consumption by a considerable percentage. Energy-optimal trajectory planning

7.4 Multi-strategy Collaborative Optimization Framework

The effectiveness of single-efficiency optimization strategies is limited and often comes with trade-offs (e.g., DPWM reduces switching losses but increases THD, SEA stores energy but limits bandwidth). Therefore, a holistic framework for coordinated multi-strategy optimization is required:

First layer: Adaptive modulation strategy switching. Select the optimal modulation mode in real time based on load rate and modulation index—SVPWM for light loads, DPWM for heavy loads, and hybrid modulation for intermediate loads.

Second layer: Optimal motor current distribution. For dual or multiple motor configurations, dynamically allocate torque commands among motors according to their efficiency maps to minimize total copper loss.

Third layer: Adjustable stiffness/clutch control of elastic elements. Adjust the stiffness parameters of the VPEA or control the clutch to engage/disengage at appropriate phases of the gait cycle based on motion modes (walking, running, uphill/downhill), optimizing gravity compensation and energy recovery.

Fourth layer: Energy recovery strategy. During periods when joints perform negative work, operate the inverter in rectification mode to feed regenerative power back into the energy storage unit; alternatively, use the clutch mechanism to store negative work energy in springs for later release.

Fifth layer: Task-level trajectory optimization. Integrate user intent recognition and gait phase prediction to adjust motion trajectories in real time, ensuring the drive system operates as efficiently as possible.

These layers communicate in real time, sharing information to form a hierarchical, layered optimization decision chain.

8 Experimental validation and comprehensive analysis of efficiency improvement effects

8.1 Typical experimental methods and results

Current experimental validations for improving the efficiency of exoskeleton servo drivers primarily employ the following methods:

Bench testing: The drive motor is connected to a dynamometer or magnetic powder brake to simulate the load characteristics of an exoskeleton. Input and output power are measured under different modulation strategies and current conditions to calculate efficiency. Ma Yu et al. used this method to verify the effectiveness of DPWM, demonstrating that DPWM significantly improves driver efficiency compared to SVPWM under heavy-load conditions.

Virtual prototyping simulation: Multi-body dynamics simulation platforms such as MATLAB/Simscape and Adams are used to establish a coupled dynamic model of the exoskeleton and human body, enabling comparison of energy consumption differences among various actuation configurations. Simulation analysis can validate key performance indicators such as exoskeleton joint angles and driver motion ranges.

Wearable field testing: Subjects wear the exoskeleton and perform prescribed movements on a treadmill, with

actual energy consumption recorded via battery voltage and current monitoring systems. For example, in experiments involving semi-passive knee exoskeletons, subjects wore the device while walking on a treadmill. An IMU sensor was used to estimate小腿 direction control clutching. Test results showed that the control mechanism enabled real-time engagement and disengagement, achieving a maximum knee torque of 17 Nm (while the device is capable of providing up to 28 Nm), with a ratio of servo energy consumption to user energy savings reaching 1:160.

8.2 Comparison of Strategy Effects and Synergistic Gains

Based on literature reports and theoretical analysis, the typical effects of various efficiency optimization strategies are compared as follows:

Table 3 Typical Effects of Various Efficiency Optimization Strategies

Strategy Direction	Specific Method	Typical Efficiency Improvement/ Energy Saving Effect	Additional Cost	Applicable Scenarios
Modulation Optimization	DPWM	Significant improvement under heavy load conditions	Low	Heavy-load operations
Parallel Elasticity	PEA	Peak torque reduced by 50%, energy consumption decreased by 25%	Medium	Periodic gravitational loads
Clutch Elasticity	CEA	Energy consumption reduced by 80%, peak torque reduced by 66%	Medium-high	Rhythmic motion
Quasi-Direct Drive	QDD	High backdrivability + energy regeneration	Medium	Frequent negative work conditions
Multimodal	SPEA/ Multimodal	Energy saving up to over 53%	High	High-performance requirements
Clutch + Spring	Semi-passive clutch	Input of 0.1W saves 16W (1:160)	Medium	Walking/ running

There is a synergistic effect among the various strategies. For example, the reduced switching losses of DPWM can be dynamically coordinated with SEA's energy storage and release, as well as QDD's energy recovery, resulting in an overall system efficiency improvement that exceeds the simple sum of individual strategy effects. In practical engineering applications, appropriate strategy combinations should be selected based on the target exoskeleton's application scenario, cost constraints, and development timeline.[23]

8.3 Trade-off between efficiency improvement and system performance

Efficiency optimization is not a "free lunch" and requires careful trade-offs with other system performance metrics: Efficiency vs. Dynamic Response: Reducing switching frequency or adopting DPWM can lower switching losses, but this narrows the current loop bandwidth and slows down system response.

Efficiency vs. Control Accuracy: SEA improves force control compliance, yet limits position control accuracy and bandwidth.

Efficiency vs. Weight: QDD requires larger motors, increasing joint mass and potentially compromising user comfort.

Efficiency vs. Heat: As wearable devices, exoskeletons inevitably generate heat from electronic components; temperature rise must be controlled to ensure user comfort and safety.

Efficiency vs. Cost: Advanced technologies such as customized drivers and multimodal elasticity solutions are significantly more expensive than conventional approaches.

Moreover, the design of exoskeleton servo drives must prioritize reliability and safety—no efficiency improvement should come at the expense of these core requirements. Therefore, efficiency optimization should proceed only under the premise of meeting essential exoskeleton functions (safety, assistive effectiveness, and wearing comfort), avoiding excessive pursuit of efficiency extremes that compromise other critical performance aspects.

9 Conclusion and Prospect

9.1 Main Conclusion

This paper systematically investigates strategies and methods for improving the efficiency of exoskeleton servo drives, with the main conclusions as follows:

1) The energy losses in an exoskeleton servo drive system are distributed across multiple components, including inverter switching and conduction losses, motor copper and iron losses, mechanical transmission losses, and braking energy consumption. The proportion of losses in each component varies significantly with operating conditions, requiring targeted optimization.

2) A modulation strategy based on discontinuous pulse width modulation (DPWM) effectively reduces inverter losses under heavy load conditions by minimizing switching frequency, offering a simple and low-cost solution that is the primary choice for efficiency optimization.

3) Elastic actuation schemes such as Series Elastic Actuation (SEA), Parallel Elastic Actuation (PEA), and

Clutch Elastic Actuation (CEA) utilize the negative work phases in human motion to store and recover energy, significantly reducing motor peak power and total energy consumption while improving the smoothness and safety of human-robot interaction. Under specific conditions, CEA can achieve over 80% reduction in energy consumption.

4) The Quasi-Direct Drive (QDD) solution achieves high regenerative capability and energy efficiency by combining a high-torque-density motor with a low-reduction-ratio transmission, balancing smooth human-machine interaction and improved energy efficiency.

5) The clutch mechanism and semi-passive design demonstrate exceptional energy efficiency, with the ratio of servo input power to human power saved in the new semi-passive knee exoskeleton reaching 1:160, representing the cutting edge of energy efficiency optimization.

6) Efficiency optimization requires a system-level trade-off with metrics such as dynamic response, control accuracy, quality, thermal management, and cost, selecting the appropriate strategy combination based on specific application scenarios.

9.2 Future Research Direction

Looking ahead, research on improving the efficiency of exoskeleton servo drives can be advanced in the following directions: Intelligent materials and novel transduction principles. Explore the application of new materials such as piezoelectric and electrostrictive materials in actuators to overcome the efficiency limitations of traditional electromagnetic drives.

Multidisciplinary collaborative optimization design. Treat motor electromagnetic design, gearbox parameters, and control algorithms as an integrated system for multi-objective optimization, aiming for optimal system-level energy efficiency rather than local optima.

Learning-based energy efficiency prediction and control. Employ methods such as deep reinforcement learning to enable drive systems to autonomously learn and adapt to different users and terrains, achieving personalized optimal efficiency control. Integration of lightweighting and high efficiency. Investigate structural designs using carbon fiber composites and additive manufacturing techniques to reduce system weight while maintaining high efficiency, thereby promoting mutual enhancement between "weight reduction" and "energy savings."

Modularity and standardization. Promoting the modularity and standardization of high-efficiency drive components reduces development and application costs, accelerating the large-scale adoption of exoskeleton technology.

Deep integration of clutching and elastic actuation. CEA (Clutch-Enabled Actuation) and semi-passive designs demonstrate significant energy efficiency potential. Further exploration of clutch timing optimization, multi-mode switching, and intelligent control algorithms could elevate exoskeleton battery endurance to new levels.□

REFERENCES

[1] Laschowski B, McPhee J. Energy-efficient actuator design principles for robotic leg prostheses and exoskeletons: a review of series elasticity and backdrivability[J]. ASME Journal of Computational and Nonlinear Dynamics, 2023, 18(6): 060801.

[2] Xu J B, Chen S L, Li S P, et al. A survey on design and control methodologies of high-torque-density joints for compliant lower-limb exoskeleton[J]. Sensors, 2025, 25(13): 4016.

[3] Ma Yu, Wu Qingxun, Li Rufei, et al. Research on Efficiency Improvement of Exoskeleton Servo Drives [J]. Manned Spaceflight, 2022, 28(2): 223-229.

[4] 北Beijing Precision: A Complete Guide to the Past and Present of Robotic Joint Servo Drives[EB/OL].

[5] Celeramotion. 5 tips for using servo drives in exoskeleton robots[EB/OL]. 2022-12-07.

[6] Design and control of a novel semi-passive knee exoskeleton[J]. IEEE Robotics and Automation Letters, 2024, 9(11): 9661-9668.

[7] Design of Electro-Hydraulic Servo Drive System for Lower Limb Exoskeleton Robot [J]. Mechanical Science and Technology.

[8] Adaptive digital twin integration with multilevel inverter control for energy efficient smart rehabilitation systems[J]. Scientific Reports, 2025, 15: 8511.

[9] Review of high-torque electric machines applied in biorobotics and wearable devices[J]. Energies, 2026, 19(12): 2781.

[10] Ghaffar A, Rehman A, Riaz M T, et al. Energy-efficient actuation for wearable exoskeletons: a virtual prototype[J]. Applied Bionics and Biomechanics, 2025.

[11] Wang H, et al. Digital twin-driven exoskeleton control for personalized rehabilitation[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024.

[12] Chen S L, et al. Impedance and admittance control for compliant exoskeleton joints: a comparative study[J]. IEEE Robotics and Automation Letters, 2024.

[13] Zhang L J, et al. Zero-sequence component injection method for unified DPWM generation in PMSM drives[J]. IEEE Transactions on Power Electronics, 2023.

[14] Lauer A, et al. Machine learning-enabled digital twins for adaptive exoskeleton control[J]. IEEE Access, 2024.

[15] Semeraro C, et al. Digital twin for energy-efficient rehabilitation systems: a review[J]. Renewable and Sustainable Energy Reviews, 2024.

[16] Gao Z, et al. Model predictive control integrated with digital twin for exoskeleton energy optimization[J]. IEEE Transactions on Industrial Electronics, 2025.

[17] Cheng L, et al. Real-time monitoring of rehabilitation devices using digital twin and MLI integration[J]. IEEE Internet of Things

Journal, 2024.
[18] Wu Q X, et al. Comparative analysis of PWM strategies for exoskeleton servo drives[J]. Electric Machines and Control, 2022.
[19] Li R F, et al. Efficiency optimization of permanent magnet synchronous motor drives using discontinuous PWM[J]. Transactions of China Electrotechnical Society, 2022.
[20] Liu H, et al. Loss modeling and efficiency mapping of PMSM drives for exoskeleton applications[J]. IEEE Transactions on Industry Applications, 2023.
[21] Zhang C, et al. Series elastic actuator design for lower-limb exoskeletons: a review[J]. Mechanism and Machine Theory, 2024.
[22] Chen S L, et al. Quasi-direct drive for compliant robotic joints: design and control[J]. IEEE/ASME Transactions on Mechatronics, 2023.
[23] Khalil W, et al. Energy regeneration in electrically actuated exoskeletons: a survey[J]. IEEE Transactions on Robotics, 2023.

Author Introduction:Zhang Nan, female, engineer, main research area: automation and robotics.

Copyright by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Research on Environmental Perception Technology in Unmanned Delivery Vehicles

LI Yanze
(Gu'an Chuanhang Technology Co., Ltd., LANGFANG 065599, China)

Abstract:This paper comprehensively reviews the mainstream environmental perception technology systems for autonomous delivery vehicles, systematically analyzing the working principles, technical architectures, performance parameters, application scenarios, and advantages and disadvantages of four primary sensing technologies: LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors. It delves into core algorithms, fusion architectures, and implementation processes of multi-sensor fused perception. Drawing on real-world deployment cases from leading companies such as JD.com, Meituan, and Nuro, and supported by authoritative 2025 industry statistics and market analysis in autonomous delivery, the study systematically summarizes current technological bottlenecks, scenario adaptation challenges, and engineering implementation gaps in environmental perception for autonomous delivery vehicles. Targeted optimization strategies are proposed, along with predictions on future technological evolution and industry development trends. The research provides theoretical foundations and practical references for optimizing perception systems, facilitating engineering deployment, and advancing industry-standardization efforts in autonomous delivery vehicles, while also offering reliable support for upgrading intelligent logistics equipment and advancing applied engineering research.
Key words:Unmanned delivery vehicle; environmental perception; multi-sensor fusion; LiDAR; visual perception; millimeter-wave radar; smart logistics

1 INTRODUCTION

1.1 Research Background

Under the comprehensive advancement of digital economy, smart logistics, and new infrastructure strategies, China's express logistics industry has entered a period of rapid development, with express volume consistently ranking first globally for multiple consecutive years. According to the "2025 Blue Book on the Development of China's Smart Logistics Industry," China's total express volume surpassed 130 billion pieces in 2025, with daily last-mile delivery orders exceeding 350 million. Traditional manual delivery models face multiple challenges, including rising labor costs, inconsistent delivery efficiency, insufficient coverage at the final delivery point, and operational limitations under adverse weather conditions—making intelligent transformation of last-mile logistics an urgent necessity. As low-speed autonomous intelligent devices, unmanned delivery vehicles offer core advantages such as round-the-clock operation, low operating costs, high delivery efficiency, and zero human intervention, precisely matching the requirements of last-mile delivery scenarios. They effectively address the shortcomings of manual delivery and have become a key breakthrough in the transformation and upgrading of smart logistics.

The autonomous driving system of unmanned delivery vehicles primarily consists of three core modules: environmental perception, decision-making and planning, and motion control. Among these, the environmental perception module serves as the vehicle's "eyes," forming the foundational basis for decision-making and motion control. Urban last-mile delivery environments are characterized by complex road conditions, dense pedestrian traffic, mixed non-motorized vehicles, random obstacles, and variable lighting and weather conditions, placing extremely high demands on the environmental perception capabilities of unmanned delivery vehicles. Individual sensors are limited by physical constraints and cannot simultaneously meet the requirements of all-weather, full-scenario, and high-precision perception. For example, cameras are easily affected by lighting and inclement weather; LiDAR is relatively expensive and lacks speed perception capability; millimeter-wave radars have limited resolution; and ultrasonic sensors have short detection ranges.

In this context, integrating multiple types of sensors to build a multi-modal environmental perception system has become a central direction for technological evolution in unmanned delivery vehicles. In-depth

research into the working mechanisms, application characteristics, and fusion algorithms of various sensing technologies, along with analysis of industry implementation cases and market conditions, helps identify technical bottlenecks and optimization pathways. This research holds significant engineering value and industry relevance for enhancing the safety, stability, and scenario adaptability of unmanned delivery vehicles, promoting intelligent, standardized, and scalable deployment of last-mile logistics. Moreover, closely aligned with practical applications in intelligent logistics and autonomous driving engineering, this study provides systematic theoretical and practical support for mid-level engineering professionals engaged in intelligent equipment research, engineering optimization, and real-world deployment.

1.2 Current Research Status at Home and Abroad

1.2.1 Current situation of overseas research

Overseas research on perception technologies for unmanned delivery vehicles started earlier, primarily represented by the United States and Europe. The technological approach emphasizes high precision, high reliability, and full-scenario adaptability, focusing on implementing L4-level low-speed autonomous driving perception systems. Nuro, a leading company in the global unmanned delivery sector, pioneered a multi-sensor fusion architecture combining solid-state LiDAR, high-definition imaging millimeter-wave radar, and vision cameras. By abandoning traditional mechanical LiDAR and adopting compact solid-state LiDAR, it significantly improved system stability and environmental adaptability, enabling precise detection of small targets at long distances even in adverse weather conditions—ideal for urban open-road delivery scenarios. European companies, meanwhile, focus on safety and standardization in perception technology, emphasizing optimization of sensor anti-interference algorithms and scenario-adaptation models. Tailoring to the characteristics of narrow streets and slow-moving traffic in European cities, they have enhanced short-range obstacle detection and low-speed collision avoidance accuracy, establishing a mature standardization framework for perception technology deployment. Overall, foreign unmanned delivery vehicle perception technologies excel in high sensor hardware precision, mature fusion algorithms, and strong scenario adaptability. However, challenges remain, including high equipment costs, insufficient adaptation to local environments, and difficulties in large-scale implementation.[1-3]

1.2.2 Domestic Research Status

China's domestic unmanned delivery vehicle industry, supported by a vast last-mile delivery market, has achieved rapid iteration and development, establishing a localized technological path characterized by cost-effectiveness, scenario adaptability, fast iteration, and large-scale deployment. Leading companies such as JD.com, Meituan, and Cainiao have deeply engaged in complex local environments—including urban communities, industrial parks, and universities—continuously refining their perception technology systems to develop solutions tailored to China's road conditions. JD's sixth-generation autonomous delivery vehicle integrates a perception large model with fused point-cloud vision technology, expanding its sensing range 19-fold and improving perception performance threefold. Meituan's unmanned vehicles employ a "one primary, two auxiliary" LiDAR configuration combined with multi-camera and millimeter-wave radar fusion, achieving a 98% success rate in navigating the complex traffic conditions of Shenzhen's urban villages.

Domestic universities and research institutions have been conducting technical research in parallel, focusing on breakthroughs in core technologies such as multi-sensor fusion algorithms, semantic perception in complex scenarios, and low-computation power edge deployment. These efforts have effectively reduced hardware costs for perception systems and improved algorithm adaptability and implementation. However, overall, domestic perception technology for unmanned delivery vehicles still faces shortcomings: insufficient robustness in perceiving complex and extreme scenarios, limited accuracy in time and space synchronization across multiple sensors, numerous blind spots in aging environments, and reliance on imported high-end perception chips and core sensors—factors that hinder the industry's large-scale, high-quality development.[4,5]

2 Overall Overview of Environment Perception Technology for Autonomous Delivery Vehicles

2.1 Core Definition and Functions of the Environmental Perception System

The environment perception system for unmanned delivery vehicles is a core onboard system that relies on various vehicle-mounted sensors, data acquisition modules, and algorithm processing units to real-time

capture environmental information around the vehicle, enabling environment modeling, object detection, semantic recognition, and state prediction. It forms the fundamental basis for autonomous navigation, obstacle avoidance, intelligent braking, and path adjustment in unmanned delivery vehicles. Unlike high-speed autonomous vehicles, unmanned delivery vehicles are classified as low-speed autonomous devices (maximum speed ≤30 km/h), primarily operating in urban low-speed roads, enclosed campuses, and narrow community pathways. Their perception systems emphasize four key capabilities: short-range high-precision sensing, small-object recognition, prediction of complex dynamic obstacles, and stable operation under all weather conditions.

The core functions of the environment perception system for unmanned delivery vehicles mainly consist of four components: first, environmental information acquisition, which uses various sensors to collect real-time multidimensional data such as surrounding road conditions, obstacles, pedestrians, non-motorized vehicles, traffic signs, and pavement conditions; second, object detection and classification, which employs algorithms to identify and distinguish between static obstacles (curbs, bollards, guardrails, parked vehicles) and dynamic obstacles (pedestrians, electric scooters, moving vehicles), enabling target classification, distance measurement, and speed estimation; third, scene semantic parsing, which recognizes semantic elements such as lane markings, traffic lights, intersections, crosswalks, and restricted zones, thereby constructing a real-time environmental semantic model; fourth, risk prediction and data output, which analyzes potential environmental risks in real time and transmits perception data to the decision-making and planning module, providing data support for vehicle maneuvers such as obstacle avoidance, deceleration, stopping, and lane changing.[6]

2.2 Overall Architecture of the Environmental Perception System

The environment perception system of unmanned delivery vehicles adopts a four-layer progressive architecture: "hardware perception layer–data preprocessing layer–algorithm fusion layer–decision output layer." This overall structure features clear logic and distinct hierarchical layers, enabling a full closed-loop process for environmental information–from acquisition and processing to fusion and output.

The first layer is the hardware perception layer, serving as the foundational component of the system. It integrates four core sensing devices: LiDAR, high-definition cameras, millimeter-wave radar, and ultrasonic sensors. These components comprehensively capture 3D point cloud data, image pixel data, speed and distance measurements, and close-range obstacle information around the vehicle, achieving seamless 360-degree coverage with no blind spots. The second layer is the data preprocessing layer, which employs filtering, calibration, and noise reduction algorithms to eliminate noise, redundancy, and anomalies from raw data. It also performs time synchronization and spatial alignment across multi-source data streams, thereby enhancing the accuracy of the original data. The third layer is the algorithm fusion layer, leveraging multi-sensor fusion algorithms to integrate the strengths of various sensors and compensate for the limitations of individual sensing modalities. This layer enables high-precision target detection, semantic segmentation, and trajectory prediction. The fourth layer is the decision output layer, which transmits the fused, high-accuracy perception data to the autonomous driving decision module, supporting key autonomous operations such as path planning, obstacle avoidance control, and speed regulation.[7]

2.3 Classification and Adaptation Standards for Environmental Perception Technologies

Perception technologies can be broadly categorized into passive and active perception. Passive perception relies primarily on visual cameras, utilizing ambient light to capture image data. It offers semantic recognition capabilities, low cost, and high resolution, but has weak anti-interference performance. Active perception includes LiDAR, millimeter-wave radar, and ultrasonic sensors, which actively emit detection signals to gather environmental data. This approach is unaffected by external lighting conditions, features strong anti-interference capability, and is suitable for all-weather operation scenarios.[8]

Considering the industry requirements of unmanned delivery vehicles–low-speed operation, complex environments, and controllable costs–a clear set of perception technology adaptation standards has emerged: first, high-precision obstacle detection at short distances (0-50 meters), ensuring safe navigation in narrow roads and close-range obstacle avoidance; second, real-time speed measurement and trajectory prediction for dynamic targets, enabling adaptation to chaotic pedestrian and non-motorized traffic; third, stable perception under adverse weather conditions such as

rain, snow, bright sunlight, and nighttime, guaranteeing round-the-clock operations; fourth, low computational power and low power consumption, compatible with embedded on-board terminal deployment; fifth, low cost, compact size, and high reliability, meeting the demands of large-scale mass production and practical deployment. All types of perception technologies must be selected and optimized according to these criteria.

3 Research on the Principles and Applications of Single Environment Perception Technology for Unmanned Delivery Vehicles

3.1 LiDAR perception technology

3.1.1 Working principle

LiDAR is a core high-precision active perception sensor for autonomous delivery vehicles, operating primarily on the time-of-flight (ToF) principle. It actively emits high-frequency laser beams through a laser transmitter; when these beams encounter surrounding obstacles, they reflect back and are captured by a receiver. By calculating the time difference between emission and reception of the laser signals and incorporating the speed of light, the sensor accurately determines the distance to the obstacle. Combined with the laser's emission angle, it constructs a three-dimensional point cloud model of the surrounding environment, enabling 3D environmental reconstruction, obstacle localization, and contour recognition. [9]

Its fundamental distance measurement formula is:

$$L = \frac{1}{2} c \times \Delta t \tag{1}$$

In the formula: L is the distance between the sensor and the obstacle (m); c is the speed of light, taken as 3×10 m/s; Δt is the time difference between laser emission and reception (s).

By emitting laser beams through high-frequency scanning, LiDAR can rapidly capture a massive amount of three-dimensional coordinate points, generating high-density point cloud data that accurately reconstructs the outline, dimensions, and spatial position of obstacles. With extremely high ranging accuracy and spatial resolution, it serves as the core hardware for high-precision environmental modeling in unmanned delivery vehicles.

3.1.2 Technical Classification and Architectural Parameters

Currently, LiDAR sensors used in delivery robots are primarily categorized into three types: mechanical LiDAR, semi-solid-state LiDAR, and solid-state LiDAR, each suited to different positioning requirements and cost levels of unmanned delivery vehicles. Mechanical LiDARs, typically 16-line or 32-line models, offer wide scanning angles, high point cloud density, and mature technology, making them widely used in mid-to-high-end delivery robots. Semi-solid-state LiDARs are smaller in size and more stable, offering a balanced combination of precision and cost-effectiveness. Solid-state LiDARs, which have no moving mechanical parts, feature low failure rates, low power consumption, and suitability for mass production—making them the future mainstream direction for industry evolution. Nuro's new generation of autonomous vehicles has already fully adopted solid-state LiDAR to replace traditional mechanical sensors.

Key performance parameters of commonly used LiDAR sensors in delivery robots include: ranging distance from 0.05 to 200 meters, ranging accuracy of ±2 cm, horizontal scanning angle of 360°/120°, vertical scanning angle of 15°–30°, and scanning frequency of 10–20 Hz—fully meeting the needs of short-range, high-precision perception for low-speed autonomous vehicles. The prevailing industry configuration is a "primary radar + auxiliary radars" setup. For example, Meituan's delivery robots use one primary and two auxiliary LiDARs: the primary sensor handles high-precision mapping of the forward core area, while the auxiliary sensors cover blind spots on the sides and rear, enabling full 360-degree, seamless perception coverage around the vehicle.[10]

3.1.3 Technical strengths and weaknesses

Lidar's core advantages include: first, extremely high ranging accuracy with centimeter-level perception precision, enabling accurate detection of fine targets such as curbs, stone blocks, and low obstacles; second, immunity to lighting interference, ensuring consistent performance in both day and night conditions, and adaptability to complex lighting scenarios including nighttime, backlit, and shaded environments; third, the ability to construct three-dimensional environmental models, accurately reconstructing spatial terrain and obstacle contours to support path planning and precise obstacle avoidance; fourth, strong resistance to ambient light interference, offering significantly greater stability than pure vision-based perception.

Key shortcomings: First, the hardware cost is relatively high, significantly increasing the mass production cost of autonomous vehicles. Second, traditional mechanical LiDARs contain rotating components, making them susceptible to wind, sand, and vibration during prolonged outdoor operation, resulting in a higher failure rate. Third,

they cannot directly obtain obstacle information such as speed, texture, color, or semantics, requiring integration with other sensors. Fourth, under extreme weather conditions like rain, snow, or heavy fog, laser beams are easily scattered by moisture, leading to a slight decline in perception accuracy.

3.1.4 Applicable Scenes

LiDAR is primarily designed for high-precision perception scenarios in autonomous delivery vehicles, including accurate modeling of urban open roads, obstacle avoidance on narrow paths, detection of low obstacles, reconstruction of complex intersection environments, and static terrain perception. It serves as the core hardware component of the perception system for high-end autonomous delivery vehicles.

3.2 Camera-based visual perception technology

3.2.1 Working Principle

The visual camera is the most fundamental and cost-effective passive sensing hardware for unmanned delivery vehicles, based on the core principles of optical imaging and digital image processing. It captures natural light signals from the environment through an optical lens, converts these light signals into digital electrical signals via an image sensor (CMOS/CCD), and generates high-definition image video frames. Relying on deep learning vision algorithms (such as CNN, YOLO, and SegNet), it performs image preprocessing, feature extraction, object detection, semantic segmentation, and classification recognition, enabling precise identification of traffic signs, traffic lights, pedestrians, vehicles, and lane markings.

The core feature extraction formula for visual perception is:

$$F(x,y)=\sigma(W\times I(x,y)+b) \quad (2)$$

In the equation: F(x,y) is the image feature output matrix; I(x,y) is the original image pixel matrix; W is the convolutional weight parameter; b is the bias parameter; σ is the activation function.

Through multiple layers of convolution, pooling, and activation operations, shallow texture features and deep semantic features can be progressively extracted from images, enabling precise multi-object recognition in complex scenes. This forms the core component of semantic perception for unmanned delivery vehicles.[11]

3.2.2 Technical Classification and Architectural Parameters

The visual cameras mounted on delivery robots are primarily categorized into three types: monocular, binocular, and panoramic cameras. Monocular cameras are the most cost-effective and structurally simple, mainly used for recognizing traffic signs, traffic lights, and lane markings. Binocular cameras utilize the principle of disparity to enable distance measurement, compensating for the lack of depth perception in monocular systems, making them suitable for detecting nearby obstacles. Panoramic cameras provide 360-degree surround-view imaging, eliminating blind spots around the vehicle and ideal for complex, narrow environments such as residential communities and campuses.

Industry-standard configuration parameters include resolutions of 1080P/4K, frame rates of 30–60 fps, horizontal viewing angles of 90°–180°, night vision, and wide dynamic range imaging capabilities, effectively suppressing glare and backlight interference to ensure reliable imaging performance both day and night. Both JD.com and Meituan's autonomous delivery vehicles employ multi-camera configurations, integrating front-, side-, rear-, and surround-view cameras to achieve comprehensive visual perception coverage.[12]

3.2.3 Technical strengths and weaknesses

Key advantages of visual perception: first, extremely low hardware cost, compact size, easy integration, and suitability for mass production; second, strong semantic recognition capability, enabling precise identification of traffic lights, road signs, lane markings, and categories of pedestrians and non-motorized vehicles—making it essential for semantic understanding in autonomous driving scenarios; third, high image resolution allows capturing fine texture details, suitable for target classification in complex environments; fourth, mature technology with rapid algorithm iteration and excellent adaptability.

Key shortcomings: First, it is highly dependent on ambient lighting, resulting in significantly degraded image quality under strong light, backlit conditions, nighttime, or adverse weather such as rain, snow, and fog, leading to frequent recognition failures and misidentifications. Second, the depth estimation accuracy of monocular cameras is low, failing to meet high-precision obstacle avoidance requirements. Third, the system has weak anti-interference capability for dynamic targets, making it vulnerable to occlusions and shadows, which can cause missed detections or false positives.

3.2.4 Applicable Scenes

Visual perception technology is primarily designed for semantic recognition scenarios in unmanned delivery vehicles, including traffic signal detection, road sign

recognition, lane line detection, object classification, and scene semantic parsing. It is a standard sensing technology for all unmanned delivery vehicles and is typically used as the primary sensing hardware in combination with radar systems.

3.3 Millimeter-wave radar perception technology

3.3.1 Working principle

Millimeter-wave radar is the core hardware for all-weather dynamic perception in unmanned delivery vehicles, functioning as an active sensing device based on frequency-modulated continuous wave (FMCW) detection principles. It emits high-frequency millimeter-wave electromagnetic signals through its antenna; when these waves encounter obstacles, they reflect back. The radar receives the echo signals and precisely calculates the distance, relative velocity, and angular position of the obstacles by analyzing the frequency and phase differences between the transmitted and received signals, enabling real-time monitoring and trajectory tracking of moving targets.

The core formula for speed measurement using millimeter-wave radar is:

$$v = \frac{c \times \Delta f}{2 f_0} \tag{3}$$

where v is the target relative velocity (m/s), c is the speed of light, Δf is the frequency difference between the transmitted and echo signals (Hz), and f_0 is the radar's center operating frequency (Hz).

Millimeter-wave radar inherently possesses velocity sensing capability, enabling real-time detection of dynamic obstacles' velocity changes and prediction of their motion trajectories, thereby compensating for LiDAR's lack of velocity measurement and low accuracy in visual velocity estimation.

3.3.2 Technical Classification and Architectural Parameters

Most unmanned delivery vehicles are equipped with two types of millimeter-wave radars: 24GHz and 77GHz. The 24GHz radar offers low cost, moderate detection range, and strong anti-interference capability, making it ideal for short-range obstacle monitoring and blind-spot warnings. In contrast, the 77GHz radar features wider bandwidth, higher resolution, and greater velocity measurement accuracy, enabling precise perception of distant dynamic targets and suitable for high-speed dynamic obstacle avoidance on open roads. Nuro's new generation of autonomous vehicles is equipped with high-definition imaging millimeter-wave radars, capable of accurately detecting small targets at long distances even in adverse weather conditions, significantly enhancing safety in complex environments.

Key performance parameters: Detection range of 0.2–150 meters, ranging accuracy of ±0.5 meters, speed measurement accuracy of ±0.2 m/s, horizontal detection angle of ±60°, and stable operation under all weather conditions without being affected by lighting, rain, snow, fog, or sandstorms.[13]

3.3.3 Technical strengths and weaknesses

Millimeter-wave radar offers several core advantages: first, it excels in all-weather operation, unaffected by lighting conditions, rain, snow, fog, or sandstorms, providing the best environmental robustness; second, it can directly and accurately measure velocity, enabling real-time tracking of dynamic targets' movements, making it ideal for scenarios involving pedestrians and non-motorized vehicles moving unpredictably; third, its hardware is compact, low-power, moderately priced, and highly stable, making it well-suited for long-term in-vehicle deployment; fourth, it has penetration capability, allowing detection of partially occluded targets, with outstanding anti-interference performance.

Key shortcomings: First, low spatial resolution prevents the identification of obstacle contours, textures, and semantic information, making it difficult to distinguish target categories. Second, detection accuracy is limited at close range and is easily affected by ground clutter interference. Third, the system has weak capability in detecting small static obstacles, failing to meet the requirements for high-precision terrain perception.

3.3.4 Application Scenarios

Millimeter-wave radar is primarily suited for dynamic risk perception scenarios in unmanned delivery vehicles, including speed measurement and trajectory prediction of moving pedestrians, non-motorized vehicles, and other vehicles, all-weather monitoring under adverse conditions, blind spot risk warnings, and dynamic obstacle detection during high-speed driving.。[14]

3.4 Ultrasonic Sensor Perception Technology

3.4.1 Working Principle

The ultrasonic sensor is the core hardware for short-range auxiliary perception in unmanned delivery vehicles, classified as an active sensing device. Its fundamental principle relies on ultrasonic distance measurement. The sensor transmits high-frequency ultrasonic signals through its probe; when these waves encounter an obstacle, they reflect back and are received by the probe. By calculating the time difference between the emission

and reception of the ultrasonic signal and combining it with the speed of sound propagation, the system determines the short-range distance between the sensor and the obstacle, enabling detection of nearby obstacles and collision avoidance warnings. [15]

The core distance calculation formula is:

$$S = \frac{1}{2} v_s \times \Delta t \tag{4}$$

where S is the detection distance (m); v_s is the ultrasonic wave propagation velocity at room temperature, taken as 340 m/s; Δt is the round-trip time difference of the ultrasonic wave (s).

3.4.2 Technical Specifications and Deployment Method

Ultrasonic sensors offer advantages such as ultra-low cost, high stability, and no blind spots, with a detection range of 0.02–5 meters and a distance measurement accuracy of ±1 cm. They respond quickly and can precisely detect small objects at close range, including low obstacles, walls, curbs, and steps. The industry-standard deployment method involves a full-vehicle surround layout, with 8 to 12 ultrasonic sensors evenly distributed on the front, rear, and both sides of the vehicle. This setup enables comprehensive, short-range collision avoidance sensing without blind spots, serving as a reliable backup to high-end perception hardware.

3.4.3 Technical strengths and weaknesses

Core Advantages: First, extremely low cost, simple structure, low failure rate, and maintenance-free operation; second, exceptionally high detection accuracy at short distances with no blind spots, effectively compensating for the near-range perception limitations of radar and cameras; third, strong all-weather stability unaffected by lighting or weather conditions; fourth, fast response time, suitable for emergency close-range collision avoidance scenarios.

Core Limitations: First, very short detection range, only suitable for short-range sensing, incapable of medium- to long-range environmental detection; second, susceptible to environmental airflow and noise interference, resulting in significant errors at longer distances; third, lacks speed measurement, semantic recognition, and 3D modeling capabilities, enabling only basic distance measurement functions.[16]

3.4.4 Applicable Scenarios

Ultrasonic sensors serve as auxiliary sensing devices, ideal for short-range, low-speed collision avoidance in unmanned delivery vehicles, including narrow road overtaking, parking alignment, close-range obstacle detection, and blind spot protection. They are essential low-cost backup sensing hardware for unmanned delivery vehicles.

4 Research on Multi-Sensor Fusion Perception Technology for Unmanned Delivery Vehicles

4.1 Core Concept and Necessity of Multi-Sensor Fusion Technology

As revealed by the analysis of the aforementioned single-sensing technologies, each type of sensor has inherent technical limitations, making a single sensing approach unsuitable for complex urban last-mile delivery scenarios: pure vision-based systems are vulnerable to interference and have high failure risks in adverse weather; LiDAR lacks semantic understanding and speed measurement capabilities; millimeter-wave radar provides no contour or semantic information; and ultrasonic sensors have limited detection ranges. Multi-sensor fusion perception technology overcomes these limitations by integrating the multidimensional strengths of LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors through hardware协同calibration and algorithmic fusion processing. This enables comprehensive perception capabilities including "high-precision 3D modeling, accurate semantic recognition, all-weather dynamic speed estimation, and close-range blind-spot-free fallback," effectively eliminating the shortcomings of individual sensing technologies. It significantly enhances the perception robustness and accuracy of unmanned delivery vehicles in complex road conditions, harsh weather, and dynamic environments, representing the only optimal solution for current unmanned delivery vehicle perception technology.[17]

4.2 Levels of Multi-Sensor Fusion and Core Algorithms

Based on data processing levels, multi-sensor fusion in unmanned delivery vehicles is categorized into three tiers: data-level fusion, feature-level fusion, and decision-level fusion. Each tier supports different algorithms and application scenarios, forming a complete fusion technology framework.

4.2.1 Data-Level Fusion

Data layer fusion refers to the integration of raw data at the foundational level, directly synchronizing, aligning, denoising, calibrating, and fusing raw point clouds, images, distance, and velocity data collected from various sensors. This approach preserves all original data characteristics, achieving the highest perception accuracy. The core adaptation algorithm is the Kalman

filter, which effectively removes noise from raw data and enables spatiotemporal synchronization and calibration of multi-source data.

Kalman filter core iterative formula:

$$
\begin{aligned}
&\text{状态预测：} \hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \\
&\text{协方差预测：} P_{k|k-1} = AP_{k-1|k-1}A^T + Q \\
&\text{增益计算：} K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \\
&\text{状态更新：} \hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}) \\
&\text{协方差更新：} P_{k|k} = (I - K_kH)P_{k|k-1}
\end{aligned}
\tag{5}
$$

In the equation: x^ is the system state value; P is the covariance matrix; K is the filter gain; A, B, and H are system parameter matrices; Q and R are noise matrices; z is the observation value. This algorithm enables real-time filtering and precise calibration of raw data from multiple sensor sources, providing a high-quality data foundation for higher-level fusion.[18,19]

4.2.2 Feature-level Fusion

Feature-level fusion is a mid-level integration of feature information, which extracts features from various sensor data separately and then fuses the multidimensional feature information, balancing accuracy and computational speed. It is the most commonly used fusion level in unmanned delivery vehicles. Visual algorithms extract image semantic features, LiDAR captures three-dimensional spatial features, millimeter-wave radar extracts velocity features, and ultrasonic sensors obtain short-range distance features. A weighted fusion algorithm associates and reconstructs these features, enabling precise target classification, localization, and speed estimation.

4.2.3 Decision-level Fusion

The decision-making layer integrates into a high-level result fusion, where various sensors independently perform target detection and risk assessment, outputting their respective decisions. These are then fused through the D-S evidence theory algorithm to generate a final perception and decision outcome. This approach features low computational load and high fault tolerance, primarily used in risk mitigation and emergency obstacle avoidance scenarios.

4.3 Multi-sensor Fusion Architecture and Implementation Process

This paper integrates mainstream industry technologies to establish a "four-dimensional integrated" multi-sensor fusion architecture tailored for unmanned delivery vehicles—comprising "LiDAR 3D modeling + visual semantic recognition + millimeter-wave dynamic speed measurement + ultrasonic close-range backup."

The overall implementation process consists of five steps: First, synchronized data acquisition from multiple sensors, with four types of sensors collecting multidimensional environmental data in parallel while ensuring hardware clock synchronization. Second, raw data preprocessing, including noise reduction via Kalman filtering, spatiotemporal calibration, and removal of abnormal data. Third, hierarchical feature extraction, separately extracting spatial, semantic, velocity, and short-range distance features. Fourth, multi-feature fusion analysis, integrating multidimensional features using weighted fusion algorithms to achieve target detection, scene modeling, and risk prediction. Fifth, perception result output, transmitting the fused high-precision environmental data to the decision and control module to support autonomous driving and obstacle avoidance operations.

4.4 Core Advantages of Fusion Technology

Multi-sensor fusion technology effectively overcomes the scenario limitations of single-sensor perception systems, offering three core advantages: First, comprehensive perception capability, covering long and short distances, moving and stationary targets, and all-day, all-weather conditions. Second, high-precision perception, integrating 3D space, semantics, and velocity information to enable accurate target localization, classification, and speed estimation. Third, high robustness—when one sensor fails, others can compensate and ensure continuous perception, significantly enhancing vehicle safety and operational stability, making it fully suitable for complex urban last-mile delivery environments.[20,21]

5 Engineering Application Case Study of Perception Technology in Unmanned Delivery Vehicles

5.1 JD.com's Unmanned Delivery Vehicle Perception Technology Case

As a leading enterprise in China's unmanned delivery industry, JD.com has been deeply committed to end-of-line unmanned delivery technology for many years. Its sixth-generation intelligent unmanned delivery vehicle is an industry benchmark product, featuring a technical solution centered on "large-scale perception models and deep integration of multiple sensors," designed to adapt to complex delivery environments across diverse scenarios such as urban open roads, narrow community

streets, rural roads, and wholesale markets.

In terms of hardware configuration, the JD sixth-generation unmanned vehicle is equipped with a three-dimensional perception system. It features high-precision lidar mounted on the roof, multiple high-definition stereo cameras and panoramic surround-view cameras distributed across the body, and standard millimeter-wave radars and surround ultrasonic sensors on the front, rear, and sides–forming a four-layer perception architecture that relies primarily on lidar, supplements with vision, supports with radar, and provides protection via ultrasonic sensors. Technologically, the vehicle integrates a self-developed perception large model and point-cloud visual fusion technology, enabling deep integration of point-cloud and image data. This expands the perception range by 19 times compared to traditional solutions and triples overall perception performance. Additionally, it innovatively adopts a "lightweight map + strong perception" approach, eliminating reliance on high-definition maps and instead using onboard perception systems to achieve real-time mapping and environmental adaptation, significantly enhancing navigation capabilities in complex settings.

Engineering deployment results: To date, JD's unmanned delivery vehicles have been deployed at scale across more than 30 cities, over 200 parks, and communities nationwide, completing over 100 million last-mile deliveries. In challenging conditions–including narrow city streets, nighttime operations, strong backlighting or glare, light rain, and light snow–the perception accuracy reaches 99.2%, with a significant improvement in passage success rates under complex road conditions. Equipment failure rates have decreased by 40% year-on-year, effectively addressing industry challenges such as perception failure and operational lag in complex environments. The system now enables stable, round-the-clock, all-scenario delivery operations.[22]

5.2 Meituan's Autonomous Delivery Vehicle Perception Technology Application Case

Meituan's unmanned delivery vehicles primarily target high-frequency short-distance delivery scenarios in urban communities, villages within cities, and commercial districts. To address the extremely complex conditions of Chinese urban villages–such as narrow roads, chaotic obstacles, dense pedestrian and non-motorized traffic, and irregular road conditions–the company has developed a customized multi-sensor fusion perception solution.

The hardware architecture adopts the industry-standard "1 primary, 2 auxiliary" LiDAR layout. The main LiDAR is responsible for high-precision 3D modeling of the core 120-degree forward driving area, while two auxiliary LiDARs cover blind spots on the vehicle's sides and rear. Combined with 11 high-definition cameras, the system enables full-scenario semantic recognition. High-precision millimeter-wave radars are integrated to detect dynamic object speeds and predict trajectories, and ultrasonic sensors are deployed around the vehicle body to provide fallback protection during close-range navigation in narrow spaces. Through feature-level multi-sensor fusion algorithms, the system achieves real-time calibration by integrating 3D point cloud data, image semantics, and velocity information, significantly enhancing detection capabilities for low obstacles, temporary roadside clutter, and sudden pedestrian crossings.

Engineering implementation results: Meituan's autonomous vehicles have been fully deployed in urban villages and old residential communities across first-tier cities such as Shenzhen, Guangzhou, and Shanghai. They achieve a 98% passage success rate in complex urban village environments, precisely adapting to unstructured road conditions and densely populated dynamic人流 scenarios, demonstrating exceptional stability day and night. By the end of 2025, Meituan's autonomous delivery service had surpassed 800 million cumulative orders, validating the scalability, adaptability, and reliability of its perception technology. It has become a benchmark solution for autonomous delivery perception in China’s complex urban environments.。

5.3 Nuro's Autonomous Delivery Vehicle Perception Technology Application Case in the United States

Nuro is a pioneer in the global low-speed autonomous delivery sector, focusing on lightweight unmanned delivery in open urban environments overseas. Its technology emphasizes high precision, high reliability, and adaptability to harsh weather conditions, with perception capabilities leading globally. Nuro's new generation of autonomous delivery vehicles abandons traditional mechanical LiDAR systems in favor of compact solid-state LiDARs, eliminating potential mechanical failure risks and enhancing long-term stability for outdoor operations. Additionally, they are equipped with high-definition imaging millimeter-wave radars, enabling long-range, high-precision detection of small objects and ensuring stable perception of minute obstacles even in extreme weather such as heavy rain, dense fog, or sandstorms.

The vehicle’s perception system employs a deeply

integrated solution combining solid-state LiDAR, imaging millimeter-wave radar, and high-definition vision. Through proprietary fusion algorithms, it achieves millisecond-level synchronization of multi-source data, with key optimizations in predicting fast-moving dynamic targets, identifying distant obstacles, and resisting interference in extreme conditions. This allows for high-precision perception across all scenarios without relying on ultrasonic sensors, delivering industry-leading robustness and safety in perception performance.。

Engineering implementation results: Nuro's autonomous vehicles have obtained open-road autonomous operation permits in multiple U.S. states, enabling legal operation on urban public roads. They achieve stable delivery operations around the clock and in all weather conditions, with a 35% improvement in perception accuracy over traditional solutions under extreme or adverse weather conditions. The equipment downtime rate is below 0.5%, fully demonstrating the practical value of high-precision multi-sensor fusion technology and providing a reference direction for the evolution of advanced sensing solutions in the industry.[23]

6 Industry statistics and market development analysis

6.1 Overall market size of the industry

With the support of smart logistics policies and continuous advancements in sensing technologies, China's unmanned delivery vehicle industry has entered a period of rapid large-scale deployment. According to authoritative data from the "Blue Book on the Development of Low-Speed Autonomous Driving Industry (2025-2026)," China's unmanned delivery vehicle market surpassed 10 billion yuan in 2024, growing by 50% year-on-year to reach 15 billion yuan in 2025. In 2025, cumulative outdoor unmanned delivery vehicle shipments in China exceeded 39,000 units, with 27,000 newly delivered throughout the year—marking a significant leap compared to previous years. Industry institutions forecast that by 2030, China's unmanned delivery vehicle market will surpass 41 billion yuan, maintaining an annual compound growth rate above 35%, indicating substantial development potential.

Globally, in 2025, L4-level unmanned delivery vehicle shipments reached 30,000 units, with China accounting for over 40%, making it the world's largest market for both consumption and R&D of unmanned delivery vehicles. The three core application scenarios are last-mile express delivery, supermarket retail, and campus logistics, which account for 52%, 28%, and 15% respectively, while the remaining 5% covers niche areas such as healthcare and sanitation.

6.2 Analysis of the Perception Technology Market Landscape

The current industry has developed three mainstream sensing technology approaches, each tailored to different market segments: the first is the high-end fully integrated approach, represented by Nuro and JD's premium models, which employ a fusion solution combining solid-state LiDAR, high-definition millimeter-wave radar, and multi-camera vision systems. This approach offers high sensing accuracy and strong stability, suitable for open-road scenarios and targeting the high-end commercial market. The second is the mid-range cost-effective approach, exemplified by Meituan's mainstream models, using a combination of mechanical LiDAR, stereo vision, and standard millimeter-wave radar. This balances precision with cost efficiency, making it ideal for complex urban community environments and currently the dominant mass-production solution in the industry. The third is the low-end pure vision approach, used in small-scale campus delivery robots, relying on multi-camera and ultrasonic sensor configurations. With extremely low costs, this solution suits simple, enclosed environments and targets lower-tier, underserved markets.

In terms of market share, as of 2025, the domestic autonomous delivery vehicle landscape has begun consolidating, with Neolix, Jiu Shi Intelligence, JD, and Meituan emerging as core leading players. Together, these four companies account for over 90% of the market. All have successfully achieved large-scale deployment of multi-sensor fusion perception technologies. Single-sensor solutions—pure vision or pure radar—are gradually being phased out of the mainstream, and multi-sensor fusion has become an essential technical requirement across the industry.

6.3 Current Status of Sensing Technology Industrialization

With continuous iteration of domestically produced sensing hardware and open-source optimization of fusion algorithms, the cost of perception systems for autonomous delivery vehicles has steadily declined, significantly accelerating industrialization and deployment. The localization rate of LiDAR sensors rose from 40% in 2022 to 75% in 2025, while overall

sensor hardware costs dropped by 52% year-on-year, laying a solid foundation for large-scale production. Meanwhile, perception algorithms continue to be lightweighted and optimized for embedded deployment, enabling compatibility with low-computational-power onboard terminals. This addresses the traditional challenges of high computational demands and difficult implementation associated with fusion algorithms, greatly enhancing engineering practicality.

Currently, the industry has established an industrialization trend characterized by "domestically sourced hardware, lightweight algorithms, standardized fusion, and full-scenario coverage." Sensing technology is no longer a bottleneck for autonomous delivery vehicle deployment. Instead, adapting to complex environments, resisting interference in extreme weather conditions, and achieving a balance between low cost and high precision have become the key focus areas for ongoing technological evolution.[24]

7 Core Challenges in Environment Perception Technology for Autonomous Delivery Vehicles

7.1 Insufficient Robustness in Perceiving Complex and Extreme Scenarios

Current perception technologies for autonomous delivery vehicles perform stably under normal conditions but still exhibit significant shortcomings in highly complex and extreme scenarios. First, adaptability to extreme weather is poor: during heavy rain, dense fog, snowstorms, or severe sandstorms, LiDAR point cloud noise increases dramatically, visual imaging becomes blurred, and millimeter-wave radar suffers from clutter interference, leading to a sharp decline in target recognition accuracy. Second, visual semantic recognition is severely affected by complex lighting conditions—such as strong backlighting, flickering streetlights at night, or alternating shadows from trees—resulting in frequent false positives and missed detections. Third, the system struggles with detecting small, scattered objects; fine obstacles such as loose stones, debris, low barriers, or overhead cables are often overlooked, creating blind spots that pose safety risks.

7.2 Limited Synchronization Accuracy in Multi-Sensor Fusion

Multi-sensor fusion relies on precise time synchronization and spatial calibration. However, current engineering implementations still suffer from synchronization errors. Differences in sampling frequencies and response delays among sensors lead to millisecond-level timing discrepancies in data acquisition from LiDAR, cameras, and millimeter-wave radars, causing misalignment when fusing multi-source data. Additionally, long-term vehicle vibrations and outdoor variations in temperature and humidity can cause sensor mounting shifts, resulting in spatial calibration drift. This ultimately reduces fused perception accuracy, leading to target localization errors and identification confusion, thereby compromising driving stability.

7.3 Difficulty Balancing Cost and Precision

High-precision perception hardware—such as solid-state LiDAR and high-definition imaging millimeter-wave radars—is expensive, significantly increasing production costs for autonomous vehicles and hindering industry-wide adoption. In contrast, low-cost sensors offer poor precision and reliability, making them unsuitable for complex urban environments. The industry currently faces a dilemma: "high precision at high cost, low cost with low precision." How to improve perception accuracy through algorithm optimization and architectural iteration while keeping hardware costs under control remains a key challenge for industrial deployment.

7.4 Weak Algorithm Generalization Capability

Current mainstream perception fusion algorithms are primarily trained on conventional road datasets, resulting in insufficient training samples for complex localized scenarios in China—such as narrow alleys in urban villages, temporary road construction, disordered non-motorized traffic, and irregular obstacles—leading to poor algorithm generalization. These systems often fail to recognize unfamiliar environments or unconventional obstacles, making them unsuitable for the diverse and complex last-mile delivery conditions in China, with algorithm iteration lagging behind real-world deployment demands.

7.5 Insufficient Hardware Reliability and Durability

Unmanned delivery vehicles operate long-term in outdoor, open-air environments exposed to extreme temperatures, humidity, wind, sand, and rain. Traditional perception sensors have limited waterproofing, dustproofing, and shock resistance, leading to hardware aging, performance degradation, and frequent failures over time. Equipment durability and adaptability to outdoor conditions require further improvement, increasing maintenance costs and downtime losses.

8 Technical Optimization Strategies and Future Development Trends

8.1 Targeted Technical Optimization Strategies

8.1.1 Enhancing Perception Algorithms for Complex Scenarios

To address challenges such as extreme weather, complex lighting conditions, and small object detection, a localized dataset for complex scenarios should be constructed, enriched with samples of rain, snow, fog, strong light, shadows, and cluttered obstacles. The structure of deep learning models should be optimized. Large-scale models can be introduced to enhance scene semantic understanding, improving recognition and prediction capabilities for irregular obstacles and unknown environments, thereby strengthening algorithm generalization and robustness.

8.1.2 Improving Multi-source Data Synchronization and Calibration Accuracy

Establish a unified hardware clock synchronization mechanism and optimize timestamp calibration algorithms to eliminate timing discrepancies among multiple sensors. Design flexible anti-vibration mounting structures suitable for vehicle vibration conditions, perform periodic spatial calibration of sensors to correct installation offset errors, and refine Kalman filter fusion algorithms to strengthen anomaly detection, thus enhancing accuracy and stability in multi-source data fusion.

8.1.3 Building a Low-cost, High-precision Perception Architecture

Replace imported hardware with domestically produced sensors to reduce procurement costs. Optimize sensor layout by eliminating redundant components and adopting a differentiated configuration of "high-precision primary sensors + low-cost auxiliary sensors." Leverage algorithm lightweighting, intelligent data completion, and virtual sensing technologies to compensate for precision limitations of low-cost hardware, achieving an optimal balance between cost and performance.

8.1.4 Strengthening Outdoor Durability of Hardware

Select automotive-grade sensors with waterproof, dustproof, and shock-resistant properties, and optimize sensor sealing and heat dissipation designs. Implement real-time hardware health monitoring modules to detect precision degradation and faults early, enabling proactive maintenance, extending hardware lifespan, and reducing operational and maintenance costs.

8.2 Future Technological Development Trends

8.2.1 Integrated and Standardized Perception Fusion

Future unmanned delivery vehicles will move beyond single-sensor perception solutions. Deep integration of multiple sensors will become industry standard, gradually establishing unified standards for hardware layout, data fusion, and algorithm processing. This modular, standardized, and universal approach will significantly reduce R&D and mass production costs, accelerating industry-wide scale adoption.

8.2.2 Deep Empowerment by Perception Large Models

Perception large models will play a pivotal role in future development, providing powerful capabilities in scene understanding, context reasoning, and dynamic adaptation. They will enable more accurate and reliable perception in complex and unpredictable environments, driving innovation across autonomous delivery systems.

8.2.3 Hardware lightweighting, solid-state evolution, and domestication

LiDAR will fully transition from mechanical to solid-state sensors, achieving miniaturization, low power consumption, and high stability; millimeter-wave radar will upgrade toward high-definition imaging and higher resolution; vision cameras will deliver high dynamic range and ultra-clear night vision. Meanwhile, core perception hardware will be entirely replaced with domestically produced alternatives, significantly reducing equipment costs and accelerating the widespread adoption of unmanned delivery vehicles.[25]

8.2.4 Vehicle-Infrastructure Cooperative Perception Fusion Upgrad

Future unmanned delivery vehicles will no longer rely solely on autonomous onboard sensing. Instead, they will deeply integrate with smart roads, roadside perception devices, and cloud-based data platforms to achieve full-domain collaborative perception through "onboard sensing + roadside sensing + cloud-based dispatching." This integration enables early access to long-range traffic conditions and risk warnings, effectively eliminating blind spots and prediction delays in single-vehicle perception, thereby greatly enhancing driving safety and traffic efficiency.

8.2.5 Low-power edge-side intelligent deployment

By leveraging algorithm distillation, lightweight pruning, and computational optimization, high-precision fusion algorithms and large-model algorithms are efficiently deployed on low-computing-power in-vehicle terminals. This reduces the power consumption of perception systems, adapts to the power supply constraints and limited computing capabilities of autonomous vehicles, and enhances device endurance and long-term stable operation.

9 Conclusion

This paper focuses on environment perception technology for unmanned delivery vehicles, systematically organizing the overall framework of such technology. It thoroughly analyzes the working principles, core formulas, performance parameters, advantages and disadvantages, and application scenarios of four single-sensor perception technologies: LiDAR, visual cameras, millimeter-wave radar, and ultrasonic sensors. Particular emphasis is placed on the layered architecture, core algorithms, implementation processes, and technical advantages of multi-sensor fusion perception. Drawing on real-world engineering deployment cases from three leading companies—JD.com, Meituan, and Nuro—and supported by the latest 2025 industry market statistics, the study comprehensively examines the current state of engineering applications, market landscape, and industrialization trends in unmanned delivery vehicle perception technology. It precisely identifies key bottlenecks in existing technological applications, proposes targeted optimization strategies, and forecasts future directions of technological evolution.

The research concludes that due to physical limitations, each individual perception technology has inherent shortcomings in scenario adaptability and cannot meet the demands of all-weather, high-precision, and high-safety perception required for complex urban last-mile deliveries. Multi-sensor fusion technology, by integrating the multidimensional strengths of various sensing hardware, enables comprehensive coverage in spatial awareness, semantic understanding, velocity estimation, and close-range detection, representing the optimal solution at present and a core trend in the industry. The sector has now entered a phase of large-scale deployment, with the primary focus of perception technology iteration shifting from "having or lacking perception" to "precise, intelligent, low-cost, and full-domain perception." Key breakthroughs in the future will hinge on adapting to complex environments, algorithm generalization capability, balancing cost and precision, and achieving integration between vehicles and infrastructure.

With continuous advancements in artificial intelligence, sensor hardware, big data, and vehicle-infrastructure coordination, the environment perception systems of unmanned delivery vehicles will rapidly evolve toward integration, intelligence, standardization, full-domain coverage, and low cost. This will fundamentally resolve challenges in perceiving complex environments, significantly enhance the safety, stability, and efficiency of autonomous delivery operations, and continuously drive the intelligent and unmanned transformation of last-mile logistics. It will provide solid technical support for high-quality development in smart logistics and digital city construction. Furthermore, the findings of this study offer systematic theoretical foundations and practical references for mid-level engineering professionals engaged in R&D, optimization, and deployment of intelligent unmanned equipment, demonstrating significant value in both engineering applications and academic research.□

REFERENCES

[1] Ge Xinjian. Research on Navigation and Obstacle Avoidance Technology for Unmanned Delivery Vehicles in Campus Environment [D]. Jinan: Shandong Jianzhu University, 2024.

[2] Ge Xinjian. A Review of Intelligent Driving Technology in Outdoor Environments [J]. Computer Measurement & Control, 2024, 32(10): 1-8.

[3] Li Yule, Zhang Xiaoyu, Liu Na. Research on an Improved Scheme for Intelligent Campus Logistics Management Based on Unmanned Delivery Vehicles [J]. Logistics Technology, 2025, 48(13): 57-61.

[4] Research on Environmental Perception Algorithms for Low-Speed Unmanned Logistics Vehicles Based on LiDAR [D]. 2020.

[5] Research on the Design of Unmanned Delivery Vehicles Based on Context Awareness [D]. Nanjing: Southeast University, 2021.

[6] Vermesan O, Bahr R, Sand H E, et al. Advancing edge AI perception platforms and sensor fusion for last-mile delivery autonomous vehicles[M]//AI and Edge Computing for Autonomous Vehicles. London: Taylor & Francis, 2026.

[7] Du H, Ren L, Wang Y D, et al. Advancements in perception system with multi-sensor fusion for embodied agents[J]. Information Fusion, 2025, 115: 102637.

[8] Sensor systems for autonomous vehicles: functionality and reliability challenges in adverse environmental conditions[J]. Measurement, 2025, 242: 116538.

[9] The automation concept of the U-Shift II vehicle[J]. SAE Technical Papers, 2025, 1: 2025-01-7023.

[10] LogisticsVISTA: 3D terminal delivery services with UAVs, UGVs and USVs based on foundation models and scenarios engineering[C]//2024 IEEE Intelligent Vehicles Symposium (IV). Macau: IEEE, 2024.

[11] In-Depth Analysis of the Core Technology System for Autonomous Delivery Vehicles: Full-Chain Technological Breakthroughs from Perception to Energy[EB/OL]. (2025-08-15)[2026-07-01]. https://four-faith.com/jishuyingyong/1653.html.

[12] Upstream Technology Analysis of the Autonomous Delivery Vehicle Industry Chain[EB/OL]. (2025-08-01)[2026-07-01]. https://www.automotiveworld.cn.

[13] Unmanned Vehicles Toward Embodied Intelligence: Perception Enhancement and End-to-End Decision-Making [J]. New Electronics Technology Magazine, 2025, (12): 1-10.

[14] Liu Yang, Lei Ting. Design of an Intelligent Multifunctional Campus Unmanned Logistics Vehicle [J]. Mechanical & Electrical

Products Development and Innovation, 2024, 37(5): 53-57.
[15] Zhang Yuzheng. Exploring the Role of Artificial Intelligence in Modern Logistics Management Models [J]. Logistics Technology, 2022, 45(10): 105-107.
[16] Xue Xiaowei. Research on the Application of Artificial Intelligence in Modern Logistics Systems [J]. China Logistics and Procurement, 2024, (9): 106-107.
[17] Autonomous Vehicles and Drones: 5G Accelerates the Arrival of Intelligent Logistics Era [J]. Logistics Technology, 2019, 42(7): 1-10.
[18] Yang Ping, Tu Guangming. Exploring Innovative Directions for Modern Logistics Management in the Age of Artificial Intelligence [J]. Logistics Engineering and Management, 2021, 43(3): 68-70.
[19] Intelligent Mobile Robot Systems Facilitate Smart Upgrading of Factory Logistics [J]. Automation Expo, 2020, 37(7): 36-40.
[20] Xiao Gangchao, Zhu Xiaoming, Li Zubiao, et al. Research and Application of a New Flexible Distribution Service Model for Huaihua Tobacco Logistics [J]. China Market, 2022, (31): 173-177.
[21] Yao Changqing. A Process-Based Strategy for Flexible Logistics Design in Tobacco Distribution Centers [J]. Chongqing and the World (Academic Edition), 2015, 32(12): 138-141.
[22] Ma Libing. Artificial Intelligence Drives Transformation of Modern Logistics Enterprises—Taking Warehouse-Based Logistics Companies as Examples [J]. China Storage and Transportation, 2019, (9): 142-144.
[23] Li B, Liu S S, Tang J, et al. Autonomous last-mile delivery vehicles in complex traffic environments[J/OL]. arXiv preprint, 2020, arXiv:2001.08096. https://arxiv.org/abs/2001.08096.
[24] Meituan Unmanned Delivery Department. Visual Global Localization Technology Based on Feature Points[EB/OL]. (2019-04-26). https://cloud.tencent.com/developer/article/1418690.
[25] Tsinghua University, Huazhong University of Science and Technology, Beijing SanKuai Online Technology Co., Ltd. (Meituan), Beijing Jingdong Ganshi Technology Co., Ltd. (JD.com), et al. Intelligent Perception and Navigation Control Technology for Heterogeneous Autonomous Unmanned Systems in Complex Multi-Domain Environments [R]. Beijing: China Association of Automation Scientific and Technological Achievements Appraisal Conference,2024.

Author Introduction:Li Yanze, male, engineer, mainly researches power systems and coating technology.

Copyright by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Application of Smart Valves in Wastewater Treatment Systems

TAN Tu
(Ocean Source Trading Co., Ltd., Hong Kong, China)

Abstract:This paper systematically outlines the overall structural composition, core classification criteria, and closed-loop operating principles of smart valves, and specifically investigates their application methods, operational logic, and practical value in five key process modules: influent control systems for wastewater treatment, biological aeration systems, sludge recirculation systems, chemical dosing systems, and remote intelligent monitoring systems. From an engineering application perspective, it deeply analyzes the core technological advantages of smart valves over traditional valves, and objectively identifies current challenges in equipment promotion, including high cost barriers, limited adaptability to complex operating conditions, shortage of professional maintenance personnel, insufficient system compatibility, and weak data security. Finally, considering industry trends such as smart water management, dual-carbon policies, and intelligent manufacturing, the paper forecasts future development directions for smart valves, including higher precision and durability, full-domain intelligence, industry standardization, and green, energy-efficient integration.
Key words:Smart valve; wastewater treatment; automatic control; process system; smart water management; Internet of Things

1 INTRODUCTION

As urbanization in China continues to accelerate and the industrial system steadily improves, the discharge of municipal wastewater and industrial effluent has been increasing year by year, intensifying pressure on water environment management. In recent years, the government has successively issued multiple policy documents, including the "Three-Year Action Plan for Improving the Quality and Efficiency of Urban Wastewater Treatment" and the "Technical Guidelines for Smart Water Management," clearly requiring the wastewater treatment industry to abandon extensive operation and maintenance models and promote automated and intelligent upgrades of treatment facilities to achieve multiple goals such as improved treatment quality, energy conservation, consumption reduction, and enhanced efficiency. As a core component of water resource recycling, aquatic ecosystem restoration, and water environment management, the operational stability, treatment accuracy, and energy consumption control of wastewater treatment directly affect regional ecological quality and water resource utilization efficiency.

A complete modern wastewater treatment process includes several interrelated and mutually influencing unit operations: influent pretreatment, biological treatment, solids-liquid separation, sludge disposal, chemical dosing, and treated effluent discharge. Operations such as fluid conveyance, air flow regulation, chemical dosing, and sludge recirculation across these units all require precise control via valve equipment. The operating status and control precision of valves directly determine the overall system's operational stability, effluent compliance rate, and energy consumption levels. For a long time, most small- and medium-sized wastewater treatment plants and older municipal stations in China have commonly used traditional manual valves or conventional electric mechanical valves as core control devices. These devices offer only basic functions such as simple pipeline on/off control and fixed opening adjustment, unable to adapt in real time to dynamic fluctuations in influent quality and flow, resulting in significant operational limitations. In practical engineering operations, this often leads to problems such as uneven aeration airflow distribution, excessive or insufficient chemical dosing, imbalanced sludge return ratios, sudden pressure fluctuations in pipelines, and abnormal tank level variations. These issues not only reduce treatment efficiency and cause effluent parameters like COD, ammonia nitrogen, and total phosphorus to exceed standards but also result in substantial waste of electricity and chemicals, increase equipment wear and maintenance costs, and hinder high-quality development of the wastewater treatment sector.

Against the backdrop of widespread adoption of intelligent manufacturing and Internet of Things (IoT) technologies, smart fluid control equipment is rapidly evolving, with intelligent valves gradually replacing traditional mechanical valves as a key component of smart wastewater treatment systems. Intelligent valves

are integrated smart devices combining mechanical valve bodies, intelligent actuators, high-precision sensor modules, embedded control chips, and IoT communication units. They overcome the functional limitations of conventional valves by integrating data perception, intelligent computation, precise actuation, fault diagnosis, remote transmission, and automatic early warning capabilities. Relying on online monitoring data and adaptive control algorithms, they enable unmanned, precise control throughout the entire wastewater treatment process, effectively adapting to complex and variable wastewater conditions and significantly improving the level of fine-grained operation and maintenance.

Currently, intelligent valves have been widely applied in various water environment management projects, including municipal domestic wastewater treatment, chemical wastewater treatment, textile wastewater treatment, and aquaculture wastewater treatment, demonstrating clear advantages in ensuring stable effluent quality, reducing operational energy consumption, lowering labor costs, and enhancing system automation. However, due to constraints such as equipment cost, operating environments, industry standards, and talent availability, the penetration rate of intelligent valves remains low in small- and medium-sized wastewater treatment plants and rural sewage stations. Moreover, their operational stability under harsh conditions involving severe corrosion, high impurities, and humid environments still presents certain shortcomings. Challenges remain regarding device compatibility, operation and maintenance systems, and data security—issues that urgently need resolution. Based on the current industry landscape, this paper systematically investigates the application of smart valves in wastewater treatment systems. It comprehensively elaborates on the structural principles of the equipment, specific application scenarios, core technological advantages, and existing challenges. By integrating industry policies and technological evolution trends, it clarifies the future upgrade directions for smart valves, providing thorough theoretical support and practical guidance for intelligent transformation projects in wastewater treatment.。[1]

2. Basic Components, Classification, and Working Principles of Smart Valves

2.1 Basic Components of Smart Valves

Smart valves are a new generation of fluid control devices that integrate mechatronics and IoT-based intelligence. Compared to traditional valves, which have simple structures and limited functions, smart valves abandon purely mechanical drive and manual operation, instead adopting an integrated architecture combining "mechanical actuation, intelligent sensing, computational control, and remote transmission." The entire system features a sophisticated design and comprehensive functionality, with independent yet coordinated modules working together to achieve precise fluid parameter regulation and intelligent equipment status management. The core components mainly consist of four parts: the mechanical valve body structure, intelligent actuator, sensor detection module, and intelligent control and communication module. Below is a detailed description of each module's specific functions and structural characteristics.

First, the mechanical valve body structure. As the fundamental mechanical framework of the smart valve, it directly contacts the sewage medium and serves as the core actuating component for pipeline on/off control and fluid regulation. It primarily comprises the valve housing, valve core, seat, sealing gasket, stem, and limit mechanism. Depending on the medium characteristics and control requirements of different wastewater treatment processes, common valve types include butterfly valves, gate valves, ball valves, globe valves, control valves, and check valves. Butterfly valves feature simple structure, compact size, and excellent linear regulation performance, with nominal diameters ranging from DN50 to DN1400 and operating pressures from 0.1 to 1.6 MPa. They are widely used in aeration air ducts and sewage conveyance pipelines. Gate valves offer high flow capacity and low resistance coefficient (≤0.05), with no throttling dead zones, making them ideal for controlling main inlet and outlet pipelines. Control valves provide high-precision adjustment and support stepless fine-tuning, primarily applied in scenarios requiring precise dosing of chemicals or sludge recirculation. To accommodate the characteristics of wastewater—such as high suspended solids, strong corrosiveness, and tendency to scale—the valve bodies are typically made of QT450 ductile iron, 304/316L stainless steel, or rubber-lined corrosion-resistant materials. Sealing components use wear-resistant and aging-resistant rubber materials such as EPDM and PTFE. Standard models operate within a temperature range of 0–80°C, while special corrosion-resistant versions can withstand temperatures from -20°C to 120°C, significantly enhancing the device's resistance to corrosion, clogging, and wear, enabling long-term

continuous operation under complex wastewater conditions.

Second, the intelligent actuator. The actuator is the power core of the smart valve, responsible for receiving control instructions from the microcontroller and driving the valve stem and core to perform opening/closing, fine adjustment, and reset actions, directly determining the valve's response speed and control accuracy. Currently, smart valve actuators used in wastewater treatment applications are mainly divided into two types: electric and pneumatic actuators. Electric actuators use servo motors as the power source, combined with precision gear reduction mechanisms, offering standard 0–100% stepless stroke adjustment with a resolution up to 0.1%. Full-stroke response time ranges from 5 to 30 seconds, providing high precision, smooth operation, convenient control, and strong adaptability, making them the most widely used type in modern wastewater treatment systems. Pneumatic actuators utilize compressed air at 0.4–0.6 MPa, featuring simple structure and explosion-proof ratings up to Ex d II BT4, with low failure rates. They are primarily used in chemical wastewater and flammable/explosive environments. Smart actuators incorporate built-in limit protection, overload protection, and torque protection mechanisms, with rated torques covering 50–20,000 N·m, effectively preventing damage caused by over-travel, jamming, or excessive loading, thereby improving operational safety and stability.

Third, the sensor detection module. Serving as the "sensory nervous system" of the smart valve, this module forms the data foundation for intelligent control. Composed of various high-precision sensors and signal acquisition units, it continuously collects real-time operational parameters from pipelines and equipment around the clock. The integrated sensors in the intelligent valve specifically designed for wastewater treatment mainly include flow sensors, pressure sensors, level sensors, temperature sensors, valve position feedback sensors, and overload sensors. All sensor elements feature an IP67 protection rating, making them suitable for damp, dusty, and water-immersed conditions in industrial plants. The flow sensor achieves a measurement accuracy of ±0.5% FS with a range covering 0.5 to 2000 m^3/h, enabling real-time monitoring of wastewater, chemical dosing, and airflow rates in pipelines. The pressure sensor has a range of 0–2.5 MPa and an accuracy of ±0.2% FS, monitoring changes in pipeline water and air pressure to predict risks such as pipe blockage, pressure loss, or bursting. The level sensor offers a distance measurement accuracy of ±1 mm, ideal for monitoring liquid levels in regulating tanks, aeration tanks, and sedimentation tanks, providing data support for influent control. The valve position feedback sensor features a resolution of 0.1%, precisely capturing the current valve opening degree in real time and converting mechanical movements into electrical signals transmitted back to the control chip, enabling closed-loop regulation. All sensors operate at sampling frequencies between 10 and 100 Hz, accurately capturing minor fluctuations in operating conditions with extremely low error rates, thus delivering precise data support for intelligent system computation and dynamic adjustment.[2]

Fourth, the intelligent control and communication module. This module serves as the "central nervous system" of the smart valve, primarily composed of an embedded high-speed PLC control chip, a high-precision signal processing unit, an optimized PID algorithm program, a fault diagnosis unit, and an IoT communication unit. The control chip receives 4–20 mA analog signals and RS485 digital signals, performing signal conversion, data computation, and logical analysis. By integrating with preset process parameter thresholds in the system, it generates precise control commands to drive actuators for adaptive regulation. Equipped with an optimized PID adaptive algorithm, it achieves an overshoot of ≤3% and a steady-state error of ≤1%, effectively resolving issues such as overshooting, lag, and oscillation common in traditional control systems, significantly enhancing control stability. Meanwhile, the communication module supports multiple protocols including 4G, 5G, LoRa, Modbus-RTU/TCP, and Profibus, with a communication delay of ≤200 ms. It enables seamless integration with wastewater treatment plant DCS distributed control systems, PLC control cabinets, and cloud-based smart water management platforms, allowing real-time data upload, remote parameter tuning, equipment status monitoring, and proactive fault alerts—fully realizing unmanned, intelligent operation and management of valve equipment. Additionally, the module features a built-in 128 MB local storage unit capable of storing over 100,000 operational data points and operation logs, with data retention lasting more than five years, meeting requirements for condition traceability, process optimization, and maintenance analysis. The device operates reliably within an ambient temperature range of -10°C to 60°C and relative humidity up to 95% without condensation, making it well-suited for the harsh operating conditions of wastewater treatment plants.

2.2 Main Categories of Smart Valves

Based on the application scenarios and control requirements of the wastewater treatment industry, smart valves can be categorized according to drive type, regulation function, communication mode, and medium compatibility. Different valve types exhibit distinct performance characteristics and are suited to specific application environments. By drive type, they can be classified into smart electric valves, smart pneumatic valves, and smart hydraulic valves. Smart electric valves are suitable for most municipal wastewater conditions; smart pneumatic valves are commonly used in explosion-proof and corrosion-resistant environments; while hydraulic valves are ideal for large-diameter, high-flow-rate, high-pressure pipelines. Based on regulation function, they fall into two categories: smart on/off valves and smart modulating valves. On/off valves are used solely for pipeline isolation and are typically applied as main influent or discharge valves; modulating valves enable fine, stepless adjustment of opening degree, making them suitable for precise control applications such as aeration, chemical dosing, and sludge recirculation. In terms of communication mode, smart valves can be divided into wired communication valves and wireless IoT valves. Wired communication offers strong stability and low latency, making it ideal for fixed equipment within treatment plants; wireless valves offer flexible deployment and lower cabling costs, making them well-suited for remote rural wastewater stations. Accurately differentiating valve types can significantly improve the precision of equipment selection in wastewater treatment systems and ensure optimal process compatibility.

2.3 Working Principle of Smart Valves

Smart valves eliminate the traditional mode of manual judgment and operation, relying instead on multi-module coordinated interaction to form a fully closed-loop intelligent control logic that includes real-time sensing, signal conversion, data processing, precise execution, feedback correction, and data uploading. This enables fully automated, self-adaptive, and high-precision fluid control, ensuring seamless operation, rapid response, and accurate regulation throughout the entire process.

During normal operation, the sensing and detection module continuously collects real-time operational parameters such as flow rate, pressure, liquid level, temperature, and valve opening from the sewage treatment pipeline network and tanks. It converts the acquired analog signals into digital electrical signals, which are then transmitted in real time via signal transmission lines to the core control chip. The control chip contains pre-set optimal threshold values for wastewater treatment processes and an optimized PID control algorithm. It rapidly processes, compares, and analyzes the massive volume of incoming real-time data to accurately determine whether the current system operating conditions are stable and optimal, identifying issues such as fluctuations in water quality or flow, abnormal pipeline pressure, or valve position deviations.

When the system detects that operating conditions deviate from preset standards—such as sudden increases in influent flow, low dissolved oxygen levels in aeration tanks, insufficient chemical dosing flow, imbalanced sludge return flow, or excessive pipeline pressure—the control chip quickly generates corresponding adjustment commands. It precisely controls the intelligent actuators to dynamically adjust valve openings or on/off states, instantly modifying fluid flow rates, air pressure, and water pressure to correct operational deviations promptly. After adjustments, the valve position feedback sensors collect updated valve positions and fluid parameters in real time and transmit them back to the control chip, forming a closed-loop feedback mechanism. Based on this feedback, the system further refines its control accuracy, avoiding over-correction or under-correction through iterative optimization until the system returns to its optimal operating range, ensuring stable and reliable performance throughout the wastewater treatment process.

Meanwhile, the intelligent communication module synchronously collects data on equipment status, operational parameters, adjustment records, and fault information, uploading it in real time to both local control panels and remote cloud monitoring platforms. This enables data visualization, cloud storage, generation of operational trend curves, and statistical reporting. When faults occur—such as valve core clogging, actuator overload, signal interruption, sensor anomalies, or excessive pipeline pressure—the system's built-in fault diagnosis program quickly identifies the type and location of the fault, automatically triggering audible and visual alarms and sending alerts via mobile notifications to maintenance personnel. This ensures early detection, prompt troubleshooting, and rapid repair, minimizing the impact of equipment failures on wastewater treatment operations. The entire operational process requires no manual intervention and can run continuously 24/7

with self-adaptive capabilities, perfectly adapting to the dynamic fluctuations typical in wastewater treatment environments.[3,4]

3. Specific Applications of Smart Valves in Various Wastewater Treatment Systems

3.1 Application in the water intake control system

The influent control system is the first and most critical process in wastewater treatment, responsible for sewage collection, homogenization and stabilization, flow distribution, and pre-treatment buffering. The operational stability of this entire system directly determines the effectiveness of subsequent biological treatment and sludge-water separation processes. The influent conditions for municipal domestic wastewater and industrial effluent are highly random and fluctuating, with significant differences in discharge volume and pollutant concentration between day and night, morning and evening, and on holidays. During daytime hours, when residential activities and industrial production are at peak, influent flow rates and pollutant concentrations are high; whereas at night, flow decreases significantly and water quality load drops. In rainy seasons, stormwater entering the sewer network causes sudden increases in influent volume and dramatic fluctuations in water quality due to dilution. Traditional wastewater treatment plants rely on manual adjustment of conventional valves to regulate influent flow, which suffers from severe lag and cannot respond in real time to changing conditions. This often leads to system overloading or underutilization, resulting in issues such as clogging in pre-treatment units, imbalance in microbial activity in biological systems, and reduced treatment efficiency.

Intelligent valves are primarily deployed at key locations within the influent control system, including main influent pipelines, branch lines to equalization tanks, distribution channels to various treatment units, and overflow lines. Combined with high-precision level sensors, flow sensors, and pressure sensors, they form an intelligent influent regulation system that delivers five core functions: constant liquid level, constant flow rate, over-limit flow restriction, anti-backflow protection, and balanced water distribution. During normal operation, the intelligent valve continuously collects data on instantaneous influent flow, equalization tank water level, and pipeline pressure, dynamically adjusting its opening via built-in algorithms. When influent flow peaks—causing rapid rise in the equalization tank level approaching warning thresholds—the valve automatically reduces its opening slightly to limit flow and stabilize pressure, slowing down the inflow rate and preventing downstream equipment such as bar screens, grit chambers, and biological reactors from becoming overloaded, clogged, or failing to meet treatment standards. Conversely, during low-flow periods, when the equalization tank level drops too low, the valve automatically increases its opening to ensure continuous and stable influent supply, maintaining uninterrupted plant operation and avoiding problems such as sludge aging and microbial die-off caused by insufficient influent.

In special conditions such as sudden heavy rainfall and abrupt increases in pipeline pressure during the rainy season, intelligent influent valves can coordinate with overflow line valves to automatically open, diverting excess wastewater to relieve pressure and prevent accidents like pipeline bursts, backflow, or flooding within the plant. Additionally, for multiple parallel treatment units, intelligent valves can precisely allocate influent flow among units, ensuring even load distribution and avoiding overloading or idling of individual units. This significantly enhances the stability and shock resistance of the pre-treatment system, providing consistent and reliable influent conditions—both in terms of flow and quality—for subsequent biological treatment processes, thereby reducing process variability at the source.

Engineering Application Case 1: Intelligent Influential Control Retrofit Project at Zhangzhou Changtai Wastewater Treatment Plant

The Zhangzhou Changtai Wastewater Treatment Plant is the first fully intelligent, upgraded primary A-level wastewater treatment facility in Fujian Province, designed to treat 150,000 tons of sewage per day. The plant faces typical challenges such as diurnal flow fluctuations, sudden increases in flow during rainy seasons, and significant water quality variations caused by mixed industrial discharges. Traditional manual valve control is slow and prone to uneven influent distribution, imbalanced tank levels, and clogging in pretreatment units. This upgrade involved installing 68 DN600–DN1200 anti-scaling intelligent electric butterfly valves across the main influent pipeline, distribution channels of each treatment unit, and overflow lines. Rated for 1.6 MPa operating pressure and IP67 protection, these valves feature a high-curvature streamline flow path and nano-scale anti-scaling coating, reducing sludge adhesion by 80% and making them ideal for handling high-impurity, high-suspended-solid wastewater conditions. The entire intelligent influent control system integrates with

level, flow, and pressure sensors, relying on ±0.2% FS high-precision data collection to achieve self-adaptive regulation of constant flow and constant liquid level. After the upgrade, influent flow distribution uniformity improved by 40%, rain-season overflows and backflow incidents were completely eliminated, and the failure rate of pretreatment unit clogging dropped by 90%. Overall influent stability has significantly increased, laying a solid foundation for stable operation of downstream biological treatment processes and ensuring continuous compliance with effluent discharge standards throughout the plant.[5,6]

3.2 Application in aeration systems

The aeration system is the core unit of wastewater biological treatment and also the most energy-consuming process in wastewater treatment plants. Its primary function is to deliver air into the aeration tank via blowers, supplying oxygen to aerobic microorganisms to ensure effective degradation and removal of pollutants such as organic matter, ammonia nitrogen, and total phosphorus in wastewater. Industry data indicate that blower energy consumption in aeration systems accounts for 50%–60% of the total operational energy use in wastewater treatment plants, making it a critical area for energy conservation and efficiency improvement. Dissolved oxygen (DO) concentration in the bioreactor is the key control parameter of the aeration system, directly affecting microbial activity and treatment performance: insufficient DO leads to inadequate oxygen supply, significantly reducing pollutant degradation efficiency and resulting in excessive COD and ammonia nitrogen levels in effluent; conversely, excessively high DO causes unnecessary energy waste and may trigger operational issues such as sludge bulking, poor settling properties, elevated suspended solids in effluent, and sludge expansion.

Traditional aeration systems typically operate with fixed valve openings and constant airflow, unable to adjust aeration rates according to variations in influent water quality and load. Regardless of influent pollutant concentrations, blowers and valves remain at fixed operating conditions—either leading to insufficient oxygen supply and substandard treatment results or causing over-aeration and significant energy waste, resulting in poor process stability and energy efficiency. In contrast, intelligent valves integrated with online dissolved oxygen monitoring devices and blower variable frequency drive (VFD) control systems form a closed-loop intelligent aeration regulation system, effectively addressing the long-standing challenges of conventional aeration methods. Smart air volume control valves are evenly distributed across the branch ducts of different zones within the aeration tank, enabling zonal, fine-grained airflow control and resolving the traditional problem of global control with localized imbalances.

During operation, dissolved oxygen sensors continuously monitor DO levels at the front, middle, and rear sections of the aeration tank, transmitting real-time data to the control system. The smart valves dynamically adjust valve openings using optimized PID algorithms to precisely regulate aeration volumes in each zone. When influent concentrations of organic matter or ammonia nitrogen rise, increasing microbial oxygen demand and lowering DO levels, the valves automatically open wider to increase airflow, rapidly replenishing oxygen and maintaining optimal microbial activity for pollutant degradation. Conversely, when influent load decreases, microbial oxygen consumption drops, and DO exceeds required levels, the valves automatically reduce opening to lower aeration rates, eliminating unnecessary energy consumption. Through dynamic control by smart valves, DO concentration in the bioreactor can be stably maintained within the optimal range of 2.0–2.8 mg/L, precisely matching the metabolic needs of microorganisms.[7]

Moreover, the intelligent valves enable differentiated aeration control across zones—increasing aeration in high-pollutant-load areas near the influent end and reducing it in low-load regions toward the effluent end—thereby solving the imbalance issue of traditional aeration systems, where aeration is insufficient at the front and excessive at the rear. Practical engineering data show that aeration systems equipped with smart valves can reduce blower energy consumption by 25%–35%, significantly improving energy efficiency in wastewater treatment. At the same time, they stabilize biological treatment conditions, minimize occurrences of sludge bulking and effluent quality fluctuations, and substantially enhance compliance with discharge standards.

Case Study 2: Precision Aeration Energy-Saving Retrofit Project at Hangzhou Qianjiang Water Treatment Plant

The Phase IV project of Hangzhou Qianjiang Water Treatment Plant employs a conventional activated sludge process. The plant's aeration system has long operated in a fixed-valve mode with constant airflow, resulting in severe over-aeration and significant energy waste. Additionally, dissolved oxygen (DO) distribution

across different zones of the biological reactor is uneven—oxygen supply is insufficient at the front end and excessive at the rear end—leading to unstable pollutant degradation efficiency and poor sludge settling performance. To address these issues, the plant implemented an intelligent precision aeration upgrade by replacing all branch air supply pipes in the aeration tanks with smart modulating butterfly valves. These valves offer a control accuracy of ±1% and a full-stroke response time of ≤5 seconds. Combined with distributed online DO monitoring equipment, the system forms a closed-loop precision aeration network integrating "DO data acquisition, intelligent valve control, and variable-frequency blower coordination." Relying on an optimized PID algorithm (with overshoot ≤3% and steady-state error ≤1%), the system dynamically adjusts valve openings based on real-time water quality and load conditions in each zone, enabling demand-based, precise oxygen supply. After the upgrade, dissolved oxygen concentration in the biological reactor is stably maintained within the optimal range of 2.0–2.5 mg/L, significantly improving DO uniformity throughout the tank. Effluent compliance rates for ammonia nitrogen and COD have reached 100%. Overall aeration energy consumption decreased by 32%, with annual electricity savings exceeding 200,000 kWh per biological reactor. The results demonstrate remarkable energy efficiency improvements and provide a benchmark model for intelligent energy-saving upgrades in large-scale wastewater treatment plant aeration systems.[8]

3.3 Application in the sludge recycle system

The sludge return system is a critical component of the activated sludge wastewater treatment process. Its core function is to convey high-concentration activated sludge settled at the bottom of the secondary clarifier back to the front end of the aeration tank via return pipelines, continuously replenishing microbial populations in the biological reactor and maintaining a stable mixed liquor suspended solids (MLSS) concentration, thereby ensuring continuous, stable, and efficient biological reactions. The sludge return ratio is the key process parameter of this system; its stability and accuracy directly determine the total microbial biomass, sludge activity, and pollutant degradation capacity within the biological system. In conventional municipal wastewater treatment processes, the optimal sludge return ratio typically ranges from 50% to 100%. If the return ratio is too low, insufficient sludge concentration in the aeration tank results in inadequate microbial population, preventing effective pollutant degradation. Conversely, an excessively high return ratio leads to sludge accumulation, shortened sludge retention time, microbial aging, increased pipeline energy consumption, and equipment wear, and may easily trigger sludge bulking and poor solid-liquid separation.

Traditional sludge return systems rely on manually fixed valve settings, where operators adjust valves periodically based solely on experience. This approach suffers from low control precision and significant lag, making it unsuitable for dynamic changes in influent quality and flow rate. During fluctuations in influent load, the return ratio cannot respond promptly, resulting in weak shock resistance of the biological system and unstable effluent quality. With the widespread application of intelligent control valves in sludge return pipelines, an adaptive return regulation system can be established by integrating data from MLSS sensors in the aeration tank, influent flow sensors, and interface monitoring devices in the secondary clarifier, enabling dynamic and precise control of the return ratio.

In actual operation, the smart valve continuously collects key parameters such as return sludge flow rate, MLSS concentration in the aeration tank, and real-time influent flow. Based on a process model, it calculates the optimal return ratio in real time and dynamically adjusts the valve opening. When influent pollutant concentrations are high and hydraulic loads increase, the system automatically increases the valve opening to boost sludge return flow, enhancing microbial biomass in the aeration tank and improving pollutant degradation efficiency, thus ensuring stable treatment performance under high-load conditions. Conversely, when influent quality is good and load is low, the system reduces return flow to prevent excessive sludge accumulation and microbial aging, minimizing unnecessary energy consumption and pipeline wear. Additionally, the smart valve employs stepless, smooth modulation, avoiding water hammer, sludge flow interruption, and flow oscillations caused by abrupt opening or closing of traditional valves, ensuring continuous and stable sludge return flow.

Moreover, the smart valve features intelligent anti-clogging control. Given that sludge media often contain impurities and are prone to sedimentation, the valve can periodically open slightly to flush the pipeline, effectively reducing clogging incidents and lowering maintenance frequency. Through precise control by the smart valve, the MLSS concentration in the aeration tank can be stably maintained within the optimal range of 3,000–4,000 mg/L, significantly improving the operational stability

and shock resistance of the biological system, reducing process failure rates, and ensuring consistent solid-liquid separation and effluent quality.

Engineering Application Case 3: Intelligent Retrofit of Sludge Recirculation System at a Sewage Treatment Plant in Coastal Fujian

A municipal wastewater treatment plant located along the coast of Fujian Province is adjacent to an industrial zone and a marine area. The wastewater contains high salinity, abundant impurities, and strong corrosiveness, making the sludge return pipelines highly susceptible to scaling, sedimentation, and corrosion. The original manually operated control valves had a regulation error of ±8% to ±10% and a response delay exceeding 40 seconds, resulting in poor control accuracy and weak anti-clogging capability. This led to long-term imbalance in the sludge return ratio, causing significant fluctuations in sludge concentration within the biological reactor, frequent occurrences of sludge bulking, and excessive suspended solids in effluent–common process failures. To address these complex operating conditions, the plant replaced all sludge return pipeline valves with intelligent, corrosion-resistant control valves equipped with self-cleaning functions. The valve bodies feature a dual-seal, 316L stainless steel corrosion-resistant design, suitable for high-salinity, high-impurity sludge media. With a nominal pressure rating of 1.0 MPa and a regulation resolution of 0.1%, the valves incorporate an adaptive sludge return control algorithm that dynamically optimizes the return ratio by integrating data from sludge concentration and influent flow sensors. Additionally, the system includes a scheduled flushing function, which automatically opens and closes the valves at set intervals daily to clean the pipelines, effectively eliminating clogging issues caused by sludge accumulation. After the upgrade, the control accuracy of the sludge return ratio improved to ±2%, the fluctuation range of mixed liquor suspended solids (MLSS) in the aeration tank decreased by 70%, and incidents of sludge bulking and pipeline blockages were nearly eliminated. The biological treatment system's resistance to shock loads significantly increased, reducing annual process-related downtime by 95% and substantially lowering operational labor costs.[9]

3.4 Application in chemical dosing systems

The chemical dosing system is a critical process for achieving wastewater treatment compliance and advanced purification, encompassing multiple sub-processes such as coagulant dosing, disinfectant dosing, acid-base neutralization, and phosphorus removal. By adding chemicals including polyaluminum chloride, polyacrylamide, sodium hypochlorite, pH adjusters, and phosphorus removal agents, the system effectively removes suspended solids, colloids, heavy metals, bacteria, viruses, and total phosphorus from wastewater, while simultaneously regulating water pH to ensure treated effluent meets discharge standards. The accuracy of chemical dosing directly determines treatment efficiency, chemical cost, and water safety: insufficient dosing fails to achieve adequate purification, disinfection, or neutralization, resulting in超标 effluent parameters; excessive dosing not only wastes significant chemical resources and increases operating costs but also leads to chemical residues, secondary pollution, and pH imbalance.

Traditional chemical dosing systems typically operate with fixed flow rates and manual time-based adjustments, making them ill-suited to dynamic fluctuations in influent water quality. During periods of severe pollution, dosing may be inadequate, while during periods of good water quality, over-dosing occurs–resulting in long-term issues such as chemical waste, unstable treatment performance, and high risks of secondary contamination. By integrating smart valves with online water quality sensors, chemical flow meters, and dosing pumps, an automated, precise dosing control system can be established, enabling a refined operational model characterized by "on-demand dosing, dynamic adaptation, and accurate dosage control."

During operation, sensors for COD, ammonia nitrogen, turbidity, and pH continuously monitor key parameters such as pollutant concentration, water turbidity, and pH levels. Based on real-time data and process models, the control system precisely calculates optimal chemical dosing rates and automatically adjusts the opening of smart control valves to continuously modulate chemical feed flow. For coagulation and sedimentation processes, the system automatically increases chemical dosing when turbidity rises, ensuring rapid flocculation and settling of suspended solids. For disinfection, it dynamically adjusts sodium hypochlorite dosing according to bacterial indicators in the effluent, minimizing chemical residue while maintaining effective disinfection. For acid-base neutralization, it precisely regulates the flow of acidic or alkaline reagents based on pH readings, stabilizing the effluent pH within the compliant range of 6–9.

Moreover, smart valves enable coordinated control

of multiple chemicals, meeting the multidimensional purification requirements of complex industrial wastewater and effectively addressing the shortcomings of traditional dosing methods—low precision, high energy consumption, and substantial waste. Engineering practice demonstrates that intelligent valve-based dosing systems can reduce chemical consumption by 15% to 20%, significantly lowering material costs in wastewater treatment. At the same time, they stabilize effluent quality and completely eliminate secondary pollution caused by overdosing, achieving dual goals of improved treatment quality and reduced resource consumption.。

Engineering Application Case 4: Smart Chemical Dosing Retrofit Project at Wuhan Beihu Wastewater Treatment Plant

Wuhan Beihu Wastewater Treatment Plant is a large-scale municipal integrated wastewater treatment facility that receives domestic sewage and a small amount of industrial effluent. The influent quality exhibits significant diurnal and seasonal fluctuations in turbidity and pollutant concentrations. Traditional fixed-flow chemical dosing methods show obvious lag, resulting in insufficient dosing during poor water quality periods—leading to non-compliant effluent—and excessive chemical usage during good water quality conditions, causing both reagent waste and residual chemicals in the treated water. To achieve precise operation and maintenance, the plant upgraded its coagulation, disinfection, and phosphorus removal chemical dosing systems with intelligent controls, installing multiple miniature smart electric control valves capable of stepless fine adjustment from 0% to 100%, with response times ≤5 seconds. These are paired with high-precision online sensors for turbidity, total phosphorus, and residual chlorine (measurement accuracy ±0.5% FS), forming a fully automatic closed-loop dosing control system. The system dynamically adjusts valve openings based on real-time water quality data, enabling stepless regulation of chemical dosing rates and achieving "precise, demand-based dosing." After the upgrade, the plant reduced overall consumption of PAC, PAM, and disinfectants by 18.5%, saving over 600,000 RMB annually in chemical costs. Effluent stability for total phosphorus and turbidity improved significantly, and the issue of chemical residue was completely resolved. Under the premise of ensuring compliant effluent discharge, the plant has achieved cost reduction, efficiency improvement, and green operation.[10]

3.5 Applications in remote monitoring systems

Traditional wastewater treatment operation and maintenance models heavily rely on manual on-site operations, scheduled inspections, and manual parameter adjustments. These systems suffer from numerous drawbacks, including low operational efficiency, high labor costs, delayed fault detection, inability to trace operating conditions in real time, and a high rate of human operational errors. Small- and medium-sized wastewater treatment plants commonly face shortages of maintenance personnel and weak technical expertise, making them highly vulnerable to issues such as untimely equipment failure response, delayed process parameter adjustments leading to effluent (exceeding discharge standards), and equipment damage. With the continuous advancement of smart water management initiatives, remote intelligent monitoring systems have become standard in modern wastewater treatment plants. Smart valves serve as the core terminal execution devices within these remote monitoring systems and form the foundational platform for achieving unattended, intelligent operation and maintenance.

Smart valves, equipped with built-in IoT communication modules, seamlessly integrate with local DCS control systems, PLC control cabinets, cloud-based smart water management platforms, and mobile monitoring systems at wastewater treatment plants, establishing a fully visualized operation and maintenance framework that spans "terminal data collection - cloud transmission - platform analysis - remote control." All smart valves can upload real-time, multi-dimensional data including valve opening degree, equipment status, pipeline flow rate, medium pressure, equipment temperature, fault codes, and adjustment records. The platform automatically generates operational curves, performance reports, and energy consumption statistics through big data analytics, enabling full visualization of all process parameters in wastewater treatment.

Operations personnel no longer need to be on-site; they can remotely monitor the operational status of valves and equipment across all process nodes via computer terminals or mobile apps, and perform remote operations such as valve opening/closing, fine-tuning of opening degrees, parameter calibration, and mode switching, achieving remote control from any location. Additionally, the smart valves are equipped with high-precision self-diagnostic systems capable of detecting various faults—including valve core blockage, actuator overload, signal interruption, sensor failure, abnormal pipeline pressure, and flow anomalies—in real time. These

systems accurately pinpoint fault locations and identify root causes, immediately sending alarm notifications to both the platform and maintenance staff, enabling early fault prediction and rapid response, thereby minimizing risks of process disruptions and equipment damage.

Moreover, the system retains all operational data and operation logs over the long term, supporting historical condition tracing, process parameter optimization, energy consumption analysis, and equipment maintenance statistics. This provides precise data support for plant process upgrades, energy-saving retrofits, and maintenance scheduling. The deep integration between smart valves and remote monitoring systems has completely transformed traditional manual operation models, significantly reducing labor costs, enhancing equipment reliability and process control accuracy, and accelerating the transition of wastewater treatment plants toward automation, intelligence, and unattended operation.

Engineering Application Case 5: Smart Remote Operation and Maintenance Upgrade Project for Guangzhou Datansha Wastewater Treatment Plant

Guangzhou Datansha Wastewater Treatment Plant, a long-established large-scale municipal wastewater facility in China, faces challenges due to aging equipment and complex piping systems. A significant number of valves are located in hazardous areas such as underground tunnels and elevated equipment levels. Traditional operation and maintenance practices rely heavily on frequent manual inspections and manual valve adjustments, resulting in high operational risks, low efficiency, and delayed fault response. To enhance intelligent operation and maintenance capabilities, the plant has implemented an electric intelligent upgrade for critical valves, replacing them in bulk with smart extended-stem electric valves equipped with IoT communication functions. These devices support the Modbus-TCP communication protocol, offer communication latency of ≤200ms, and feature IP66 protection rating. They can upload real-time data—including valve position, pressure, flow rate, and operational faults—enabling seamless integration with the plant's DCS control system and cloud-based remote monitoring platform. After the upgrade, operators can remotely open or close valves, fine-tune positions, and calibrate parameters from the central control room or mobile devices. Valve response speed has improved by over 90%, completely eliminating the need for manual operations in high-risk areas such as elevated platforms and underground corridors. Additionally, the valves are equipped with high-precision self-diagnostic modules that accurately detect faults such as blockages, overloads, and signal anomalies, triggering pop-up alerts proactively and enabling predictive maintenance. As a result, equipment failure rates have dropped by 65%, and on-site maintenance workload has been reduced by 50%. The plant has successfully transitioned from "manual on-site monitoring" to a "remote intelligent control and unmanned operation" model, providing a proven case study for intelligent retrofitting of aging wastewater treatment facilities.[11]

4. Technical Advantages of Smart Valves in Wastewater Treatment

4.1 Excellent control accuracy significantly enhances process operation stability

Traditional mechanical valves can only achieve fixed-step or coarse manual adjustments, with typical regulation errors ranging from ±5% to ±10% and response delays lasting 30 to 60 seconds. They are unable to provide fine, stepless control and struggle to adapt to the dynamic fluctuations in wastewater treatment flow rates and water quality. In contrast, smart valves leverage high-precision sensor data acquisition and optimized PID closed-loop control algorithms to enable smooth, stepless adjustment across 0–100% opening. Key performance metrics include a regulation accuracy error of ≤±1%, steady-state response time of ≤5 seconds, and an algorithm overshoot of ≤3%. These valves offer fast response, excellent linearity, and strong adaptability to varying operating conditions, enabling precise detection of minor process fluctuations and rapid parameter correction. In critical, precision-driven processes such as aeration, chemical dosing, and sludge recirculation, smart valves maintain key parameters—including dissolved oxygen concentration, chemical dosage, and sludge return ratio—within optimal operational ranges, effectively resolving industry challenges such as human-induced lag, large errors, and system oscillations. Furthermore, the closed-loop feedback correction mechanism continuously iteratively refines control accuracy, significantly enhancing the wastewater treatment system's resilience to shock loads. This ensures stable effluent quality under complex conditions such as influent fluctuations and sudden changes in water quality, substantially reducing the risk of effluent non-compliance.

4.2 Significant energy savings and consumption

reduction, lowering overall operating costs

Energy consumption and material costs are the most critical operating expenses for wastewater treatment plants, with aeration blowers, pumps, and chemical reagents accounting for the majority of operational expenditures. Traditional fixed-orifice valves operate in a coarse, inefficient manner, resulting in significant energy waste and excessive chemical usage, leading to extremely low resource utilization. Smart valves enable dynamic, fine-grained control throughout the entire process, precisely matching aeration rates, recycle flow, chemical dosing, and influent load requirements. This fundamentally eliminates wasteful practices such as over-aeration, excessive chemical dosing, and unnecessary hydraulic losses. In aeration systems, precise oxygen control can reduce blower energy consumption by 25% to 35%. In chemical dosing systems, on-demand dosing reduces reagent consumption by 15% to 20%. In pipeline conveyance systems, stable flow and pressure minimize frequent pump starts and stops and prevent overload operation, thereby reducing pump energy use and equipment wear. Additionally, consistent and stable operating conditions significantly lower equipment failure rates and maintenance costs, extending the service life of valves, pumps, blowers, and other supporting equipment. This comprehensive approach substantially reduces overall operational and maintenance costs for wastewater treatment plants, enabling them to achieve low-carbon, energy-efficient operations.

4.3 Highly intelligent with a fully upgraded operation and maintenance model

The intelligent valve integrates multiple functions including perception, computation, actuation, early warning, communication, and data storage, enabling a leapfrog upgrade from "passive execution" to "proactive intelligent control." It completely replaces the traditional, labor-intensive operation model involving manual inspections, manual valve adjustments, and on-site troubleshooting. The device can operate automatically 24/7 without requiring real-time human intervention, significantly reducing staffing requirements and minimizing process failures caused by human operational errors. Additionally, it features self-diagnosis of faults, proactive alarms, data traceability, and intelligent analysis capabilities, allowing for early prediction of issues such as equipment aging, pipeline blockages, and parameter anomalies, thus enabling preventive maintenance and shifting away from the conventional reactive "repair after failure" approach. Combined with a remote monitoring platform, it enables centralized control, remote commissioning, and unattended operation across entire facilities, greatly enhancing the automation and intelligence level of wastewater treatment systems and meeting the evolving demands of modern smart water management.[12]

4.4 Strong adaptability to operating conditions, with outstanding versatility and reliability

Wastewater treatment operations involve complex and variable conditions, with media characterized by high impurity content, strong corrosiveness, tendency to scale, and significant flow fluctuations. Additionally, plant environments are often damp, subject to large temperature variations, and dusty. Intelligent valves can be customized with wear- and corrosion-resistant materials such as stainless steel, rubber-lined components, or anti-corrosion coatings based on specific process scenarios and medium characteristics. They feature optimized valve core designs to prevent clogging, advanced sealing structures, and protective sensor probe configurations, enabling effective adaptation to challenging conditions involving sewage, sludge, chemical reagents, and humid air. These valves exhibit excellent resistance to corrosion, blockage, and aging. Moreover, they support various intelligent control modes—including PID adaptive regulation, split-range control, interlocking control, and threshold protection—allowing flexible integration into all wastewater treatment processes such as influent regulation, aeration oxygen control, sludge recirculation, chemical dosing, and effluent discharge. Suitable not only for municipal domestic wastewater treatment, these valves are also ideal for industrial applications in chemical processing, textile dyeing, aquaculture, and food production. Their versatility, environmental adaptability, and operational reliability significantly surpass those of conventional valves.

4.5 Data traceability supports process optimization and standardized management

Conventional valves lack data collection and storage capabilities, making it impossible to quantitatively record process operation parameters. Maintenance and operations management rely heavily on manual experience, making standardized and data-driven control difficult to achieve. Smart valves, in contrast, can continuously collect, store, and upload all operational data and adjustment records, forming a complete operational condition database. Operations personnel

can review historical data and analyze parameters to precisely identify shortcomings in process performance, energy waste points, and frequent equipment failure locations, enabling targeted optimization of process parameters and operating strategies. Moreover, comprehensive data recording meets the requirements for environmental compliance monitoring, process verification, and project acceptance by providing traceable data, facilitating the transition of wastewater treatment operations from "experience-driven" to "data-driven" management and supporting wastewater plants in achieving standardized, refined, and digitalized operations.

5. Issues and shortcomings in smart valve applications

5.1 Equipment procurement and modification costs are relatively high.

Smart valves are highly integrated mechatronic intelligent devices that combine advanced technologies such as high-precision sensing, embedded control, IoT communication, and precision mechanical processing. Their manufacturing process is complex, and the cost of core components is relatively high, resulting in a significantly higher purchase price per unit compared to traditional mechanical valves. For large-scale new smart wastewater treatment plants, sufficient funds for equipment modernization allow for widespread adoption of complete sets of smart valves. However, for numerous small- and medium-sized township wastewater treatment stations and older facilities undergoing upgrades, limited operating budgets and tight financial constraints are common. The initial costs associated with bulk replacement—including equipment procurement, installation, commissioning, and system integration—are prohibitively high, creating substantial financial pressure. Currently, cost barriers remain the key factor limiting the large-scale deployment of smart valves at grassroots wastewater treatment sites, leading to an uneven industry landscape where intelligent equipment is widely adopted in large plants but lags behind in smaller facilities. This imbalance severely hinders the overall intelligent transformation of the wastewater treatment sector.

5.2 Insufficient equipment stability and durability under complex wastewater conditions

The environment in wastewater treatment is complex and harsh, with sewage containing large amounts of silt, suspended solids, colloidal particles, and fibrous impurities. Additionally, acidic, alkaline, and chemical wastewater are highly corrosive, making long-term operation particularly prone to wear and failure of intelligent valve equipment. The sensor probes, valve core gaps, and sealing structures of smart valves are precision and vulnerable components. Prolonged exposure to high-impurity wastewater easily leads to issues such as impurity accumulation, scaling, clogging, corrosion, and wear, resulting in distorted sensor data, sluggish valve adjustment, incorrect opening positions, and poor sealing causing leaks. Moreover, wastewater treatment plants are typically located in open-air or underground damp environments characterized by significant temperature fluctuations, high humidity, and abundant dust. Although current mainstream smart valves generally offer protection ratings of IP65 to IP67, suitable for typical damp and dusty conditions, their electrical control modules and communication modules remain susceptible to moisture aging, wire corrosion, signal interference, and module damage when exposed to prolonged high humidity, strong electromagnetic interference, and fluctuating temperatures. This ultimately leads to reduced device accuracy, shortened service life, and failure of intelligent functions. Compared to traditional purely mechanical valves, which offer high durability and low failure rates, the precise electromechanical structure of smart valves has lower tolerance to harsh operating conditions, and their ability to operate continuously and stably still shows significant shortcomings.

5.3 The technical requirements for operations and maintenance are high, and there is a severe shortage of skilled professionals at the grassroots level

Smart valves are no longer simple mechanical devices, but rather integrated intelligent equipment combining mechanics, electrical systems, automatic control, the Internet of Things (IoT), big data, and communication technologies. Their selection, installation, parameter calibration, algorithm tuning, system integration, fault diagnosis, and routine maintenance all require specialized technical support, placing extremely high demands on the professional competence of operation and maintenance personnel. Currently, the overall professional level of grassroots wastewater treatment operation teams in China remains relatively low. Most operators have long been engaged in traditional equipment operations, mastering only basic tasks such as opening and closing mechanical valves or starting and stopping pumps, while lacking expertise

in automatic control, IoT, and smart device debugging. When smart valves encounter issues such as signal anomalies, regulation deviations, communication interruptions, algorithm failures, or sensor malfunctions, frontline maintenance staff are unable to quickly identify root causes or complete repairs and adjustments. As a result, many installed smart valve systems fail to function properly, remaining idle or operating at partial capacity for extended periods. This leads to wasted equipment resources and prevents achieving the expected outcomes of intelligent upgrades. The shortage of skilled professionals has thus become a significant human factor limiting the efficient application of smart valves.

5.4 Industry standards are inconsistent, and equipment system compatibility is poor

Currently, China's intelligent fluid control equipment industry is still in a stage of rapid development, and has not yet established unified and standardized production standards, communication protocols, or interface specifications. The market features numerous smart valve brands with inconsistent product parameters, and hardware interfaces, communication protocols, control logic, and data formats vary significantly across manufacturers, resulting in poor interoperability. Many older wastewater treatment plants have legacy control systems, PLCs, and monitoring platforms that were built years ago, featuring outdated system architectures that are incompatible with new smart valves. This leads to significant challenges in device integration and debugging, high retrofitting costs, and some old systems being unable to support new smart valve equipment at all. Additionally, smart valves from different batches and brands cannot be centrally managed, and data cannot be shared or exchanged, causing fragmentation in plant control systems and making it difficult to build an integrated, standardized smart management platform, which severely affects the overall effectiveness of intelligent integration in wastewater treatment systems.[13]

6. Trends in the Development of Smart Valves in the Wastewater Treatment Industry

6.1 Hardware materials upgraded, enabling high precision, high reliability, and long service life of equipment

To address the current challenges of smart valves—such as susceptibility to corrosion, clogging, rapid loss of accuracy, and poor durability under complex wastewater conditions—future smart valve development will focus on technological advancement through material upgrades and structural optimization. On the material front, advanced materials such as new anti-corrosion and wear-resistant composites, stainless steel alloys, ceramic sealing materials, and anti-scaling coatings will be widely adopted. These materials will significantly enhance the corrosion resistance, clog resistance, wear resistance, and anti-scaling performance of valve bodies, valve cores, sealing structures, and sensor probes, enabling reliable operation in harsh wastewater environments characterized by high impurity levels, strong corrosiveness, and humidity. On the structural side, improvements will include streamlined valve core designs, self-cleaning mechanisms, and protective structures for sensor probes, along with features such as automatic flushing, anti-sedimentation, and self-sealing functions, effectively addressing issues related to impurity blockage and scaling-induced measurement inaccuracies. Additionally, upgraded high-precision digital sensing modules and anti-interference chips will improve the device's control accuracy and operational stability under complex electromagnetic environments and extreme water quality fluctuations, extending service life, reducing failure rates and maintenance costs, and ensuring highly reliable, long-term stable operation.[14]

6.2 Artificial intelligence deeply empowers predictive intelligent control and unmanned operation

Current smart valves are mostly passive closed-loop control systems, capable only of adjusting parameters based on real-time operational data, without the ability to anticipate changes in operating conditions. In the future, leveraging artificial intelligence, big data analytics, and digital twin technologies, smart valves will undergo a transformative leap from "passive regulation" to "proactive prediction." By accumulating vast amounts of long-term wastewater treatment operation data, predictive models for water quality and flow fluctuations, as well as process optimization models, can be developed. The system can accurately forecast trends in water quality and flow variations across day-night cycles, seasons, and rainy periods, enabling early adjustment of valve operating parameters for precise, proactive control, significantly enhancing the system's resilience to shock loads. Meanwhile, integrated with a digital twin-based smart water management platform, a virtual simulation system for all plant equipment can be established, enabling visualization of valve operating status, process simulation and optimization, intelligent

fault diagnosis, and autonomous problem resolution. This fully realizes unmanned, intelligent, self-operated maintenance throughout the entire wastewater treatment process, greatly advancing the level of smart wastewater treatment.

6.3 Industry standardization is being implemented, enabling widespread adoption of universal and low-cost equipment

With the rapid development of the smart valve industry, national and industry associations will gradually establish unified industry standards for manufacturing, interface specifications, communication protocols, control standards, and testing and acceptance procedures. This will fundamentally address current industry issues such as fragmented equipment brands, poor compatibility, and lack of data interoperability. A standardized system will enable seamless integration among smart valves, control systems, and monitoring platforms from different manufacturers, significantly reducing costs associated with equipment integration and upgrades. Meanwhile, as production technologies mature and economies of scale emerge, the manufacturing cost of smart valves will continue to decline, greatly improving their cost-effectiveness. In the future, smart valves will progressively overcome financial barriers, fully meeting the retrofit needs of small- and medium-sized wastewater treatment plants and rural sewage stations, achieving comprehensive coverage across all scales of wastewater treatment facilities and driving overall intelligent transformation in the wastewater treatment sector.

6.4 Green, energy-efficient, and integrated, meeting the development needs of the dual-carbon industry

Against the backdrop of national dual-carbon strategy and the industry's push for energy-efficient, high-quality wastewater treatment, green, low-carbon, energy-saving, and consumption-reducing approaches will become the core development direction for smart valves. In the future, smart valves will be equipped with ultra-low-power actuators and energy-saving control algorithms to optimize equipment operation logic. While reducing the valve's own energy consumption, they will further refine process control strategies to maximize energy efficiency in aeration, water conveyance, and chemical dosing, thereby supporting energy conservation and carbon reduction in wastewater treatment plants. Meanwhile, equipment will evolve toward modularity, integration, and compact design, incorporating multiple functions such as sensing, control, communication, early warning, and data storage. This will streamline device structure, reduce size, simplify installation, and lower maintenance complexity, enabling rapid deployment and modular integration. Such advancements will meet the standardized and lightweight construction needs of intelligent wastewater treatment systems, driving the sector toward green, low-carbon, and high-quality development.

7. CONCLUSIONS

As a core terminal execution device in intelligent wastewater treatment systems, smart valves integrate advanced mechanical structures, high-precision sensing technologies, adaptive automatic control algorithms, and IoT communication technologies to establish a fully closed-loop intelligent regulation system encompassing "perception, computation, actuation, feedback, and traceability." This innovation completely transforms the traditional mode of manual operation and coarse control, effectively addressing industry challenges such as low regulation accuracy, delayed response, high energy consumption, complex maintenance, and poor adaptability to operating conditions. Through systematic research and engineering case analysis, this paper demonstrates that smart valves can precisely meet the fine-tuning requirements of five key process systems in wastewater treatment: influent control, biological aeration, sludge recirculation, chemical dosing, and remote monitoring. They significantly stabilize treatment processes, improve effluent quality compliance rates, reduce system energy consumption and labor costs, and offer multiple technical advantages including broad operational adaptability, high intelligence, data traceability, and strong process optimization capabilities. As such, smart valves serve as a pivotal enabling technology for the intelligent, precise, and low-carbon transformation of modern wastewater treatment industries.

Currently, smart valves in engineering applications still face practical challenges such as high initial investment costs, insufficient stability under complex operating conditions, a shortage of skilled maintenance personnel at the grassroots level, lack of unified industry standards, poor system compatibility, and weak data security. These issues to some extent hinder the large-scale adoption and efficient application of the equipment. However, with continuous advancements in intelligent manufacturing technologies, artificial

intelligence algorithms, and Internet of Things (IoT) technologies, along with ongoing improvements in industry standardization, smart valves are expected to evolve toward higher reliability, greater precision, lower cost, full automation, green energy efficiency, and enhanced safety and stability, gradually overcoming existing limitations.

Against the backdrop of widespread adoption of smart water management, deepening implementation of the dual-carbon strategy, and continuously upgraded standards for water environment governance, the application value and development potential of smart valves will continue to grow. They will fully replace traditional mechanical valves, enabling precise, intelligent, and unmanned operation throughout the entire wastewater treatment process. This will drive continuous improvement in quality and efficiency within China's wastewater treatment sector, support its transition toward green and low-carbon operations, and provide solid technological and equipment foundations for water ecological protection and water resource recycling.□

REFERENCES

[1] Sun Hongchang, Su Yunfei, Jiang Yongxiang, et al. Fluid Control of Sewage Treatment Pipeline Based on IGWO-BP-PID [J]. Sensors and Microsystem, 2025(6): 91-94.

[2] Ma M, Xing K, Mei L L, et al. Development of a smart wastewater valve to optimize reverse-osmosis membrane-cleaning cycles for enhanced water purification efficiency[J]. Processes, 2024, 12(12): 2865.

[3] Jiangsu Environmental Engineering Technology Co., Ltd., Nanjing Yunpu Environmental Technology Co., Ltd. Intelligent Gate Control System for Rainwater and Wastewater in Chemical Industrial Parks [C]//China Environmental Protection Industry Association. Typical Cases of New Quality Productivity in Environmental Protection. Beijing, 2025.

[4] Aerospace Kaitian Environmental Protection Technology Co., Ltd. "One Plant, One Pipeline" Model for Wastewater Collection in Hengshui High-Tech Zone and AI-Based Smart Operation of Wastewater Treatment Plants [J]. China Water Network, 2025.

[5] Wuhan Holding Drainage Company Beihu Wastewater Treatment Plant: Case Study on Automatic Control Retrofit for Biological Tank Condensate Discharge [J]. Wuhan Water Affairs, 2026.

[6] COVNA Group. Application Solutions of Automated Valves in Water Treatment [C]//2026 Shanghai International Water Exhibition. Shanghai, 2026.

[7] Ni Wei, Ni Chao. An Intelligent Wastewater Treatment System and Method: CN114409192A[P]. 2022-04-29.

[8] Zhuang Wenli. Intelligent Sewage Control Device: CN108279291A[P]. 2019-01-01.

[9] DeZURIK. APCO SmartCHECK Pump Control Valve (CPC) Technical Manual[Z]. 2024.

[10] Application of CAN-based Intelligent Valve Network Control System in Wastewater Treatment [J]. Automation Technology and Application, 2004(11).

[11] Zhang Haonan, Wang Guofeng, Zhao Qiaonan, et al. Numerical Simulation and Optimization of V-Port Ball Valve Core [J]. Journal of Shenyang Institute of Engineering (Natural Science Edition), 2019, 15(2): 112-117.

[12] Chen Chaobo, Hu Haitao, Gao Song. Research on Parameter Self-Adaptation of Fractional-Order PID Controller for Artificial Bee Colony[J]. Control Engineering, 2020, 27(6): 956-961.

[13] Gao Xingquan, Huang Dongdong, Ding Sanmao. PID Parameter Optimization Using Penalty Function Combined with Genetic Algorithm [J]. Journal of Jilin Institute of Chemical Technology, 2021, 38(3): 57-60.

[14] Xi'an Zhongke Optoelectronic Precision Engineering Co., Ltd. Ground-track inverted (dual-machine) welding robotTR-iTR21D[EB/OL].

Author Introduction:Tan Tu, Male, Engineer, currently engaged in engineering automation design and sales.

Copyright by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

JOURNAL OF DIGITAL INTELLIGENT INDUSTRY

EDITORIAL BOARD

MEMBERS

Chen Hongzhong Sinopec Nanjing Engineering Co., Ltd. Senior Engineer
Dong Dianchen Shaanxi Yanchang Petroleum Yan'an Energy & Chemical Co., Ltd. Senior Engineer
Gao Jian Hangzhou Wolunsen Electric Co., Ltd. General Manager
Li Jianfeng Schneider Electric (China) Co., Ltd. Industry Lead for Energy & Chemicals
Meng Qinghe Dongying Hebang Chemical Co., Ltd. Instrument Supervisor
Tan Xingqiang Sinopec Yizheng Chemical Fiber Co., Ltd. Senior Engineer
Wang Yuhu Tianjin Shihua Engineering Design Co., Ltd. I&C Chief Engineer
Xu Yongcun Senier (Nanjing) Process Control Co., Ltd. General Manager
Zhang Xizhao Beijing Yiqibang Technology Information Co., Ltd General Manager

JOURNAL OF DIGITAL INTELLIGENCE FOR INDUSTRY

Bimonthly
Vol.1 No.1
October 12, 2026

Chief Editor	Chen Hongzhong chiefeditor@jdii.org
Executve Editor	Cui Ming Wang Xiang
Designer	Jing Jing
Edited & Published by	HONG KONG OCEANWISH PRESS LIMITED
Address	UNIT 9, 1/F, THE CLOUD, 111 TUNG CHAU STREET, TAI KOK TSUI, HONG KONG
Tel.	90454293
E-mail	journal@digital-intelligentindustry.com
Website	https://jdii.org
ISSN	3137-0837 (Online), ISSN 3137-0829 (Print)