Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 6, 2025

Power-Material Storage & Scheduling Optimization via 6G-CPS and Deep Deterministic Policy Gradient

Jun Zhao[1,#], Zengchao Wang[1,#,*], Ming Su[1], Dongliang Wang[2], Feng Hao[1], Kefeng Li[1], Zihan Li[1]

[1]State Grid Hebei Procurement Co, Shijiazhuang, Hebei Province, 050000, China
[2]NARI Nanjing Control System Ltd., Nanjing, Jiangsu Province, 210032, China
[#]These authors are the co-first authors.
[*]Corresponding author's email: hv_wzc@163.com

Abstract. The current material warehousing and outbound management of the power system has problems such as low efficiency, rigid planning, and frequent path conflicts, which lead to material accumulation, regional congestion, and reduced operating efficiency, seriously restricting the stable operation of the power supply chain. In response to this situation, this paper proposes a power material storage and scheduling optimization method based on the deep deterministic policy gradient (DDPG) algorithm and the 6G-enabled cyber-physical system (CPS). In the warehousing stage of power materials, combined with the high bandwidth and low latency characteristics of the 6G network, the CPS system realizes real-time perception and dynamic storage adjustment of material information; in the outbound stage, the DDPG algorithm is used to construct a continuous state-action space, optimize the path planning and dynamic obstacle avoidance of automatic guided vehicles (AGV) equipment, and improve the scheduling efficiency of power emergency materials. Experimental results show that after integrating the 6G-CPS system, the material scanning time is shortened from the traditional 45 seconds/item to 25 seconds/item, and the efficiency is improved by 44%; the optimized path distances of DDPG in simple and complex environments are 7.2m and 8.6m respectively, which are better than the comparison algorithms such as A2C (Advantage Actor-Critic) and A3C (Asynchronous Advantage Actor-Critic), and the convergence speed and reward value are optimal. The research results provide an efficient solution for smart grid material management, which can effectively support the material dispatching needs in scenarios such as new energy access and power emergency repair.

Keywords. Deep Deterministic Policy Gradient, Cyber-Physical System, Path Optimization, Warehouse Scheduling, 6G Network

1. Introduction

Power material storage and dispatching [1,2] plays a key role in the supply chain management of the power system, directly affecting the timeliness of material supply and the stability of the operation process. With the expansion of the scale of the power grid and the improvement of the level of intelligence, the material storage system faces more complex management needs. This not only meets the daily operation and maintenance, but also needs to cope with various scenarios such as emergency repairs and new energy [3] access, which puts higher requirements on the accuracy and efficiency of material dispatching. The efficient dispatch of power material storage directly supports the stability of the renewable energy supply chain and the agile response of the smart grid. In the scenario of large-scale access to new energy, the rapid deployment of energy storage equipment, cables and other materials can shorten the construction period of power stations; and in emergency repairs of power grids, dynamic path planning can accelerate the distribution of key equipment and improve fault recovery efficiency.

Traditional power material storage systems [4] rely on fixed rules for warehousing and outbound scheduling. In the case of diverse material types and frequent demand fluctuations, they lack intelligent adjustment capabilities, resulting in limited warehousing and outbound operation processes and reduced overall operating efficiency. In terms of warehousing management, the material storage layout is not optimized, the storage space utilization rate is low, and the storage and access path design is unreasonable, resulting in reduced operating efficiency. Some materials occupy fixed positions for a long time and fail to make intelligent adjustments based on the frequency of delivery, which increases the operating distance of the handling equipment. In addition, the warehouse environment is complex, and the traditional management model mainly relies on manual or barcode scanning to record data, which makes it difficult to achieve real-time perception of the status of materials, affecting the accuracy of the scheduling system. During the outbound process, equipment scheduling lacks efficient planning, and path conflicts are prone to occur during task execution, resulting in reduced work efficiency. Traditional warehousing systems are still mainly based on static scheduling, and fail to dynamically optimize based on equipment status and environmental changes, resulting in lengthy equipment travel paths and extended operation time. There are many transport equipment inside the warehouse, and the paths frequently cross. There is a lack of efficient path optimization methods, which further affects the scheduling efficiency. In this context, it is urgent to introduce intelligent scheduling technology to optimize the storage method of incoming materials, improve the ability to plan outgoing paths, and ensure the efficiency and stability of warehouse management.

For the optimization of power material storage scheduling, there are methods based on the Internet of Things (IoT) [5,6] and big data analysis [7], which collect storage environment data through sensors and optimize material scheduling in combination with data analysis technology. However, In some complex warehousing environments, the data transmission rate of traditional IoT architecture may be unable to meet the needs of some high-real-time applications, affecting the system's perception and response efficiency. In addition, big data analysis relies on historical data for optimization and lacks adaptive adjustment capabilities for dynamic environments, which affects the scheduling optimization effect. In the management of inbound and outbound routes, some studies have used classic algorithms such as A* [8] and Dijkstra [9] to optimize the driving routes of warehousing equipment. These methods can provide better paths in static environments, but they are difficult to adjust in real time in complex dynamic scenarios and cannot adapt to changes in the equipment, materials, and environmental conditions in the storage system. Reinforcement learning methods [10] have also been introduced for path optimization. These methods have achieved certain results in path optimization, but in the storage scheduling tasks in high-dimensional continuous space, the reinforcement learning methods in discrete action spaces have great limitations and are difficult to optimize path decisions efficiently. Some other methods use rule-based warehouse scheduling strategies to adjust equipment task allocation through predefined rules. However, such methods are difficult to adapt to real-time changes in the warehouse environment and cannot effectively solve equipment scheduling conflicts. Multi-agent reinforcement learning can be introduced to try to optimize scheduling strategies in multi-equipment collaborative scheduling scenarios. Due to the complex task environment in the storage system and the difficulty in accurately modeling the collaborative relationship between equipment, these methods still face great challenges in practical applications. In addition, the power storage environment is highly dynamic, including complex factors such as randomly appearing obstacles (such as temporary stacking equipment), sudden tasks (such as emergency repair material allocation), and multi-AGV collaborative scheduling, which are difficult to fully model using traditional mathematical programming methods. Although methods such as mixed integer linear programming (MILP) have the advantage of interpretability, they need to be frequently remodeled and solved in dynamic environments, and are difficult to meet the needs in scenarios with strict real-time requirements. In this context, DDPG is a good choice. DDPG can directly process continuous action spaces (such as the precise steering angle and speed adjustment of AGVs), avoiding the loss of accuracy caused by discretization, which is particularly important for the precise handling of large power equipment. Furthermore, the application of 6G-CPS also provides inspiration for this study. The real-time environmental perception data provided by the 6G-CPS system perfectly matches the online learning characteristics of DDPG, enabling the system to continuously optimize strategies.

This paper adopts a warehouse intelligent scheduling optimization method based on 6G integrated cyber-physical system (CPS) and deep deterministic policy gradient (DDPG) algorithm. First, CPS and 6G network are used to realize real-time perception of warehouse status and data interaction, and dynamically adjust the material storage location to improve space utilization. Then, the DDPG reinforcement learning algorithm is used to perform path planning and obstacle avoidance optimization in the continuous action space during the outbound stage. The contributions of this paper are as follows:

(1) A power material warehouse scheduling optimization method based on 6G network and CPS (cyber-physical system) is constructed. The ultra-low latency and high bandwidth characteristics of 6G are used to realize real-time perception and dynamic storage adjustment of material information, greatly improving the warehouse efficiency and intelligence level.

(2) This paper adopts the deep deterministic policy gradient (DDPG) algorithm for path planning and dynamic obstacle avoidance, successfully overcoming the limitations of traditional path optimization methods in dynamic environments, especially in complex warehouse scenarios. The path planning and scheduling efficiency of AGV equipment has been significantly improved.

(3) The innovation of this paper is primarily reflected in three aspects: First, for the specific scenario of power material storage, we deeply integrate the real-time sensing capabilities of 6G-CPS systems with DDPG algorithms to achieve millisecond-level updates of environmental status information. This enables adaptive adjustment of reward factor weights according to dynamic changes in storage environments, whereas

existing studies predominantly focus on static or quasi-static scenarios. Second, in reward mechanism design, we adopt a task urgency-based dynamic priority adjustment strategy. By dynamically reconstructing reward functions using real-time task information obtained through 6G networks, this approach demonstrates unique value in power emergency material dispatch scenarios. Finally, we establish an energy consumption model considering the characteristics of power equipment, incorporating energy consumption patterns of heavy-duty equipment like transformers into reward functions. This fundamentally differs from energy optimization approaches in conventional logistics scenarios.

Article organization: Chapter 1 is an introduction: This chapter introduces the main problems existing in the storage and dispatch of power materials, such as low efficiency, frequent path conflicts, and the optimization needs for these problems. This article proposes a scheduling optimization method based on 6G network and CPS system, and briefly describes the research background and objectives.

Chapter 2 is related work: This chapter reviews the research progress related to the storage and dispatch of power materials, involving intelligent storage systems, path optimization algorithms, reinforcement learning and other fields, analyzes the advantages and disadvantages of existing methods, and provides a theoretical basis for the method proposed in this article.

Chapter 3 is a method: This chapter describes the proposed optimization method in detail, including the challenges of power material storage and dispatch, CPS system modeling, the integration of 6G network and CPS, and the application of DDPG algorithm in path planning. The method section explains how to use advanced technology to solve the efficiency problem in the storage system.

Chapter 4 is an experimental design: This chapter introduces the design of the experiment, including the construction of the experimental environment, the setting of evaluation indicators, and the comparative analysis of different algorithms. The purpose of the experimental design is to verify the effectiveness of the proposed method in actual storage dispatch.

Chapter 5 is the results: This chapter shows the experimental results, analyzes the performance of different algorithms in terms of path optimization, time consumption, and operation efficiency, and verifies the advantages and effects of the 6G-CPS and DDPG methods proposed in this paper.

Chapter 6 is the conclusion: This chapter summarizes the research results of this paper, emphasizes the intelligence and efficiency of the power material storage scheduling optimization method based on 6G network and CPS system, and discusses future research directions and potential applications.

2. Related Works

For the optimization of power material storage scheduling, researchers mainly study two aspects: intelligent storage system and path optimization algorithm. In the research of intelligent warehousing system,Pal S [11] et al. used convolutional neural network to optimize storage space layout and improve storage efficiency. The results show that CNN (Convolutional Neural Networks) can accurately analyze spatial relationships and realize intelligent storage strategy optimization, thereby improving storage operation efficiency. In practical applications, the generalization ability of the model, real-time computing requirements, and adaptability to complex warehousing environments still need further research and optimization. Zakaria I H [12] et al. explores the pivotal role of smart warehousing solutions in enhancing logistics efficiency and driving sustainable development. The study clarifies the definition, technical components, and sustainability practices of intelligent storage systems. Through case analysis and comparative research, it reveals smart warehouses' significant advantages in resource optimization, cost reduction, and supply chain resilience enhancement. However, challenges remain including high costs, complex system integration, and data security concerns. Sodiya E O [13] and others studied the profound impact of AI (Artificial Intelligence) in warehouse automation and explored how AI-driven systems can optimize warehouse operations through technologies such as machine learning, computer vision, and reinforcement learning, improve accuracy, speed, and cost-effectiveness, and promote the efficiency, accuracy, and adaptability of warehouse automation systems. However, they still face challenges such as real-time data processing and adaptation to environmental changes.Research on intelligent warehousing systems demonstrates the application potential of the technology, but also exposes challenges in real-time computing and environmental adaptability, which promotes further exploration of path optimization algorithms.

As a key technology, CPS has a wide range of applications in the field of logistics. Domestic and foreign scholars have explored the scheduling optimization path of intelligent warehousing under the CPS architecture. Liu B et al. analyzed the opportunities and challenges of the CPS scheduling system in the logistics park and proposed the need to strengthen the intelligent collaboration between perception, decision-making and execution [14]; Aron C et al. proposed the "cloud material handling system" to achieve dynamic resource allocation and digital interoperability, enhancing the flexibility and intelligence of warehousing scheduling [15]; Lu Y et al. focused on the role of information-physical fusion in promoting the environmental adaptability of warehousing systems in green port scenarios [16]; Piardi L et al. constructed a CPS architecture for intelligent warehouses based on

MAS to achieve distributed warehousing management and scheduling optimization [17]. These studies not only expand the theoretical basis of intelligent warehousing systems, but also further reveal the key value of path optimization algorithms in scheduling systems.

In the research of path optimization algorithms, Li K [18] used reinforcement learning and soft computing methods to solve the scheduling problem of stackers and automatic guided vehicles (AGVs). By establishing a Q-learning-based model and introducing fuzzy control methods, he optimized the warehouse logistics scheduling process and improved the efficiency of the system. Çelik M [19] studied the time-constrained storage replenishment routing problem in parallel channel warehouses and proposed a heuristic method to optimize the replenishment path and timeliness, reduce the total travel time and ensure timely availability of goods. Rijal A [20] proposed an efficient order processing solution by comprehensively planning warehouse operations and transportation scheduling, combining the expansion of temporary storage space and the extension of delivery time windows, which significantly reduced overall costs and improved distribution efficiency. Liu S et al. [21] proposed the A*-GWO algorithm by improving the Gray Wolf Optimization (GWO) algorithm and combining it with the A* algorithm. This algorithm solved the problem of random changes in static obstacles in the path planning of a mobile charging robot in a parking lot and improved the optimization effect of the number of iterations and path length.The above path optimization algorithms demonstrate different solutions, but there are also problems such as convergence speed, energy efficiency and real-time performance in complex environments. Moreover, the existing intelligent warehousing systems and path optimization algorithms mainly focus on static optimization and do not fully consider the challenges of dynamic scheduling.

In the field of power material storage, dynamic changes in the environment and emergencies are inevitable in actual operations, which requires the dispatching system to have higher flexibility and real-time performance. Dynamic dispatching strategy has become an important direction in current research and urgently needs to be explored and developed. In the face of dynamic scheduling strategies in the system, Some studies have proposed a multi-agent manufacturing system based on the Internet of Things to realize the dynamic optimization of flexible workshop scheduling problems. It has advantages and wide applicability in dynamic event scheduling problems, but it may still face challenges of computing efficiency and real-time performance in complex environments [22,23]. In recent years, with the widespread application of AGVs in intelligent warehousing systems, research on dynamic path planning has become increasingly in-depth. Bai Y et al. proposed a two-level path planning method that combines Kohonen neural network and Q learning to achieve dual optimization of path length and planning time in multi-AGV dynamic path planning, effectively improving system efficiency [24]; Zhang L et al. constructed a scheduling optimization architecture based on deep reinforcement learning to address the energy consumption problem in AGV battery replacement scheduling, and achieved dual-objective optimization of delay and energy consumption through a novel dueling dual Q network algorithm, significantly improving the cost-effectiveness of the flexible manufacturing process [25]; Fan F et al. proposed a hybrid deep reinforcement learning model for intensive manufacturing scenarios, optimized the robot's spatiotemporal path tracking, and effectively balanced the efficiency and safety of the system [26]. In addition, Leon J F et al. combined simulation and reinforcement learning to develop a hybrid algorithm for optimizing warehouse storage location allocation, which helps to deal with problems such as environmental variable uncertainty and worker interaction in warehouse operations [27]. Zhang L et al. further introduced digital twin technology into AGV scheduling and proposed an integrated optimization framework that combines static path planning and deep reinforcement learning to effectively avoid dynamic congestion and deadlock problems and significantly improve system stability and efficiency [28]. Although existing research has made significant progress in intelligent warehousing systems and path optimization algorithms, they still face challenges in practical applications such as insufficient generalization ability, high real-time computing requirements, and poor environmental adaptability.

Existing research has laid a solid foundation for the intelligent and efficient scheduling of power material storage from the construction of intelligent storage systems, the application of CPS architecture to the design of path optimization algorithms. In terms of intelligent storage systems, researchers have improved the automation level and system intelligence of storage management by introducing convolutional neural networks, artificial intelligence, robotic systems and CPS architectures, and promoted the optimization of storage resource allocation and scheduling methods. In the field of path optimization, a variety of algorithms such as reinforcement learning, ant colony algorithms, deep neural networks and intelligent agent collaborative control have been applied to AGV, drone and mobile robot scheduling, significantly improving the accuracy and efficiency of path planning. At the same time, some studies have combined simulation environments, digital twin platforms and multi-agent collaboration mechanisms to show good potential in dynamic scheduling strategies, energy consumption management and system adaptation. These achievements demonstrate the broad application prospects of intelligent technology in power storage scheduling and expand the depth and breadth of theoretical research and engineering practice. Although current research has made many breakthroughs, there are still some problems that cannot be ignored. Most methods have excellent optimization performance in static environments, but their generalization and real-time response capabilities are still insufficient in actual scenarios with frequent dynamic events and strong environmental uncertainty. In addition, the algorithm has high computational complexity and high resource consumption, which limits its deployment efficiency in large-scale storage systems. Some studies do not consider the complexity of system integration, data security, and edge computing coordination mechanisms, which leads to certain obstacles in its application in industrial scenarios. Therefore, it is urgent to propose an intelligent scheduling solution that takes into account efficiency, real-time performance, and environmental adaptability to meet the growing demand for intelligence in power material storage systems.Based on this, this paper proposes an optimization scheme for power material storage scheduling. By using the DDPG algorithm to optimize path selection, the performance bottleneck of traditional algorithms in complex environments is effectively improved, and further exploration and innovation are made in the intelligence, automation and efficiency of warehouse scheduling systems.

3. Methods

Figure 1 illustrates the intelligent grid material storage and dispatch optimization framework, which, through the deep integration of 6G networks and cyber-physical systems, establishes a comprehensive perception-decision-execution closed-loop system. The 6G network provides ultra-low latency and ultra-high bandwidth communication, enabling the cyber-physical system's perception layer to collect real-time data on power material status. This real-time data is used for dynamic optimization of storage layouts and as input to the DDPG algorithm, driving the intelligent path planning of AGVs. The DDPG agent continuously optimizes its decision-making strategy using a six-factor reward mechanism, ensuring the shortest path while also managing energy consumption and obstacle avoidance safety.

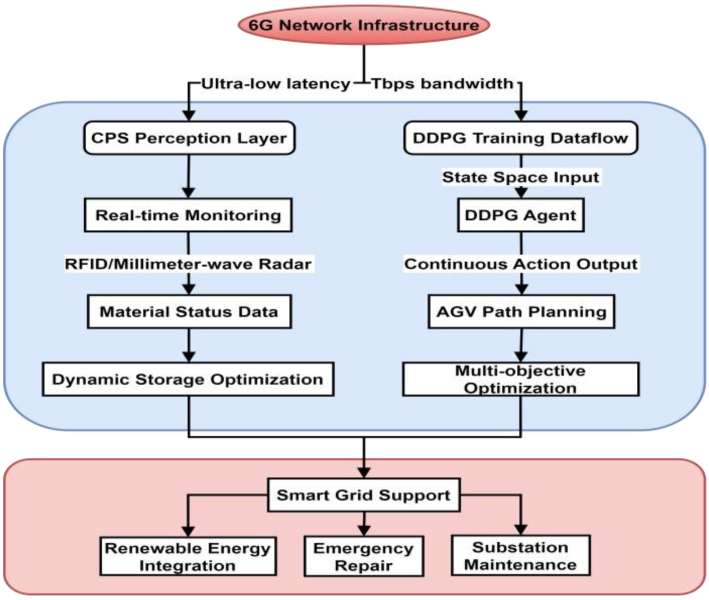

Figure 1. Optimal framework of smart grid material storage and scheduling.

A. Problem Description

Power material storage management is a key link to ensure the normal operation of the power system, among which warehousing management is the core issue affecting warehousing efficiency and the stability of the material supply chain. In the warehousing link, due to the unreasonable arrangement of material storage locations, it often results in low space utilization and long paths, resulting in low material storage and access efficiency. Traditional warehousing management methods usually rely on fixed rules and manual judgment, lacking flexible optimization mechanisms, which increases the time and transportation costs of robot handling. In addition, the types and storage requirements of materials are often complex and diverse, and the unreasonable order of material storage may lead to scheduling conflicts in the subsequent outbound link, affecting the efficiency of the overall warehousing system. In the outbound link, the order of material removal and path planning have always been difficult to optimize in warehousing management. Traditional outbound route planning often ignores dynamic factors such as changes in the warehouse's internal environment, material outbound priorities, and storage conditions. This can lead to unreasonable route planning and frequent equipment scheduling conflicts, resulting in route congestion and long equipment waiting times, which seriously affect operational efficiency.In addition, the scheduling of internal storage and transport equipment lacks a real-time feedback mechanism, which cannot be flexibly adjusted according to the actual environment, greatly reducing the overall scheduling capacity of the system. The existence of these problems leads to the lack of efficient and nimble scheduling strategy for the management of power materials in and out, serious waste of resources, low operational efficiency, and negative impact on the timely response and emergency treatment of the power system.

B. CPS System Modeling

Cyber-physical systems (CPS) [29,30] rely on big data, networks, and massive computing, integrate technologies such as intelligent perception, analysis, prediction, and optimization, achieve the organic coordination of computing, communication, and control (3C), and promote the deep integration of cyber space and physical space and its objects, environment, and groups. CPS can collect, store, model, analyze, mine, evaluate, predict, optimize, and coordinate big data of physical space, environment, and activities, and combine it with the design, testing, and operational performance characterization of objects, so that cyberspace and physical space can be deeply integrated, interact in real time, be coupled, and update each other. It can promote the comprehensive intelligence of industrial assets through self-perception, self-memory, self-cognition, self-decision-making, self-reconstruction, and intelligent support.

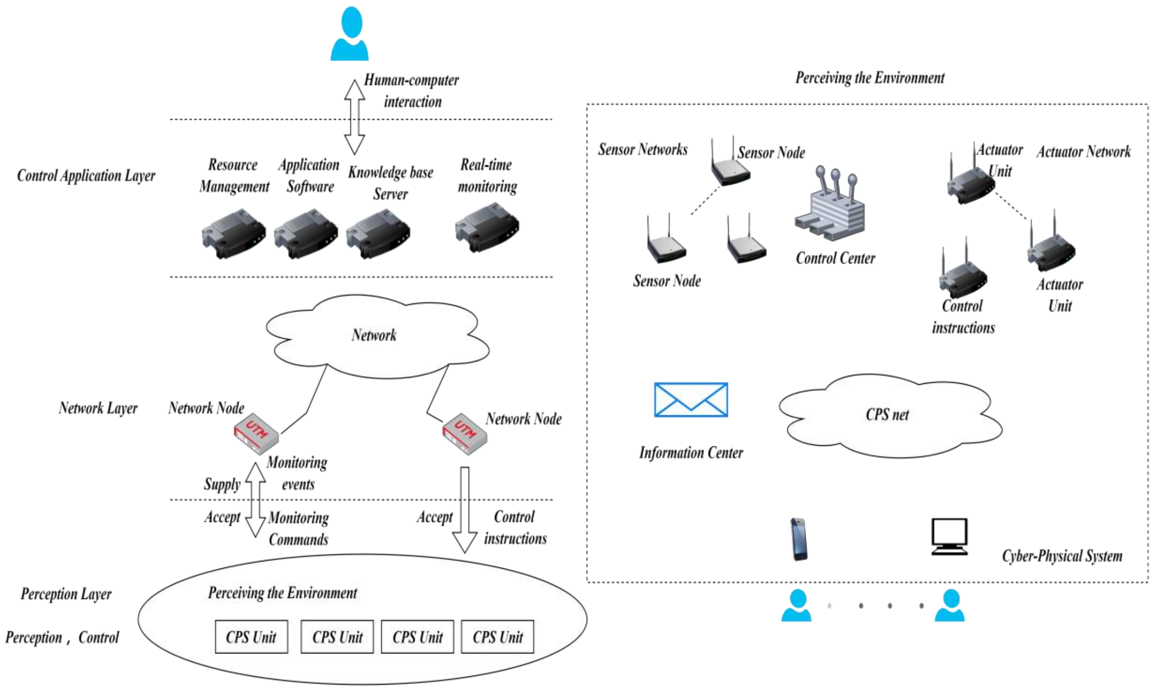

Figure 2. CPS system flow.

The cyber-physical system is mainly divided into three parts (Figure 2), namely the perception layer, network layer and control layer.

The perception layer is mainly composed of sensors, controllers, collectors and other devices. The sensors in the perception layer are the terminal devices in the cyber-physical system. They mainly collect specific information about the environment. The perception layer mainly obtains information data about the environment through sensors and sends it to the server at regular intervals. After receiving the data, the server processes it accordingly and returns the corresponding information to the physical terminal device. After receiving the data, the physical terminal device needs to make corresponding changes.

The network layer is mainly a bridge connecting the information world and the physical world. It mainly realizes data transmission, provides real-time network services for the system, and ensures the real-time reliability of network packets. The control layer mainly analyzes the data sent back by the physical device based on the cognitive results of the cognitive layer, and returns the corresponding results to the client to present them to the customer in a visual interface.

C. CPS Integrates 6G Network to Optimize Storage Management

In the scenario of power material storage, the irreplaceability of 6G compared to 5G is mainly reflected in three aspects: 1) Ultra-low latency can ensure real-time obstacle avoidance of AGV between dense shelves, avoiding the risk of collision caused by delays in the 5G environment; 2) The THz frequency band can achieve centimeter-level material positioning, which meets the precise storage management of large power equipment such as cables and transformers, and improves the meter-level accuracy compared to 5G; Intelligent slicing technology allocates exclusive bandwidth for emergency material dispatch. When disasters such as typhoons occur, it can ensure that the dispatch instructions of emergency repair materials are transmitted first, while 5G static slicing is difficult to cope with sudden traffic. This paper introduces 6G network [31,32] to greatly improve the real-time performance, data transmission capability and intelligence level of CPS, providing a better solution for power material storage and dispatching.

Table 1. Parameter comparison of different networks(Source: Authors' own work).

Network	4G	5G	6G
Peak rate	1 Gbps	10-20 Gbps	1 Tbps
User-Experience-Rate	10-100 Mbps	100 Mbps-1 Gbps	10 Gbps and above
Latency	30-50 ms	1 ms	Less than 0.1 ms
Spectrum range	2-6 GHz	24-100 GHz	1 THz
Device connection density	10^5 Equipment/km²	10^6 Equipment/km²	10^7 Equipment/km²
Position-accuracy	10m	1m	1m to 1cm
Energy consumption optimization	High-energy-consumption	Relative optimization	Very low power consumption

Table 1 shows the parameter comparison of different networks. Compared with 4G and 5G, 6G network has greatly improved in speed, latency, connectivity, intelligent computing, etc. The speed of 6G network can reach 1Tbps, which enables CPS to have stronger data processing capabilities. Ultra-low time delay, the delay is reduced to 0.1 milliseconds, ensuring real-time control and collaborative computing capabilities, and improving the accuracy of task scheduling. It also supports concurrent connection of multiple devices, enhancing the adaptability of CPS in large-scale IoT environments. The 6G network also has THz frequency bands and intelligent sensing technology, which enables CPS to have environmental perception capabilities and better locate and detect materials.In the power material storage scenario, 6G's millisecond-level latency advantage is directly reflected in key links. The transmission delay of AGV control instructions is reduced from 1ms in 5G to below 0.1ms, which increases the emergency obstacle avoidance response speed by more than ten times. When multiple AGVs are coordinated and dispatched, 6G's 0.1ms-level synchronization accuracy can reduce path conflicts. High-frequency inventory status updates ensure accurate tracking of time-sensitive materials such as new energy equipment. This feature is particularly important in power emergency repair scenarios.

Ultra-high-precision RFID (Radio Frequency Identification) [33,34], millimeter-wave radar, and holographic imaging sensors are used at the perception layer to achieve accurate perception of the status of materials and equipment. By attaching tags, real-time monitoring of information such as material type, quantity, and storage environment can be achieved, and ultra-high-definition monitoring data can be transmitted back in real time to provide support for intelligent analysis.

At the network layer, the ultra-low latency data transmission capability of 6G networks can avoid data congestion and ensure the real-time nature of key CPS decisions. Through 6G's intelligent slicing technology [35,36], different levels of service quality can be assigned to different tasks of CPS to ensure the reliable transmission of high-priority materials.

In the control layer, edge computing combined with 6G networks can effectively process key tasks locally, reduce cloud dependence, and improve decision-making speed. By combining 6G networks, the CPS system can effectively manage and optimize the storage and dispatch of power materials, greatly improve work efficiency, and reduce unnecessary troubles and computing consumption.

6G networks are superior to 5G in terms of speed, latency and intelligence, and are particularly suitable for large-scale IoT, real-time scheduling and complex decision-making needs in dynamic environments. In power material storage, although 5G can meet most real-time needs, there are still bottlenecks in equipment coordination, path replanning and exception handling. After 6G is integrated with CPS, it provides 0.1ms ultra-low latency and Tbps bandwidth, supports intelligent communication and perception, realizes real-time monitoring of materials, prediction of environmental changes and optimization of equipment autonomous decision-making, and significantly improves warehouse management efficiency. After 6G integrates CPS technology, it not only provides 0.1ms ultra-low latency and Tbps-level large bandwidth, but also realizes real-time perception of material status, prediction and feedback of storage environment changes, and autonomous optimization of equipment decision-making behavior through intelligent communication and perception empowerment. It should be noted that 5G network can meet the real-time perception requirements of CPS in most static storage scenarios. However, for special scenarios such as emergency material dispatching of power, when it is necessary to process high-definition sensing data of more than 1000 nodes at the same time, 6G's Tbps level bandwidth advantage is significant. In the power storage system with extremely high requirements for efficiency, safety and intelligent response, the deep integration of 6G and CPS has become a necessary choice to break through traditional bottlenecks and improve the overall intelligence level of the system, rather than a simple performance stacking.

Although the integration of 6G and CPS requires high initial investment, such as sensor upgrades, infrastructure transformation, and edge computing node deployment, the long-term returns are significant. The ultra-high bandwidth and low latency of the 6G network reduce communication waiting and resource conflicts, improve warehousing efficiency, and save manpower and energy consumption. The intelligent scheduling supported by 6G enables the CPS system to dynamically respond to environmental changes and reduce operating costs. As the technology matures and costs decrease, the intelligent warehousing management solution based on 6G+CPS has long-term economic feasibility and promotes the large-scale application of intelligent energy storage systems.

D. DDPG Model Optimizes Outbound Management

This paper introduces the DDPG model algorithm to optimize the material outbound management of AGV (automatic guided vehicle) equipment. AGV is an intelligent transportation equipment that does not require manual driving and relies on a navigation system to move autonomously. It can complete material handling tasks efficiently and safely. The specific flow chart is shown in Figure 3.

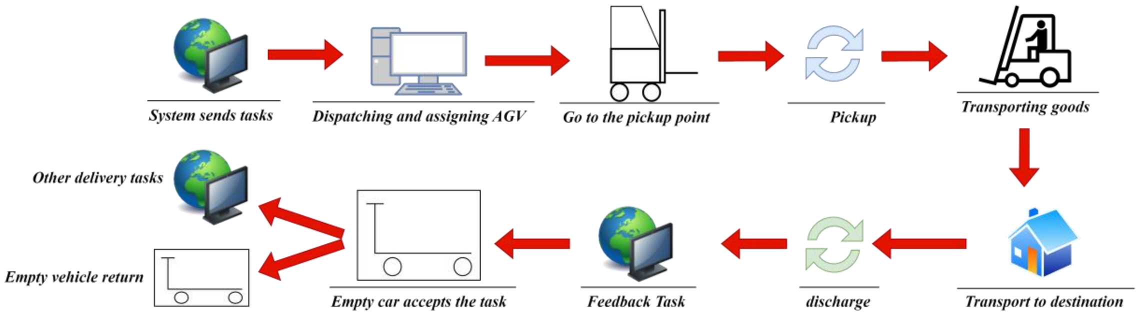

Figure 3. AGV delivery outbound process.

As can be seen from Figure 3, after the superior management system issues instructions, the dispatching system formulates and dispatches AGV equipment, and the AGV goes according to the address information obtained, and after the goods are scanned and identified, they transport the materials. It effectively transports the materials to the designated location, scans and identifies them, and unloads the materials after the identification is confirmed. After the materials are transported out of the warehouse, feedback can be given to the system. After receiving the feedback information, the system can screen and find out whether there are still materials that need to be shipped out, and send instructions to the AGV. If not, it can return to rest. However, in the warehouse management system for power material storage and scheduling, the path planning and scheduling methods of AGV equipment when performing material retrieval tasks have certain limitations. Traditional outbound route planning often fails to fully consider key factors such as the dynamic changes in the warehouse's internal environment, the priority of material outbound delivery, and storage conditions, resulting in a lack of flexibility in planning and prone to scheduling conflicts. Due to unreasonable path selection, AGVs may encounter route congestion, long waiting times, and other problems during operation, thereby reducing overall operating efficiency.

DDPG is a deep neural network optimization method based on reinforcement learning, which is suitable for decision-making tasks in continuous action space. DDPG combines the advantages of Deep Q Network (DQN) and deterministic policy gradient, adopts Actor-Critic architecture and uses target network and experience replay to improve training stability.

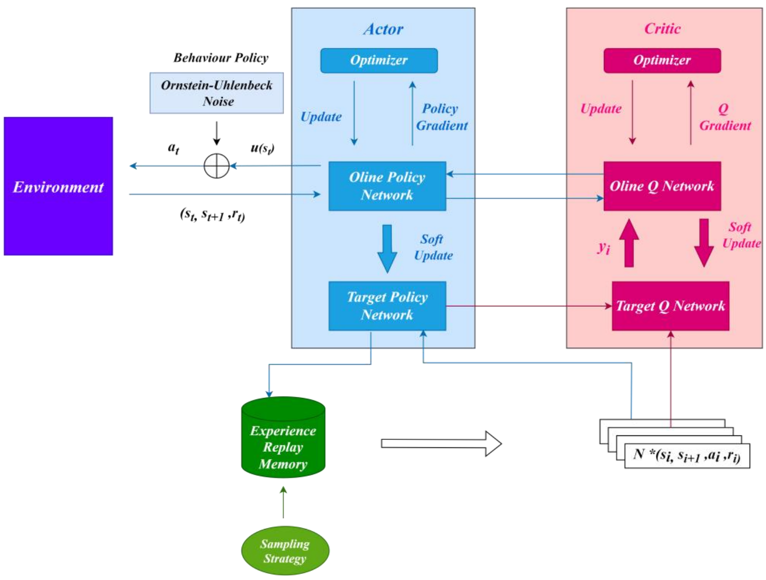

Figure 4. DDPG optimization process.

Figure 4 shows the DDPG algorithm framework, which uses the Actor-Critic architecture. DDPG learns the optimal strategy through the interaction of agents in the environment to maximize the cumulative reward. The Actor-Critic framework mainly includes four neural networks. The Actor network is responsible for inputting states and outputting deterministic actions. The Critic network is responsible for inputting states, actions, and outputs. The target Actor network & target Critic network are responsible for delayed updates and stable training.

In order to improve the scheduling and path planning efficiency of AGV equipment in the outbound management of power materials, this paper models the outbound management problem as a reinforcement learning optimization problem in a continuous action space. The optimization goal is to maximize the cumulative reward of the overall outbound process of AGV, comprehensively considering multiple indicators such as the shortest path, the shortest time, the successful obstacle avoidance, the smooth path, the lowest energy consumption and the satisfaction of task priority, so as to maximize the outbound efficiency, improve the system operation stability and optimize the resource utilization. The optimization target expression is shown in formula (1).

$$\max_{\pi} E\left[\sum_{t=0}^{T} \varrho^t r_t\right] \quad (1)$$

Where π represents the strategy output by the policy network, ϱ^t represents the reward discount factor, and T represents the upper limit of the time step of task execution.

In this paper, the decision variable is the AGV mobile action selection, such as forward, turning angle, speed adjustment and other continuous actions.

The constraint condition formula is shown in (2).

$$\begin{cases} \varsigma_{AGV,\,obstacle}(t) \geq \varsigma_{safe} \\ Eg_{total} \leq Eg_{\max} \\ Path(S_{start}S_{goal}) \in Path_{\forall} \\ T_i^{finish} \leq T_i^{deadline} \\ \left|\theta_{t+1} - \theta_t\right| \leq \varphi_{\max} \text{ and } 0 \leq v_t \leq v_{\max} \end{cases} \quad (2)$$

Where $\varsigma_{AGV,\,obstacle}$ represents the distance between the AGV and the obstacle at any time, and ς_{safe} represents the safety distance. Eg_{total} represents the total energy consumption of the AGV in the outbound task, and $Eg_{\max}$ represents the maximum allowable energy consumption threshold. $T_i^{deadline}$ represents the deadline of the task, $\varphi_{\max}$ represents the maximum turning rate, and $v_{\max}$ represents the maximum speed.

The definition of the state space is shown in formula (3).

$$S_t = \left[x_t, y_t, \theta_t, v_t, g_t, \varsigma_{AGV,\,obstacle}(t), Eg_{remain}(t), p_t, T_i^{clapsed}\right] \quad (3)$$

Where x_t and y_t represent the current position coordinates of the AGV, θ_t represents the current orientation angle of the AGV, v_t represents the current speed, g_t represents the target position coordinates, $Eg_{remain}(t)$ represents the current remaining power, p_t represents the priority information of the current task, and $T_i^{clapsed}$ represents the elapsed time.

The definition of the action space is shown in formula (4).

$$A_t = \left[a_v(t), a_\theta(t), a_f(t)\right] \quad (4)$$

Where $a_v(t)$ represents linear speed adjustment, $a_\theta(t)$ represents angular speed adjustment, and $a_f(t)$ represents direction.

Framework process steps:

Actor selects one a_t according to the behavior strategy and sends it to the environment to execute a_t:

$$a_t = \mu(s_t \mid \theta^\mu) + \mathcal{N}_t \quad (5)$$

The behavior strategy is a random process generated by the current online strategy and random noise, and the value sampled from this random process.

In reinforcement learning, the state-action value function $Q^\pi(s,a)$ is used to measure the expected value of future cumulative rewards after taking action a from state s under the strategy:

$$Q^\pi(s,a) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \middle| s_0 = s, a_0 = a\right] \quad (6)$$

γ is the discount factor.

Under the deterministic strategy $\mu(s)$, the optimal Q value satisfies the Bellman equation [31]:

$$Q(s,a) = r(s,a) + \gamma \mathbb{E}_{s' \sim P}\left[Q(s', \mu(s'))\right] \quad (7)$$

$s^{'}$ is the next state.

Deterministic Policy Gradient: DDPG uses a policy gradient method to optimize the policy μ(s).

$$\nabla_{\theta^{\mu}} j = \mathbb{E}_{s \sim \rho^{\beta}} \left[\nabla_a Q(s, a \left| \theta^{Q} \right.)_{a=\mu(s)} \nabla_{\theta^{\mu}} \mu(s \left| \theta^{\mu} \right.) \right] \quad (8)$$

Among them, j is the expected return of the strategy, and θ^{μ} is the parameter of the policy network.

Target value of the Critic network: The Critic network is used to evaluate the action selection of the Actor network, and its target value is calculated by the target Q network.

$$y = r + \gamma Q^{'}(s^{'}, \mu^{'}(s^{'} \left| \theta^{\mu^{'}} \right.) \left| \theta^{Q^{'}} \right.) \quad (9)$$

Among them, $Q^{'}$ and $\mu^{'}$ are the target Critic network and the target Actor network respectively; $\theta^{\mu^{'}}$ and $\theta^{Q^{'}}$ are the parameters of the target network.

The Critic network uses the mean square error loss function to minimize the Q value estimation error. The optimization purpose is to make the estimated Q value as close to the target value as possible.

$$L(\theta^{Q}) = \mathbb{E}\left[(Q(s, a \left| \theta^{Q} \right.) - y)^2 \right] \quad (10)$$

Policy update of actor network: The parameter update of actor network is based on policy gradient.

$$\nabla_{\theta^{\mu}} j = \mathbb{E}_{s \sim \mathcal{D}} \left[\nabla_{\theta^{\mu}} \mu(s \left| \theta^{\mu} \right.) \nabla_a Q(s, a \left| \theta^{Q} \right.) \left| a = \mu(s) \right. \right] \quad (11)$$

Among them, $\mathcal{D}$ is the experience playback buffer.

In order to improve the training stability, DDPG adopts a soft update strategy to update the parameters of the main network to the target network:

$$\theta^{Q^{'}} \leftarrow \tau \theta^{Q} + (1-\tau) \theta^{Q^{'}} \quad (12)$$

$$\theta^{\mu^{'}} \leftarrow \tau \theta^{\mu} + (1-\tau) \theta^{\mu^{'}} \quad (13)$$

Among them, τ is a soft update parameter.

DDPG uses the experience replay mechanism to store the experience of the agent's interaction with the environment and randomly samples for training to reduce data correlation and improve training stability.

The stored data format is:

$$(s_t, a_t, r_t, s_{t+1}) \quad (14)$$

Since DDPG uses a deterministic strategy and is prone to falling into local optimality, it is necessary to add noise for exploration:

$$a_t = \mu(s_t \left| \theta^{\mu} \right.) + \mathcal{N}_t \quad (15)$$

N is the noise term, and the Ornstein-Uhlenbeck (OU) process is usually used to simulate time-related noise and improve exploration efficiency.

OU process:

$$dx_t = \theta(\mu - x_t)dt + \sigma dW_t \quad (16)$$

Among them, μ is the long-term mean.

Based on the traditional DDPG reward mechanism, this paper innovatively designs a six-factor reward structure, including path length reward, time consumption penalty, obstacle avoidance reward, path smoothness reward, energy consumption optimization penalty and task priority reward. Through multi-objective weighted optimization, the design guides AGV to perform efficient path planning in complex warehouse environments, reduce redundant points, reduce the number of turns and energy consumption, and optimize task priorities. In terms of obstacle avoidance rewards and energy consumption optimization, continuous quantitative indicators are introduced to avoid extreme negative incentives and ensure a smooth training process. The task priority is dynamically adjusted in the reward mechanism, and the reward weight is adjusted in real time according to the timeliness of the task, thereby optimizing the decision-making strategy in multi-task scheduling and improving scheduling efficiency and resource utilization.

$$r_t = w_1 R_d + w_2 R_t + w_3 R_s + w_4 R_c + w_5 R_p \quad (17)$$

Among them:

- r_t : The reward obtained by the agent at time t.
- R_d : Path length reward.
- R_t : Time reward.
- R_s : Path smoothness reward, reduce sharp turns.
- R_c : Obstacle avoidance reward, avoid collisions or high-risk areas.

- R_p : Task priority reward, high-priority tasks receive additional incentives.

The weight parameters are determined by grid search: other weights are fixed first, and different combinations are tested in the range of [w1, w2,...]∈{0.1,0.3,0.5}, and the weights that make the AGV completion time, path length and other indicators optimal are selected.

Obstacle avoidance reward: To improve safety, AGV should avoid collisions with obstacles.

$$R_c = \begin{cases} -\lambda_3, & 1 \\ 0, & 0 \end{cases} \quad (18)$$

Among them, λ_3 is the collision penalty coefficient, 1 means collision occurs, and 0 means no collision. The design penalty can also be performed by adjusting the distance between obstacles.

$$R_c = -\lambda_4 \sum_{i=1}^{N} \frac{1}{d_{obs,i} + \varepsilon} \quad (19)$$

The above reward factors comprehensively consider the internal and external factors of AGV machines and equipment, conform to the actual navigation situation, improve the quality and efficiency of DDPG path planning, and enhance environmental adaptability.

The system interaction diagram of 6G+CPS and DDPG is shown in Figure 5.

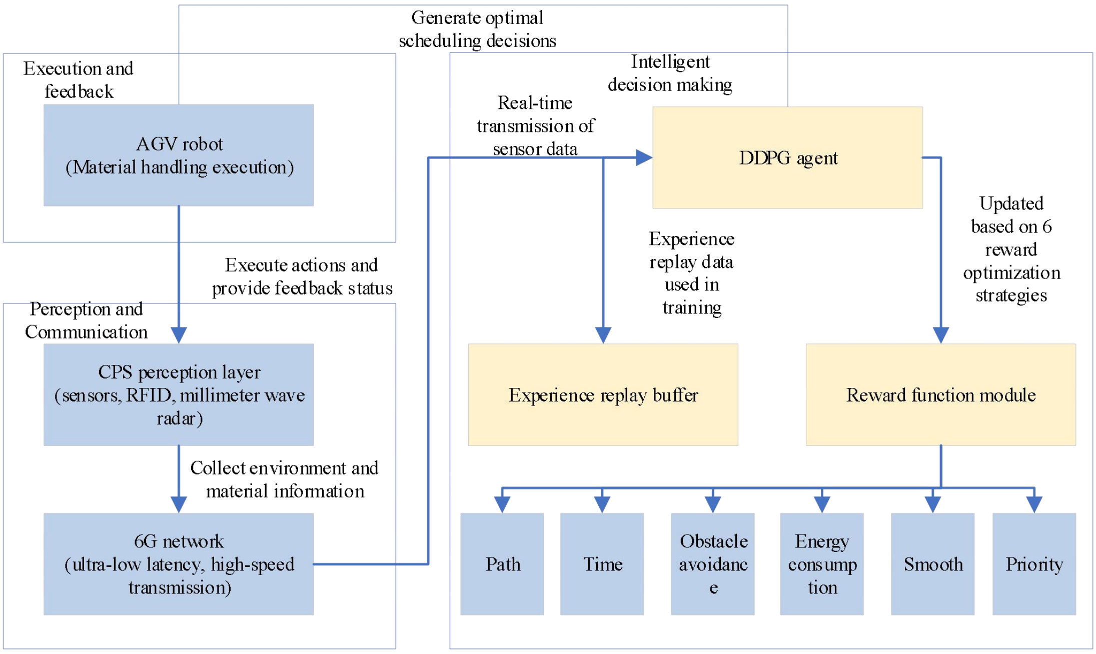

Figure 5. 6G+CPS and DDPG system interaction diagram.

In Figure 5, the CPS perception layer collects warehouse environment and material status information in real time through ultra-high precision sensors, and efficiently transmits the perception data to the DDPG agent with the help of the ultra-low latency and high-speed transmission capabilities of the 6G network. Based on the received environmental status, DDPG combines a multi-dimensional reward mechanism to generate the optimal outbound dispatch decision and command the AGV equipment to perform the handling task. After completing the action, the AGV transmits the feedback status back to the CPS, forming a closed loop of perception-decision-execution-feedback, continuously improving the system's autonomous adaptability and optimization capabilities, and realizing the intelligent and efficient management of power material in and out of the warehouse.

Compared with general material storage and dispatching, there are several distinct and unique features in the storage and dispatching of power materials.

(1) Power materials are complex in type, high in value, and some materials have timeliness and urgency requirements, which require extremely high storage and retrieval speed and outbound response time.

(2) The power industry's requirements for safety and stability far exceed those of general logistics. The inbound and outbound processes of materials need to be strictly controlled to avoid dispatch conflicts and equipment waiting.

(3) The storage environment requirements for power materials are diverse, such as high temperature, low temperature, explosion-proof, and moisture-proof, which requires highly dynamic perception and adaptation of storage layout and dispatching.

(4) Sudden emergency events, such as natural disasters and equipment failures, pose extremely high challenges to the rapid dispatch response of materials.

The above characteristics are directly reflected in modeling and methodology, such as introducing the CPS system to realize real-time environmental perception and intelligent feedback, using 6G network to improve the timeliness of perception and control, using reinforcement learning DDPG to dynamically optimize AGV path planning and scheduling strategies, and incorporating multi-dimensional indicators such as path smoothness, energy consumption optimization and task priority into the reward mechanism to ensure the scheduling system is efficient, stable and intelligent in response to highly complex environments.

4. Experimental Design

A. Dataset Collection

During the experiment, warehouse location information was collected and a high-precision 3D warehouse model was built based on this data. The model was built using Autodesk Revit professional 3D modeling software, which completely restored the warehouse's internal layout, shelf distribution, and channel information, providing support for subsequent warehouse management optimization and AGV path planning experiments.The display diagram of the simulated power material warehouse is shown in Figure 6. In the warehouse scene in Figure 6, the first picture shows that the robot is mainly used to move goods from the goods placement ground to the assembly line, the second and fourth pictures show that the robot is mainly used to stack items, and the third picture shows that the robot is mainly used to move goods from the assembly line to the shelf.

This study used Autodesk Revit software to build a high-precision three-dimensional warehouse model. As a professional software widely used in building information modeling (BIM), Revit has the advantages of precise modeling, strong detail restoration ability, and good compatibility. It can accurately reproduce the internal layout of the warehouse, shelf distribution, and channel information. This paper chooses Revit mainly because it supports complex structure modeling, space optimization analysis, and can provide an accurate three-dimensional environment foundation for subsequent warehouse management optimization and AGV path planning simulation experiments, thereby improving the transparency and repeatability of the research.

In the warehousing management of electric power materials, the warehousing task refers to the process of receiving, inspecting, coding, and storing newly arrived materials in accordance with established warehousing rules. The warehousing task requires that materials be placed in the optimal location quickly and accurately to maximize warehouse space utilization and facilitate subsequent outbound operations. The outbound task is to extract specified materials from the warehouse according to scheduling instructions and transport them to the shipping or use area. The outbound task must not only consider the outbound priority and distribution order of materials, but also optimize route planning, avoid transportation conflicts, and improve overall outbound efficiency. As the two core links of warehousing operations, warehousing and outbound directly affect the operating efficiency and service response speed of the warehousing system.

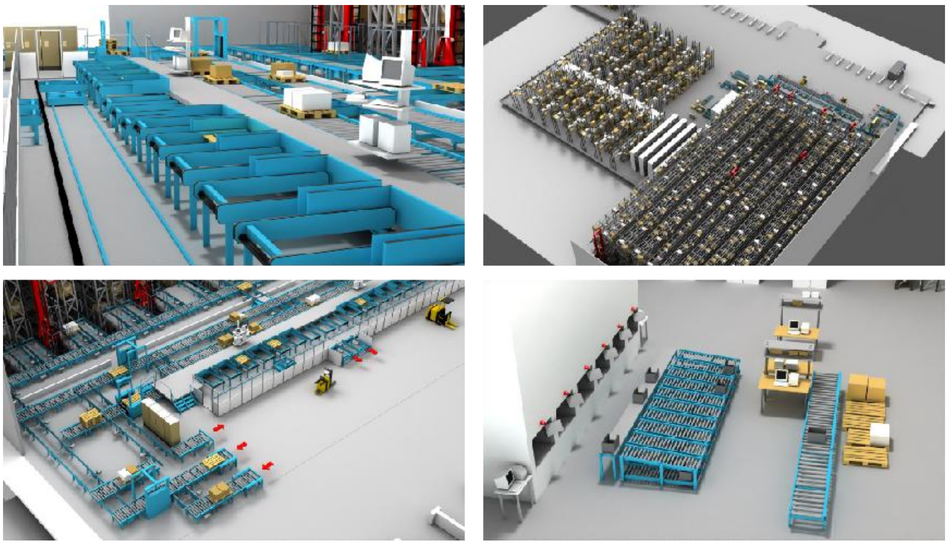

Figure 6. Display of simulated power material warehouse.

As can be seen from Figure 6, different robots are performing material flow work at their respective positions. Goods are effectively identified by robot equipment and distributed to different shelf locations for effective management and storage. Different goods are stored in different locations, which requires robots to have high-intensity computing capabilities and strategies. This paper uses the CPS system combined with the 6G network to effectively place and manage the incoming materials. At the same time, multiple AGV machines and equipment use superior instructions to effectively carry out the outbound delivery of materials. This paper introduces the DDPG algorithm to optimize the path planning efficiency and operation performance of AGV in the material handling process.

Table 2. Warehouse shelf placement information.

Shelf number	Latitude(°)	Longitude(°)	Storage category	Shelf capacity	Current inventory
S01	115.5368	30.4321	Power transmission &distribution equipment	100	80
S02	115.4789	30.3012	Cable	50	40
S03	115.5997	30.2894	Power distribution components	200	150
S04	115.5213	30.4118	Insulation materials	150	120
S05	115.6812	30.2743	Electric motor	80	60
S06	115.4608	30.3539	Battery equipment	120	100
S07	115.5091	30.4204	Cable	90	70
S08	115.6289	30.3592	Insulation materials	60	50
…	…	…	…	…	…

Table 2 shows the placement of warehouse shelves. The location and commodity categories of each shelf are different. They can be divided into (power transmission and distribution equipment, cables/wires, distribution components, insulation materials, motors, storage batteries) and other categories. The incoming materials need to be stored according to the actual categories to facilitate subsequent outbound operations.

Table 3. AGV machine equipment location.

Equipment No.	Latitude(°)	Longitude(°)	Load capacity (kg)	Running speed (m/s)	State
AGV-01	121.6237	31.4962	500	1.2	Leisure
AGV-02	121.4591	31.2299	600	1.5	In transit
AGV-03	121.5374	31.3745	550	1.3	Charging
AGV-04	121.6982	31.3126	700	1.4	Under maintenance
AGV-05	121.4859	31.2988	500	1.2	Leisure
AGV-06	121.6002	31.2354	650	1.6	In transit
AGV-07	121.5346	31.3901	550	1.3	In transit
AGV-08	121.6905	31.2783	1000	1.0	Leisure
AGV-09	121.5229	31.4154	700	1.4	In transit
…	…	…	…	…	…

Table 3 shows the parameter indicators of each AGV machine equipment, its location, the weight of goods it can bear, the delivery speed and the status of the machine equipment.

B. Data Ppreprocessing

The collected data is preprocessed to better train the subsequent model and check whether there are any missing values. In the case of missing data, the mean filling method is used to fill.

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (20)$$

For the location longitude and latitude in common data tables, they can be converted to decimal degrees.

$$\mathrm{DD} = \mathrm{Degrees} + \frac{\mathrm{Minutes}}{60} + \frac{\mathrm{Seconds}}{3600} \quad (21)$$

Table 2 contains the category marks of different stored commodities, which can be optimized using the label encoding formula.

$$l_i = \text{Label Encoding}(\mathrm{c_i}) \quad (22)$$

Finally, the data is normalized and scaled to a specific range.

$$Z = \frac{X-\mu}{\sigma} \quad (23)$$

μ is the average value of the data.

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (24)$$

C. Simulation Experiment

In order to verify the effectiveness of the DDPG fusion algorithm for inbound and outbound scheduling management, this paper conducts a simulation experiment. A workspace is created in the specified directory and the working environment is configured. This paper sets the environment variables and adds dependencies. The simulation algorithm is a map-free navigation based on DQN [32], DDPG, A2C [33,34], A3C, GWO, REINFORCE (Monte Carlo policy gradient) [35]. It includes two simulation environments: simple environment and complex environment. The relevant experimental parameters are shown in Table 4.

In this paper, the hyperparameter setting of the DDPG algorithm is mainly based on the trade-off between the convergence speed and stability of the algorithm. A smaller learning rate (0.0001 for Actor network and 0.001 for Critic network) helps to improve the stability of training and avoid violent gradient fluctuations; the discount factor of 0.99 is used to emphasize long-term returns, making the strategy more forward-looking in decision-making; the sampling batch size of 128 balances the training efficiency and update quality; the soft update rate of 0.01 ensures smooth updates of the target network and reduces the accumulation of estimation errors; the exploration noise variance is set to 0.1 to guide the agent to have sufficient exploration in the early stage, while controlling the noise amplitude to accelerate convergence; the experience pool size is set to 100,000 to ensure rich sample diversity, reduce the correlation between samples, and improve the training effect. The overall parameter setting aims to take into account the training stability, exploration ability and final strategy performance in the simulation environment.

Table 4. Initialization parameter settings.

Training parameters	Value
Learning Rate	0.0001
Discount Factor	0.99
Sampling batch	128
Soft update rate	0.01
Exploring Noise Variance	0.1
Experience pool size	100 000
Critic network learning rate	0.001
Actor Network Learning Rate	0.0001

Table 4 shows the initial parameters designed for the experimental process, including learning rate, discount factor, sampling batch, soft update rate, exploration noise variance, experience pool size, and learning rate functions of the two target networks (Critic network, Actor network).

5. Results

A. Warehouse Scheduling Optimization

By combining CPS with the 6G network, each incoming material is scanned and identified, and the material category label is added to it, so that the incoming materials can be effectively managed.In Table 5, the storage efficiency is calculated based on the average change in the processing time of a single item.

Table 5. Scheduling optimization effect.

Evaluation Metric	Traditional Solution (No Optimization)	6G Network Slicing Only	DDPG Algorithm Only (5G)	Full Solution (6G+DDPG)	Contribution Breakdown
Warehousing efficiency	45s/piece	32s/piece	38s/piece	25s/piece	6G: 29%↓ (45→32)
					DDPG: 15%↓ (45→38)
Category Classification	73%	88%	85%	93%	6G: 15%↑ (73%→88%)
					DDPG: 12%↑ (73%→85%)
Placement Accuracy	75%	82%	90%	96%	6G: 7%↑ (75%→82%)
					DDPG: 15%↑ (75%→90%)

Table 5 presents the performance comparison of warehouse scheduling optimization solutions before and after implementation. In traditional configurations, warehouse efficiency stood at 45 seconds per item with 73% category classification accuracy and 75% placement precision. The adoption of the comprehensive optimization solution (6G+DDPG) achieved a 44% improvement in warehouse efficiency to 25 seconds per item, a 20% increase in category classification accuracy to 93%, and a 21% enhancement in placement precision to 96%. Technical contribution analysis reveals that 6G network slicing primarily boosted warehouse efficiency (from 45 to 32 seconds, contributing 29%), attributed to its ultra-low latency that significantly accelerated material scanning and data transmission. The DDPG algorithm further optimized warehouse efficiency (from 32 to 25 seconds, contributing 15%) while substantially improving placement accuracy. Both 6G and DDPG contributed to improvements in category classification. Results demonstrate that 6G networks excel in real-time data transmission, while DDPG algorithms demonstrate superior spatial optimization capabilities. Their synergistic effect achieved significant overall performance enhancement.

B. Route Planning

The algorithm used in this paper and other optimization algorithms have a path optimization effect in a simple simulation scenario, where there are fewer obstacles.

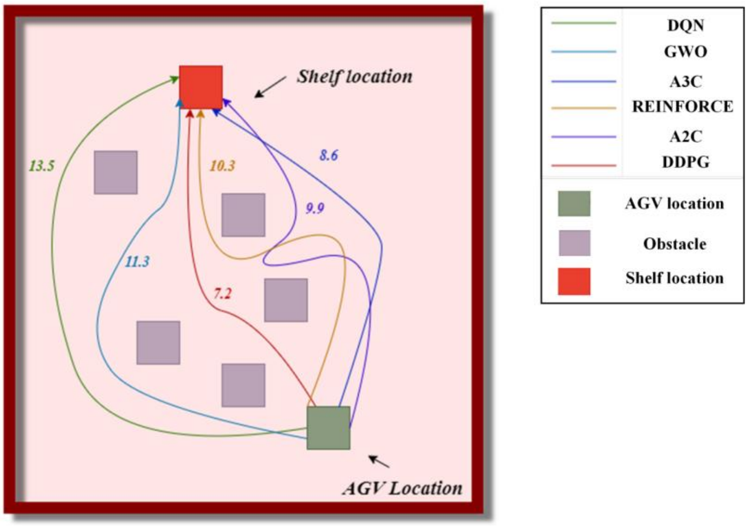

Figure 7. Path planning results in a simple environment.

Figure 7 shows the results of path simulation by different path optimization algorithms in a simple environment. The simple environment represents the number of obstacles (tables, stools) in the warehouse (greater than or equal to 5), and the path planning is performed by extracting power materials from the AGV equipment. The green line represents the DON optimization algorithm, which uses a deep neural network to represent the objective function and constraints of the optimization problem. The optimization approximates the optimal solution through a deep learning model. However, the DQN algorithm is sensitive to the initialization parameter settings. If the initialization is improper, the model may fall into a local optimal solution. The planned route distance is poor, and the path distance is 13.5m. The A3C algorithm can train multiple strategies in parallel to improve training efficiency. It generates and updates model parameters through asynchronous training strategies to accelerate algorithm convergence, but it requires multi-core hardware support and consumes a lot of resources. Its path distance is 8.6m. The A2C algorithm is a synchronous version of A3C. Both training and updating are relatively more stable. By introducing the advantage function, it effectively reduces the variance of the traditional policy gradient method and improves training efficiency. Due to the conservative policy update, A2C may have a low exploration degree, resulting in slow training or falling into a local optimum, with an optimization path distance of 9.9m. The REINFORCE algorithm uses the Monte Carlo method to update the policy, does not rely on function estimation, and directly optimizes the policy. However, the policy update can lead to a large variance of the policy gradient, and the training process is prone to instability and slow convergence. Its optimization path distance is 10.3m. The GWO algorithm is a swarm intelligence optimization algorithm that simulates the cooperative hunting behavior of gray wolf groups and searches for the optimal solution through different wolves distributed in the search space. However, it has poor adaptability and performs poorly in dynamic environments and cannot make rapid adjustments to the environment. Its optimized path distance is 11.3m. The DDPG algorithm Actor and Critic network used in this paper can efficiently generate path planning. The optimized path distance of the machine equipment is 7.2m, and the optimization effect is the best, which is significantly better than other algorithms. The optimized path distance of the machine equipment is 7.2m, and the optimization effect is the best, which is significantly better than other algorithms, indicating that it helps to optimize the path of AGV machine equipment and enable it to efficiently complete material outbound delivery.

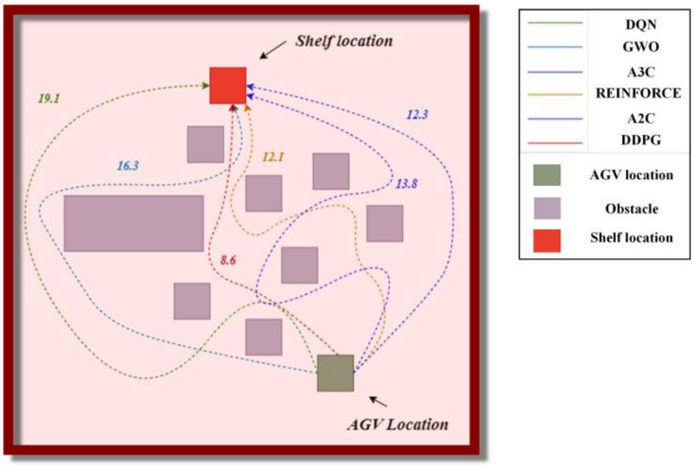

Figure 8. Path planning results in complex environments.

In order to verify the path optimization effect in different environments, a complex environment (Figure 8) was designed again for experiments. The number of obstacles (greater than or equal to 8) increased significantly, and the obstacle area also changed. By planning the path for AGV equipment to extract power materials in a complex environment, the DDPG algorithm also has the best path optimization effect, with an optimized distance of 8.6m, which is not significantly different from that in a simple environment. The DQN algorithm has the worst optimization effect with an optimized path distance of 19.1m in an environment with more and larger obstacles. The path distance of the A3C algorithm in a complex environment is 12.3m. The path distance of the A2C algorithm also reached 13.8m. The path optimization distance of the REINFORCE algorithm was 12.1m. The GWO algorithm performed poorly in complex environments, with a path distance of 16.3m. By verifying the material outbound management experiment in complex environments, it can be concluded that the DDPG algorithm has the best optimization effect in different environments. It can dynamically help the AGV robot to extract materials out of the warehouse, and effectively avoid obstacles, reducing unnecessary troubles and energy consumption.

Table 6. shows the optimization results of path planning under dynamic obstacle environment.

DDPG	9.2	22	26.5	1.3
A3C	12.7	30	22.3	1.9
A2C	14.1	33	21.1	2.1
REINFORCE	13.3	35	20.5	2.2
GWO	17.2	38	17.8	2.4
DQN	20.3	40	16.5	2.5

Table 6 presents the optimization results of path planning for different algorithms in a dynamic obstacle environment. The DDPG algorithm performed exceptionally well, achieving an optimized path length of 9.2 meters, significantly outperforming other algorithms. It also had the fastest convergence rate, reaching the optimal solution in just 22 training sessions. In terms of reward values, DDPG achieved an average reward of 26.5, demonstrating strong learning capabilities and optimization effects, with a standard deviation of 1.3, indicating high stability. In contrast, other algorithms such as A3C, A2C, and REINFORCE had longer optimized path lengths, at 12.7 meters, 14.1 meters, and 13.3 meters, respectively, and slower convergence rates, requiring 30,33, and 35 training sessions. Additionally, the GWO and DQN algorithms performed poorly in dynamic obstacle environments, with longer optimized path distances and lower reward values. These results highlight DDPG's superiority in complex dynamic environments, particularly in path optimization and algorithm stability.

C. *Reward Value Verification*

In this experiment, the same reward mechanism is designed for each algorithm to verify the reward value change trend of each algorithm in different environments.

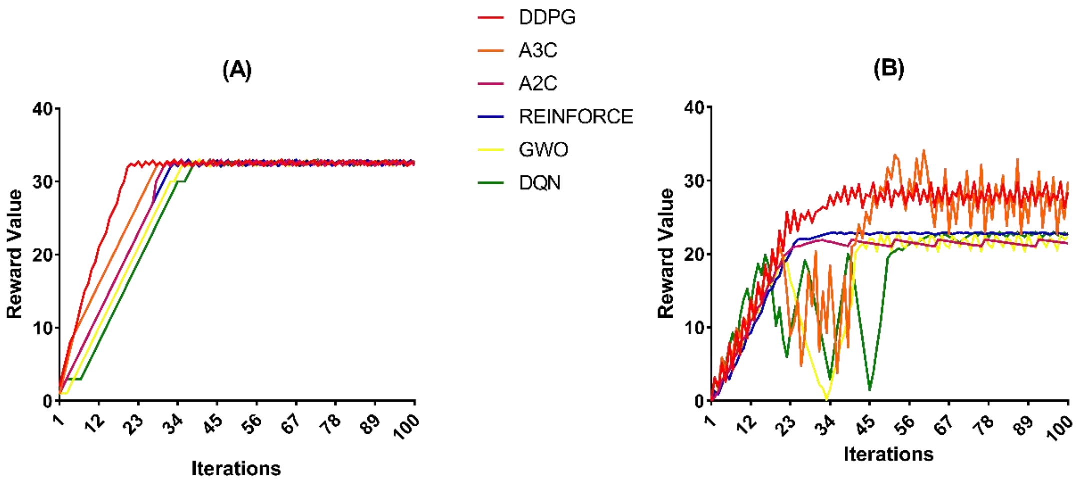

Figure 9. Comparison of reward values. (A). Reward value in a simple environment (B). Reward value in a complex environment.

Figure 9 shows the reward value results of path planning performed by each algorithm in different environments. The X-axis represents the number of iterations of the algorithm, and the Y-axis represents the reward value. Figure 9A shows the reward values of each algorithm in a simple environment. It can be seen that the red line represents the DDPG algorithm used in this paper. When the number of training times is 20, the line tends to converge, and the reward value tends to be stable at 32. This shows that the DDPG algorithm can make the most accurate route planning in the early training. The A3C algorithm started to converge and gradually stabilize after 28 training times. The convergence speed was lower than that of the DDPG algorithm. The convergence time of the A2C algorithm was similar to that of the A3C algorithm. It gradually converged after 30 training times. The REINFORCE algorithm started to converge after 32 training times. The GWO algorithm did not have a good convergence effect on the reward value during its path optimization process. The algorithm converged after 35 training times. The DQN algorithm has the worst reward value convergence effect and the slowest convergence speed. The algorithm stabilizes after 38 trainings. Figure 9B shows the reward value change trend of each model in a complex environment. The average reward value of the DDPG algorithm in a complex environment can reach 24.12, and stabilizes in the 28 range in the later stage of training iterations. The average reward value of the A3C algorithm is 21.26, but in the early stage of training (1-45), the reward value fluctuates greatly and lacks stability. The average reward value of the A2C algorithm is 18.98, and its reward value performs well, rising steadily and stabilizing in the 21 range in the later stage. The average reward value of the REINFORCE algorithm is 19.75. The reward value is stable in the 22 range in the later stage. The average reward value of the GWO algorithm is 16.53. Its reward value curve showed huge fluctuations in the early stage of training, lacking stability. The average reward value of the DQN algorithm was 16.76, which also showed huge fluctuations. Under the influence of complex environments, it was difficult for the algorithm to effectively calculate the optimal path, lacking stable performance. Through the analysis of the reward results, it can be concluded that the DDPG algorithm used in this paper is superior to other excellent algorithms in terms of convergence speed and reward value comparison, which means that it can effectively help AGV equipment complete the resource outbound scheduling optimization task.

D. Significance Test

Based on the experimental design, this paper collects 50 running results of different algorithms in each environment. For each algorithm, the average reward value and standard deviation of multiple runs are calculated in each environment. The experiment sets up two hypotheses: Null hypothesis (H0): There is no significant difference in the reward value of the two algorithms in the same environment. Alternative hypothesis (H1): There is a significant difference in the reward value of the two algorithms in the same environment. If the p value is less than the significance level (0.05), the null hypothesis is rejected and it is considered that there is a significant difference between the two algorithms. The results of the significance test are shown in Table 7.

Table 7. Significance test results.

Environment	Algorithm	Average reward value	Standard deviation	T-test results with DDPG	P value
Simple environment	DDPG	28.5	1.25	-	-
	A3C	25.3	1.45	2.35	0.032
	A2C	24.8	1.2	3.40	0.01
	REINFORCE	22.5	1.6	4.60	0.002
	GWO	20.9	2.1	5.20	0.002
	DQN	18.5	2.3	6.70	0.001
Complex environment	DDPG	22.28	1.4	-	-
	A3C	19.14	1.8	2.50	0.024
	A2C	16.38	1.9	3.60	0.008
	REINFORCE	17.15	2	2.85	0.015
	GWO	14.24	2.3	4.90	0.01
	DQN	14.49	2.1	5.10	0.01

In Table 7, it can be seen that in simple environments, the average reward value of the DDPG algorithm is the highest, which is 28.5, with a standard deviation of 1.25. Among other algorithms, the average reward values of A3C, A2C, REINFORCE, GWO, and DQN are 25.3, 24.8, 22.5, 20.9, and 18.5, respectively. After the T test with DDPG, the p values are all less than 0.05, indicating that compared with DDPG, these algorithms have significant differences in performance in simple environments. Among them, the DQN algorithm has the largest t value with DDPG, which is 6.70, and the p value is 0.001, indicating that the difference is the most significant. In complex environments, the average reward value of DDPG is 22.28, which is still higher than other algorithms; the average reward values of A3C, A2C, REINFORCE, GWO, and DQN drop to 19.14, 16.38, 17.15, 14.24, and 14.49, respectively. In complex environments, the T-test p values of all algorithms and DDPG are still all less than 0.05, indicating that the DDPG algorithm is also significantly better than other algorithms in complex environments. Whether in simple or complex environments, the DDPG algorithm shows obvious advantages in average reward value and stability.

Compared with other intelligent warehouse optimization studies, this method performs better in terms of reward value improvement and convergence speed. The average reward value of existing warehouse scheduling algorithms based on reinforcement learning in simple environments is usually less than 25, and the training convergence requires more than 30 times, while the DDPG algorithm has a reward value of 28.5 under the same conditions and only requires 20 times to converge. In complex environments, the traditional method has large fluctuations in reward values (standard deviation>2) due to insufficient adaptability to dynamic obstacles. This method stabilizes the reward value above 22 (standard deviation 1.4) through the coordination of 6G-CPS real-time perception and DDPG dynamic decision-making, verifying its robustness advantage in complex scenarios.

6. Conclusions

This study proposes an intelligent storage and dispatch optimization scheme for power materials based on deep reinforcement learning DDPG algorithm. By integrating CPS and 6G network technology, the response efficiency of the power material supply chain is significantly improved. In the warehousing stage, the system shortens the material scanning time from 45 seconds to 25 seconds (efficiency improvement of 44%), and the classification accuracy rate reaches 93%, providing real-time data support for the access of new energy equipment and the allocation of emergency repair materials; in the outbound stage, the DDPG algorithm makes the path optimization distance of AGV in the power storage environment (7.2m in simple environment/8.6m in complex environment) significantly better than the comparison algorithm, effectively ensuring the rapid outbound of power repair materials. Experiments show that the scheme has the best convergence speed (stable after 20 trainings) and reward value (28 points in complex environment), and can adapt to the storage and dispatch needs of special power materials such as substation equipment and cables. The research results provide key technical support for the precise dispatch of materials in the construction of smart grids. This method not only improves storage efficiency through 6G-CPS real-time perception and DDPG dynamic path planning, but can also be extended to new energy equipment logistics scenarios, such as coordinating the in-and-out priorities of energy storage batteries, or adapting to the emergency material dispatching needs of smart grids, further supporting low-carbon electricity.

There are still some shortcomings in this paper, which need to be further improved: I. The experiment did not fully consider complex situations such as dynamic obstacles and sudden task adjustments, which affected the generalization ability of the algorithm in real application scenarios. II. The algorithm may have problems such as slow training convergence and limited real-time performance in large-scale warehousing systems. In the future, lightweight model optimization strategies or edge computing technologies can be combined to further improve the computational efficiency and online scheduling capabilities of the algorithm. III. The optimization strategies in this study are mainly evaluated based on simulation experiments and have not yet been tested on a large scale in real warehouse systems. In the future, AGV can be deployed for field verification in combination with the actual warehouse environment to further evaluate the adaptability and stability of the method, optimize the scheduling strategy, and improve the practical application value of the system.

Acknowledgement

This work was sponsored in part by State Grid Hebei Electic Power Co.: Key technologies and applications of integrated coordination of power material detection, storage and delivery(kj2021-053)

In writing this literature review, the study drew on the research findings of many predecessors across various fields. Special thanks go to the scholars who have provided crucial theoretical and technical support for power material storage scheduling, reinforcement learning applications, CPS systems, and 6G networks. The author is deeply grateful for the inspiration from the data, algorithms, and methods used in related work. Additionally, the author appreciates the contributions of other authors to the research on intelligent warehousing and path optimization, which have provided valuable references and insights for this study.

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Author Contribution

[Jun Zhao]: Developed and planned the study, performed experiments, and interpreted results. Edited and refined the manuscript with a focus on critical intellectual contributions.

[Zengchao Wang, Ming Su]: Participated in collecting, assessing, and interpreting the date. Made significant contributions to date interpretation and manuscript preparation.

[Dongliang Wang, Feng Hao, Kefeng Li, Zihan Li]: Provided substantial intellectual input during the drafting and revision of the manuscript.

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] J.L. Cox, W.T. Hamilton, A.M. Newman, M.J. Wagner, A.J. Zolan. Real-time dispatch optimization for concentrating solar power with thermal energy storage. Optimization and Engineering, 2023, 24(2), 847-884. DOI: 10.1007/s11081-022-09711-w

[2] X.R. Li. On Warehouse and distribution operation optimization of electric power materials. Academic Journal of Business & Management, 2023, 5(5), 146-151. DOI: 10.25236/AJBM.2023.050521

[3] T. Egeland-Eriksen, A. Hajizadeh, S. Sartori. Hydrogen-based systems for integration of renewable energy in power systems: Achievements and perspectives. International Journal of Hydrogen Energy, 2021, 46(63), 31963-31983. DOI: 10.1016/j.ijhydene.2021.06.218

[4] J. Mitali, S. Dhinakaran, A.A. Mohamad. Energy storage systems: A review. Energy Storage and Saving, 2022, 1(3), 166-216. DOI: 10.1016/j.enss.2022.07.002

[5] X.S. Mu, M.F. Antwi-Afari. The applications of Internet of Things (IoT) in industrial management: a science mapping review. International Journal of Production Research, 2024, 62(5), 1928-1952. DOI: 10.1080/00207543.2023.2290229

[6] S.A. Hashmi, C.F. Ali, S. Zafar. Internet of things and cloud computing-based energy management system for demand side management in smart grid. International Journal of Energy Research, 2021, 45(1), 1007-1022. DOI: 10.1002/er.6141

[7] A.Z. Arsad, M.A. Hannan, A.Q. Al-Shetwi, M. Mansur, K.M. Muttaqi, Z.Y. Dong, F. Blaabjerg. Hydrogen energy storage integrated hybrid renewable energy systems: A review analysis for future research directions. International Journal of Hydrogen Energy, 2022, 47(39), 17285-17312. DOI: 10.1016/j.ijhydene.2022.03.208

[8] Wang J, Zhao N, Mi C. ART path planning method based on the 3D steering angle weighted A-star algorithm. International Journal of Vehicle Information and Communication Systems, 2024, 9(3), 256-275. DOI: 10.1504/IJVICS.2024.139778

[9] A. Ubaidillah, H. Sukri. Application of Odometry and Dijkstra Algorithm as Navigation and Shortest Path Determination System of Warehouse Mobile Robot. Journal of Robotics and Control (JRC), 2023, 4(3), 413-423. DOI: 10.18196/jrc.v4i3.18489

[10] Z.F. Zong, T. Feng, J.W. Wang, T. Xia, Y. Li. Deep Reinforcement Learning for Demand-Driven Services in Logistics and Transportation Systems: A Survey. ACM Transactions on Knowledge Discovery from Data, 2025, 19(4), 1-42. DOI: 10.48550/arXiv.2108.04462

[11] S. Pal. Intelligent warehouse space optimization using convolutional neural networks. World Journal of Advanced Engineering Technology and Sciences, 2023, 10(2), 030-036. DOI: 10.30574/wjaets.2023.10.2.0278

[12] I.H. Zakaria, J. Lyn, P.M. Azizul, F. Anis, A.bakri. Smart Warehousing Solutions: Enhancing Efficiency and Sustainability in Logistics and Transportation. Journal on Technical and Vocational Education, 2024, 9(2), 409-420. http://upikpolimas.edu.my/ojs/

[13] E.O. Sodiya, U.J. Umoga, O.O Amoo, A. Atadoga. AI-driven warehouse automation: A comprehensive review of systems. GSC Advanced Research and Reviews, 2024, 18(2), 272-282. DOI: 10.30574/gscarr.2024.18.2.0063
[14] B.T. Liu, Y. Liu, S.Y. Hu, W.M. Zhe. Opportunities and challenges of scheduling in logistics industrial park cyber-physical systems. IEEE Transactions on Industrial Cyber-Physical Systems, 2023, 1(1), 322-334. DOI: 10.1109/TICPS.2023.3330118
[15] C. Aron, F. Sgarbossa, E. Ballot, D. Ivanov. Cloud material handling systems: A cyber-physical system to enable dynamic resource allocation and digital interoperability. Journal of Intelligent Manufacturing, 2024, 35(8), 3815-3836. DOI: 10.1007/s10845-023-02262-6
[16] Y. Lu, S.D. Fang, G.H. Chen, T. Niu and R.J. Liao. Cyber-physical integration for future green seaports: Challenges, state of the art and future prospects. IEEE Transactions on Industrial Cyber-Physical Systems, 2023, 1(1), 21-43. DOI: 10.1109/TICPS.2023.3283234
[17] L. Piardi, P. Costa, A. Oliveira, A. Oliveira, P. Leitão. MAS-based Distributed Cyber-physical System in Smart Warehouse. IFAC-PapersOnLine, 2023, 56(2), 6376-6381. DOI: 10.1016/j.ifacol.2023.10.826
[18] K.G. Li. Optimizing warehouse logistics scheduling strategy using soft computing and advanced machine learning techniques. Soft Computing, 2023, 27(23), 18077-18092. DOI: 10.1007/s00500-023-09269-4
[19] M. Çelik, C. Archetti, H. Süral. Inventory routing in a warehouse: The storage replenishment routing problem. European Journal of Operational Research, 2022, 301(3), 1117-1132. DOI: 10.1016/j.ejor.2021.11.056
[20] A. Rijal, M. Bijvank, R. De Koster. Dynamics between warehouse operations and vehicle routing. Production and Operations Management, 2023, 32(11), 3575-3593. DOI: 10.1111/poms.14051
[21] S.J.N. Liu, S.H. Liu, H.P. Xiao. Improved gray wolf optimization algorithm integrating A* algorithm for path planning of mobile charging robots. Robotica, 2024, 42(2), 536-559. DOI: 10.1017/S0263574723001625
[22] Q.W. Nie, D.B. Tang, H.H. Zhu, H.W. Sun. A multi-agent and internet of things framework of digital twin for optimized manufacturing control. International Journal of Computer Integrated Manufacturing, 2022, 35(10-11), 1205-1226. DOI: 10.1080/0951192X.2021.2004619
[23] X. Jing, X.Y. Yao, M. Liu, J.J. Zhou. Multi-agent reinforcement learning based on graph convolutional network for flexible job shop scheduling. Journal of Intelligent Manufacturing, 2024, 35(1), 75-93. DOI: 10.1007/s10845-022-02037-5
[24] Y. Bai, X. Ding, D. Hu, Y. Jiang. Research on dynamic path planning of multi-AGVs based on reinforcement learning. Applied Sciences, 2022, 12(16), 8166-8188. DOI: 10.3390/app12168166
[25] L.X. Zhang, Y. Yan, Y.G. Hu. Deep reinforcement learning for dynamic scheduling of energy-efficient automated guided vehicles. Journal of Intelligent Manufacturing, 2024, 35(8), 3875-3888. DOI: 10.1007/s10845-023-02208-y
[26] Fan F, Xu G, Feng N, et al. Spatiotemporal path tracking via deep reinforcement learning of robot for manufacturing internal logistics. Journal of Manufacturing Systems, 2023, 69(1), 150-169.DOI: 10.1016/j.jmsy.2023.06.011
[27] J.F. Leon, Y.D. Li, X.A. Martin, L. Calvet, J. Panadero, A.A. Juan. A hybrid simulation and reinforcement learning algorithm for enhancing efficiency in warehouse operations. Algorithms, 2023, 16(9), 408-429. DOI: 10.3390/a16090408
[28] L.X. Zhang, C. Yang, Y. Yan, Z. Cai, Y.G. Hu. Automated guided vehicle dispatching and routing integration via digital twin with deep reinforcement learning. Journal of Manufacturing Systems, 2024, 72(1), 492-503. DOI: 10.1016/j.jmsy.2023.12.008
[29] W.L. Duo, M.C. Zhou M.C, A. Abusorrah. A survey of cyber attacks on cyber physical systems: Recent advances and challenges. IEEE/CAA Journal of Automatica Sinica, 2022, 9(5), 784-800. DOI: 10.1109/JAS.2022.105548
[30] Z.H. Yu, H.X. Gao, X.Y. Cong, N.Q. Wu, H.H. Song. A survey on cyber–physical systems security. IEEE Internet of Things Journal, 2023, 10(24), 21670-21686. DOI: 10.1109/JIOT.2023.3289625
[31] S. Naser, L. Bariah, S. Muhaidat, E. Basar. Zero-energy devices empowered 6G networks: Opportunities, key technologies, and challenges. IEEE Internet of Things Magazine, 2023, 6(3), 44-50. DOI: 10.1109/IOTM.001.2200235
[32] Y. Zhong, X.H. Ge. Toward 6G carbon-neutral cellular networks. IEEE Network, 2024, 38(5), 174-181. DOI: 10.1109/MNET.2024.3411599
[33] S. Abbas, M. Abu Talib, Q. Nasir, S. Idhis, M. Alaboudi, A. Mohamed. Radio frequency fingerprinting techniques for device identification: a survey. International Journal of Information Security, 2024, 23(2), 1389-1427. DOI: 10.1007/s10207-023-00801-z
[34] C. Proctor-Parker, R. Stopforth. Radio frequency identification (RFID) stock control and geo-location data system from a moving vehicle. Journal of Engineering, Design and Technology, 2024, 22(1), 159-181. DOI: 10.1108/JEDT-10-2021-0548
[35] W. Wu, C.H. Zhou, M.S. Li, H.Q. Wu, H.B. Zhou, N. Zhang, et al. AI-native network slicing for 6G networks. IEEE Wireless Communications, 2022, 29(1), 96-103. DOI: 10.1109/MWC.001.2100338
[36] M.K. Hasan, N. Jahan, M.Z.A. Nazri, S. Islam, M. Attique Khan, A. I. Alzahrani, et al. Federated learning for computational offloading and resource management of vehicular edge computing in 6G-V2X network. IEEE Transactions on Consumer Electronics, 2024, 70(1), 3827-3847. DOI: 10.1109/TCE.2024.3357530