ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/364320158

Deep Reinforcement Learning in Smart Manufacturing: A Review and Prospects

Article *in* CIRP Journal of Manufacturing Science and Technology · February 2023

DOI: 10.1016/j.cirpj.2022.11.003

CITATIONS
43

READS
4,118

5 authors, including:

Chengxi Li
The Hong Kong Polytechnic University
17 PUBLICATIONS **330** CITATIONS
SEE PROFILE

Pai Zheng
The Hong Kong Polytechnic University
228 PUBLICATIONS **6,471** CITATIONS
SEE PROFILE

Yue Yin
The Hong Kong Polytechnic University
10 PUBLICATIONS **222** CITATIONS
SEE PROFILE

Baicun Wang
Zhejiang University
88 PUBLICATIONS **2,500** CITATIONS
SEE PROFILE

All content following this page was uploaded by Pai Zheng on 28 October 2022.

The user has requested enhancement of the downloaded file.

Deep Reinforcement Learning in Smart Manufacturing: A Review and Prospects

Chengxi Li[1,2], Pai Zheng[1,2]*, Yue Yin[1], Baicun Wang[3**], Lihui Wang[4]

[1]*Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong Special Administrative Region, China*

[2]*Laboratory for Artificial Intelligence in Design, Hong Kong Special Administrative Region, China*

[3]*State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China*

[4]*Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden*

Abstract

To facilitate the personalized smart manufacturing paradigm with cognitive automation capabilities, Deep Reinforcement Learning (DRL) has attracted ever-increasing attention by offering an adaptive and flexible solution. DRL takes the advantages of both Deep Neural Networks (DNN) and Reinforcement Learning (RL), by embracing the power of representation learning, to make precise and fast decisions when facing dynamic and complex situations. Ever since the first paper of DRL was published in 2013, its applications have sprung up across the manufacturing field with exponential publication growth year by year. However, there still lacks any comprehensive review of the DRL in the field of smart manufacturing. To fill this gap, a systematic review process was conducted, with 264 relevant publications selected to date (20-Oct-2022), to gain a holistic understanding of the development, application, and challenges of DRL in smart manufacturing along the whole engineering lifecycle. First, the concept and development of DRL are summarized. Then, the typical DRL applications are analyzed in the four engineering lifecycle stages: design, manufacturing, logistics, and maintenance. Finally, the challenges and future directions are illustrated, especially emerging DRL-related technologies and solutions that can improve the manufacturing system's deployment feasibility, cognitive capability, and learning efficiency, respectively. It is expected that this work can provide an insightful guide to the research of DRL in the smart manufacturing field and shed light on its future perspectives.

Keywords: Deep reinforcement learning; smart manufacturing; engineering life cycle; artificial intelligence; review

* *Corresponding author*: pai.zheng@polyu.edu.hk

1 Introduction

Smart manufacturing is an advanced manufacturing paradigm that profoundly integrates the new generation of information technology such as Internet-of-Things, cloud computing and artificial intelligence, and cutting-edge manufacturing technology into the production process [1]. These technologies are adopted to promote the efficiency of manufacturing systems by empowering them with the abilities of autonomous perception, optimized decision-making, and precise execution, bringing new vitality to smart manufacturing [2]. Such an intelligent trend has brought up new opportunities for the transformation and upgrade of the global manufacturing industry, realizing more flexible, adaptive, and personalized production processes, which is of great significance for the development of smart manufacturing.

Reinforcement learning (RL), as an important branch of AI algorithms, originally owns an outstanding capability of sequential decision-making. In the past decade, with the rapid development of high-performance computing power and the advancement of deep learning (DL) techniques, algorithms combining RL with deep neural networks, i.e., DRL have not only improved environment perception, but also enabled RL algorithms to make decisions with better performance, adaptability, and time efficiency [3]. Recently, DRL has been not only widely applied in games [4], recommendation systems [5], finance [6], network communication systems [7], and robot control [8][9][10], but it has also attracted ever-increasing attention from the industry field. As the manufacturing paradigm is shifting toward mass personalization, manufacturing systems need to respond to orders with a shorter lead-time and a higher quality, which requires the production process to be more flexible and adaptable. Under these circumstances, DRL has great potential owing to its self-learning capabilities to make precise and fast decisions when facing dynamic and complex situations. To date, hundreds of papers have discussed DRL applications in various engineering domains with a sharp increase year by year, including Cyber-Physical System [11], Energy System [12], Process Control [13][14], and Production Scheduling [15]. Nevertheless, in the manufacturing domain, there is only one systematic review on DRL published [16], which mainly focuses on production system applications. In addition, to the best of our knowledge, the engineering lifecycle implementations and technical trends of DRL in futuristic smart manufacturing have been little reported.

To address the gap, this systematic review attempts to summarize the current status of DRL applications in the typical four stages along the engineering lifecycle, i.e., design, manufacturing, distribution, and maintenance. Meanwhile, the challenges faced are analyzed to clearly explore the potential of DRL methods to address smart manufacturing demands. Lastly, the corresponding enabling technologies are proposed to shape future research directions. The rest of this paper is organized as follows: Section 2 introduces our systematic review strategy and provides an overview of the review result. Section 3 briefs the basic concept of DRL and lists the current progress of classical algorithms in DRL. Sections 4 to 7 further summarize existing methods and applications along the four different engineering lifecycle stages, respectively. Section 8 points out the challenges and future perspectives, and Section 9 discusses and concludes the paper.

2 The Systematic Review Process

Compared to the other supervised learning applications (e.g., computer vision, natural language processing, etc.), DRL incorporates trial-and-error principles to self-optimize through interacting

with the environment without any manual-labeled data. Such self-learning and labeled data characteristics of DRL significantly reduce human involvement and allow DRL to be easily adopted and deployed. Meanwhile, represented by DeepMind's AlphaGo series application [17], they demonstrate DRL's ability of instant decision making and generalization to past experience in the face of complex situations. In engineering fields, such as autonomous driving [18], IoT [19] and robot systems [20], researchers have become aware of the advantage, and comprehensive reviews are given. With the explosive growth of DRL applications in smart manufacturing, there still lacks sufficient work to comprehensively describe the research status and discuss under-solved challenges. Therefore, a state-of-the-art review of DRL along the entire engineering lifecycle of smart manufacturing is imperative.

2.1 Literature Selection Method

This subsection provides an overview of the basic literature review process for DRL applications in smart manufacturing applications. A systematic literature search is mainly conducted from two well-known academic databases, namely *Web of Science* and *Scopus*, since they cover most of the peer-reviewed interdisciplinary research papers, where a broad sum of studies on DRL can be identified. In our systematic literature review process, the keywords adopted in the search were: "manufacturing", "production", and "reinforcement learning", the time span was 2013-2022 (until 2022. Oct. 20), and only English literature was included. Although the focus algorithm type in our review is about DRL, "reinforcement learning" is still chosen as the keyword, to search relevant literature on and manually filter the appropriate target literature. This is because researchers did not clearly distinguish between the terms DRL and RL, when it was preliminarily introduced in the early development of DRL. Regarding the time span of the search, we chose 2013 as the starting point, because the representative work on DRL i.e., Deep Q-Network (DQN) [21] was warmly discussed in that year. In summary, the search phrase can be duplicated with the following search sentence:

- **Web of Science:**
  - TS= ("reinforcement learning" AND ("production" OR "manufacturing")) AND PY= (2013-2022)
- **Scopus:**
  - TITLE-ABS-KEY (("manufacturing" OR "production") AND "reinforcement learning") AND PUBYEAR > 2012

After preliminary searching and obtaining relevant literature from databases, the Web of Science and Scopus provide 913 and 1508 papers, respectively. Meanwhile, inclusion and exclusion criteria are defined to systematically narrow the scope and ensure a high-quality review. First, we excluded working papers, preprints, and other non-peer-reviewed publications. Then, we screened all the peer-reviewed journals and conferences in English and filtered out work that did not fit the objective domain or out of scope based on abstract and article browsing. Until this step, there were 573 papers correlated to our topic. Finally, only papers that leverage Deep Learning-based methods in model-free RL are considered instead of the other approach, leaving only 264 papers as the reference for our review.

2.2 Descriptive Analysis

Following our manner, applications of DRL in real or simulated manufacturing environments are reviewed and grouped into four phases of the product lifecycle (i.e., Design, Manufacturing,

Logistics, and Maintenance) based on the respective tasks solved. As the quantitative analysis in *Figure 1* shows, the majority of DRL applications are still located in the manufacturing stage.

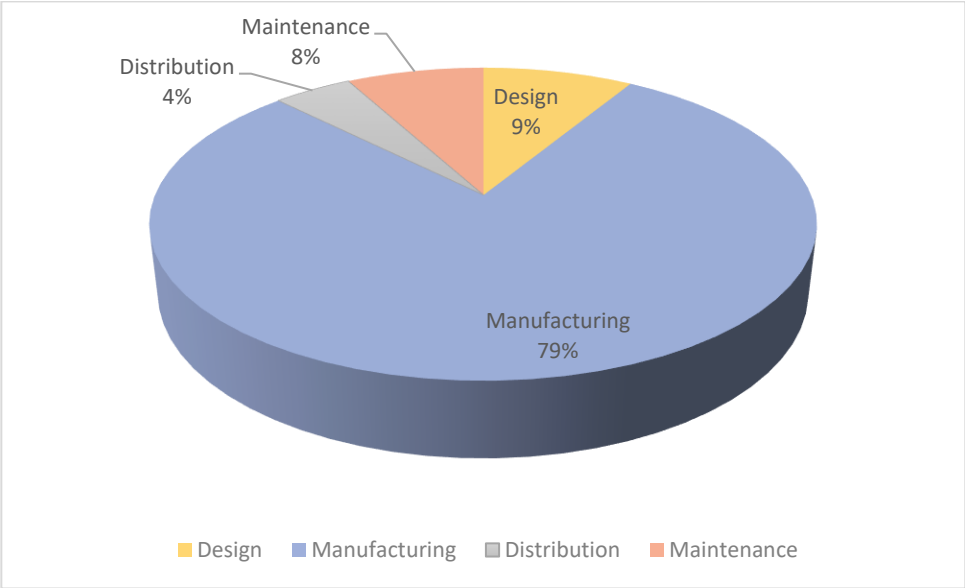

Figure 1- Overview of DRL Applications Addresses in Manufacturing

Furthermore, an intuitive realization and accurate understanding of the manufacturing application areas and feasibility of DRL at the four typical stages along the engineering lifecycle is a prerequisite. To achieve that, according to the specific task, the searched results are further classified into a concrete sub-subject under the engineering lifecycle in terms of the phases of design, manufacturing, distribution, and maintenance, as depicted in *Figure 2*. The corresponding sections and the amount of reviewed literature are indicated as well.

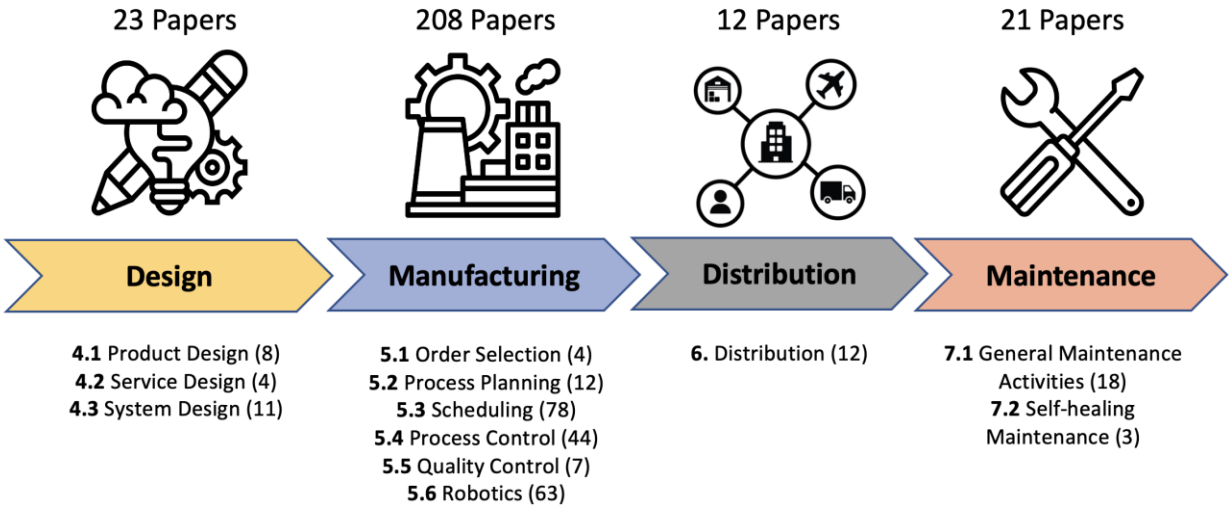

Figure 2 – Overview of DRL Application Distribution in Engineering Life Cycle

In addition to these categories, a simple statistical analysis of time is performed, and the results are displayed in *Figure 3*. It is intuitive to find that the RL utilization shows an exploding increase over the last decade years with an expected increase in potential DRL applications that have attracted more attention in recent years. Especially after 2018, by comparing the proportion of DRL papers, it can be found that DRL is attracting research interest in the manufacturing field and gradually generating broader applications.

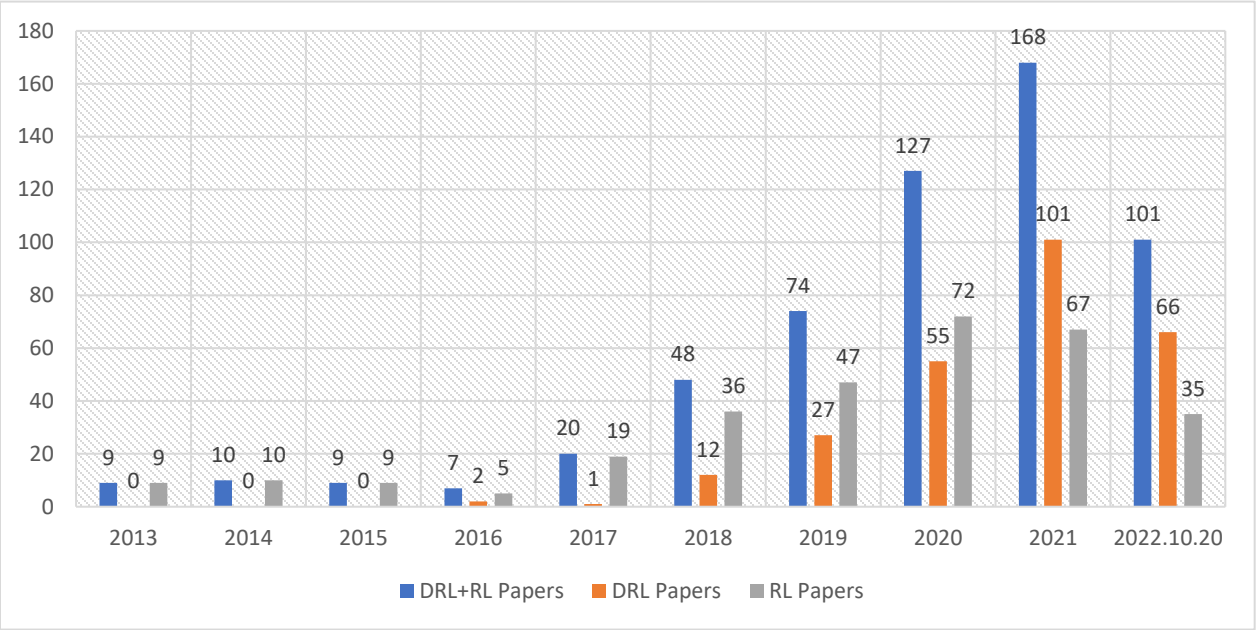

Figure 3 - Number of RL & DRL Publications (2013-2022).

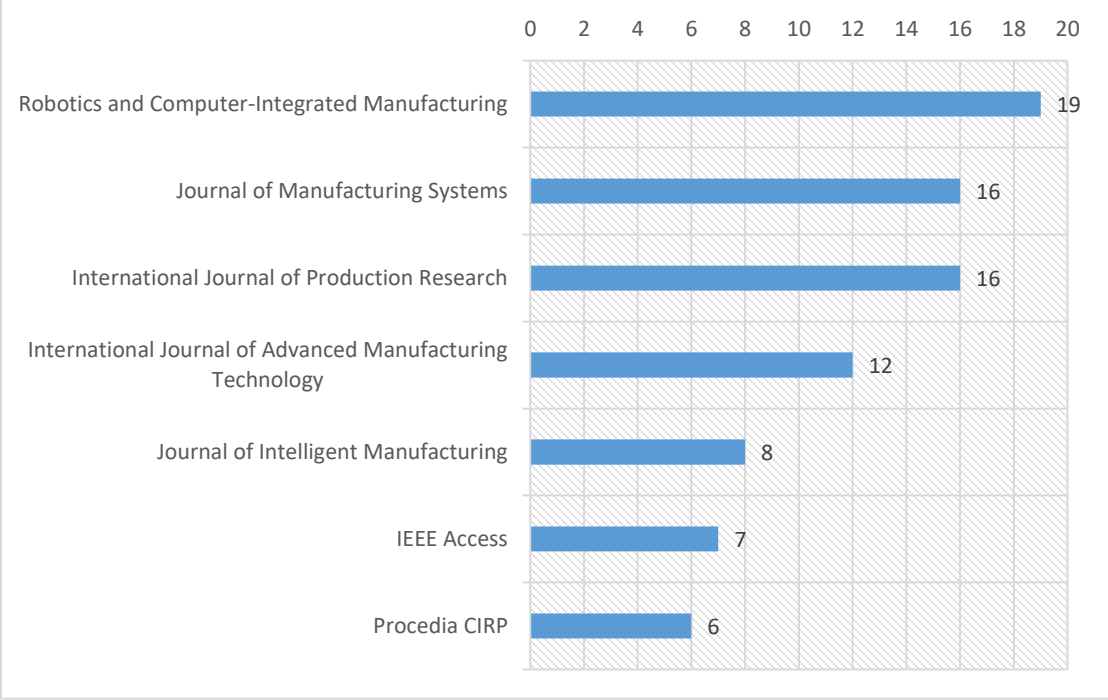

Figure 4 - Number of DRL in Manufacturing Publication (2013-2022)

Furthermore, the journals and conferences from the manufacturing domain that have published more than five papers about DRL from 2013 to 2022 are listed in *Figure 4*. One can see that Robotics and Computer-integrated Manufacturing, Journal of Manufacturing Systems, International Journal of Production Research, and International Journal of Advanced Manufacturing Technology rank in the Top 3 places and contribute the most publications, and the rest are peer-recognized in other high-quality manufacturing journals and conferences as well.

They not only reveal the recognition of DRL in the manufacturing field, but also its promising prospect of being widely used in various fields related to manufacturing systems.

3 Fundamentals of Deep Reinforcement Learning

In this section, the basics of RL and the general model-free DRL learning framework, where the value-based and policy-based methods are introduced sequentially. Moreover, the derivation of DRL algorithms and classical DRL algorithms are presented as well.

3.1 Reinforcement Learning

RL is an algorithm family used for optimizing the performance of sequential decision processes with Markov Property (i.e., Markov Process Decision) [22]. RL can be divided into Model-Based RL and Model-Free RL according to whether the model (i.e., state transition probability) of the system is known or not. In most practical manufacturing control areas, the model of the system is unknown and complex, and the control task is done under the unknown model. Therefore, the model-free RL is commonly adopted and is the main concern in this work.

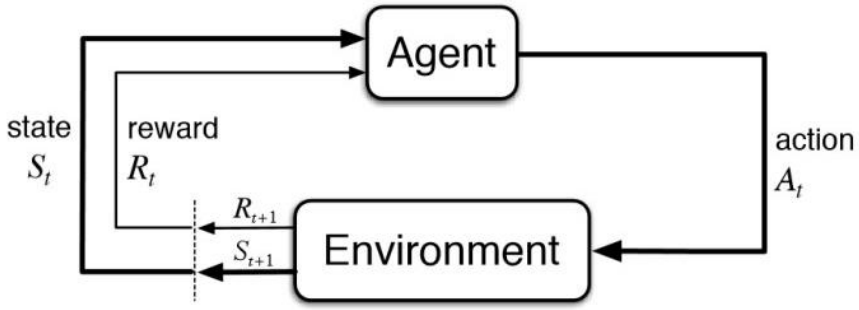

Figure 5 - Learning Framework of Reinforcement Learning [23]

As shown in *Figure 5*., an agent interacts with the environment, and the RL algorithm reinforces the agent to choose an action a according to the located state S_t of the environment to obtain the larger return reward. The reward $r_t(s_t, a_t)$ is based on the coming state S_{t+1} after the action a is executed in the next time step. Sometimes, the return reward is delayed instead of immediate in some long-horizon tasks, thus the immediate reward r_t could not claim the real performance of the agent. To determine that, the reward accumulates the value of all possible subsequent actions and resulting states multiplied by a discount factor γ, to obtain the cumulative reward R_t as:

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} \cdots = \sum_{t=0}^{\mathrm{T}-1} \gamma^t r_{t+1} \tag{1}$$

Where the discount factor γ is to express the uncertainty of reward and leads to the reward decay over time. Because of the randomness of the environment, the policy cannot be ensured the same future reward can be obtained with the same action. The more distant the future, the greater the uncertainty. Therefore, it is common practice to use decaying future rewards instead of certain future rewards. In a practice sense, it means the long-term reward leads a less influence on the current state. Thus, the goal of the RL algorithm is to acquire the optimal policy π^* to maximize the cumulative reward can be defined as:

$$\pi^* = arg\max_{\pi} E_{\pi}\left[\sum_{t=0}^{t_{max}} \gamma^t r_t\right] \tag{2}$$

In RL, the perception and representation of the environment state are one of the key problems that must be solved before an agent choose the corresponding action. Before DL, in RL-based decision tasks, the environment features were often extracted manually through a human expert usually based on prior knowledge of the task. Hence, such a pattern leads to RL algorithms whose performance was highly dependent on manual features and limited to solving low-dimensional state problems. With advances in DL [24], DNN can automatically extract compact high-dimensional representations (features) to overcome dimensional catastrophe (e.g., images, text, and audio) using the powerful representation learning properties. Therefore, DRL refers to a series of RL algorithms that utilize the representation learning capabilities of DNN to enhance decision-making capabilities, and the algorithm framework of DNN-based RL is shown in *Figure 6*. In DRL, DNN takes responsibility for extracting different environment information and inferring the optimal policy π^* in an end-to-end manner. Depending on the algorithm type, the DNN could drive the RL algorithm to output the Q-value (value-based) for each state-action pair or the probability distribution of the output action (policy-based). Lastly, a brief summary of the model-free DRL algorithm is provided in *Figure 7*.

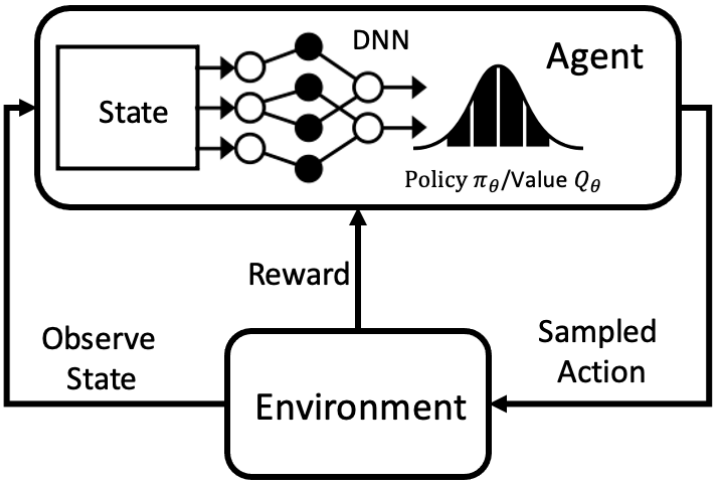

Figure 6 – Workflow Illustration of Deep Reinforcement Learning

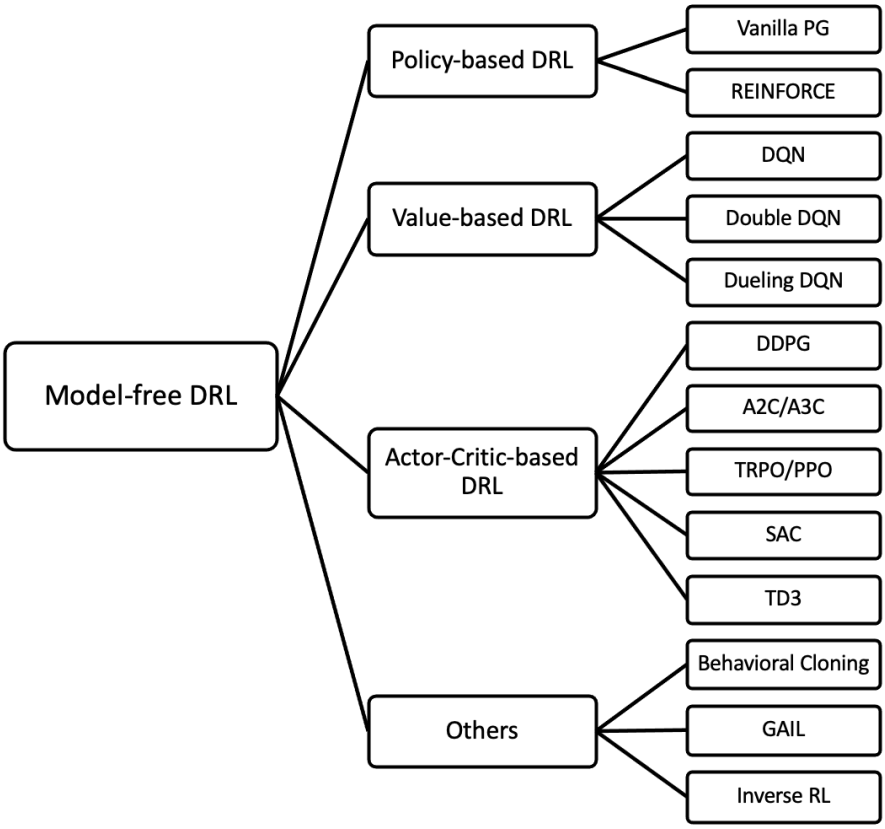

Figure 7- Summary of the Model-free DRL Algorithms

3.2 Value-based Deep Reinforcement Learning

For environments with discrete action spaces, traditional Q-Learning or SARSA learn to estimate the value of state-action pair by temporal-difference learning or Monte Carlo methods and store the values in a value table. When exploiting, the agent queries the table to obtain the value of each feasible state-action pair and selects the one with the highest value. As the environment becomes complex and the number of states becomes large, the value table is too large to store information. Hence, value-based DRL introduces DNN to extract features describing the states and to output actions to drive the agent to interactively sample the environment. Then, with the sampled data, the agent could update the weights of the DNN to accurately approximate and estimate the values of the corresponding state-action pairs. Here, the weight of DNN is updated by the gradient descent approach and temporal-difference learning. Therefore, with the help of DNN, the value of each state-action pair can be estimated instead of the table storing and querying, which greatly improves the efficiency of the algorithm. Among these algorithms, the classical value-based DRL algorithms are DQN and various variants such as Double DQN [25], Dueling DQN [26], etc.

3.3 Policy-based Deep Reinforcement Learning

As stated, the scope of Value-based DRL applications is mainly in the discrete action space. Since Value-based DRL is not applicable in the case of large spatial scale and continuous behavior. Even though the DQN is compatible with continuous action space (i.e., the continuous action space is discretized) but the action space becomes extremely large. Moreover, the division of each

continuous action space into a couple of discrete actions cannot achieve fine-tuning, and such division itself also brings a loss of performance and leads to poor learning ability. In addition to that, the optimal policy is a stochastic policy in part of applications, which requires selecting different actions and their corresponding probabilities. However, value-based DRL uses the greedy strategy when deploying. The requirement for performing various potential actions according to probability could not be met. For example, players could not always begin with the same strategy in chess.

In DRL, policy-based learning can be performed to solve continuous action space problems, i.e., the policy is viewed as a policy function with parameters for state-action pair as formulated in Equation 3, where DNN is employed to estimate the function:

$$\pi_\theta = P(a|s,\theta) \approx \pi(a|s) \tag{3}$$

The objective function as shown in Equation 4:

$$J(\pi) = E_\pi \left[\sum_{t=0}^{t_{max}} \gamma^t r_t | s, \pi \right] \tag{4}$$

To improve the accumulated reward, the DNN-based policy function π_θ is derived and searches a set of parameter vectors θ of to maximize the performance function $J(\pi)$. During the optimization process, the gradient ascent is adopted to update the parameters of DNN models. It takes the advantage of the reward generated by the interaction of the agent with the environment and can omit the process of learning the value of states. The actions can be generated directly for the continuous behavior space. The drawbacks of the policy-based DRL (e.g., Vanilla Policy Gradient (PG) [27], REINFORCE[28]) lie in the computational difficulty, and long iteration time in solving some complex problems, with high variance estimation of policy.

To avoid above mentioned drawbacks, the actor-critic mechanism is proposed [29]. Essentially, the Actor-Critic algorithm is a class of RL algorithms that combines state-action pair value function and policy-based learning algorithms, in which two DNNs are initialized simultaneously to approximate the value function and the policy function, respectively. The network for the policy function, called the Actor-network, generates behaviors to interact with the environment; the network for the value function, called the Critic network, evaluates the actor's performance and guides the actor's subsequent actions. This algorithm evaluates and optimizes the strategy based on the value function on the one hand, and the optimized strategy function, on the other hand, makes the value function more accurately reflect the value of the state, and the two networks promote each other to approach the optimal policy. To date, the actor-critic based algorithms are mostly used and the classical algorithms include Deep Deterministic Policy Gradient (DDPG) [30], Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic method (A3C) [31], Trust Region Policy Optimization (TRPO) [32], Proximal Policy Optimization (PPO) [33], Soft Actor-Critic (SAC) [34], Twin Delayed DDPG (TD3) [35], etc.

3.4 Other Algorithms

In the above-mentioned DRL algorithm, to obtain a better cumulative reward, the agent continuously interacts with the environment to improve interaction strategies. Such a trial and error-based learning approach can perform well in environments, where interaction costs are low.

However, in some sequential decision applications, the agent cannot obtain feedback frequently (sparse reward). Thus, the solution search space is large and hard to converge.

One possible solution to the above problem is Imitation Learning (IL). In IL, the agent learns from behavioral data provided by the demonstrator rather than trying to learn from sparse rewards or manually specified reward functions from the environment. Where the data is typically well-performed decision data provided by human experts, each data point contains a trajectory of states and actions. Usually, the DNN is employed as the brain of an agent. The DNN model attempts to learn the optimal policy by imitating the expert's behavioral patterns and the training goal of the model is to match the distribution of state-action trajectories generated by the model with the distribution of the input trajectories. The classical IL algorithms are Behavioral Cloning [36], Generative Adversarial Imitation Learning (GAIL) [37], etc.

One step further IL, the algorithm enables the agent to obtain an efficient and reliable performance function rather than only imitating the behavior pattern of provided demonstration, which is called inverse reinforcement learning (IRL) [38]. With the gained behavior pattern and reward function, the agent not only can jointly adopt IRL and RL together to improve the precision of the reward function and the learning process of the policy but also could generalize them to other cases to improve the efficiency.

4 Deep Reinforcement Learning in Design Stage

In this section, the existing applications of DRL in the design phase are firstly summarized, which are further divided into *product design*, *service design*, and *system design*, as detailed below.

4.1 Product Design

In the product design phase, DRL has been mainly leveraged to 1) reduce the reliance on expert knowledge and 2) improve the design efficiency and generalization, as summarized in *Table 1*.

To reduce the dependence on expert knowledge in the design process, Park et al. [39] proposed a DRL-based decoupling capacitors design method for silicon interposer-based 2.5-D/3-D integrated circuits, which doesn't require complex analytical models specialized in the field of power integrity. Yang et al. [40] developed a DRL-driven automated fixture design approach by learning through interaction with the working environment instead of case-based or rule-based reasoning. Son et al. [41] used PPO with Long Short-term Memory (LSTM) layers to explore an optimal 3-D cross-point array structure for component packing by only considering the signal integrity index. Similarly, for 3D printing products, Yang et al. [42] integrated DRL in finite element-based material simulations to search and design 3-D printed periodic microstructures, which enhance the microstructural architecture and the mechanical or thermal performance of engineering components.

In terms of design efficiency and commonality/generalization, the strong exploration ability of DRL enables simplifying and accelerating the product optimization process. Zimmerling et al. [43] proposed a DRL-based model to infer the estimated optimal process parameters for variable component geometries by extracting reusable information and deriving it into new, non-generic components. Moreover, due to the increased product diversity with a shorter lead-time in clothing

manufacture [44], DRL is adopted to improve the existing clothes design and shorten design circles based on users' preferences and feedback.

Table 1– Literature of DRL in Product Design

Product Design in Design Stage				
DRL Objectives	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Reduce the demand of manual involvement and expert knowledge • Improve design efficiency and generalization • Automated and accelerate the design process	Silicon interposer-based 2.5-D/3-D IC component decoupling capacitors design	Double DQN	2020	[39]
	Automated machining/measurement fixture design generation	A2C	2020	[40]
	X-Point circuit array structure design considering signal integrity	PPO+LSTM	2022	[41]
	3-D printed periodic microstructures design	A2C	2021	[42]
	Optimum part design in tailoring manufacturing	DRL	2022	[43]
	Garment designs recommendation	DRL	2020	[44]
	Carbon fiber reinforced plastic design	PPO	2022	[45]
	Carbon fiber reinforced material component design for electromobility	DDPG	2022	[46]

4.2 Service Design

In the service design phase, DRL was mainly adopted to 1) simplify the service design process and 2) optimize the service quality and the applications are listed in *Table 2*.

Zhang et al. [47], Liang et al. [48] and Liu et al. [49] contributed to addressing the problem of overcomplex service design combination in cloud manufacturing using DQN and DDPG. The addition of DQN allows the Quality-of-Service index to be taken into account, thus accelerating the service solutions exploration process to find the optimal service composition and satisfy consumers' requirements. Moreover, to optimize the service quality, Moghaddam et al. [50] combined multi-agent reinforcement learning（MARL) with matching markets theories to optimize the design interaction processes and governance service policies for smart manufacturing marketplaces on the micro-service architecture.

Table 2 – Literature of DRL in Service Design

Service Design in Design Stage				
DRL Objectives	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Simplify design flow and tuning process • Optimize service design performance index	Cloud manufacturing service composition solution design	DQN	2020	[47]
		DQN	2021	[48]
		DDPG	2022	[49]
	Design and governance of SM marketplaces strategy identification	DQN/MARL	2019	[50]

4.3 System Design

In the system design phase, most DRL applications are aimed at optimizing design system performance, including response efficiency, system generalization, and adaptability, as summarized in *Table 3*.

In terms of system response efficiency, Moon et al. [51] applied the DQN integrated with transfer learning to the multi-access edge computing structure so that cooperative scheduling can be performed independently among edge devices with low delay time and high security. In order to enhance the resilience of the manufacturing system, Bauer et al. [52] proposed a DRL-based system control loop integrated with IoT devices and CPS, which could respond to events from the supply network effectively and efficiently.

To improve the system design efficiency, flexibility, and generalization among varied environments, Zou et al. [53] proposed a DRL-enhanced simulation allocation learning framework to improve the policy learning efficiency in a CPS. She et al. [54], Xia et al. [55], and Ren et al. [56] used digital transformation techniques to build DT-based manufacturing systems, which could realize the information synchronization between physical production units and the virtual manufacturing system. Moreover, by integrating DRL in the DT system, a few key issues of manufacturing systems can be addressed significantly including design assistance, automated production control, maintenance, scheduling, and life-cycle management. Lastly, Blasi et al. [57] not only designed a DT system integrated with DRL for designing but also deployed the system in physical space with the addition of domain randomization.

Table 3 – Literature of DRL in System Design

System Design in Design Stage				
DRL Objectives	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Enhance system functions and abilities • Improve system response efficiency and the system resilience • Improve system design performance (efficiency, adaptability, generalization)	Edge computing-enabled collaborative manufacturing design	DQN	2021	[51]
	Delivery reliability control in the supply network	DRL	2021	[52]
	Industrial distributed control simulation system design	DQN	2020	[53]
	DT-based industrial automation control system design	DRL	2021	[54]
	DT-based architecture design for full life cycle management	DQN	2021	[55]
	DT-enabled production life cycle management system design	DRL	2021	[56]
	Simulation-based DRL physical manufacturing automation system design	PPO	2021	[57]
	3D printing-based repair system design for structural and electrical restoration	DQN	2022	[58]
	Virtual prototyping system design for cyber-physical products	DQN	2022	[59]

Control framework design for the optimization of manufacturing system	MARL	2022	[60]
Multi-stage production systems control method design for machines' cycle time	A2C	2022	[61]

5 Deep Reinforcement Learning in Manufacturing Stage

In this section, DRL applications in the manufacturing stage of the engineering life cycle are comprehensively reviewed, including *order selection, process planning*, *scheduling, process control,* and *quality control.* Considering that DRL-based robot control has recently become a prevailing topic in academia and industry, we add the subcategory *robotics* applications. The subcategories are distributed following the order of publication number, in which scheduling, robotics, and process control are dominant, as shown in *Figure 8*.

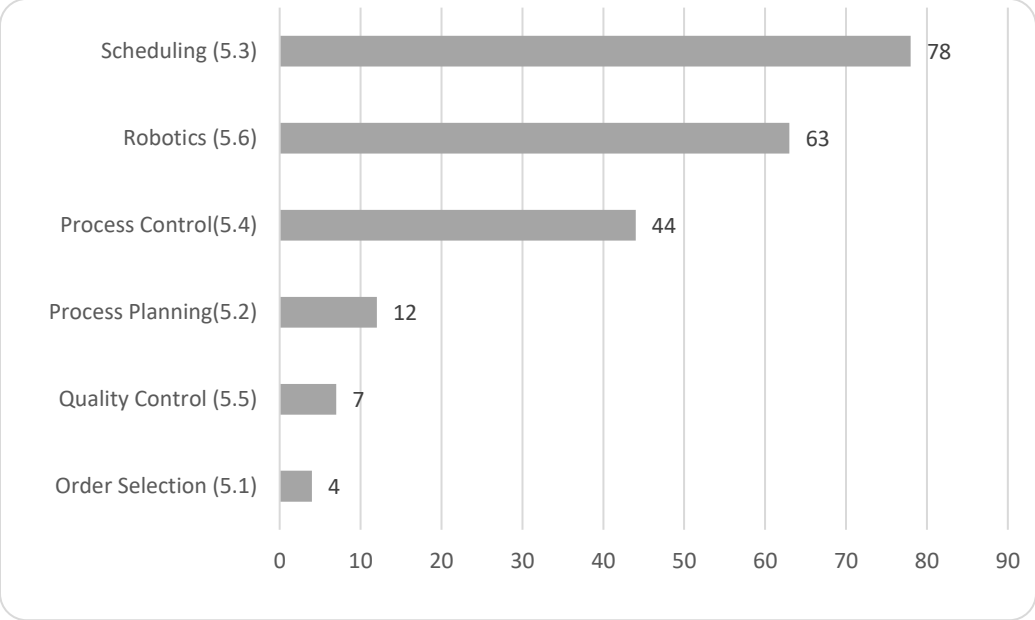

Figure 8 - Number of Publications in the Manufacturing Stage

5.1 Order Selection

Reacting flexibility to customers' requirements is significant in mass personalization, in which order selection is critical. *Table 4* lists the relevant applications of order selection enabled by DRL. Pahwa et al. [62] and Zhang et al. [63] formulated the order selection process as an MDP sequential decision-making problem, in which DRL is used to improve the order selection policy to avoid loss due to short-sightedness. Dittrich et al. [64] proposed a cooperative multi-agent order decision-making framework, which drives agents to collect the ordering experience in a decentralized manner and feeds data to a DQN-based central control unit for order selection optimization. Leng et al. [65] proposed a DRL-based order acceptance model in the print circuit board production system to optimize order selection, comprehensively considering the production cost, completion time, and carbon consumption.

Table 4 - Literature of DRL in Order Selection

Order Selection in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize order selection performance index • Multi-objective optimization • Enhance the adaptability of policy	Suppliers service marketplace orders acceptation	DRL	2021	[62]
	Modular production enterprises order acceptation	DQN	2020	[63]
	Multi-agent system-based cooperative order acceptation	DQN	2021	[64]
	PCB industry energy-efficient sustainable order acceptation	DRL	2021	[65]

5.2 Process Planning

According to order demand, an optimal manufacturing process plan is essential to improve the quality, production efficiency, and cost-effectiveness of machined parts. Thus, DRL is broadly adopted in process planning to improve planning efficiency while increasing flexibility for various scenes, as shown in *Table 5*.

To improve sequential process planning efficiency, Wu et al. [66] and Mueller-Zhang et al. [67] adopted DRL in Computer-Aided Process Planning (CAPP), which improves the planning efficiency by leveraging the advantages of DRL in terms of experience reusability and generalization. Meanwhile, Sugisawa et al. [68] proposed an IRL-based method to enhance the CAPP via extracting implicit decision rules from human experts. In the textile industry, He et al. [69] transformed the textile chemical manufacturing process into MDPs and optimized the process control policy via DQN. In part production, Ghorbel et al. [70] and Wu et al. [71] used the DRL-based approach to improve the planning of manufacturing and assembly processes, respectively, under the constraints of dynamic resources. He et al. [72] further introduced a DQN combined with a MARL framework to optimize the combination of multiple objectives, including quality, productivity, and cost.

Moreover, the applications of DRL are not restricted to sequential process planning but are also used in spatial planning. Klar et al. [73] adopted Double DQN to generate the factory layout while considering the functional units' distribution and transportation time. Kim et al. [74] and Woo et al. [75] both applied A3C-based spatial management algorithms in the ship-building process. The proposed algorithms can minimize the load unbalancing and rearrangement cost caused by workshop physical constraints and unpredictable time and space circumstances.

Table 5 - Literature of DRL in Process Planning

Process Planning in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Improve the sequential process planning performance index • Reduce manual involvement and automate the decision process • Reduce the reliance on expert knowledge	CAPP in mass-customized production	Actor-Critic	2021	[66]
		DQN	2020	[67]
	Expert knowledge-enhanced CAPP	Inverse RL	2021	[68]
	Textile chemical process planning decision support	DQN	2021	[69]
	Multi-objective process planning optimization in textile manufacturing	DQN+MARL	2021	[72]
	Heterarchical agents' network for parts production planning	PPO	2021	[70]
	Agile part assembly planning to dynamic resources changes	Actor-Critic	2021	[71]
	Factory scene layout planning	Double DQN	2021	[73]
	Ship block arrangement of stockyards	A3C	2021	[74]
	Multi-Ship building load balancing	A3C	2020	[75]
	Mixed model assembly lines balancing and sequencing	Actor-Critic	2022	[76]
	Part assembly planning optimization for time and sample efficiency	Actor-Critic/ DQN/ Rainbow	2022	[77]

5.3 Scheduling

The scheduling process in a complex manufacturing system is often faced with great uncertainties, such as flexible on-demand product orders or unexpected operational unit failures. Compared with traditional rule-based or other heuristic methods, the DRL-based solutions for scheduling problems bring strong effectiveness, flexibility, and generalization abilities, and effectively reduce human involvement/workload. The existing works can be mainly categorized into cloud manufacturing and reconfigurable manufacturing paradigm, as shown in *Table 6* and *Table 7*. Then, *Table 8* and *Table 9* wrap up the scheduling work on the most discussed scenario and industry: job shops and semi-conductor industries. Lastly, the varied works on different manufacturing systems are briefly listed in *Table 10*.

Cloud manufacturing - In cloud-based manufacturing service scheduling, DRL can help simplify the system modeling process, reduce computational costs, improve system performance, and balance multi-objective optimization problems. Dong et al. [78] and Liu et al. [79] utilized DRL-based scheduling algorithms to minimize task execution time, for tasks with precedence relationship to cloud servers and online single-task, respectively. Zhu et al. [80] transformed the multi-resource cloud manufacturing scheduling problems into optimization targets and employed a DRL-based approach for solving them.

* *Corresponding author*: pai.zheng@polyu.edu.hk

Table 6 - Literature of DRL in Cloud Manufacturing Scheduling

Cloud Manufacturing Scheduling in Manufacturing Stage				
DRL Objectives	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Simplify system modeling process • Reduce system computational cost and improve performance index • Multi-objective optimization	Task scheduling with precedence relationship to cloud servers	DQN	2020	[78]
	Multiple resource scheduling problems	DRL	2019	[79]
	Online single-task scheduling	DRL	2020	[80]
	Decentralized robot manufacturing services scheduling	DQN	2022	[81]
	Services distributed real-time scheduling towards dynamic and customized orders	Dueling DQN	2022	[82]
	Cloud manufacturing group services dynamic scheduling	DQN	2022	[83]

Reconfigurable Manufacturing - In Reconfigurable Manufacturing Systems (RMSs), DRL could enhance the system adaptability and improve the system working efficiency to satisfy the growing complex and diverse production requirements. To improve the adaptability of RMS, Tang et al. [84], Yang et al. [85], and Hofmann et al. [86] used Double DQN, DQN, and A2C, respectively, in reconfigurable task scheduling to manage the completion status of the assigned orders while minimizing the reconfiguration action cost. Moreover, the modular manufacturing paradigm also provides strong support for establishing RMS; thus, Schwung et al. [87] [88] adopted actor-critic and MADDPG algorithms to develop control approaches with self-learning and plug-and-play features for modular manufacturing units. In RMSs, Automated Guided Vehicles (AGVs) provide transportation services to link the modular workstations and coordinate the production service. Mayer et al. [89] used PPO to control AGVs' operation to achieve the orchestration between products, workstations, and vehicles, further realizing production maximization. Gankin et al. [90] and Li et al. [91] also used the MARL method to implement the scheduling of AGVs to link the modular workstations.

Table 7 -- Literature of DRL in Reconfigurable Manufacturing Systems

Reconfigurable Manufacturing System in Manufacturing Stage				
DRL Objectives	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Optimize the adaptability toward the dynamic and complex environment • Reduce system computational cost and improve system performance index	Conduct production plan while minimizing the reconfiguration	Double DQN	2021	[84]
		A2C	2021	[85]
	Modular-based manufacturing unit scheduling control	DQN	2020	[86]
		Actor-Critic	2018	[87]
		MADDPG	2019	[88]
	AGV-based modular manufacturing task scheduling	DQN	2021	[90]
		PPO	2021	[89]

	AGV-based material handling in flexible shop floors	MADDPG	2021	[91]

Job shops - The facing production tasks in job shops are often characterized by random order, smaller batches, and mass variety. Hence, it is challenging to manage manufacturing resources and improve production efficiency. The selected papers show that DRL-based scheduling methods in job shops can significantly reduce production costs and improve processing efficiency, which is divided into three categories: *order dispatching, job-shop control,* and *rescheduling*.

For order dispatching, reducing the reliance on domain expert knowledge can increase the generality of the algorithm and simplify the system design process, thus improving scheduling efficiency. A. Kuhnle et al. [92] [93] and Rummukainen et al. [94] designed and implemented an effective and adaptive order dispatching system by introducing the DRL-based approach with limited domain knowledge. Moreover, Kim et al. [95] used MARL architecture with GNN to enhance the autonomy of decision-making and interactivity. Also to improve the order dispatching efficiency, A.Gannouni et al. [96] explored the applicability and scalability of the neural combinatorial optimization methods for minimizing the manufacturing resource costs. Han et al. [97] proposed the CNN-based DRL-enabled framework based on the disjunctive graph to find the dispatching solutions with higher productivity and lower delay. Furthermore, J. C. S. Ruiz et al. [98] and Hu et al. [99] both leveraged the DT technologies to extract system states and support the DRL algorithm for solving the dynamic scheduling problems involving order arrivals, shared resources, and route flexibility.

In job shop control, DRL is mainly used to improve job-shop control efficiency and solution searching under resource constraints. In terms of optimizing job-shop control performance, researchers (Lin et al. [100], Zhou et al. [101], Kim et al. [102], Lang et al. [103], Zhao et al. [104], Samsonov et al. [105] and Thomas et al. [106]) all formulated the control problems into MDPs and adopted value-based DQN-series algorithms for optimizing different single or composite optimization goals such as time efficiency, throughput, and universality. In addition to the value-based DRL algorithm, Zeng et al. [107] proposed an evolutionary job scheduling algorithm initialized by DRL. Zhao et al. [108] introduced an actor-critic algorithm to optimize the processing time and makespan for job-shop production control by collaborating with multiple sub-manufacturing systems. Due to their effective self-exploration abilities, DRL-based approaches are suitable for addressing the more challenging resource-constrained job-shop scheduling problems. For instance, Luo et al. [109] combined a multimodal hybrid neural network, and the action masked pruning method with the PPO algorithm to learn dynamic shop floor scheduling. Similarly, with traffic constraints, Kang et al. [110] proposed a DQN-based dynamic routing strategy to shorten delivery time delay. And Li et al. [111] used DQN to address the job shop scheduling problem, which can minimize the makespan and total energy consumption under insufficient transportation resources.

With the sharp increase in the uncertainty and complexity of the manufacturing process, rescheduling approaches for dealing with unexpected events have become a key issue of real-time disruption management strategy. Facing dynamic conditions, DRL is employed to improve adaptability and robustness without sacrificing cost-effectiveness, product quality, and delivery efficiency. Similar to conventional job-shop scheduling control, the majority of the applications (Palombarini et al. [112], Luo [113], Shi et al. [114], Yang et al. [115], Seito et al. [116], Zhou et al. [117], Liu et al. [118]) leveraged DQN-series value-based algorithms with relevant DRL tricks

to improve the adaptability of job shops or production lines. Except for value-based algorithms, the policy-based PPO algorithm is also adopted by Wang et al. [119] to find the optimal re-scheduling policy for the problem complexity increment. Park et al. [120] proposed a DT control model combining with actor-critic algorithms to generate the re-entrancy and dispatching rules caused by stochastic arrivals. While facing the challenges of new job insertions and machine breakdowns, Luo et al. [121] proposed to use MARL architecture with PPO to address such dynamic partial-no-wait multi-objective scheduling problems. Similarly, Zhou et al. [122] built a distributed manufacturing architecture integrating the PPO method, which can solve the problem of inefficiency and unreliability caused by over-dependence on central controllers and limited communication channels during low-volume-high-mix orders online scheduling. Palombarini et al. [123] [124] designed a control policy to learn schedule policies via PPO using a color-rich Gantt chart and negligible prior knowledge directly from high-dimensional sensory inputs.

Table 8 - – Literature of DRL in Job Shop Scheduling

Job Shop Scheduling in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenario**	**Algorithm**	**Year**	**Reference**
• Reduce reliance on domain knowledge and manual work • Improve scheduling performance index (time, productivity, etc.) • Improve solution searching efficiency under resource constraints • Enhance rescheduling solution adaptability	Order Dispatching	Job shop order dispatching towards small batch sizes, large product variety, and varied material flow tasks in a changeable environment	DRL	2019	[92]
			DRL	2021	[93]
			PPO	2019	[94]
			DRL	2020	[96]
			Dueling Double DQN	2020	[97]
			MARL+DQN	2020	[95]
			DRL	2022	[125]
		DT-based job shop order dispatching model construction	DRL	2021	[98]
			DQN+GNN	2020	[99]
	Job Shop Flow Scheduling	Dynamic complex job shop scheduling and control towards time efficiency, cost reduction, resource constraints, and processing efficiency	DQN	2019	[100]
			DRL	2020	[101]
			Dueling Double DQN	2021	[102]
			DQN	2020	[103]
			DQN	2021	[104]
			DQN	2021	[105]
			DRL	2018	[106]
			A2C	2021	[107]
			Actor-Critic	2021	[108]
			PPO	2021	[109]
			DQN+RNN	2019	[110]
			MADDPG	2022	[126]
			DQN/PPO	2022	[127]
			DQN	2022	[128]
			Double DQN	2022	[129]

	Limited AGV-transportation resource-based job shop scheduling	DQN	2022	[111]
Process Rescheduling	Job-shop dynamic rescheduling control policy generation towards unforeseen events (e.g., machine breakdown, job rework) or tasks (e.g., on-demand orders)	DQN	2019	[112]
		Double DQN	2020	[113]
		DQN	2020	[114]
		Double DQN	2021	[115]
		DRL	2020	[116]
		DQN	2021	[117]
		Double DQN	2021	[118]
		PPO	2021	[119]
		Actor-Critic	2021	[120]
		PPO+MARL +Hierarchical RL	2021	[121]
		DQN+MARL	2021	[122]
		PPO	2022	[123]
		DQN	2019	[124]
		MARL+Double DQN	2022	[130]
		PPO	2022	[131]
		Double DQN	2022	[132]
		REINFORCE	2022	[133]
		DQN	2022	[134]
		MADDPG	2022	[135]
		DQN	2022	[136]
		PPO	2022	[137]

From the industry application perspective, DRL applications in the semi-conductor industry are mostly discussed. The DRL is adopted to improve the wafer manufacturing process, including *control productivity, system stability, and decision-time efficiency*, which can be found in *Table 9*.

To improve productivity in terms of task management and makespan, Scholars (Waschneck et al. [138] [139], Chien et al. [140], Sakr et al. [141], Lee et al. [142]) adopted DQN-series value-based DRL algorithms as optimizer. For those abstracted front-end-of-line production facilities and practical manufacturers, the DRL-based methods could gain better production performance than the heuristic and rule-based approaches. Liu et al. [143] combined the A3C and composite dispatching rules involving scheduling knowledge to maximize productivity and average daily movement and minimize mean cycle time. For process equipment, Lee et al. [144] designed a scheduling method based on DQN that controls the wafer transport robot inside the production equipment to improve equipment utilization. Hong et al. [145] proposed a condition-based cleaning approach aiming to maximize productivity while maintaining wafers quality by adopting MARL algorithms and further optimizing the scheduling cluster tools. To improve production efficiency and system stability, Wang et al. [146] and Wang et al. [147] adopted fuzzy

hierarchical RL and LSTM-based DRL algorithms separately to control the cycle time by adjusting the priority of each wafer lot. Kuhnle et al. [148] proposed an explainable-DRL control policy to increase the plausibility of control methods in the semi-conductor production system. For process equipment scheduling, Park et al. [149] presented a scheduling method for semi-conductor packaging facilities using DDPG in a centralized manner.

Table 9 - Literature of DRL in Semi-Conductor Industry Scheduling

Semi-Conductor Industry Scheduling in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Enhance system control productivity, adaptability, and stability • Improve solution searching efficiency under resource constraints	Dynamic semi-conductor manufacturing task scheduling towards short lead-time, higher throughout, higher productivity	DQN	2018	[138]
		DQN	2018	[139]
		DQN+GA	2020	[140]
		DQN	2021	[141]
		DQN	2022	[142]
		Hierarchical DRL	2021	[146]
		DRL	2018	[147]
		A3C	2020	[143]
		DRL	2021	[148]
		Actor-Critic	2022	[150]
	Wafer transport robot control for equipment utilization	DQN	2021	[144]
	Semi-conductor packaging facilities scheduling	DDPG	2022	[149]
	Equipment maintenance scheduling to improve productivity	A3C+MARL	2019	[145]

Except for the above applications, other DRL-enabled applications also have been applied in different manufacturing scenarios to improve productivity, as shown in *Table 10*. For instance, in the injection mold industry, Lee et al. [151] formulated the mold scheduling problem as an MDP framework, and the DQN algorithm is employed to find the optimal scheduling policy. The policy could minimize the total weighted tardiness caused by mold product diversity. In the automotive industry, Leng et al. [152] used the Color-Histogram model to deal with the color-batching resequencing problem, combining DQN to minimize color transition cost. Gros et al. [153] explored the effectiveness of DRL in real-time decision-making for unforeseeable events in the automotive manufacturing process. Overbeck et al. [154] intergraded DRL into DT to learn the production control logic, then used it in task assignments between workers on a production line.

Table 10 - Literature of DRL in Other Industry Applications of Scheduling

Other Industry Applications of Scheduling in Manufacturing Stage				
DRL Objectives	**Application**	**Algorithm**	**Year**	**Reference**
• Optimize performance index (time delay, cost, productivity, etc.)	Injection mold manufacturing scheduling	DQN	2020	[151]
	Order resequencing and regroup for automotive car painting	DQN	2020	[152]
	Real-time decision-making for the car manufacturing process	DQN	2020	[153]
	DT-based control logic for tasks distribution between the different workers	DRL	2021	[154]
	Resequencing scheduling of automotive manufacturing for both paint shop and assembly shop	DQN	2022	[155]

5.4 Process Control

DRL-based applications in the manufacturing process control stage fall into two main categories according to different performance targets. One is sustainable manufacturing-oriented process control, which aims for *energy saving or emissions reduction* as summarized in *Table 11*. The other is manufacturing entity control towards equipment-level, such as *machining, welding, additive manufacturing*, and *other industrial processing equipment control*, as listed in *Table 12*.

DRL is a promising optimization approach for manufacturing process control that can support sustainable manufacturing systems by providing various energy-saving strategies and services for industrial production facilities and machines. For instance, Kohne et al. [156], Huang et al. [157] Schwung et al. [158] [159], and Zhu et al. [160] all proposed actor-critic methods to determine the optimal energy reduction control scheduling policies while meeting certain manufacturing constraints like demanded throughput. Also, Schwung et al. [161] presented a distributed optimization in production systems that adopted DRL and particularly emphasized energy-efficient production, in which the Teacher-Student distillation method is used to initialize the Programmable Logic Controller. Meanwhile, Bakakeu et al. [162], Roesch et al. [163], and Lu et al. [164] all adopted MARL to establish a demand response scheme for energy management in manufacturing systems, which can learn the optimal time or collaboration scheduler for different agents to ensure energy efficiency and load management.

Regarding emission-reduction applications, DRL mainly takes responsibility for improving system performance and controlling emissions. To reduce the emission, Zhao et al. [165] proposed an actor-critic scheduling method for the blast furnace gas system in the by-product gas system of the steel industry. For industrial energy supply systems, Weigold et al. [166] proposed a model-based optimization approach in conjunction with PPO to reduce the CO_2 emissions and reduce electricity costs by transferring electricity demand to times of lower electricity prices. Fu et al. [167] adopted LSTM neural network layers to build a denitrification efficiency prediction model for coal-fired power plants and used DRL to control selective catalytic reduction denitrification efficiency to decrease the emission of greenhouse gas.

* *Corresponding author*: pai.zheng@polyu.edu.hk

Table 11 - Literature of DRL in Sustainable-Oriented Process Control

Sustainable-Oriented Process Control in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize process performance index (energy cost, resource utilization, etc.) • Optimize energy control policy adaptability	Energy Consumption Management	Energy-saving control policy design towards consumption schedule and cost reduction	PPO	2020	[156]
			Actor-Critic	2019	[157]
			Actor-Critic	2018	[158]
			Actor-Critic	2019	[159]
			DDPG	2021	[161]
			Double DQN	2022	[168]
			DRL	2022	[169]
			SAC	2022	[160]
		Heterogeneous cluster of flexible manufacturing machines control	MARL+AC	2020	[162]
		Industrial load management for energy-oriented rescheduling	PPO	2019	[163]
		Demand response for energy management of discrete manufacturing systems	MADDPG	2020	[164]
	Emission Control	Blast furnace gas emission control	Actor-Critic	2021	[165]
		Cooling towers emission control	PPO	2021	[166]
		Denitrification efficiency control	A3C	2021	[167]
		Grid multi-energy system control	DDPG	2022	[170]

DRL and other artificial intelligence technologies are increasingly involved in smart manufacturing process control, significantly enhancing adaptability, reducing reliance on expert knowledge, and improving machine performance index. To enhance the adaptability of the machining control policy, Xiao et al. [171] integrated meta-RL to adaptively determine the optimal machining parameters for various machines, workpieces, and tools. Huang et al. [172] presented an integrated modeling method based on GNN and MARL, which can collaboratively control the adjustment of individual machining process parameters such as spindle speed and cutting depth. Meanwhile, Dornheim et al. [173] investigated a DRL approach to generate the optimal structure formation by considering the processing paths for the target material structure space, thus reducing the reliance of the optimization process on expert knowledge. Li et al. [174] proposed a DRL-based trajectory smoothing method, which outputs servo commands directly based on the current tool path and running state in every cycle. Furthermore, Zhang et al. [175] and Samsonov et al. [176] optimized the tool orientation planning and found an optimal clamping position for a workpiece with the help of DRL. In terms of improving machine performance via DRL, Schoop et al. [177] used PPO to design control policies to maximize the cutting tools' life while ensuring machining quality. Gulde et al. [178] applied DRL for vibration compensation to the machine tool axis to obtain higher machining precision and a longer component lifetime. Jiang et al. [179] modeled the internal CNC data consisting of feeding axis tracking error as an LSTM network; then,

an LSTM-based DQN control strategy was proposed to reduce contour prediction error and enhance the machining compensation.

Table 12– Literature of DRL in Process Control Applications

Process Control Applications in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize process performance index (time efficiency, quality, etc.) • Enhance control policy adaptability • Reduce reliance on expert knowledge and manual work	Machining	Machining process parameter tuning for energy-saving machining	Actor-Critic	2021	[171]
		Adaptive control local process machining parameters	MARL+A2C+ GNN	2021	[172]
		Optimization of machining processing paths	DQN	2021	[173]
			DRL	2021	[174]
		Machining tool orientation planning	SAC	2022	[175]
		Workpiece optimal clamping position searching	SAC	2021	[176]
		Machining process parameters tuning towards tool-wear	PPO	2021	[177]
		Machine tool control vibration compensation	PPO	2019	[178]
		CNC machine tool contour error compensation	DQN+LSTM	2021	[179]
		Motion control of laser machining	PPO	2022	[180]
		Milling parameters optimization for surface roughness and material removal rate	Double DQN	2022	[181]
	Welding	Weld pool width control	DQN	2019	[182]
		Real-time govern welding power	Actor-Critic	2016	[183]
		Optimize the stencil printing parameters of PCBs	DQN	2021	[184]
	Additive Manufacturing	Generating toolpaths in 3D printing	Actor-Critic	2020	[185]
		Versatile control strategy for powder-based additive manufacturing	PPO	2021	[186]

In welding, the actual welding quality depends on various factors, such as welding current, arc voltage, and welding speed. Thus, DRL implementations in welding mainly focus on how to optimize the control parameters to ensure good welding performance. Jin et al. [182] leveraged the actor-critic algorithm to control the width of the weld pool, thus improving the quality of gas tungsten arc welding and gas metal arc welding. Günther et al. [183] adopted DL and DRL to learn context-appropriate control policies, and then govern welding power in real-time to address the control difficulties of laser welding. Similarly, Khader et al. [184] also adopted DQN to optimize

the stencil printing parameters in real-time, which can better control the solder paste volume TE and increase the first-pass yields of printed circuit boards within the spec limits.

In additive manufacturing, DRL has also been employed to tune control parameters to improve system performance. Patrick et al. [185] used the actor-critic algorithm to optimize the printing path control function so as to improve the printing quality and reduce printing time. Ogoke et al. [186] proposed a DRL-based control framework for minimizing the potential defects in powder-based 3D printing.

As shown in *Table 13*, except for the above broad categories of manufacturing process control applications, DRL is also adopted in the following areas: polishing process control [187], PLC fault recovery [188], fixture locators exploration [189], metal fabrication industry [190] [191], furnace heating process of tempered glass [192], injection molding industry [193], textile industry [194], wire manufacturing [195], fiber production [196], plugging operations [197], and cold foaming [198].

Table 13 – Literature of DRL in Other Industry Applications of Process Control

Other Industrial Applications of Process Control in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize the application performance index • Enhance the control adaptability • Reduce expert knowledge reliance and manual work	Chemical mechanical polishing	DDPG	2020	[187]
	PLC fault recovery control	PPO/DQN	2021	[188]
	Fixture locators tuning of mental assembly	PPO	2019	[189]
	Metal sheet deep drawing	DQN	2020	[190]
	Industrial scale forging process	Double DQN	2021	[191]
	Heating process of tempered glass in the industrial electric furnace	DQN	2021	[192]
	Injection molding process parameters settings	DQN	2019	[193]
	Textile forming material draw-in optimization	Actor-Critic	2020	[194]
	Roll gab control and wire hot rolling process	DDPG	2021	[195]
	Fiber drawing control	DRL	2021	[196]
	Pipe isolation tool energy-saving control	DQN	2021	[197]
	Shear and tensile loading tuning of cold-forming tool geometries	DQN	2021	[198]
	Strip rolling process control	PPO	2022	[199]

5.5 Quality Control

Smart manufacturing requires automated and high-level quality inspection approaches. However, the increasing complexity of the fabrication process makes quality inspection more challenging. In this context, adopting DRL can improve the quality inspection capabilities of manufacturing systems due to its exploration ability while reducing the dependence on expert knowledge or manual inspection. The DRL applications are listed in *Table 14*.

To reduce the reliance on expert knowledge, Jorge et al. [200] used DRL to identify and limit systematic errors in expert systems used for geometry assurance, or even to reject biased advice from expert systems. In addition, Brito et al. [201] used collaborative robots with human-in-the-loop control to support smart inspection and corrective action, in which DRL is adopted for the robot inspection policy learning. Meanwhile, to reduce manual labour, Lončarević et al. [202] generated a robot inspection moving path according to the Computer-aided Design (CAD) model of inspected part. Then, the DRL algorithm is employed to control the robot's moving speed without image quality degradation. Similarly, Landgraf et al. [203] presented a DRL-based approach to determine a high-quality set of sensor view poses for arbitrary workpieces based on the 3D CAD model.

With the exploration ability of DRL, Luo et al. [204] proposed a model-driven adaptive PPO for production assembly. Facing the complicated and high-dimensional spaces of the assembly environment, this method can enable the assembly system autonomously to rectify the bolt posture error, thus improving assembly efficiency and stability. Cheng et al. [205] proposed the formulation of the multi-light source lighting strategy to improve the inspection capability and result quality, where RL is adopted to improve the lighting strategy to extract the diverse defects.

Table 14 - Literature of DRL in Quality Control

Quality Control in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Improves automatic quality inspection performance • Simplify the inspection settings • Reduce expert knowledge reliance and manual work	Geometry assurance	DDPG	2018	[200]
	Autonomously rectify the bolt posture error	PPO	2021	[204]
	Multi-lights source lighting strategy of Automated Optical Inspection	DDPG	2021	[205]
	Smart inspection and corrective actions for robot quality control systems	Actor-Critic	2020	[201]
	Robot inspection path generation and parameter tuning	DRL	2021	[202]
	Workpieces sensor view poses generation	PPO	2021	[203]
	Wafer probing coverage path planning	DRL	2022	[206]

5.6 Robotics in Manufacturing

With AI and advanced control technologies empowering industrial robots with higher-level cognitive and execution capabilities, they no longer solely take charge of repetitive and heavy physical work in manufacturing. Especially with the DRL algorithm, industrial robots tend to perform flexible, high-precision dexterous tasks in complex and unpredictable manufacturing scenes. In this section, the DRL algorithm-driven robot applications in manufacturing can be divided into four main categories: *manipulation, motion planning, scheduling, cloud robotics,* and *human-robot interaction*.

Manipulation - Industrial assembly tasks require robots to embrace contact-rich manipulation skills with high adaptability, which are challenging for traditional classical control and motion planning approaches but more suitable for adopting DRL algorithms. DRL can be used to improve control performance and manipulation adaptability and reduce the reliance on expert knowledge. The applications are summarized and listed in *Table 15*.

Lutter et al. [207] and Hebecker et al. [208] adopted IL and PPO, respectively, to learn assembly skills and automate the process of contact-rich-compliant assembly. Lan et al. [209] used MARL architecture integrated into DQN to optimize the coordination for a multi-robot pick and place system. Chen et al. [210] proposed a meta-RL control policy, which can enhance the adaptability of collaborative robots when facing new tasks through task modularization and efficient transfer. Moosmann et al. [211] proposed a DRL-based approach to separate entangled workpieces and minimized the setup effort. Zhang et al. [212] proposed the DRL method incorporating classical force control to find the optimal compensation term, which can satisfy the needs in robot tracking scenarios when facing unknown curved workpieces. Similarly, Zhang et al. [213] proposed a DRL-based force control algorithm for the impact and processing stages of robotic constant-force grinding. Liang et al. [214] proposed a method of inner/outer loop impedance control based on natural gradient actor-critic RL to reduce vibration, thus improving rubber unstacking performance. As for elastic and soft textile objects (such as shoe tongues and shoe textile uppers), Tsai et al. [215] and Li et al. [216] both used DRL to generate robot task control policies. One is to enable a robotic arm to learn a shoe tongue's specific image feature points through iterative training to improve manufacturing accuracy. The other one is to generate multi-point punching paths, in which the inspired path planning algorithm is conducted by DRL to get optimal results.

Meanwhile, to reduce the reliance on expert knowledge and manual labor, Thomas et al. [217] combined DRL with CAD design files to obtain the task's geometric information and then guided the robot along the computed geometric motion plan to complete high-precision assembly. Regarding tightening applications, Luo et al. [218] presented a transfer learning-based DRL method to extract the mathematical mapping between model agents and subjective knowledge. The proposed algorithm can enable agents to learn from human knowledge more efficiently, in which an inverse RL method based on prior knowledge is presented to acquire reward functions. Maldonado-Ramirez et al. [219] proposed a DRL approach that provides the robot agent with sufficient exploration and observation variability through a virtual environment and domain randomization, by which the robot can learn to track welding paths adaptively. In robot batching, Hildebrand et al. [220] proposed a DRL-based training approach and designed a Unity simulation framework incorporating existing commercial robot batching units to train control policies.

Table 15 – Literature of DRL in Robotics Control - Manipulation

Robot Manipulation in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize the application performance index • Enhance the control adaptability towards uncertainty • Reduce expert knowledge reliance and manual work	Assembly	Robot assembly involves contact-rich dynamics	DRL	2018	[217]
		Expert-based assembly skill-generating	DRL+IL	2021	[207]
		Contact rich compliant assembly	PPO	2021	[208]
		Automotive assembly	DRL	2022	[221]
		Assembly skill-transfer	SAC	2022	[222]
	Pick&Place	Multi-robot pick&place system coordination optimization	DQN+MARL	2021	[209]
		Enhance task adaptability of collaborative robots	Meta RL	2021	[210]
		Separate entangled workpieces	DQN	2021	[211]
	Tightening	Robot tightening assembly and inspection	Inverse DRL	2020	[218]
	Curve Tracing	Robot end-effector tracking unknown curved-surface workpieces	A2C	2020	[212]
	Overshoot Prevention	Robotic constant-force grinding	DRL	2019	[213]
	Elastic/Soft Objects	Unpredictable and time-variable adhesion rubber force control	Actor-Critic	2021	[214]
		Automated manufacturing of soft fabric shoe tongues	DQN	2020	[215]
		Punching of textile uppers in shoemaking	A3C	2018	[216]
	Welding	Welding robot path following	DRL	2021	[219]
	Batching	Robot-batching control optimization	PPO	2020	[220]

Motion planning - For smart manufacturing, integrating self-learning capabilities into the current fixed, repetitive, task-oriented industrial manipulators, thus leading them to an intelligent manner, is a promising direction. Among that, it is essential to develop more cognitive and flexible motion planning strategies. DRL is beneficial in optimal motion planning solution searching due to its exploration capability. The summary of the motion planning applications is listed in *Table 16.*

The application scenarios can be split into single-robot motion planning and multi-robot motion coordination. For single-robot motion planning, Zeng et al. [223], Pane et al. [224], Meyes et al. [225], Matulis et al. [226], Li et al. [227], [228] , and Kim et al. [229] all adopted the DRL algorithm to help the industrial robot search and generate the path to the target position through DT models, simulators or the physical robot itself. Lu et al. [230] proposed a hybrid particle swarm optimization (PSO) and RL approach, combining DRL and particle swarm optimization, which could provide higher accuracy by analyzing the movement trajectory and speed. Hua et al. [231] proposed a motion planning algorithm for the redundant robot by leveraging a series of hardware designs and DRL-based training to handle the skills. Zheng et al. [232] stepwise introduced an industrial knowledge graph-based MARL method for achieving multi-robot arms motion planning.

Table 16 - Literature of DRL in Robotics Control - Motion Planning

Robot Motion Planning in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Improve solution searching efficiency • Optimize the application performance index	Industrial Robot	Industrial robot arm motion and task planning	DDPG	2020	[223]
			Actor-Critic	2019	[224]
			Actor-Critic	2018	[225]
			PPO	2021	[226]
			TD3	2020	[229]
			DQN	2021	[230]
			PPO	2022	[227][228]
	Redundant Robot	Collision-free path planning for the duct-enter task	DDPG	2020	[231]
	Muti-Industrial Robot	Multi-robot arm motion planning in cognitive manufacturing	SAC+MARL	2021	[232]

Insertion – The insertion process is a classical task in robot manipulation. Plenty of works have discussed insertion due to the diversity of operational targets, the complexity of contact forces, and a wide range of application scenarios. DRL is adopted in insertion to enhance the robot's precision control flexibility, adaptability, and learning efficiency. The detail of the robot-based insertion applications is listed in *Table 17*.

To improve the insertion precision and learning efficiency, Inoue et al. [233] firstly performed a peg-in-hole task with a tight clearance through training a recurrent neural network, which adopted DRL to observe the robot sensors and estimate the system state, then take the optimal action. However, there still exist challenges in applying DRL to the contact-needed assembly tasks since the exertion of excessive force may cause danger or task failure in the random search process of DRL. Thus, Aschersleben et al. [234] integrated position control and force sensor signals into the DRL algorithm to compensate for positioning inaccuracies. Kim et al. [235] used the neural-network-based movement primitive to generate a continuous trajectory for the contact task by transmitting it to the force controller and learned the policy via DDPG and IL. Another approach is to learn from human demonstration. Cho et al. [236] and Vecerik et al. [237] enabled the robot

to learn uncertain shape entity insertion with efficiency and robustness by combining the DRL approach with a small number of human demonstrations.

Facing the challenges of sample inefficiency, safety issue, observation lack, and sparse reward signals in insertion control policy learning, DRL is employed to solve these problems and improve adaptability. Schoettler et al. [238], Zhao et al. [239], Beltran-Hernandez et al. [240], and Li et al. [241] [242] all tried to help robots robustly learn assembly skills while minimizing real-world interaction sample amount requirements, which is more flexible and suitable for realistic assembly scenarios. Their proposed algorithms include combining RL with prior knowledge, generating virtual data to argument transition samples, and bootstrapping the training speed using several transfer-learning techniques. Lastly, Luo et al. [243] defined the criteria for industry-oriented DRL from the demonstration. And they performed a thorough comparison between the proposed criteria with a NIST benchmark recently established by a professional industrial integrator.

Table 17– Literature of DRL in Robotics Control - Insertion

Robot Insertion Applications in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize the insertion performance index (time, precision, success rate) • Improve solution searching efficiency • Enhance the control adaptability towards uncertainty • Simplify the task settings and reduce reliance on expert knowledge	System Integration	High precision peg-in-hole assembly	DQN	2017	[233]
			DQN	2020	[234]
			DDPG	2020	[235]
	Sim2Real	Industrial insertion tasks with visual inputs and different natural reward specifications	SAC+TD3	2020	[238]
		Efficiently learn assembly policy	SAC	2020	[239]
		Complex, high-precision assembly in an unstructured environment	SAC	2020	[240]
		DT-enabled flexible assembly	DDPG	2022	[241]
			DDPG	2021	[242]
	Imitation	Generalize motor skills in different shapes of pegs and holes	DRL	2020	[236]
		Narrow-clearance peg-insertion task/Deformable clip-insertion task	DDPG	2019	[237]
	Standard	Randomly moving target assembly benchmark	DRL+LFD	2021	[243]

Scheduling - As the demand for rapid product iteration becomes increasingly fluctuant and customized, industrial robots need to cooperate, thus bringing new challenges, including dynamic reconfiguration, ubiquitous sensing, and communication with time constraints. With the learning efficiency and exploration capabilities, DRL can be practical to schedule and coordinate among multiple robots to advance autonomy and increase manufacturing efficiency. Applications are listed in *Table 18*.

Utilizing DRL's outstanding exploration capability and learning efficiency, Tan et al. [244] established an industrial robot assembly process model and a MARL-based approach for planning and scheduling multi-industrial robot-based assembly. Arviv et al. [245] proposed a dual Q-learning functions-based DRL, which assigns different reward functions component to robots to minimize the robot idle time and job waiting time. Furthermore, aiming for flexible manufacturing, Schwung et al. [246] embedded a learning module into the manufacturing cell, which allows the robots to learn to solve the given task and find an optimal cooperative behavior policy simultaneously. As for the AGVs, Agrawal et al. [247] provided a standardized framework and designed an integrated MARL-based job scheduling approach for an autonomous mobile robot-driven shop floor.

Due to the complexity arising from rapid environmental changes and the tight coupling between dispatching, path planning, and route execution, dispatching transport is difficult in dynamic production environments. By leveraging DRL's learning efficiency, Malus et al. [248] proposed an order dispatching algorithm based on MARL, where the AGV agent learns to bid on orders based on their observations. Hameed et al. [249] provided a curiosity-based DRL algorithm, using intrinsic motivation as a reward, on a flexible robot manufacturing cell and AGVs to alleviate scheduling problems. Chang et al. [250] used DQN to learn an AGV's dispatching policy. In the implementation, a target production line as a virtual simulated grid-shaped workspace is modeled to develop DQN; then an optimal dispatching policy can be automatically generated without requiring human control or prior expert knowledge.

Table 18 - Literature of DRL in Robotics Control – Scheduling

Robot Scheduling in Manufacturing Stage					
DRL Objectives	**Objects**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize the scheduling performance • Improve solution searching efficiency	Multiple Industrial Robots	Planning and scheduling algorithm for industrial robot assembly	MARL	2019	[244]
		Cooperative flexible robot manufacturing units	Double DQN	2016	[245]
			DRL	2019	[246]
			MARL	2021	[247]
			MARL	2022	[251]
			Double DQN	2022	[252]
	AGVs	Shop floor-based AGV navigation control and job scheduling	TD3+MARL	2021	[248]
			Actor-Critic	2021	[249]
			DQN+MARL	2021	[250]

Cloud robotics - Industrial cloud robotics combines cloud computing and industrial robotics, which embraces the benefits of resource sharing, easy access, and high efficiency. Applications are listed in *Table 19*. In current manufacturing shops, most disconnected industrial robots use resource-limited onboard processors and memory, which leads to limitations in information sharing across multiple robots. Thus, DRL is integrated into cloud services to improve the control flexibility and adaptability of cloud-based robots. Du et al. [253] proposed a framework with a cloud-based knowledge-sharing mechanism and a DRL-based service scheduling collaborative

optimization approach for cloud robots. Liu et al. [254] proposed a framework for industrial robot skill training in cloud manufacturing with DRL to learn various manipulation skills.

Table 19 - Literature of DRL in Cloud Robotics

Cloud Robotics in Manufacturing Stage				
DRL Objectives	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Enhance the policy learning efficiency • Improve the control policy adaptability	Knowledge sharing for multi-robot collaborative optimization	DRL	2019	[253]
	Industrial robot skill training for cloud manufacturing	DRL	2020	[254]

Human-robot interaction - The intuitive interaction between humans and industrial robots is essential to the road of smart manufacturing, while automated robots improve efficiency and precision, and human participation ensures flexibility. Taking human factors into consideration in interaction strategies and thus optimizing human-robot collaboration has become a prevailing trend. However, unpredictable human behaviors challenge the task planning and decision-making of the robot. Thus, DRL is employed to develop the control approach to deal with human uncertainty and own adaptability. The applications of DRL-driven robot control policies in HRI scenarios range from essential safety measures to cognitive robot assistance for workers, as listed in *Table 20.*

Safety insurance owns the highest priority during manufacturing operations. To achieve safety control in human-robot interaction (HRI), DRL is mainly used to generate collision avoidance motion planning or navigation control strategies (Xiong et al. [255], Zhu et al. [256], Liu et al. [257], Terra et al. [258]) for industrial robot arms or AGVs.

Furthermore, assembly is the most discussed application in production activities. DRL is mostly adopted as an adapted learning approach to support robots in assembly assistance. Robots can learn to collaborate with various human operators with the help of DRL to accomplish a high-precision assembly task (Liu et al. [259], Meng et al.[260], Wang et al. [261]). DRL can also be used to improve the efficiency of manufacturing processes, such as optimizing the assembly sequence and balancing job distribution. DRL is capable of seeking scheduling policy for symbiosis human-robot collaboration even without experts' knowledge. With the help of DRL, robots could own the real-time decision-making ability when facing a dynamic environment, thus improving the efficiency and flexibility in finding an optimal policy for task scheduling (Yu et al. [262], Yu et al. [263], Zhang et al. [264], Lv et al. [265], Zhang et al. [266]).

Lastly, human workers' flexibility can sometimes be seen as a disturbance within the HRI system, making system modeling and optimization more challenging. Thus, it is a significant issue that enables robots to adapt their behavior according to variations in human performance proactively. Oliff et al. [267] presented a methodology that can effectively model the HRI system and developed a DRL agent capable of autonomous decision-making. The proposed method enables robots to change their actions based on the observed information and improves the interaction between robots and their human partners. Moreover, Alonso et al. [268] used Double DQN and Recurrent Neural Network (RNN) to detect anomalous behavior patterns. By predicting the worker’s attention level, fatigue, and distraction, the algorithm can prevent workers from hazardous situations in automated and robotized agile-production environments.

Table 20 - Literature of DRL in Robotics Control – Human-Robot Interaction

Human-Robot Interaction in Manufacturing Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Optimize the HRI performance (safety, response, task productivity, working efficiency) • Enhance the control adaptability towards uncertainty	Safety Control	Industrial robot real-time collision avoidance and 3D motion planning	DDPG	2019	[255]
			DRL	2021	[256]
			DDPG	2021	[257]
		Risk mitigation modules for human-robot collaboration	DQN	2020	[258]
	Assembly Assistance	Human-robot collaborative assembly/ hybrid assembly tasks	DQN	2021	[259]
			PPO	2021	[260]
			Inverse RL	2019	[261]
		HRC task planning/assignment	DQN+MARL	2021	[262]
			DRL	2020	[263]
			DDPG	2022	[264]
			DQN	2022	[265]
			SAC	2022	[266]
	Worker Behavior	Prevent hazardous situations for workers	DQN	2021	[267]
		Robot adaptation to changed observed information	Double DQN	2021	[268]
		Electromyography-based human intention prediction	DDPG	2022	[269]

6 Deep Reinforcement Learning in Distribution Stage

In the manufacturing distribution stage, due to the growing market demands and globalized production networks, the DRL applications in the distribution stage are mainly split into *inventory management* and *supply chain optimization*, as summarized in *Table 21*.

In the inventory management aspect, to improve the adaptability of management policy, Dittrich et al. [270] and Perez et al. [271] both utilized the DRL-based inventory management framework to support manufacturers obtain satisfied management control behaviors, such as lower cost and higher stability/balance. Similarly, for drug suppliers (Zwaida et al. [272]) and semi-conductor components suppliers (Chien et al. [273]), DRL-based methods could effectively predict uncertain demand and provide corresponding supply control policies.

Furthermore, DRL is adopted as an effective optimization approach to improve supply chain performance. To solve the multi-period capacitated supply chain optimization problems under uncertain demand, Peng et al. [274] proposed a PG-based DRL to control the number of products that need to be produced and delivered to each retailer. Achamrah et al. [275] proposed a DRL approach that combines Genetic Algorithm, which can minimize inventory cost and sales losses in a two-level supply chain by considering transshipment and substitution. With the increased involvement of humans in storage operation management, Niu et al. [276] utilized a MARL

approach in human-robot collaborative order picking tasks. The proposed method aimed to improve working efficiency while considering human comfort as a criterion, which provides a practical approach to human-centric manufacturing.

Table 21 – Literature of DRL in Distribution Stage

Inventory Optimization/Supply/HRI Applications in Distribution Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Improve the adaptability of management policy • Optimize the application performance index (productivity, cost, time efficiency)	Inventory Optimization	Global-level self-optimizing inventory control	DQN	2021	[270]
		Inventory policies design to cope with network disruptions	PPO	2021	[271]
		Sustainable inventory management for hospital supply chain	DRL	2021	[272]
		Optimal demand forecast model selection	DQN	2020	[273]
		Cost optimization of the serial supply chain network	DQN	2021	[277]
		Mobile robot scheduling in automated warehouse	MARL	2022	[278]
	Supply Chain Management	Multi-period capacitated supply chain optimization under demand uncertainty	Vanilla PG	2019	[274]
		Combinatorial complexity of dynamic and stochastic inventory routing	DQN	2021	[275]
		Human-robot collaboration of warehousing order assignment	DQN+MARL	2021	[276]
		Multi-item stochastic capacitated lot-sizing problem optimization	PPO	2022	[279]
		Deliver vehicle routing optimization	DRL	2022	[280]
		Pollution and returns optimization for green closed-loop supply chain cycle	DQN	2022	[281]

7 Deep Reinforcement Learning in Maintenance Stage

DRL applications in the maintenance phase of manufacturing systems can be classified into two categories: *general maintenance activities* and *self-healing maintenance*, as shown in *Table 22*.

7.1 General Maintenance Activities

The general maintenance activities can be categorized into reactive maintenance (including remanufacturing), preventive maintenance, and predictive maintenance according to the

maintenance timing. In general maintenance activities, DRL mainly benefits maintenance activities by improving productivity, flexibility, adaptability, and reducing human labor.

Uncertainty, such as the type and conditions of returned products for repair, always exists in reactive maintenance. To address the resulting high volatility in reactive maintaining disassembly, Wurster et al. [282] and Mao et al. [283] both used Petri-Net to transform the disassembly sequence planning into DRL-solvable MDPs. In addition, Wurster et al. used DQN in the automated workstation to generate the material flow control policy, which can balance the materials entry and disassembly velocity to maximize the working efficiency. For products with unpredictable inner structures and non-predefined disassembly steps, Mao et al. also introduced DQN in assisting disassembly planning generation and integrated the system into Virtual Reality (VR) for maintenance training. To improve the production system performance, scholars (Huang et al. [284], Kuhnle et al. [285], Su et al. [286], Yan et al. [287], Nguyen et al. [288], Yan et al. [289] and Valet [290]) applied DRL for the preventive maintenance policies design and optimization. Applying the DRL algorithm allows for joint consideration of production resource loss and delivery strategy constraints. Meanwhile, it simplifies the system modeling process, so that the final generated maintenance policy can embrace lower downtime, increased throughput, and reduced cost. Hoong Ong et al. [291] and Rabbanian et al. [292] adopted DRL to generate maintenance policies. With the integration of IoT, real-time production data can be collected and feedback to the DRL system, which realizes the continuous maintenance policy optimization and obtains better foreseeable decision-making.

Not limited to production systems, the DRL-based maintenance of tools also takes a significant role in ensuring the machining quality and improving the productivity of automatic systems. In an end-to-end training mode, Wang et al. [293] combined CNN with an improved actor-critic algorithm for bearings and tool fault recognition. This work could well distinguish compound faults under heavy background noise. Yao et al. [294] integrated a transfer learning-based DRL method into an LSTM network to predict tools wear and remaining useful life. For chemical vapor deposition tools, Liao et al. [295] also used DQN and supervised LSTM to predict the predictable elements used in calculating the Predictive Overall Equipment Effectiveness, and stochastic dynamics in production and quality.

7.2 Self-healing Maintenance

Manufacturing systems may operate in non-optimal conditions due to aging equipment failures and raw material changes. Under this circumstance, the adoption of DRL can improve system adaptability and enable self-tuning or self-repair to maintain optimal operating efficiency. Verma et al. [296] proposed a DRL-based damage-aware control architecture for robots incorporating domain randomization, which can conduct diagnosis in the damaged space using LTSM and relearn the control policy in a single shot before robots have gait selection. After inferring deficient components from the variation in product quality, Epureanu et al. [297] established a DRL-based multiple-level self-repair strategies to maintain the normal operation of the manufacturing system. Qin et al. [298] proposed a DRL-based intelligent non-optimality self-recovery method for batch processes.

Table 22 - Literature of DRL in Maintenance Stage

General / Self-healing Maintenance Applications in Maintenance Stage					
DRL Objectives	**Category**	**Application Scenarios**	**Algorithm**	**Year**	**Reference**
• Improve maintenance performance index (cost, productivity, adaptability) • Reduce reliance on expert knowledge and human labor • Enable the self-tuning/self-repair to maintain operating efficiency	General Maintenance	Condition-based control for hybrid disassembly systems	DQN	2022	[282]
		Adaptive disassembly sequence planning for the VR maintenance training	DQN	2022	[283]
		Manufacturing system preventive maintenance planning	Double DQN	2020	[284]
			PPO	2019	[285]
			Actor-Critic+MARL	2022	[286]
			NN-based Q-learning	2022	[287]
			MARL	2022	[288]
			DQN	2022	[289]
			DQN	2022	[290]
		Manufacturing equipment preventive maintenance policy design	Double DQN	2020	[291]
			DQN	2021	[292]
			MARL	2022	[299]
			PPO	2022	[300]
			DRL	2022	[301]
		Tool wear and equipment fault diagnosis and prediction	Actor-Critic	2021	[293]
			DQN	2021	[294]
			DQN	2018	[295]
			PPO	2022	[302]
	Self-healing Maintenance	Robot damage-aware control	DRL	2020	[296]
		Manufacturing system deficient components self-repair strategies	DRL	2020	[297]
		Non-optimality self-recovery method for batch process	Actor-Critic	2018	[298]

8 Challenges & Prospects

From the above review, the outstanding performance of DRL methods makes us believe that manufacturing systems based on DRL undoubtedly occupy an important place in the future smart manufacturing paradigm. However, it is still challenging to design, invoke and deploy the DRL algorithms due to the practical system issues, such as over-complexity of manufacturing systems, security limitations of manufacturing systems, and high cost of data acquisition. Therefore, the authors extract the technical pain points that may exist in the pipeline of building DRL-enabled manufacturing system. Meanwhile, the emerging and critical DRL technologies adapt them from algorithm programming, algorithm type, and algorithm setting views are correspondingly listed.

Furthermore, how they can improve the deployment feasibility, cognitive capability, and learning efficiency of the manufacturing systems are also discussed.

8.1 Universal Interface

For the practice of DRL in manufacturing, there exist difficulties in algorithm integration and software environment compatibility like the OpenAI gym environment and programming interface [303] [304]. In the application summary, it can be seen that process control, scheduling, and the robot-based manufacturing applications still remain dominant and challenging. However, there lacks a universal programming and evaluation interface for integrating DRL into manufacturing systems. Currently, the learning environment is not uniform nor compatible with existing mature algorithm libraries, leading to significant difficulties and extra effort in deploying ready-made algorithms. Meanwhile, due to the lack of uniform standards and interfaces, it is too complicated for users to make comparisons for algorithm selection and employ advanced algorithms and parameter optimization. When facing a novel manufacturing scenario, the scholar may stack in adopting a DRL algorithm instead of developing their idea. It may cause an inaccurate presentation of the research significance and a lack of reference significance. Finally, some research in manufacturing solely focuses on algorithms, instead of system design, exploration strategies, and reward shaping. In the above cases, the unified deployment interface with mature algorithm implementation could significantly improve the usage efficiency and provide further guidance for similar research problems.

8.2 Generalization

With the trend of small-batch and customization in smart manufacturing, manufacturing systems are faced with more diverse and time-sensitive tasks. The potential research in DRL should consider how to quickly adapt to production demands for various manufacturing equipment and order inputs, especially in the design, manufacturing, and logistics stages. Thus, meta-learning [305], hierarchical learning (curriculum learning) [306][307] series algorithms could accumulate prior knowledge and enhance adaptability and can be thought of as promising directions. First, meta-reinforcement learning methods can take advantage of the multi-objective optimization feature to accumulate prior knowledge of manufacturing tasks and thus form a spiritual learning network. As a changed task input, the learning network can adapt to it in a few shots to increase sample utilization and shorten the learning process. Except for meta-learning, a mass production task can be decomposed into sub-tasks using expert knowledge, and the decomposed task can be learned in steps using hierarchical RL and curriculum learning methods. Thus, leveraging the appropriate task settings, learning approaches, and reward shaping could decrease the appearance of unsatisfied performance and unstable strategies owing to local optima in the learning process.

8.3 Simulation to Reality

A major difficulty regarding the deployment of DRL in manufacturing systems currently lies in the transfer of simulation results to real-world scenarios, especially in process control and robotics application. In reviewed applications, most of the applications are represented in a simplified way on the simulator. The policy is acceptable for manufacturing applications with low time and environmental requirements like design, maintenance, scheduling, etc. However, in robot control or machining control, the virtual environments could not model the physical environment precisely and had to deal with many unconsidered parameters and preprocessed datasets when deploy. It makes the control methods obtained by DRL in simulated environments may lead to performance

degradation after the actual transfer, even cannot be directly deployed in physical environments. Therefore, the main research problem in the Sim2Real domain lies in how to close the gap between the simulated training environment and the real physical environment and achieve more effective knowledge transfer.

Simulation to Reality (Sim2Real) refers to the technology of migrating and deploying knowledge learned in simulators to the real world [308], i.e., setting corresponding control tasks and driving an agent to learn in a virtual environment provided by a simulator physics engine, and then deploying the policies obtained from the training to a real physical environment to achieve control of the physical agent. Currently, Sim2Real mainly focuses on solving the real problem of policy deployment for deep reinforcement learning, especially in robot control.

In the field of smart manufacturing, DT technology is dedicated to establishing a comprehensive, accurate, and real-time connection between the physical and digital worlds, thus enabling monitoring, control, and prediction of physical entities through virtual models. The knowledge migration laws from the simulated to the real environment explored by Sim2Real can reduce the difficulty of twin design and reach the goal of using simple but parameter-accurate models to achieve efficient and accurate DT/CPS systems. At the same time, the DT has a natural and close relationship with Augmented Reality/VR/Mix Reality and other technologies to build a richer holomorphic anthropomorphic display model. Combining Sim2Real technology can enhance the intuitiveness of interaction and the accuracy of control of the virtual-physical fusion [309].

8.4 Exploration & Offline RL

Due to the trial-and-error principle of DRL exploration, DRL requires continuous interaction with the environment to reinforce the agent to gain a better performance. However, random exploration is almost impossible to exist in manufacturing due to safety limitations and potential risks. With the growth of the manufacturing system scale, the changes in the environmental state become more diverse and dynamic, and the safe exploration deployment issue of DRL should be taken into more consideration. Here comes the following two approaches:

DRL Exploration Policy*:* In the above summary, the use of search capabilities of DRL to optimize manufacturing problems is mentioned in almost all application categories. Moreover, the need for exploring the unknown environment in a highly efficient manner in order to improve the efficiency of using the samples is a key technique in DRL and is also hotly debated. Many researchers have studied exploration strategies in DRL from different perspectives [310]. However, there are still few appears the work applies exploration strategies in manufacturing. Therefore, in the coming research, it is expected that the exploration strategy in DRL can be used in manufacturing to improve the search capability and performance of existing applications.

Meanwhile, the DRL control policy lacks interpretability, which makes it difficult to guarantee the reliability of DRL strategies with unknown safety hazards. Therefore, to avoid the danger caused by unreliable models, DRL is mostly limited to assisting human decision-making in safety-sensitive tasks, such as robot-assisted surgery, assisted driving, etc. Similarly in manufacturing, the goal of safe DRL exploration is to create a DRL algorithm exploration policy that is under manufacturing safe constraints during both learning and deployment to avoid learning high-reward high-risk policies. Maximizing the expected reward while ensuring reasonable system performance and/or respecting safety constraints [311]. Examples include the case of data center cooling, where temperature and pressure must always be kept below their respective thresholds,

or robots that cannot exceed speed, angle, and torque limits; Usually, such safety exploration is done mainly by means of changing optimization criteria in combination with world standards of manufacturing systems or by combining expert experience and thus changing exploration policy.

Off-line DRL/Batch DRL. Unlike supervised learning, which uses large amounts of labeled data for learning, the learning mechanism of RL is to collect feedback and thus update the intelligence through interactive trial and error, i.e., the learning process requires constant interaction with the external environment and obtaining new data collection. However, such an online learning approach is prohibitive in many real-world settings, such as autonomous driving and robot control, where iterative experimental data collection can be costly, time-consuming, and even non-legally compliant. Instead, Offline Reinforcement Learning, also known as Batch Reinforcement Learning [312], is a variant of RL that drives the agent to learn from old data sequences collected, possibly with data that is not of expert level. This algorithm aims to use stored data (e.g., from previous experiments or human demonstrations) to learn behavior, reduce the number of interactions with the environment, maximize the use of static data sets to optimize RL intelligence, and avoid the time and cost drain.

In smart manufacturing, the nature of offline learning makes its integration with manufacturing systems much more possible, without any extensive exploration or interaction with the external environment. Especially in those manufacturing applications where data collection is costly (e.g., process control, production scheduling) or hazardous (e.g., robotics/processing equipment), the paradigm of offline learning promises to address the key challenge of bringing reinforcement learning algorithms from restricted laboratory environments/simulators into the real world.

8.5 Multi-agent Reinforcement Learning

MARL is a class of methods that apply reinforcement learning algorithms to individual intelligence to solve the control tasks of multi-intelligent systems. In such systems, each agent needs to have basic learning, reasoning, and planning capabilities. By using the MARL algorithm, an intelligent agent achieves complex intelligence through the collaboration of multiple individuals with simple intelligence, which effectively improves the robustness, reliability, and flexibility of the whole system. Currently, MARL has been applied to robot navigation, transportation scheduling, power system optimization, distributed sensing networks, and other fields, and its excellent performance proves that MARL is an effective method to control multi-intelligent systems.

With the advent of Industry 5.0, the scale of manufacturing systems is larger, and the system state is more complex, which will lead to a more difficult control of manufacturing systems. In response to those problems, MARL systems should attempt to improve this situation in the following ways:

1) **Generalization**. Apply MARL combined with the representational learning capability of deep learning to such multi-intelligent manufacturing systems with higher degrees of freedom and more complex environments in order to increase the generalization capability of the agent.
2) **Collaboration Efficiency**. When the scale of the manufacturing system increases and the communication between individuals in the system is limited, how to design the objective function, state representation, and communication mechanism in the DRL algorithm to achieve efficient coordination and collaboration of multiple agents with limited communication.
3) **Human-in-the-loop**. DRL gives machines the ability to understand, learn, and make decisions autonomously. However, in the case of sudden changes in the external environment, the agents may not be able to respond in time. In the face of such problems, it is worthwhile to study how to integrate human intelligence and machine intelligence and improve the ability of human-

machine interaction based on MARL methods and the introduction of human judgment and experience in a timely manner [313].

9 Conclusion

As a critical and emerging technology, DRL has great potential in smart manufacturing lifecycle stages. Consequently, it has attracted increasing attention by providing an adaptive and flexible solution for smart manufacturing systems, thereby facilitating a more cognitive and personalized manufacturing paradigm. To systematically reveal its essence, this work provided a systematic literature review of 264 selected items in the past decade from an engineering product lifecycle perspective, and further emphasized the challenges and future directions of DRL in smart manufacturing. It is also hoped that this comprehensive review can serve as a reference to attract more in-depth research and discussion on DRL and its further adoption in smart manufacturing.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partially supported by the grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15210222), National Natural Science Foundation of China (No. 52005424), and the Laboratory for Artificial Intelligence in Design (Project Code: RP2-1), Hong Kong Special Administrative Region, China.

Reference

[1] B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, and T. Freiheit, "Smart Manufacturing and Intelligent Manufacturing: A Comparative Review," *Engineering*, vol. 7, no. 6. 2021. doi: 10.1016/j.eng.2020.07.017.

[2] A. Vatankhah Barenji, X. Liu, H. Guo, and Z. Li, "A digital twin-driven approach towards smart manufacturing: reduced energy consumption for a robotic cell," *Int J Comput Integr Manuf*, vol. 34, no. 7–8, 2021, doi: 10.1080/0951192X.2020.1775297.

[3] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," in *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2018.

[4] A. Goldwaser and M. Thielscher, "Deep reinforcement learning for general game playing," in *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 2020. doi: 10.1609/aaai.v34i02.5533.

[5] Z. Yuyan, S. Xiayao, and L. Yong, "A Novel Movie Recommendation System Based on Deep Reinforcement Learning with Prioritized Experience Replay," in *International Conference on Communication Technology Proceedings, ICCT*, 2019. doi: 10.1109/ICCT46805.2019.8947012.

[6] Y. J. Hu and S. J. Lin, "Deep Reinforcement Learning for Optimizing Finance Portfolio Management," in *Proceedings - 2019 Amity International Conference on Artificial Intelligence, AICAI 2019*, 2019. doi: 10.1109/AICAI.2019.8701368.

[7] N. C. Luong *et al.*, "Applications of Deep Reinforcement Learning in Communications and Networking: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 21, no. 4. 2019. doi: 10.1109/COMST.2019.2916583.

[8] M. Q. Mohammed, K. L. Chung, and C. S. Chyi, "Review of deep reinforcement learning-based object grasping: Techniques, open challenges, and recommendations," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3027923.

[9] K. Zhu and T. Zhang, "Deep reinforcement learning based mobile robot navigation: A review," *Tsinghua Sci Technol*, vol. 26, no. 5, pp. 674–691, Oct. 2021, doi: 10.26599/TST.2021.9010012.

[10] H. Nguyen and H. La, "Review of Deep Reinforcement Learning for Robot Manipulation," *Proceedings - 3rd IEEE International Conference on Robotic Computing, IRC 2019*, pp. 590–595, Mar. 2019, doi: 10.1109/IRC.2019.00120.

[11] X. Liu, H. Xu, W. Liao, and W. Yu, "Reinforcement learning for cyber-physical systems," *Proceedings - IEEE International Conference on Industrial Internet Cloud, ICII 2019*, pp. 318–327, Nov. 2019, doi: 10.1109/ICII.2019.00063.

[12] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A Review of Deep Reinforcement Learning for Smart Building Energy Management," *IEEE Internet Things J*, vol. 8, no. 15, pp. 12046–12063, Aug. 2021, doi: 10.1109/JIOT.2021.3078462.

[13] R. Nian, J. Liu, and B. Huang, "A review On reinforcement learning: Introduction and applications in industrial process control," *Comput Chem Eng*, vol. 139, p. 106886, Aug. 2020, doi: 10.1016/J.COMPCHEMENG.2020.106886.

[14] V. Samsonov, K. ben Hicham, and T. Meisen, "Reinforcement Learning in Manufacturing Control: Baselines, challenges and ways forward," *Eng Appl Artif Intell*, vol. 112, p. 104868, Jun. 2022, doi: 10.1016/J.ENGAPPAI.2022.104868.

* *Corresponding author*: pai.zheng@polyu.edu.hk

[15] B. Cunha, A. M. Madureira, B. Fonseca, and D. Coelho, “Deep Reinforcement Learning as a Job Shop Scheduling Solver: A Literature Review,” *Advances in Intelligent Systems and Computing*, vol. 923, pp. 350–359, 2020, doi: 10.1007/978-3-030-14347-3_34/FIGURES/1.
[16] M. Panzer and B. Bender, “Deep reinforcement learning in production systems: a systematic literature review,” *https://doi.org/10.1080/00207543.2021.1973138*, 2021, doi: 10.1080/00207543.2021.1973138.
[17] D. Silver *et al.*, “Mastering the game of Go with deep neural networks and tree search,” *Nature*, vol. 529, no. 7587, 2016, doi: 10.1038/nature16961.
[18] B. R. Kiran *et al.*, “Deep Reinforcement Learning for Autonomous Driving: A Survey,” *IEEE Transactions on Intelligent Transportation Systems*, 2021, doi: 10.1109/TITS.2021.3054625.
[19] W. Chen, X. Qiu, T. Cai, H. N. Dai, Z. Zheng, and Y. Zhang, “Deep Reinforcement Learning for Internet of Things: A Comprehensive Survey,” *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3. 2021. doi: 10.1109/COMST.2021.3073036.
[20] H. Jiang, H. Wang, W. Y. Yau, and K. W. Wan, “A Brief Survey: Deep Reinforcement Learning in Mobile Robot Navigation,” in *Proceedings of the 15th IEEE Conference on Industrial Electronics and Applications, ICIEA 2020*, 2020. doi: 10.1109/ICIEA48937.2020.9248288.
[21] V. Mnih, D. Silver, and M. Riedmiller, “Playing Atari with Deep Q Learning,” *Nips*, 2013.
[22] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction,” *IEEE Trans Neural Netw*, vol. 9, no. 5, 1998, doi: 10.1109/tnn.1998.712192.
[23] R. S. Sutton and A. G. Barto, “Reinforcement Learning: An Introduction(2nd Ediction Draft),” *Kybernetes*, 2017.
[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, vol. 2.
[25] H. van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double Q-Learning,” in *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, 2016.
[26] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Frcitas, “Dueling Network Architectures for Deep Reinforcement Learning,” in *33rd International Conference on Machine Learning, ICML 2016*, 2016, vol. 4.
[27] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation,” in *Advances in Neural Information Processing Systems*, 2000.
[28] R. J. Williams, “Simple statistical gradient-following algorithms for connectionist reinforcement learning,” *Mach Learn*, vol. 8, no. 3, pp. 229–256, 1992.
[29] S. Bhatnagar, R. S. Sutton, M. Ghavamzadeh, and M. Lee, “Natural actor-critic algorithms,” *Automatica*, vol. 45, no. 11, 2009, doi: 10.1016/j.automatica.2009.07.008.
[30] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” Sep. 2015.
[31] V. Mnih *et al.*, “Asynchronous methods for deep reinforcement learning,” in *33rd International Conference on Machine Learning, ICML 2016*, 2016, vol. 4.
[32] J. Schulman, S. Levine, P. Moritz, M. Jordan, and P. Abbeel, “Trust region policy optimization,” in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 3.
[33] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” Jul. 2017.

[34] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *35th International Conference on Machine Learning, ICML 2018*, 2018.
[35] S. Fujimoto, H. van Hoof, and D. Meger, “Addressing Function Approximation Error in Actor-Critic Methods,” in *35th International Conference on Machine Learning, ICML 2018*, 2018, vol. 4.
[36] P. Munro *et al.*, “Behavioral Cloning,” in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 93–97. doi: 10.1007/978-0-387-30164-8_69.
[37] J. Ho and S. Ermon, “Generative Adversarial Imitation Learning,” Jun. 2016.
[38] A. Ng and S. Russell, “Algorithms for inverse reinforcement learning,” *Proceedings of the Seventeenth International Conference on Machine Learning*, vol. 0, 2000.
[39] H. Park *et al.*, “Deep reinforcement learning-based optimal decoupling capacitor design method for silicon interposer-based 2.5-D/3-D ICs,” *IEEE Trans Compon Packaging Manuf Technol*, vol. 10, no. 3, 2020, doi: 10.1109/TCPMT.2020.2972019.
[40] D. W. W. Low, D. W. K. Neo, and A. S. Kumar, “A study on automatic fixture design using reinforcement learning,” *International Journal of Advanced Manufacturing Technology*, vol. 107, no. 5–6, 2020, doi: 10.1007/s00170-020-05156-6.
[41] K. Son *et al.*, “Reinforcement-Learning-Based Signal Integrity Optimization and Analysis of a Scalable 3-D X-Point Array Structure,” *IEEE Trans Compon Packaging Manuf Technol*, vol. 12, no. 1, pp. 100–110, Jan. 2022, doi: 10.1109/TCPMT.2021.3129502.
[42] J. Yang, S. Harish, C. Li, H. Zhao, B. Antous, and P. Acar, “Deep Reinforcement Learning for Multi-Phase Microstructure Design,” *Computers, Materials and Continua*, vol. 68, no. 1, 2021, doi: 10.32604/cmc.2021.016829.
[43] C. Zimmerling, C. Poppe, O. Stein, and L. Kärger, “Optimisation of manufacturing process parameters for variable component geometries using reinforcement learning,” *Mater Des*, vol. 214, p. 110423, Feb. 2022, doi: 10.1016/j.matdes.2022.110423.
[44] E. Papachristou, A. Chrysopoulos, and N. Bilalis, “Machine learning for clothing manufacture as a mean to respond quicker and better to the demands of clothing brands: a Greek case study,” *International Journal of Advanced Manufacturing Technology*, vol. 115, no. 3, 2021, doi: 10.1007/s00170-020-06157-1.
[45] M. Szarski and S. Chauhan, “Instant flow distribution network optimization in liquid composite molding using deep reinforcement learning,” *J Intell Manuf*, 2022, doi: 10.1007/S10845-022-01990-5.
[46] M. Römer, J. Bergers, F. Gabriel, and K. Dröder, “Temperature Control for Automated Tape Laying with Infrared Heaters Based on Reinforcement Learning,” *Machines*, vol. 10, no. 3, Mar. 2022, doi: 10.3390/MACHINES10030164.
[47] H. Zhang, Y. Liu, H. Liang, L. Wang, and L. Zhang, “Service composition in cloud manufacturing: A DQN-based approach,” in *International Series in Operations Research and Management Science*, vol. 289, 2020. doi: 10.1007/978-3-030-43177-8_12.
[48] H. Liang, X. Wen, Y. Liu, H. Zhang, L. Zhang, and L. Wang, “Logistics-involved QoS-aware service composition in cloud manufacturing with deep reinforcement learning,” *Robot Comput Integr Manuf*, vol. 67, no. April 2020, p. 101991, 2020, doi: 10.1016/j.rcim.2020.101991.
[49] Y. Liu *et al.*, “Logistics-involved service composition in a dynamic cloud manufacturing environment: A DDPG-based approach,” *Robot Comput Integr Manuf*, vol. 76, Aug. 2022, doi: 10.1016/J.RCIM.2022.102323.

[50] M. Moghaddam, A. Jones, and T. Wuest, “Design of marketplaces for smart manufacturing services,” in *Procedia Manufacturing*, 2019, vol. 39. doi: 10.1016/j.promfg.2020.01.312.
[51] J. Moon, M. Yang, and J. Jeong, “A novel approach to the job shop scheduling problem based on the deep Q-network in a cooperative multi-access edge computing ecosystem,” *Sensors*, vol. 21, no. 13, 2021, doi: 10.3390/s21134553.
[52] D. Bauer, T. Bauernhansl, and A. Sauer, “Improvement of delivery reliability by an intelligent control loop between supply network and manufacturing,” *Applied Sciences (Switzerland)*, vol. 11, no. 5, 2021, doi: 10.3390/app11052205.
[53] M. Zou, E. Huang, B. Vogel-Heuser, and C. H. Chen, “Efficiently learning a distributed control policy in cyber-physical production systems via simulation optimization,” in *IEEE International Conference on Automation Science and Engineering*, 2020, vol. 2020-January. doi: 10.1109/CASE48305.2020.9249228.
[54] M. She, “Deep Reinforcement Learning-Based Smart Manufacturing Plants with a Novel Digital Twin Training Model,” *Wirel Pers Commun*, 2021, doi: 10.1007/s11277-021-09072-0.
[55] K. Xia *et al.*, “A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence,” *J Manuf Syst*, vol. 58, 2021, doi: 10.1016/j.jmsy.2020.06.012.
[56] Z. Ren and J. Wan, “Strengthening Digital Twin Applications based on Machine Learning for Complex Equipment,” in *Proceedings -Design, Automation and Test in Europe, DATE*, 2021, vol. 2021-February. doi: 10.23919/DATE51398.2021.9474133.
[57] S. de Blasi, S. Klöser, A. Müller, R. Reuben, F. Sturm, and T. Zerrer, “KIcker: An Industrial Drive and Control Foosball System automated with Deep Reinforcement Learning,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 102, no. 1, 2021, doi: 10.1007/s10846-021-01389-z.
[58] Y. Zhang, J. Qiao, G. Zhang, H. Tian, and L. Li, “Artificial Intelligence-Assisted Repair System for Structural and Electrical Restoration Using 3D Printing,” *Advanced Intelligent Systems*, p. 2200162, Oct. 2022, doi: 10.1002/AISY.202200162.
[59] Z. Liu, L. Hu, W. Hu, and J. Tan, “Petri Nets-Based Modeling Solution for Cyber–Physical Product Control Considering Scheduling, Deployment, and Data-Driven Monitoring,” *IEEE Trans Syst Man Cybern Syst*, 2022, doi: 10.1109/TSMC.2022.3170489.
[60] J. Huang, J. Su, and Q. Chang, “Graph neural network and multi-agent reinforcement learning for machine-process-system integrated control to optimize production yield,” *J Manuf Syst*, vol. 64, pp. 81–93, Jul. 2022, doi: 10.1016/J.JMSY.2022.05.018.
[61] C. Li and Q. Chang, “Hybrid feedback and reinforcement learning-based control of machine cycle time for a multi-stage production system,” *J Manuf Syst*, vol. 65, pp. 351–361, Oct. 2022, doi: 10.1016/J.JMSY.2022.09.020.
[62] D. Pahwa and B. Starly, “Dynamic matching with deep reinforcement learning for a two-sided Manufacturing-as-a-Service (MaaS) marketplace,” *Manuf Lett*, vol. 29, 2021, doi: 10.1016/j.mfglet.2021.05.005.
[63] H. Zhang, J. Leng, H. Zhang, G. Ruan, M. Zhou, and Y. Zhang, “A deep reinforcement learning algorithm for order acceptance decision of individualized product assembling,” in *Proceedings 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence, DTPI 2021*, 2021. doi: 10.1109/DTPI52967.2021.9540190.

[64] M. A. Dittrich and S. Fohlmeister, “Cooperative multi-agent system for production control using reinforcement learning,” *CIRP Annals*, vol. 69, no. 1, 2020, doi: 10.1016/j.cirp.2020.04.005.
[65] J. Leng *et al.*, “A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0,” *J Clean Prod*, vol. 280, p. 124405, 2021, doi: 10.1016/j.jclepro.2020.124405.
[66] W. Wu, Z. Huang, J. Zeng, and K. Fan, “A fast decision-making method for process planning with dynamic machining resources via deep reinforcement learning,” *J Manuf Syst*, vol. 58, 2021, doi: 10.1016/j.jmsy.2020.12.015.
[67] Z. Mueller-Zhang, P. O. Antonino, and T. Kuhn, “Integrated planning and scheduling for customized production using digital twins and reinforcement learning,” in *IFAC-PapersOnLine*, 2021, vol. 54, no. 1. doi: 10.1016/j.ifacol.2021.08.046.
[68] Y. Sugisawa, K. Takasugi, and N. Asakawa, “Machining sequence learning via inverse reinforcement learning,” *Precis Eng*, vol. 73, 2022, doi: 10.1016/j.precisioneng.2021.09.017.
[69] Z. He, K. P. Tran, S. Thomassey, X. Zeng, J. Xu, and C. Yi, “A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process,” *Comput Ind*, vol. 125, 2021, doi: 10.1016/j.compind.2020.103373.
[70] H. Ghorbel *et al.*, “SOON: Social Network of Machines to Optimize Task Scheduling in Smart Manufacturing,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2021, vol. 2021-September. doi: 10.1109/PIMRC50174.2021.9569644.
[71] W. Wu, Z. Huang, J. Zeng, and K. Fan, “A decision-making method for assembly sequence planning with dynamic resources,” *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1937748.
[72] Z. He, K. P. Tran, S. Thomassey, X. Zeng, J. Xu, and C. Yi, “Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning,” *J Manuf Syst*, 2021, doi: 10.1016/j.jmsy.2021.03.017.
[73] M. Klar, M. Glatt, and J. C. Aurich, “An implementation of a reinforcement learning based algorithm for factory layout planning,” *Manuf Lett*, vol. 30, 2021, doi: 10.1016/j.mfglet.2021.08.003.
[74] B. Kim, Y. Jeong, and J. G. Shin, “Spatial arrangement using deep reinforcement learning to minimise rearrangement in ship block stockyards,” *Int J Prod Res*, vol. 58, no. 16, 2020, doi: 10.1080/00207543.2020.1748247.
[75] J. H. Woo, B. Kim, S. H. Ju, and Y. I. Cho, “Automation of load balancing for Gantt planning using reinforcement learning,” *Eng Appl Artif Intell*, vol. 101, 2021, doi: 10.1016/j.engappai.2021.104226.
[76] Y. Lv, Y. Tan, R. Zhong, P. Zhang, J. Wang, and J. Zhang, “Deep reinforcement learning-based balancing and sequencing approach for mixed model assembly lines,” *IET Collaborative Intelligent Manufacturing*, vol. 4, no. 3, pp. 181–193, Sep. 2022, doi: 10.1049/CIM2.12061.
[77] M. Neves and P. Neto, “Deep reinforcement learning applied to an assembly sequence planning problem with user preferences,” *International Journal of Advanced Manufacturing Technology*, Oct. 2022, doi: 10.1007/S00170-022-09877-8.

[78] T. Dong, F. Xue, C. Xiao, and J. Li, “Task scheduling based on deep reinforcement learning in a cloud manufacturing environment,” *Concurr Comput*, vol. 32, no. 11, 2020, doi: 10.1002/cpe.5654.

[79] Y. Liu, L. Zhang, L. Wang, Y. Xiao, X. Xu, and M. Wang, “A framework for scheduling in cloud manufacturing with deep reinforcement learning,” in *IEEE International Conference on Industrial Informatics (INDIN)*, 2019, vol. 2019-July. doi: 10.1109/INDIN41052.2019.8972157.

[80] H. Zhu, M. Li, Y. Tang, and Y. Sun, “A Deep-Reinforcement-Learning-Based Optimization Approach for Real-Time Scheduling in Cloud Manufacturing,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2964955.

[81] Y. Liu, Y. Ping, L. Zhang, L. Wang, and X. Xu, “Scheduling of decentralized robot services in cloud manufacturing with deep reinforcement learning,” *Robot Comput Integr Manuf*, vol. 80, p. 102454, Apr. 2023, doi: 10.1016/J.RCIM.2022.102454.

[82] L. Zhang, C. Yang, Y. Yan, and Y. Hu, “Distributed Real-Time Scheduling in Cloud Manufacturing by Deep Reinforcement Learning,” *IEEE Trans Industr Inform*, vol. 18, no. 12, pp. 8999–9007, May 2022, doi: 10.1109/TII.2022.3178410.

[83] X. Wang *et al.*, “Dynamic scheduling of tasks in cloud manufacturing with multi-agent reinforcement learning,” *J Manuf Syst*, vol. 65, pp. 130–145, Oct. 2022, doi: 10.1016/J.JMSY.2022.08.004.

[84] J. Tang and K. Salonitis, “A Deep Reinforcement Learning Based Scheduling Policy for Reconfigurable Manufacturing Systems,” in *Procedia CIRP*, 2021, vol. 103. doi: 10.1016/j.procir.2021.09.089.

[85] S. Yang and Z. Xu, “Intelligent scheduling and reconfiguration via deep reinforcement learning in smart manufacturing,” *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1943037.

[86] C. Hofmann, C. Krahe, N. Stricker, and G. Lanza, “Autonomous production control for matrix production based on deep Q-learning,” in *Procedia CIRP*, 2020, vol. 88. doi: 10.1016/j.procir.2020.05.005.

[87] D. Schwung, J. N. Reimann, A. Schwung, and S. X. Ding, “Self Learning in Flexible Manufacturing Units: A Reinforcement Learning Approach,” in *9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications, IS 2018 - Proceedings*, 2018. doi: 10.1109/IS.2018.8710460.

[88] D. Schwung, M. Modali, and A. Schwung, “Self-optimization in smart production systems using distributed reinforcement learning,” in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2019, vol. 2019-October. doi: 10.1109/SMC.2019.8914088.

[89] S. Mayer, T. Classen, and C. Endisch, “Modular production control using deep reinforcement learning: proximal policy optimization,” *J Intell Manuf*, vol. 32, no. 8, 2021, doi: 10.1007/s10845-021-01778-z.

[90] D. Gankin, S. Mayer, J. Zinn, B. Vogel-Heuser, and C. Endisch, “Modular Production Control with Multi-Agent Deep Q-Learning,” in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2021, vol. 2021-September. doi: 10.1109/ETFA45728.2021.9613177.

[91] M. Li *et al.*, “Decentralized Multi-AGV Task Allocation based on Multi-Agent Reinforcement Learning with Information Potential Field Rewards,” in *2021 IEEE 18th*

International Conference on Mobile Ad Hoc and Smart Systems (MASS), 2021, pp. 482–489.

[92] A. Kuhnle, L. Schäfer, N. Stricker, and G. Lanza, "Design, implementation and evaluation of reinforcement learning for an adaptive order dispatching in job shop manufacturing systems," in *Procedia CIRP*, 2019, vol. 81. doi: 10.1016/j.procir.2019.03.041.

[93] A. Kuhnle, J. P. Kaiser, F. Theiß, N. Stricker, and G. Lanza, "Designing an adaptive production control system using reinforcement learning," *J Intell Manuf*, vol. 32, no. 3, 2021, doi: 10.1007/s10845-020-01612-y.

[94] H. Rummukainen and J. K. Nurminen, "Practical reinforcement learning - Experiences in lot scheduling application," in *IFAC-PapersOnLine*, 2019, vol. 52, no. 13. doi: 10.1016/j.ifacol.2019.11.397.

[95] Y. G. Kim, S. Lee, J. Son, H. Bae, and B. do Chung, "Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system," *J Manuf Syst*, vol. 57, 2020, doi: 10.1016/j.jmsy.2020.11.004.

[96] A. Gannouni, V. Samsonov, M. Behery, T. Meisen, and G. Lakemeyer, "Neural Combinatorial Optimization for Production Scheduling with Sequence-Dependent Setup Waste," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2020, vol. 2020-October. doi: 10.1109/SMC42975.2020.9282869.

[97] B. A. Han and J. J. Yang, "Research on adaptive job shop scheduling problems based on dueling double DQN," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3029868.

[98] J. C. S. Ruiz, J. M. Bru, and R. P. Escoto, "Smart digital twin for ZDM-based job-shop scheduling," in *2021 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2021 - Proceedings*, 2021. doi: 10.1109/MetroInd4.0IoT51437.2021.9488473.

[99] L. Hu, Z. Liu, W. Hu, Y. Wang, J. Tan, and F. Wu, "Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network," *J Manuf Syst*, vol. 55, no. December 2019, pp. 1–14, 2020, doi: 10.1016/j.jmsy.2020.02.004.

[100] C. C. Lin, D. J. Deng, Y. L. Chih, and H. T. Chiu, "Smart Manufacturing Scheduling with Edge Computing Using Multiclass Deep Q Network," *IEEE Trans Industr Inform*, vol. 15, no. 7, 2019, doi: 10.1109/TII.2019.2908210.

[101] L. Zhou, L. Zhang, and B. K. P. Horn, "Deep reinforcement learning-based dynamic scheduling in smart manufacturing," in *Procedia CIRP*, 2020, vol. 93. doi: 10.1016/j.procir.2020.05.163.

[102] S. J. Kim and B. W. Kim, "Dueling double Q-learning based reinforcement learning approach for the flow shop scheduling problem," *Transactions of the Korean Institute of Electrical Engineers*, vol. 70, no. 10, 2021, doi: 10.5370/KIEE.2021.70.10.1497.

[103] S. Lang, F. Behrendt, N. Lanzerath, T. Reggelin, and M. Muller, "Integration of Deep Reinforcement Learning and Discrete-Event Simulation for Real-Time Scheduling of a Flexible Job Shop Production," in *Proceedings - Winter Simulation Conference*, 2020, vol. 2020-December. doi: 10.1109/WSC48552.2020.9383997.

[104] Y. Zhao, Y. Wang, Y. Tan, J. Zhang, and H. Yu, "Dynamic Jobshop Scheduling Algorithm Based on Deep Q Network," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3110242.

[105] V. Samsonov *et al.*, “Manufacturing control in job shop environments with reinforcement learning,” *ICAART 2021 - Proceedings of the 13th International Conference on Agents and Artificial Intelligence*, vol. 2, pp. 589–597, 2021, doi: 10.5220/0010202405890597.
[106] T. E. Thomas, J. Koo, S. Chaterji, and S. Bagchi, “Minerva: A reinforcement learning-based technique for optimal scheduling and bottleneck detection in distributed factory operations,” in *2018 10th International Conference on Communication Systems and Networks, COMSNETS 2018*, 2018, vol. 2018-January. doi: 10.1109/COMSNETS.2018.8328189.
[107] D. Zeng, J. Zhan, W. Peng, and Z. Zeng, “Evolutionary job scheduling with optimized population by deep reinforcement learning,” *Engineering Optimization*, 2021, doi: 10.1080/0305215X.2021.2013479.
[108] Y. Zhao and H. Zhang, “Application of machine learning and rule scheduling in a job-shop production control system,” *International Journal of Simulation Modelling*, vol. 20, no. 2, 2021, doi: 10.2507/IJSIMM20-2-CO10.
[109] P. C. Luo, H. Q. Xiong, B. W. Zhang, J. Y. Peng, and Z. F. Xiong, “Multi-resource constrained dynamic workshop scheduling based on proximal policy optimisation,” *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1975057.
[110] Y. Kang, S. Lyu, J. Kim, B. Park, and S. Cho, “Dynamic vehicle traffic control using deep reinforcement learning in automated material handling system,” in *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 2019. doi: 10.1609/aaai.v33i01.33019949.
[111] Y. Li, W. Gu, M. Yuan, and Y. Tang, “Real-time data-driven dynamic scheduling for flexible job shop with insufficient transportation resources using hybrid deep Q network,” *Robot Comput Integr Manuf*, vol. 74, 2022, doi: 10.1016/j.rcim.2021.102283.
[112] J. A. Palombarini and E. C. Martinez, “Automatic Generation of Rescheduling Knowledge in Socio-technical Manufacturing Systems using Deep Reinforcement Learning,” in *2018 IEEE Biennial Congress of Argentina, ARGENCON 2018*, 2019. doi: 10.1109/ARGENCON.2018.8646172.
[113] S. Luo, “Dynamic scheduling for flexible job shop with new job insertions by deep reinforcement learning,” *Applied Soft Computing Journal*, vol. 91, 2020, doi: 10.1016/j.asoc.2020.106208.
[114] D. Shi, W. Fan, Y. Xiao, T. Lin, and C. Xing, “Intelligent scheduling of discrete automated production line via deep reinforcement learning,” *Int J Prod Res*, vol. 58, no. 11, 2020, doi: 10.1080/00207543.2020.1717008.
[115] S. Yang and Z. Xu, “Intelligent Scheduling for Permutation Flow Shop with Dynamic Job Arrival via Deep Reinforcement Learning,” in *IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021. doi: 10.1109/IAEAC50856.2021.9390893.
[116] T. Seito and S. Munakata, “Production scheduling based on deep reinforcement learning using graph convolutional neural network,” in *ICAART 2020 - Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, 2020, vol. 2. doi: 10.5220/0009095207660772.
[117] T. Zhou, D. Tang, H. Zhu, and L. Wang, “Reinforcement Learning with Composite Rewards for Production Scheduling in a Smart Factory,” *IEEE Access*, vol. 9, pp. 752–766, 2021, doi: 10.1109/ACCESS.2020.3046784.

[118] W. Liu, S. Wu, H. Zhu, and H. Zhang, “An Integration Method of Heterogeneous Models for Process Scheduling Based on Deep Q-Learning Integration Agent,” in *Proceedings of the 16th IEEE Conference on Industrial Electronics and Applications, ICIEA 2021*, 2021. doi: 10.1109/ICIEA51954.2021.9516381.

[119] L. Wang *et al.*, “Dynamic job-shop scheduling in smart manufacturing using deep reinforcement learning,” *Computer Networks*, vol. 190, 2021, doi: 10.1016/j.comnet.2021.107969.

[120] K. T. Park, S. W. Jeon, and S. do Noh, “Digital twin application with horizontal coordination for reinforcement-learning-based production control in a re-entrant job shop,” *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1884309.

[121] S. Luo, L. Zhang, and Y. Fan, “Real-Time Scheduling for Dynamic Partial-No-Wait Multiobjective Flexible Job Shop by Deep Reinforcement Learning,” *IEEE Transactions on Automation Science and Engineering*, 2021, doi: 10.1109/TASE.2021.3104716.

[122] T. Zhou, D. Tang, H. Zhu, and Z. Zhang, “Multi-agent reinforcement learning for online scheduling in smart factories,” *Robot Comput Integr Manuf*, vol. 72, 2021, doi: 10.1016/j.rcim.2021.102202.

[123] J. A. Palombarini and E. C. Martínez, “End-to-end on-line rescheduling from Gantt chart images using deep reinforcement learning,” *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.2002963.

[124] J. A. Palombarini and E. C. Martinez, “Closed-loop rescheduling using deep reinforcement learning,” in *IFAC-PapersOnLine*, 2019, vol. 52, no. 1. doi: 10.1016/j.ifacol.2019.06.067.

[125] M. Schneckenreither, S. Haeussler, and J. Peiró, “Average reward adjusted deep reinforcement learning for order release planning in manufacturing,” *Knowl Based Syst*, vol. 247, Jul. 2022, doi: 10.1016/J.KNOSYS.2022.108765.

[126] X. Jing, X. Yao, M. Liu, and J. Zhou, “Multi-agent reinforcement learning based on graph convolutional network for flexible job shop scheduling,” *J Intell Manuf*, 2022, doi: 10.1007/S10845-022-02037-5.

[127] C. B. Gil and J. H. Lee, “Deep Reinforcement Learning Approach for Material Scheduling Considering High-Dimensional Environment of Hybrid Flow-Shop Problem,” *Applied Sciences (Switzerland)*, vol. 12, no. 18, Sep. 2022, doi: 10.3390/APP12189332.

[128] L. Liu, K. Guo, Z. Gao, J. Li, and J. Sun, “Digital Twin-Driven Adaptive Scheduling for Flexible Job Shops,” *Sustainability (Switzerland)*, vol. 14, no. 9, May 2022, doi: 10.3390/SU14095340.

[129] R. Liu, R. Piplani, and C. Toro, “Deep reinforcement learning for dynamic scheduling of a flexible job shop,” *Int J Prod Res*, vol. 60, no. 13, pp. 4049–4069, 2022, doi: 10.1080/00207543.2022.2058432.

[130] D. Johnson, G. Chen, and Y. Lu, “Multi-Agent Reinforcement Learning for Real-Time Dynamic Production Scheduling in a Robot Assembly Cell,” *IEEE Robot Autom Lett*, vol. 7, no. 3, pp. 7684–7691, Jul. 2022, doi: 10.1109/LRA.2022.3184795.

[131] M. Zhang, Y. Lu, Y. Hu, N. Amaitik, and Y. Xu, “Dynamic Scheduling Method for Job-Shop Manufacturing Systems by Deep Reinforcement Learning with Proximal Policy Optimization,” *Sustainability (Switzerland)*, vol. 14, no. 9, May 2022, doi: 10.3390/SU14095177.

[132] J. Chang, D. Yu, Y. Hu, W. He, and H. Yu, “Deep Reinforcement Learning for Dynamic Flexible Job Shop Scheduling with Random Job Arrival,” *Processes*, vol. 10, no. 4, Apr. 2022, doi: 10.3390/PR10040760.

[133] Z. Dong, T. Ren, J. Weng, F. Qi, and X. Wang, "Minimizing the Late Work of the Flow Shop Scheduling Problem with a Deep Reinforcement Learning Based Approach," *Applied Sciences (Switzerland)*, vol. 12, no. 5, Mar. 2022, doi: 10.3390/APP12052366.
[134] T. Zhou *et al.*, "Reinforcement learning for online optimization of job-shop scheduling in a smart manufacturing factory," *Advances in Mechanical Engineering*, vol. 14, no. 3, Mar. 2022, doi: 10.1177/16878132221086120.
[135] X. Sun, B. Vogel-Heuser, F. Bi, and W. Shen, "A deep reinforcement learning based approach for dynamic distributed blocking flowshop scheduling with job insertions," *IET Collaborative Intelligent Manufacturing*, vol. 4, no. 3, pp. 166–180, Sep. 2022, doi: 10.1049/CIM2.12060.
[136] X. Chang, X. Jia, S. Fu, H. Hu, and K. Liu, "Digital twin and deep reinforcement learning enabled real-time scheduling for complex product flexible shop-floor," *Proc Inst Mech Eng B J Eng Manuf*, 2022, doi: 10.1177/09544054221121934.
[137] Y. Zhang, H. Zhu, D. Tang, T. Zhou, and Y. Gui, "Dynamic job shop scheduling based on deep reinforcement learning for multi-agent manufacturing systems," *Robot Comput Integr Manuf*, vol. 78, Dec. 2022, doi: 10.1016/J.RCIM.2022.102412.
[138] B. Waschneck *et al.*, "Deep reinforcement learning for semiconductor production scheduling," in *2018 29th Annual SEMI Advanced Semiconductor Manufacturing Conference, ASMC 2018*, 2018. doi: 10.1109/ASMC.2018.8373191.
[139] B. Waschneck *et al.*, "Optimization of global production scheduling with deep reinforcement learning," in *Procedia CIRP*, 2018, vol. 72. doi: 10.1016/j.procir.2018.03.212.
[140] C. F. Chien and Y. bin Lan, "Agent-based approach integrating deep reinforcement learning and hybrid genetic algorithm for dynamic scheduling for Industry 3.5 smart production," *Comput Ind Eng*, vol. 162, 2021, doi: 10.1016/j.cie.2021.107782.
[141] A. H. Sakr, A. Aboelhassan, S. Yacout, and S. Bassetto, "Simulation and deep reinforcement learning for adaptive dispatching in semiconductor manufacturing systems," *J Intell Manuf*, 2021, doi: 10.1007/s10845-021-01851-7.
[142] Y. H. Lee and S. Lee, "Deep reinforcement learning based scheduling within production plan in semiconductor fabrication," *Expert Syst Appl*, vol. 191, 2022, doi: 10.1016/j.eswa.2021.116222.
[143] J. Liu, F. Qiao, and Y. Ma, "Real time production scheduling based on Asynchronous Advanced Actor Critic and composite dispatching rule," in *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, 2020. doi: 10.1109/CAC51589.2020.9327198.
[144] C. Lee and S. Lee, "A Practical Deep Reinforcement Learning Approach to Semiconductor Equipment Scheduling," in *Proceedings of the IEEE International Conference on Industrial Technology*, 2021, vol. 2021-March. doi: 10.1109/ICIT46573.2021.9453533.
[145] C. Hong and T. E. Lee, "Multi-agent Reinforcement Learning Approach for Scheduling Cluster Tools with Condition Based Chamber Cleaning Operations," in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 2019. doi: 10.1109/ICMLA.2018.00143.
[146] J. Wang, P. Gao, P. Zheng, J. Zhang, and W. H. Ip, "A fuzzy hierarchical reinforcement learning based scheduling method for semiconductor wafer manufacturing systems," *J Manuf Syst*, vol. 61, 2021, doi: 10.1016/j.jmsy.2021.08.008.
[147] J. Wangl, J. He, and J. Zhang, "A reinforcement learning method to optimize the priority of product for scheduling the large-scale complex manufacturing systems," in *Proceedings of*

International Conference on Computers and Industrial Engineering, CIE, 2018, vol. 2018-December.

[148] A. Kuhnle, M. C. May, L. Schäfer, and G. Lanza, "Explainable reinforcement learning in production control of job shop manufacturing system," *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1972179.

[149] I. B. Park and J. Park, "Scalable Scheduling of Semiconductor Packaging Facilities Using Deep Reinforcement Learning," *IEEE Trans Cybern*, 2021, doi: 10.1109/TCYB.2021.3128075.

[150] J. Liu, F. Qiao, M. Zou, J. Zinn, Y. Ma, and B. Vogel-Heuser, "Dynamic scheduling for semiconductor manufacturing systems with uncertainties using convolutional neural networks and reinforcement learning," *Complex and Intelligent Systems*, 2022, doi: 10.1007/S40747-022-00844-0.

[151] S. Lee, Y. Cho, and Y. H. Lee, "Injection Mold Production Sustainable Scheduling Using Deep Reinforcement Learning," *Sustainability*, vol. 12, no. 20, p. 8718, 2020, doi: 10.3390/su12208718.

[152] J. Leng, C. Jin, A. Vogl, and H. Liu, "Deep reinforcement learning for a color-batching resequencing problem," *J Manuf Syst*, vol. 56, 2020, doi: 10.1016/j.jmsy.2020.06.001.

[153] T. P. Gros, J. Gros, and V. Wolf, "Real-Time Decision Making for a Car Manufacturing Process Using Deep Reinforcement Learning," in *Proceedings - Winter Simulation Conference*, 2020, vol. 2020-December. doi: 10.1109/WSC48552.2020.9383884.

[154] L. Overbeck, A. Hugues, M. C. May, A. Kuhnle, and G. Lanza, "Reinforcement Learning Based Production Control of Semi-automated Manufacturing Systems," in *Procedia CIRP*, 2021, vol. 103. doi: 10.1016/j.procir.2021.10.027.

[155] J. Leng *et al.*, "A multi-objective reinforcement learning approach for resequencing scheduling problems in automotive manufacturing systems," *Int J Prod Res*, 2022, doi: 10.1080/00207543.2022.2098871.

[156] T. Kohne, H. Ranzau, N. Panten, and M. Weigold, "Comparative study of algorithms for optimized control of industrial energy supply systems," *Energy Informatics*, vol. 3, 2020, doi: 10.1186/s42162-020-00115-7.

[157] X. Huang, S. H. Hong, M. Yu, Y. DIng, and J. Jiang, "Demand Response Management for Industrial Facilities: A Deep Reinforcement Learning Approach," *IEEE Access*, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2924030.

[158] D. Schwung, A. Schwung, and S. X. Ding, "On-line Energy Optimization of Hybrid Production Systems Using Actor-Critic Reinforcement Learning," in *9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications, IS 2018 - Proceedings*, 2018. doi: 10.1109/IS.2018.8710466.

[159] D. Schwung, A. Schwung, and S. X. Ding, "Actor-critic reinforcement learning for energy optimization in hybrid production environments," *International Journal of Computing*, vol. 18, no. 4, 2019, doi: 10.47839/ijc.18.4.1607.

[160] D. Zhu, B. Yang, Y. Liu, Z. Wang, K. Ma, and X. Guan, "Energy management based on multi-agent deep reinforcement learning for a multi-energy industrial park," *Appl Energy*, vol. 311, Apr. 2022, doi: 10.1016/J.APENERGY.2022.118636.

[161] D. Schwung, S. Yuwono, A. Schwung, and S. X. Ding, "Decentralized learning of energy optimal production policies using PLC-informed reinforcement learning," *Comput Chem Eng*, vol. 152, 2021, doi: 10.1016/j.compchemeng.2021.107382.

[162] J. Bakakeu, D. Kisskalt, J. Franke, S. Baer, H. H. Klos, and J. Peschke, “Multi-Agent Reinforcement Learning for the Energy Optimization of Cyber-Physical Production Systems,” in *Canadian Conference on Electrical and Computer Engineering*, 2020, vol. 2020-August. doi: 10.1109/CCECE47787.2020.9255795.

[163] M. Roesch, C. Linder, C. Bruckdorfer, A. Hohmann, and G. Reinhart, “Industrial load management using multi-agent reinforcement learning for rescheduling,” in *Proceedings - 2019 2nd International Conference on Artificial Intelligence for Industries, AI4I 2019*, 2019. doi: 10.1109/AI4I46381.2019.00033.

[164] R. Lu, Y. C. Li, Y. Li, J. Jiang, and Y. Ding, “Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management,” *Appl Energy*, vol. 276, 2020, doi: 10.1016/j.apenergy.2020.115473.

[165] J. Zhao, T. Wang, W. Pedrycz, and W. Wang, “Granular Prediction and Dynamic Scheduling Based on Adaptive Dynamic Programming for the Blast Furnace Gas System,” *IEEE Trans Cybern*, vol. 51, no. 4, 2021, doi: 10.1109/TCYB.2019.2901268.

[166] M. Weigold, H. Ranzau, S. Schaumann, T. Kohne, N. Panten, and E. Abele, “Method for the application of deep reinforcement learning for optimised control of industrial energy supply systems by the example of a central cooling system,” *CIRP Annals*, vol. 70, no. 1, 2021, doi: 10.1016/j.cirp.2021.03.021.

[167] J. Fu, H. Xiao, H. Wang, and J. Zhou, “Control Strategy for Denitrification Efficiency of Coal-Fired Power Plant Based on Deep Reinforcement Learning,” *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2985233.

[168] L. Yi, P. Langlotz, M. Hussong, M. Glatt, F. J. P. Sousa, and J. C. Aurich, “An integrated energy management system using double deep Q-learning and energy storage equipment to reduce energy cost in manufacturing under real-time pricing condition: A case study of scale-model factory,” *CIRP J Manuf Sci Technol*, vol. 38, pp. 844–860, Aug. 2022, doi: 10.1016/J.CIRPJ.2022.07.009.

[169] J. J. Wang and L. Wang, “A Cooperative Memetic Algorithm with Learning-Based Agent for Energy-Aware Distributed Hybrid Flow-Shop Scheduling,” *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 3, pp. 461–475, Jun. 2022, doi: 10.1109/TEVC.2021.3106168.

[170] D. Qiu, Z. Dong, X. Zhang, Y. Wang, and G. Strbac, “Safe reinforcement learning for real-time automatic control in a smart energy-hub,” *Appl Energy*, vol. 309, Mar. 2022, doi: 10.1016/J.APENERGY.2021.118403.

[171] Q. Xiao, C. Li, Y. Tang, and L. Li, “Meta-Reinforcement Learning of Machining Parameters for Energy-Efficient Process Control of Flexible Turning Operations,” *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 1, 2021, doi: 10.1109/TASE.2019.2924444.

[172] J. Huang, J. Zhang, Q. Chang, and R. X. Gao, “Integrated process-system modelling and control through graph neural network and reinforcement learning,” *CIRP Annals*, vol. 70, no. 1, 2021, doi: 10.1016/j.cirp.2021.04.056.

[173] J. Dornheim, L. Morand, S. Zeitvogel, T. Iraki, N. Link, and D. Helm, “Deep reinforcement learning methods for structure-guided processing path optimization,” *J Intell Manuf*, 2021, doi: 10.1007/s10845-021-01805-z.

[174] B. Li, H. Zhang, P. Ye, and J. Wang, “Trajectory smoothing method using reinforcement learning for computer numerical control machine tools,” *Robot Comput Integr Manuf*, vol. 61, 2020, doi: 10.1016/j.rcim.2019.101847.

[175] Y. Zhang, Y. Li, and K. Xu, “Reinforcement learning–based tool orientation optimization for five-axis machining,” *The International Journal of Advanced Manufacturing Technology*, Jan. 2022, doi: 10.1007/s00170-022-08668-5.
[176] V. Samsonov, C. Enslin, H. G. Köpken, S. Baer, and D. Lütticke, “Using reinforcement learning for optimization of a workpiece clamping position in a machine tool,” in *ICEIS 2020 - Proceedings of the 22nd International Conference on Enterprise Information Systems*, 2020, vol. 1. doi: 10.5220/0009354105060514.
[177] J. Schoop, H. A. Poonawala, D. Adeniji, and B. Clark, “AI-enabled dynamic finish machining optimization for sustained surface integrity,” *Manuf Lett*, vol. 29, 2021, doi: 10.1016/j.mfglet.2021.04.002.
[178] R. Gulde, M. Tuscher, A. Csiszar, O. Riedel, and A. Verl, “Reinforcement learning approach to vibration compensation for dynamic feed drive systems,” in *Proceedings - 2019 2nd International Conference on Artificial Intelligence for Industries, AI4I 2019*, 2019. doi: 10.1109/AI4I46381.2019.00015.
[179] Y. Jiang, J. Chen, H. Zhou, J. Yang, P. Hu, and J. Wang, “Contour error modeling and compensation of CNC machining based on deep learning and reinforcement learning,” *International Journal of Advanced Manufacturing Technology*, 2021, doi: 10.1007/s00170-021-07895-6.
[180] Y. Xie, M. Praeger, J. A. Grant-Jacob, R. W. Eason, and B. Mills, “Motion control for laser machining via reinforcement learning,” *Opt Express*, vol. 30, no. 12, p. 20963, Jun. 2022, doi: 10.1364/OE.454793.
[181] Z. Wang, J. Lu, C. Chen, J. Ma, and X. Liao, “Investigating the multi-objective optimization of quality and efficiency using deep reinforcement learning,” *Applied Intelligence*, vol. 52, no. 11, pp. 12873–12887, Sep. 2022, doi: 10.1007/S10489-022-03326-5.
[182] Z. Jin, H. Li, and H. Gao, “An intelligent weld control strategy based on reinforcement learning approach,” *International Journal of Advanced Manufacturing Technology*, vol. 100, no. 9–12, 2019, doi: 10.1007/s00170-018-2864-2.
[183] J. Günther, P. M. Pilarski, G. Helfrich, H. Shen, and K. Diepold, “Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning,” *Mechatronics*, vol. 34, 2016, doi: 10.1016/j.mechatronics.2015.09.004.
[184] N. Khader and S. W. Yoon, “Adaptive optimal control of stencil printing process using reinforcement learning,” *Robot Comput Integr Manuf*, vol. 71, 2021, doi: 10.1016/j.rcim.2021.102132.
[185] S. Patrick, A. Nycz, and M. Noakes, “Reinforcement learning for generating toolpaths in additive manufacturing,” in *Solid Freeform Fabrication 2018: Proceedings of the 29th Annual International Solid Freeform Fabrication Symposium - An Additive Manufacturing Conference, SFF 2018*, 2020.
[186] F. Ogoke and A. B. Farimani, “Thermal control of laser powder bed fusion using deep reinforcement learning,” *Addit Manuf*, vol. 46, 2021, doi: 10.1016/j.addma.2021.102033.
[187] J. Yu and P. Guo, “Run-to-Run Control of Chemical Mechanical Polishing Process Based on Deep Reinforcement Learning,” *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 3, 2020, doi: 10.1109/TSM.2020.3002896.
[188] J. Zinn, B. Vogel-Heuser, and M. Gruber, “Fault-Tolerant control of programmable logic controller-based production systems with deep reinforcement learning,” *Journal of*

Mechanical Design, Transactions of the ASME, vol. 143, no. 7, 2021, doi: 10.1115/1.4050624.
[189] C. Cronrath, A. R. Aderiani, and B. Lennartson, "Enhancing digital twins through reinforcement learning," in *IEEE International Conference on Automation Science and Engineering*, 2019, vol. 2019-August. doi: 10.1109/COASE.2019.8842888.
[190] J. Dornheim, N. Link, and P. Gumbsch, "Model-free Adaptive Optimal Control of Episodic Fixed-horizon Manufacturing Processes Using Reinforcement Learning," *Int J Control Autom Syst*, vol. 18, no. 6, 2020, doi: 10.1007/s12555-019-0120-7.
[191] N. Reinisch, F. Rudolph, S. Günther, D. Bailly, and G. Hirt, "Successful pass schedule design in open-die forging using double deep Q-learning," *Processes*, vol. 9, no. 7, 2021, doi: 10.3390/pr9071084.
[192] C. el Mazgualdi, T. Masrour, I. el Hassani, and A. Khdoudi, "A deep reinforcement learning (drl) decision model for heating process parameters identification in automotive glass manufacturing," in *Advances in Intelligent Systems and Computing*, 2021, vol. 1193. doi: 10.1007/978-3-030-51186-9_6.
[193] F. Guo, X. Zhou, J. Liu, Y. Zhang, D. Li, and H. Zhou, "A reinforcement learning decision model for online process parameters optimization from offline data in injection molding," *Applied Soft Computing Journal*, vol. 85, 2019, doi: 10.1016/j.asoc.2019.105828.
[194] C. Zimmerling, C. Poppe, and L. Kärger, "Estimating optimum process parameters in textile draping of variable part geometries - A reinforcement learning approach," in *Procedia Manufacturing*, 2020, vol. 47. doi: 10.1016/j.promfg.2020.04.263.
[195] O. Gamal, M. I. P. Mohamed, C. G. Patel, and H. Roth, "Data-Driven Model-Free Intelligent Roll Gap Control of Bar and Wire Hot Rolling Process Using Reinforcement Learning," *International Journal of Mechanical Engineering and Robotics Research*, vol. 10, no. 7, 2021, doi: 10.18178/ijmerr.10.7.349-356.
[196] S. Kim, D. D. Kim, and B. Anthony, "Dynamic Control of a Fiber Manufacturing Process using Deep Reinforcement Learning," *IEEE/ASME Transactions on Mechatronics*, 2021, doi: 10.1109/TMECH.2021.3070973.
[197] T. Wu, H. Zhao, B. Gao, and F. Meng, "Energy-Saving for a Velocity Control System of a Pipe Isolation Tool Based on a Reinforcement Learning Method," *International Journal of Precision Engineering and Manufacturing - Green Technology*, 2021, doi: 10.1007/s40684-021-00309-8.
[198] C. Zirngibl, F. Dworschak, B. Schleich, and S. Wartzack, "Application of reinforcement learning for the optimization of clinch joint characteristics," *Production Engineering*, 2021, doi: 10.1007/s11740-021-01098-4.
[199] J. Deng, S. Sierla, J. Sun, and V. Vyatkin, "Reinforcement learning for industrial process control: A case study in flatness control in steel industry," *Comput Ind*, vol. 143, Dec. 2022, doi: 10.1016/J.COMPIND.2022.103748.
[200] E. Jorge *et al.*, "Reinforcement learning in real-time geometry assurance," in *Procedia CIRP*, 2018, vol. 72. doi: 10.1016/j.procir.2018.03.168.
[201] T. Brito, J. Queiroz, L. Piardi, L. A. Fernandes, J. Lima, and P. Leitão, "A machine learning approach for collaborative robot smart manufacturing inspection for quality control systems," in *Procedia Manufacturing*, 2020, vol. 51. doi: 10.1016/j.promfg.2020.10.003.
[202] Z. Lončarević *et al.*, "Specifying and optimizing robotic motion for visual quality inspection," *Robot Comput Integr Manuf*, vol. 72, 2021, doi: 10.1016/j.rcim.2021.102200.

[203] C. Landgraf, B. Meese, M. Pabst, G. Martius, and M. F. Huber, “A reinforcement learning approach to view planning for automated inspection tasks,” *Sensors*, vol. 21, no. 6, 2021, doi: 10.3390/s21062030.
[204] W. Luo, J. Zhang, P. Feng, H. Liu, D. Yu, and Z. Wu, “An adaptive adjustment strategy for bolt posture errors based on an improved reinforcement learning algorithm,” *Applied Intelligence*, vol. 51, no. 6, 2021, doi: 10.1007/s10489-020-01906-x.
[205] C. K. Cheng and H. Y. Tsai, “Enhanced detection of diverse defects by developing lighting strategies using multiple light sources based on reinforcement learning,” *J Intell Manuf*, 2021, doi: 10.1007/s10845-021-01800-4.
[206] H. Shi, J. Li, M. Liang, M. Hwang, K. S. Hwang, and Y. Y. Hsu, “Path Planning of Randomly Scattering Waypoints for Wafer Probing Based on Deep Attention Mechanism,” *IEEE Trans Syst Man Cybern Syst*, 2022, doi: 10.1109/TSMC.2022.3184155.
[207] M. Lutter, D. Clever, R. Kirsten, K. Listmann, and J. Peters, “Building Skill Learning Systems for Robotics,” in *IEEE International Conference on Automation Science and Engineering*, 2021, vol. 2021-August. doi: 10.1109/CASE49439.2021.9551562.
[208] M. Hebecker, J. Lambrecht, and M. Schmitz, “Towards real-world force-sensitive robotic assembly through deep reinforcement learning in simulations,” in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 2021, vol. 2021-July. doi: 10.1109/AIM46487.2021.9517356.
[209] X. Lan, Y. Qiao, and B. Lee, “Towards Pick and Place Multi Robot Coordination Using Multi-agent Deep Reinforcement Learning,” in *2021 International Conference on Automation, Robotics and Applications, ICARA 2021*, 2021. doi: 10.1109/ICARA51699.2021.9376433.
[210] Q. Chen, B. Heydari, and M. Moghaddam, “Leveraging task modularity in reinforcement learning for adaptable industry 4.0 automation,” *Journal of Mechanical Design, Transactions of the ASME*, vol. 143, no. 7, 2021, doi: 10.1115/1.4049531.
[211] M. Moosmann *et al.*, “Separating Entangled Workpieces in Random Bin Picking using Deep Reinforcement Learning,” in *Procedia CIRP*, 2021, vol. 104. doi: 10.1016/j.procir.2021.11.148.
[212] T. Zhang, M. Xiao, Y. biao Zou, J. dong Xiao, and S. yan Chen, “Robotic Curved Surface Tracking with a Neural Network for Angle Identification and Constant Force Control based on Reinforcement Learning,” *International Journal of Precision Engineering and Manufacturing*, vol. 21, no. 5, 2020, doi: 10.1007/s12541-020-00315-x.
[213] T. Zhang, M. Xiao, Y. Zou, and J. Xiao, “Robotic constant-force grinding control with a press-and-release model and model-based reinforcement learning,” *International Journal of Advanced Manufacturing Technology*, vol. 106, no. 1–2, 2020, doi: 10.1007/s00170-019-04614-0.
[214] L. Liang, Y. Chen, L. Liao, H. Sun, and Y. Liu, “A novel impedance control method of rubber unstacking robot dealing with unpredictable and time-variable adhesion force,” *Robot Comput Integr Manuf*, vol. 67, 2021, doi: 10.1016/j.rcim.2020.102038.
[215] Y. T. Tsai *et al.*, “Utilization of a reinforcement learning algorithm for the accurate alignment of a robotic arm in a complete soft fabric shoe tongues automation process,” *J Manuf Syst*, vol. 56, 2020, doi: 10.1016/j.jmsy.2020.07.001.
[216] W. Li, D. Chen, J. Dai, and J. Le, “The Study of a Textile Punching Robot Based on Combined Deep Reinforcement Learning,” in *International Conference on Cloud*

Computing, Big Data and Blockchain, ICCBB 2018, 2018. doi: 10.1109/ICCBB.2018.8756483.
[217] G. Thomas, M. Chien, A. Tamar, J. A. Ojea, and P. Abbeel, "Learning Robotic Assembly from CAD," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2018. doi: 10.1109/ICRA.2018.8460696.
[218] W. Luo, J. Zhang, P. Feng, D. Yu, and Z. Wu, "A deep transfer-learning-based dynamic reinforcement learning for intelligent tightening system," *International Journal of Intelligent Systems*, vol. 36, no. 3, 2021, doi: 10.1002/int.22345.
[219] A. Maldonado-Ramirez, R. Rios-Cabrera, and I. Lopez-Juarez, "A visual path-following learning approach for industrial robots using DRL," *Robot Comput Integr Manuf*, vol. 71, 2021, doi: 10.1016/j.rcim.2021.102130.
[220] M. Hildebrand, R. S. Andersen, and S. Bogh, "Deep reinforcement learning for robot batching optimization and flow control," in *Procedia Manufacturing*, 2020, vol. 51. doi: 10.1016/j.promfg.2020.10.203.
[221] L. Leyendecker, M. Schmitz, H. A. Zhou, V. Samsonov, M. Rittstieg, and D. Lütticke, "Deep Reinforcement Learning for Robotic Control in High-Dexterity Assembly Tasks - A Reward Curriculum Approach," *Int J Semant Comput*, vol. 16, no. 3, pp. 381–402, Sep. 2022, doi: 10.1142/S1793351X22430024.
[222] Q. Yang, J. A. Stork, and T. Stoyanov, "MPR-RL: Multi-Prior Regularized Reinforcement Learning for Knowledge Transfer," *IEEE Robot Autom Lett*, vol. 7, no. 3, pp. 7652–7659, Jul. 2022, doi: 10.1109/LRA.2022.3184805.
[223] R. Zeng, M. Liu, J. Zhang, X. Li, Q. Zhou, and Y. Jiang, "Manipulator Control Method Based on Deep Reinforcement Learning," in *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020*, 2020. doi: 10.1109/CCDC49329.2020.9164440.
[224] Y. P. Pane, S. P. Nageshrao, J. Kober, and R. Babuška, "Reinforcement learning based compensation methods for robot manipulators," *Eng Appl Artif Intell*, vol. 78, 2019, doi: 10.1016/j.engappai.2018.11.006.
[225] R. Meyes, C. Scheiderer, and T. Meisen, "Continuous Motion Planning for Industrial Robots based on Direct Sensory Input," in *Procedia CIRP*, 2018, vol. 72. doi: 10.1016/j.procir.2018.03.067.
[226] M. Matulis and C. Harvey, "A robot arm digital twin utilising reinforcement learning," *Computers and Graphics (Pergamon)*, vol. 95, 2021, doi: 10.1016/j.cag.2021.01.011.
[227] C. Li, P. Zheng, S. Li, Y. Pang, and C. K. M. Lee, "AR-assisted digital twin-enabled robot collaborative manufacturing system with human-in-the-loop," *Robot Comput Integr Manuf*, vol. 76, Aug. 2022, doi: 10.1016/J.RCIM.2022.102321.
[228] C. Li, P. Zheng, Y. Yin, Y. M. Pang, and S. Huo, "An AR-assisted Deep Reinforcement Learning-based approach towards mutual-cognitive safe human-robot interaction," *Robot Comput Integr Manuf*, vol. 80, p. 102471, Apr. 2023, doi: 10.1016/J.RCIM.2022.102471.
[229] M. S. Kim, D. K. Han, J. H. Park, and J. S. Kim, "Motion planning of robot manipulators for a smoother path using a twin delayed deep deterministic policy gradient with hindsight experience replay," *Applied Sciences (Switzerland)*, vol. 10, no. 2, 2020, doi: 10.3390/app10020575.
[230] X. Lu, Y. Chen, and Z. Yuan, "A full freedom pose measurement method for industrial robot based on reinforcement learning algorithm," *Soft comput*, vol. 25, no. 20, 2021, doi: 10.1007/s00500-021-06190-6.

[231] X. Hua, G. Wang, J. Xu, and K. Chen, “Reinforcement learning-based collision-free path planner for redundant robot in narrow duct,” *J Intell Manuf*, vol. 32, no. 2, 2021, doi: 10.1007/s10845-020-01582-1.

[232] P. Zheng, L. Xia, C. Li, X. Li, and B. Liu, “Towards Self-X cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach,” *J Manuf Syst*, vol. 61, no. April, pp. 16–26, 2021, doi: 10.1016/j.jmsy.2021.08.002.

[233] T. Inoue, G. de Magistris, A. Munawar, T. Yokoya, and R. Tachibana, “Deep reinforcement learning for high precision assembly tasks,” in *IEEE International Conference on Intelligent Robots and Systems*, 2017. doi: 10.1109/IROS.2017.8202244.

[234] F. Aschersleben, R. Griemert, F. Gabriel, and K. Dröder, “Reinforcement learning for robotic assembly of fuel cell turbocharger parts with tight tolerances,” *PRODUCTION ENGINEERING-RESEARCH AND DEVELOPMENT*, 2020, doi: 10.1007/s11740-020-00968-7.

[235] Y. L. Kim, K. H. Ahn, and J. B. Song, “Reinforcement learning based on movement primitives for contact tasks,” *Robot Comput Integr Manuf*, vol. 62, 2020, doi: 10.1016/j.rcim.2019.101863.

[236] N. J. Cho, S. H. Lee, J. B. Kim, and I. H. Suh, “Learning, improving, and generalizing motor skills for the peg-in-hole tasks based on imitation learning and self-learning,” *Applied Sciences (Switzerland)*, 2020, doi: 10.3390/APP10082719.

[237] J. Luo* *et al.*, “Robust Multi-Modal Policies for Industrial Assembly via Reinforcement Learning and Demonstrations: A Large-Scale Study,” 2021. doi: 10.15607/rss.2021.xvii.088.

[238] G. Schoettler *et al.*, “Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards,” in *IEEE International Conference on Intelligent Robots and Systems*, 2020. doi: 10.1109/IROS45743.2020.9341714.

[239] X. Zhao, H. Zhao, P. Chen, and H. Ding, “Model accelerated reinforcement learning for high precision robotic assembly,” *Int J Intell Robot Appl*, 2020, doi: 10.1007/s41315-020-00138-z.

[240] C. C. Beltran-Hernandez, D. Petit, I. G. Ramirez-Alpizar, and K. Harada, “Variable compliance control for robotic peg-in-hole assembly: A deep-reinforcement-learning approach,” *Applied Sciences (Switzerland)*, 2020, doi: 10.3390/app10196923.

[241] J. Li, D. Pang, Y. Zheng, X. Guan, and X. Le, “A flexible manufacturing assembly system with deep reinforcement learning,” *Control Eng Pract*, vol. 118, 2022, doi: 10.1016/j.conengprac.2021.104957.

[242] J. Li, D. Pang, Y. Zheng, and X. Le, “Digital Twin Enhanced Assembly Based on Deep Reinforcement Learning,” in *2021 11th International Conference on Information Science and Technology, ICIST 2021*, 2021. doi: 10.1109/ICIST52614.2021.9440555.

[243] M. Vecerik, O. Sushkov, D. Barker, T. Rothorl, T. Hester, and J. Scholz, “A practical approach to insertion with variable socket position using deep reinforcement learning,” in *Proceedings - IEEE International Conference on Robotics and Automation*, 2019, vol. 2019-May. doi: 10.1109/ICRA.2019.8794074.

[244] Q. Tan, Y. Tong, S. Wu, and D. Li, “Modeling, planning, and scheduling of shop-floor assembly process with dynamic cyber-physical interactions: a case study for CPS-based smart industrial robot production,” *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 9, 2019, doi: 10.1007/s00170-019-03940-7.

[245] K. Arviv, H. Stern, and Y. Edan, “Collaborative reinforcement learning for a two-robot job transfer flow-shop scheduling problem,” *Int J Prod Res*, vol. 54, no. 4, 2016, doi: 10.1080/00207543.2015.1057297.
[246] A. Schwung, D. Schwung, and M. S. Abdul Hameed, “Cooperative robot control in flexible manufacturing cells: Centralized vs. distributed approaches,” in *IEEE International Conference on Industrial Informatics (INDIN)*, 2019, vol. 2019-July. doi: 10.1109/INDIN41052.2019.8972060.
[247] A. Agrawal, S. J. Won, T. Sharma, M. Deshpande, and C. McComb, “A multi-agent reinforcement learning framework for intelligent manufacturing with autonomous mobile robots,” in *Proceedings of the Design Society*, 2021, vol. 1. doi: 10.1017/pds.2021.17.
[248] A. Malus, D. Kozjek, and R. Vrabič, “Real-time order dispatching for a fleet of autonomous mobile robots using multi-agent reinforcement learning,” *CIRP Annals*, vol. 69, no. 1, 2020, doi: 10.1016/j.cirp.2020.04.001.
[249] M. S. Abdul Hameed, M. M. Khan, and A. Schwung, “Curiosity Based RL on Robot Manufacturing Cell,” in *Proceedings of the IEEE International Conference on Industrial Technology*, 2021, vol. 2021-March. doi: 10.1109/ICIT46573.2021.9453577.
[250] K. Chang, S. H. Park, and J. G. Baek, “AGV dispatching algorithm based on deep Q-network in CNC machines environment,” *Int J Comput Integr Manuf*, 2021, doi: 10.1080/0951192X.2021.1992669.
[251] X. Wang, L. Zhang, T. Lin, C. Zhao, K. Wang, and Z. Chen, “Solving job scheduling problems in a resource preemption environment with multi-agent reinforcement learning,” *Robot Comput Integr Manuf*, vol. 77, Oct. 2022, doi: 10.1016/J.RCIM.2022.102324.
[252] K. Bhatta, J. Huang, and Q. Chang, “Dynamic Robot Assignment for Flexible Serial Production Systems,” *IEEE Robot Autom Lett*, vol. 7, no. 3, pp. 7303–7310, Jul. 2022, doi: 10.1109/LRA.2022.3182822.
[253] H. Du, W. Xu, B. Yao, Z. Zhou, and Y. Hu, “Collaborative optimization of service scheduling for industrial cloud robotics based on knowledge sharing,” in *Procedia CIRP*, 2019, vol. 83. doi: 10.1016/j.procir.2019.03.142.
[254] Y. Liu *et al.*, “A framework for industrial robot training in cloud manufacturing with deep reinforcement learning,” in *ASME 2020 15th International Manufacturing Science and Engineering Conference, MSEC 2020*, 2020, vol. 2. doi: 10.1115/MSEC2020-8355.
[255] B. Xiong, Q. Liu, W. Xu, B. Yao, Z. Liu, and Z. Zhou, “Deep reinforcement learning-based safe interaction for industrial human-robot collaboration,” in *Proceedings of International Conference on Computers and Industrial Engineering, CIE*, 2019, vol. 2019-October.
[256] X. Zhu, Y. Liang, H. Sun, X. Wang, and B. Ren, “Robot obstacle avoidance system using deep reinforcement learning,” *Industrial Robot*, 2021, doi: 10.1108/IR-06-2021-0127.
[257] Q. Liu, Z. Liu, B. Xiong, W. Xu, and Y. Liu, “Deep reinforcement learning-based safe interaction for industrial human-robot collaboration using intrinsic reward function,” *Advanced Engineering Informatics*, vol. 49, 2021, doi: 10.1016/j.aei.2021.101360.
[258] A. Terra, H. Riaz, K. Raizer, A. Hata, and R. Inam, “Safety vs. Efficiency: AI-Based Risk Mitigation in Collaborative Robotics,” in *2020 6th International Conference on Control, Automation and Robotics, ICCAR 2020*, 2020. doi: 10.1109/ICCAR49639.2020.9108037.
[259] Z. Liu, Q. Liu, L. Wang, W. Xu, and Z. Zhou, “Task-level decision-making for dynamic and stochastic human-robot collaboration based on dual agents deep reinforcement learning,” *International Journal of Advanced Manufacturing Technology*, vol. 115, no. 11–12, 2021, doi: 10.1007/s00170-021-07265-2.

[260] Y. Meng, J. Su, and J. Wu, “Reinforcement learning based variable impedance control for high precision human-robot collaboration tasks,” in *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics, ICARM 2021*, 2021. doi: 10.1109/ICARM52023.2021.9536100.
[261] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, “Facilitating Human-Robot Collaborative Tasks by Teaching-Learning-Collaboration from Human Demonstrations,” *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, 2019, doi: 10.1109/TASE.2018.2840345.
[262] T. Yu, J. Huang, and Q. Chang, “Optimizing task scheduling in human-robot collaboration with deep multi-agent reinforcement learning,” *J Manuf Syst*, vol. 60, 2021, doi: 10.1016/j.jmsy.2021.07.015.
[263] T. Yu, J. Huang, and Q. Chang, “MASTERING the WORKING SEQUENCE in HUMAN-ROBOT COLLABORATIVE ASSEMBLY BASED on REINFORCEMENT LEARNING,” *arXiv*. 2020. doi: 10.1109/access.2020.3021904.
[264] R. Zhang, Q. Lv, J. Li, J. Bao, T. Liu, and S. Liu, “A reinforcement learning method for human-robot collaboration in assembly tasks,” *Robot Comput Integr Manuf*, vol. 73, 2022, doi: 10.1016/j.rcim.2021.102227.
[265] Q. Lv *et al.*, “A strategy transfer approach for intelligent human-robot collaborative assembly,” *Comput Ind Eng*, vol. 168, Jun. 2022, doi: 10.1016/J.CIE.2022.108047.
[266] R. Zhang, J. Lv, J. Li, J. Bao, P. Zheng, and T. Peng, “A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations,” *J Manuf Syst*, vol. 63, pp. 491–503, Apr. 2022, doi: 10.1016/J.JMSY.2022.05.006.
[267] H. Oliff, Y. Liu, M. Kumar, M. Williams, and M. Ryan, “Reinforcement learning for facilitating human-robot-interaction in manufacturing,” *J Manuf Syst*, vol. 56, pp. 326–340, Jul. 2020, doi: 10.1016/J.JMSY.2020.06.018.
[268] R. S. Alonso, “Deep tech and artificial intelligence for worker safety in robotic manufacturing environments,” in *Advances in Intelligent Systems and Computing*, 2021, vol. 1242 AISC. doi: 10.1007/978-3-030-53829-3_27.
[269] T. Zhang, H. Sun, and Y. Zou, “An electromyography signals-based human-robot collaboration system for human motion intention recognition and realization,” *Robot Comput Integr Manuf*, vol. 77, Oct. 2022, doi: 10.1016/J.RCIM.2022.102359.
[270] M. A. Dittrich and S. Fohlmeister, “A deep q-learning-based optimization of the inventory control in a linear process chain,” *Production Engineering*, vol. 15, no. 1, 2021, doi: 10.1007/s11740-020-01000-8.
[271] H. D. Perez, C. D. Hubbs, C. Li, and I. E. Grossmann, “Algorithmic approaches to inventory management optimization,” *Processes*, vol. 9, no. 1, 2021, doi: 10.3390/pr9010102.
[272] T. A. Zwaida, C. Pham, and Y. Beauregard, “Optimization of inventory management to prevent drug shortages in the hospital supply chain,” *Applied Sciences (Switzerland)*, vol. 11, no. 6, 2021, doi: 10.3390/app11062726.
[273] C. F. Chien, Y. S. Lin, and S. K. Lin, “Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor,” *Int J Prod Res*, vol. 58, no. 9, 2020, doi: 10.1080/00207543.2020.1733125.
[274] Z. Peng, Y. Zhang, Y. Feng, T. Zhang, Z. Wu, and H. Su, “Deep Reinforcement Learning Approach for Capacitated Supply Chain optimization under Demand Uncertainty,” in

Proceedings - 2019 Chinese Automation Congress, CAC 2019, 2019. doi: 10.1109/CAC48633.2019.8997498.

[275] F. E. Achamrah, F. Riane, and S. Limbourg, "Solving inventory routing with transshipment and substitution under dynamic and stochastic demands using genetic algorithm and deep reinforcement learning," *Int J Prod Res*, 2021, doi: 10.1080/00207543.2021.1987549.

[276] Y. Niu, F. Schulte, and R. R. Negenborn, "Human Aspects in Collaborative Order Picking - Letting Robotic Agents Learn about Human Discomfort," in *Procedia Computer Science*, 2021, vol. 180. doi: 10.1016/j.procs.2021.01.338.

[277] A. Oroojlooyjadid, M. Nazari, L. v. Snyder, and M. Takáč, "A Deep Q-Network for the Beer Game: Deep Reinforcement Learning for Inventory Optimization," *Manufacturing & Service Operations Management*, 2021, doi: 10.1287/msom.2020.0939.

[278] H. Lee, J. Hong, and J. Jeong, "MARL-Based Dual Reward Model on Segmented Actions for Multiple Mobile Robots in Automated Warehouse Environment," *Applied Sciences (Switzerland)*, vol. 12, no. 9, May 2022, doi: 10.3390/APP12094703.

[279] L. van Hezewijk, N. Dellaert, T. van Woensel, and N. Gademann, "Using the proximal policy optimisation algorithm for solving the stochastic capacitated lot sizing problem," *Int J Prod Res*, 2022, doi: 10.1080/00207543.2022.2056540.

[280] S. Hansuwa, M. R. Velayudhan Kumar, and R. Chandrasekharan, "Analysis of box and ellipsoidal robust optimization, and attention model based reinforcement learning for a robust vehicle routing problem," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 47, no. 2, Jun. 2022, doi: 10.1007/S12046-022-01833-2.

[281] J. Qi *et al.*, "Research on a collaboration model of green closed-loop supply chains towards intelligent manufacturing," *Multimed Tools Appl*, 2022, doi: 10.1007/S11042-021-11727-W.

[282] M. Wurster, M. Michel, M. C. May, A. Kuhnle, N. Stricker, and G. Lanza, "Modelling and condition-based control of a flexible and hybrid disassembly system with manual and autonomous workstations using reinforcement learning," *J Intell Manuf*, vol. 33, no. 2, pp. 575–591, Feb. 2022, doi: 10.1007/s10845-021-01863-3.

[283] H. Mao, Z. Liu, and C. Qiu, "Adaptive disassembly sequence planning for VR maintenance training via deep reinforcement learning," *International Journal of Advanced Manufacturing Technology*, 2021, doi: 10.1007/s00170-021-08290-x.

[284] J. Huang, Q. Chang, and J. Arinez, "Deep reinforcement learning based preventive maintenance policy for serial production lines," *Expert Syst Appl*, vol. 160, p. 113701, 2020, doi: 10.1016/j.eswa.2020.113701.

[285] A. Kuhnle, J. Jakubik, and G. Lanza, "Reinforcement learning for opportunistic maintenance optimization," *Production Engineering*, vol. 13, no. 1, 2019, doi: 10.1007/s11740-018-0855-7.

[286] J. Su, J. Huang, S. Adams, Q. Chang, and P. A. Beling, "Deep multi-agent reinforcement learning for multi-level preventive maintenance in manufacturing systems[Formula presented]," *Expert Syst Appl*, vol. 192, Apr. 2022, doi: 10.1016/j.eswa.2021.116323.

[287] Q. Yan, H. Wang, and F. Wu, "Digital twin-enabled dynamic scheduling with preventive maintenance using a double-layer Q-learning algorithm," *Comput Oper Res*, vol. 144, Aug. 2022, doi: 10.1016/J.COR.2022.105823.

[288] V. T. Nguyen, P. Do, A. Vosin, and B. Iung, "Artificial-intelligence-based maintenance decision-making and optimization for multi-state component systems," *Reliab Eng Syst Saf*, vol. 228, Dec. 2022, doi: 10.1016/J.RESS.2022.108757.

[289] Q. Yan, W. Wu, and H. Wang, "Deep Reinforcement Learning for Distributed Flow Shop Scheduling with Flexible Maintenance," *Machines*, vol. 10, no. 3, Mar. 2022, doi: 10.3390/MACHINES10030210.
[290] A. Valet, T. Altenmüller, B. Waschneck, M. C. May, A. Kuhnle, and G. Lanza, "Opportunistic maintenance scheduling with deep reinforcement learning," *J Manuf Syst*, vol. 64, pp. 518–534, Jul. 2022, doi: 10.1016/J.JMSY.2022.07.016.
[291] K. S. Hoong Ong, D. Niyato, and C. Yuen, "Predictive Maintenance for Edge-Based Sensor Networks: A Deep Reinforcement Learning Approach," in *IEEE World Forum on Internet of Things, WF-IoT 2020 - Symposium Proceedings*, 2020. doi: 10.1109/WF-IoT48130.2020.9221098.
[292] S. S. Rabbanian, M. Nemati, and G. M. Knapp, "A Deep Reinforcement Learning Approach for Maintenance Planning," in *IIE Annual Conference. Proceedings*, 2021, pp. 932–937.
[293] Z. Wang and J. Xuan, "Intelligent fault recognition framework by using deep reinforcement learning with one dimension convolution and improved actor-critic algorithm," *Advanced Engineering Informatics*, vol. 49, 2021, doi: 10.1016/j.aei.2021.101315.
[294] J. Yao, B. Lu, and J. Zhang, "Tool remaining useful life prediction using deep transfer reinforcement learning based on long short-term memory networks," *International Journal of Advanced Manufacturing Technology*, 2021, doi: 10.1007/s00170-021-07950-2.
[295] D. Y. Liao *et al.*, "Recurrent Reinforcement Learning for Predictive Overall Equipment Effectiveness," in *e-Manufacturing and Design Collaboration Symposium 2018, eMDC 2018 - Proceedings*, 2018.
[296] S. Verma, H. S. Nair, G. Agarwal, J. Dhar, and A. Shukla, "Deep reinforcement learning for single-shot diagnosis and adaptation in damaged robots," in *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 2020. doi: 10.1145/3371158.3371168.
[297] B. I. Epureanu, X. Li, A. Nassehi, and Y. Koren, "Self-repair of smart manufacturing systems by deep reinforcement learning," *CIRP Annals*, 2020, doi: 10.1016/j.cirp.2020.04.008.
[298] Y. Qin, C. Zhao, and F. Gao, "An intelligent non-optimality self-recovery method based on reinforcement learning with small data in big data era," *Chemometrics and Intelligent Laboratory Systems*, vol. 176, 2018, doi: 10.1016/j.chemolab.2018.03.010.
[299] X. Wang, G. Zhang, Y. Li, and N. Qu, "A heuristically accelerated reinforcement learning method for maintenance policy of an assembly line," *Journal of Industrial and Management Optimization*, vol. 0, no. 0, p. 0, 2022, doi: 10.3934/JIMO.2022047.
[300] M. L. Ruiz Rodríguez, S. Kubler, A. de Giorgio, M. Cordy, J. Robert, and Y. le Traon, "Multi-agent deep reinforcement learning based Predictive Maintenance on parallel machines," *Robot Comput Integr Manuf*, vol. 78, Dec. 2022, doi: 10.1016/J.RCIM.2022.102406.
[301] C. Liu *et al.*, "Probing an intelligent predictive maintenance approach with deep learning and augmented reality for machine tools in IoT-enabled manufacturing," *Robot Comput Integr Manuf*, vol. 77, Oct. 2022, doi: 10.1016/J.RCIM.2022.102357.
[302] Z. Wang, J. Xuan, and T. Shi, "Alternative multi-label imitation learning framework monitoring tool wear and bearing fault under different working conditions," *Advanced Engineering Informatics*, vol. 54, Oct. 2022, doi: 10.1016/J.AEI.2022.101749.
[303] V. Vidyadhar, R. Nagaraj, and D. v. Ashoka, "NetAI-Gym: Customized Environment for Network to Evaluate Agent Algorithm using Reinforcement Learning in Open-AI Gym

Platform," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021, doi: 10.14569/IJACSA.2021.0120423.

[304] L. Mönch, J. W. Fowler, and S. J. Mason, "Production Planning and Control for Semiconductor Wafer Fabrication Facilities," vol. 52, 2013, doi: 10.1007/978-1-4614-4472-5.

[305] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *34th International Conference on Machine Learning, ICML 2017*, 2017, vol. 3.

[306] S. Pateria, B. Subagdja, A. H. Tan, and C. Quek, "Hierarchical Reinforcement Learning: A Comprehensive Survey," *ACM Computing Surveys*, vol. 54, no. 5. 2021. doi: 10.1145/3453160.

[307] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum learning for reinforcement learning domains: A framework and survey," *Journal of Machine Learning Research*, vol. 21, 2020.

[308] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey," in *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, 2020. doi: 10.1109/SSCI47803.2020.9308468.

[309] C. Li, P. Zheng, S. Li, Y. Pang, and C. K. M. Lee, "AR-assisted digital twin-enabled robot collaborative manufacturing system with human-in-the-loop," *Robot Comput Integr Manuf*, vol. 76, p. 102321, Aug. 2022, doi: 10.1016/J.RCIM.2022.102321.

[310] T. Yang, "Exploration in Deep Reinforcement Learning: A Comprehensive Survey.," *CoRR*, vol. abs/2109.06668, 2021, Accessed: Jun. 27, 2022. [Online]. Available: https://arxiv.org/abs/2109.06668

[311] K. Srinivasan, B. Eysenbach, S. Ha, J. Tan, and C. Finn, "Learning to be Safe: Deep RL with a Safety Critic," Oct. 2020.

[312] S. Levine, A. Kumar, G. Tucker, and J. Fu, "Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems," May 2020.

[313] B. Wang, P. Zheng, Y. Yin, A. Shih, and L. Wang, "Toward human-centric smart manufacturing: A human-cyber-physical systems (HCPS) perspective," *J Manuf Syst*, vol. 63, pp. 471–490, Apr. 2022, doi: 10.1016/J.JMSY.2022.05.005.

View publication stats