Systematic Review of Smart Robotic Manufacturing in the Context of Industry 4.0

Lu Anh Duy Phan[1,2] and Ha Quang Thinh Ngo[1,2](✉)

[1] Department of Mechatronics, Faculty of Mechanical Engineering, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh, Vietnam
[2] Vietnam National University-Ho Chi Minh City (VNU-HCM), Linh Trung Ward, Thu Duc City, Ho Chi Minh, Vietnam
nhqthinh@hcmut.edu.vn

Abstract. In the rapidly evolving landscape of industrial revolution, smart robotic manufacturing has emerged as a game-changing phenomenon, revolutionizing traditional production processes, and unlocking unprecedented levels of productivity and safety. At the core of this transformative paradigm lies the seamless integration of artificial intelligence (AI) models, empowering autonomous robotic systems to carry out complex tasks with unparalleled precision and efficiency. This systematic review endeavors to explore the synergistic relationship between smart robotic manufacturing and AI technologies, delving into the various advancements, challenges, and potential implications for industrial sectors. By shedding light on the cutting-edge innovations and practical insights, this work aims to provide a valuable resource for researchers, industrialists, and policymakers seeking to leverage the transformative potential of AI-driven smart robotic manufacturing in the era of Industry 4.0.

Keywords: Intelligent Robot · Collaborative Robot · Smart Manufacturing · Predictive Model · Advanced Interaction

1 Introduction

In recent years, robotics has witnessed remarkable advancements in various domains, including manufacturing and production. Integrating machine learning techniques, particularly reinforcement learning, with robotics has been pivotal in developing intelligent robotic manufacturing systems. These systems aim to enhance the efficiency of industrial processes, flexibility, and adaptability by enabling robots to learn from their environment and make intelligent decisions.

This analysis focuses on applying reinforcement learning techniques in two specific cases of manufacturing: robotic manipulation [1] and collaborative robotics [2]. Both approaches have revolutionized traditional manufacturing practices by introducing a new generation of robots to perform complex tasks with precision and agility. In some

P. Cong Vinh and N. Thanh Tung (Eds.): ICCASA 2023, LNICST 579, pp. 19–42, 2024.
https://doi.org/10.1007/978-3-031-58878-5_2

factories with high rate of automation, only robots appear in their workplace and reach to self-assessment and decision-making. Though, in the lower level of automation, it requires that human co-exists and supports to manipulate.

In the high level, robotic manipulation is a fundamental aspect of manufacturing, involving precisely handling and manipulating objects. It is very crucial to manipulate robot based on our desired missions. Formerly, traditional robotic manipulation systems were limited in adapting to object shape, size, and position variations. However, with the advent of learning technologies, robots can now acquire new skills and autonomously adjust to changes in their environment. Several methods such reinforcement learning [3], deep learning [4], and computer vision [5] have been instrumental in enhancing the capability of robots to grasp, manipulate, and assemble objects with better accuracy and efficiency.

For lower level, collaborative robotics, on the other hand, focuses on the interaction between robots and workers. Henceforth, robots and humans worked in separate, isolated spaces in traditional manufacturing settings due to safety concerns. Nevertheless, collaborative robots, also known as cobots, are designed to work alongside humans, facilitating teamwork and cooperation. This approach involves robots learning to understand human gestures, intentions, and actions, enabling them to aid, share tasks, and ensure safe and efficient collaboration. Machine learning algorithms, such as human-robot interaction models and motion planning techniques, have been instrumental in developing effective collaborative robotic systems that enhance productivity and workplace safety.

The remainder of this study is arranged as follows. In Sect. 2, it provides an introduction to the background and concepts related to smart robotic manufacturing, machine learning, and robot learning. Section 3 delves into robotic manipulation, exploring the advancements and techniques used in this domain. Section 4 discusses human-robot collaboration in manufacturing applications, highlighting the key aspects and real-world circumstances. Finally, Sect. 5 presents open problems and future research directions for potential exploration and advancement in this field.

2 Background and Concepts

This section provides an overview of the background and fundamental concepts relevant to the robot learning scheme toward smart robotic manufacturing. Understanding these concepts is essential for comprehending the advancements and applications discussed in subsequent sections.

2.1 Smart Manufacturing

Smart manufacturing represents a modern and highly integrated manufacturing approach that combines the latest advancements in information technology, including the Internet of Things (IoT), cloud computing, and artificial intelligence (AI), with cutting-edge manufacturing processes [6]. By utilizing these technologies, smart manufacturing seeks to improve the productivity of production processes through autonomous perception, improved decision-making, and accurate execution, ushering in a new era of vitality for manufacturing [7]. This intelligent paradigm presents exciting opportunities for the global manufacturing industry, paving the way for more flexible, adaptive, and personalized production processes. The transformation and advancement of smart manufacturing hold significant implications for the overall development of manufacturing on a global scale.

2.2 Machine Learning

Machine learning is a multidisciplinary field that combines computation and statistics with connections to information theory, signal processing, algorithms, control theory, and optimization theory [8]. Machine learning has emerged as an exceptionally captivating study area in artificial intelligence. The fundamental concept of machine learning revolves around constructing models that can approximate data-based functions. Once trained on available data, these models can then be utilized to approximate and predict outcomes for new data instances. This process of transitioning from an initially weak model to a more robust one by leveraging the available data is termed "learning," analogous to the learning capabilities and processes observed in living organisms. Given that this process is predominantly accomplished by machines, i.e. computers, the term such machine learning was coined to represent this field of study.

2.3 Robot Learning

Robot learning refers to integrating various machine learning technologies within the field of robotics [9]. It specifically focuses on applying machine learning techniques to enable robots to learn and make informed decisions. Unlike traditional machine learning, robot learning emphasizes generating actions as output while perceiving and understanding the environment as input. For example, deep learning techniques enhance a robot's ability to navigate and interact with unstructured environments effectively. On the other hand, reinforcement learning provides formal frameworks to govern machine behaviors and decision-making processes. By synthesizing these machine learning technologies, robot learning enables robots to acquire knowledge, adapt to their surroundings, and engage in intelligent actions within their environment.

2.4 Reinforcement Learning

Reinforcement learning (RL) is a specialized branch within machine learning dedicated to instructing agents in making a series of decisions in an environment with the objective of maximizing a cumulative reward. RL is inspired by behavioral psychology, where an agent learns through trial and error by interacting with its environment. As shown in Fig. 1, an agent interacts with an environment and takes actions based on its current state. The working environment provides feedback to the agent in the form of rewards, which indicate the desirability of the actions of an agent. The goal of agent is to learn a policy a mapping from states to actions that maximizes the expected cumulative reward over time.

A lot of recent research choose to categorize RL algorithms into model-based approach and model-free approach, despite the fact that this is a challenging task given their

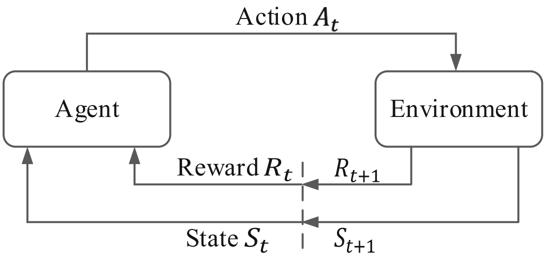

Fig. 1. Diagram for Learning Framework of Reinforcement Learning

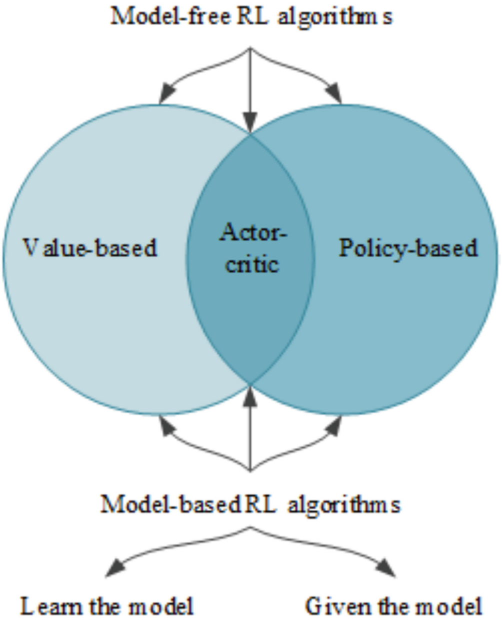

Fig. 2. Description of classifications for reinforcement learning algorithms [10].

extensive modularity. In Fig. 2, these researches, in turn, are divided into three main types: value-based, policy-based and a hybrid method called actor-critic algorithms [10].

A. Value-based RL.

Value-based algorithms generate the value function for each state or state-action pair until their values converge rather than storing any explicit policies. Instead, they achieve this via temporal difference learning. This makes it possible to decrease the variation in estimates of the expected returns, but it also necessitates a time-consuming optimization process. By acting greedily (choosing the action with the best value) on the calculated function, the optimum policy may be easily deduced from the value function. In this study, some of the algorithms of interest are as follows:

Q-Learning. Q-Learning [11] is one of the most used RL algorithms that learns the optimal action-value function (Q-function) through iterative updates. It uses the Bellman equation to update Q-values based on observed rewards and estimated future rewards.

Deep Q-Learning (DQN). DQN [12] is an extension of Q-learning that incorporates deep neural networks to approximate the Q-function. It uses experience replay to store and sample experiences, improving data efficiency and stabilizing learning.

Double Deep Q-Learning (Double DQN). Double DQN [13] addresses the issue of overestimation bias in Q-learning. It uses two sets of Q-values to decouple action selection and value estimation, reducing the tendency to overestimate the values of actions.

Dueling Deep Q-Learning (Dueling DQN). Dueling DQN [14] separates the estimation of the value and advantage functions, allowing the agent to learn the value of being in a particular state independently of the advantage of each action. This architecture enables better learning efficiency and generalization.

B. Policy-based RL.

In contrast to value-based algorithms, policy-based algorithms directly create the policy responsible for assigning the best possible action to each state. This parametrized function describing the approach is kept in memory throughout the learning process. These algorithms thereby improve the policy without relying on value function estimations. With this method, they can develop a smooth range of activities, but the variability may rise. Either gradient-based or gradient-free parameter estimation techniques can derive these algorithms. Several significant algorithms in this study include:

Vanilla Policy Gradient (VPG). VPG method is a direct application of the Policy Gradient Theorem [15]. The main idea behind VPG is to update the policy of agent in the direction that increases the expected reward. It achieves this by estimating the gradient of the expected reward with respect to the policy parameters and using this gradient to update the policy.

Trust Region Policy Optimization (TRPO). TRPO [16] is a policy optimization algorithm that seeks to improve the policy while ensuring small policy updates to maintain stability. It employs a trust region constraint to bind the maximum policy update step based on the Kullback-Leibler divergence.

Proximal Policy Optimization (PPO). PPO [17] is an advanced reinforcement learning algorithm designed to overcome the computational overhead of TRPO. While TRPO offers stability and convergence guarantees, it can be computationally expensive due to the need for solving complex constrained optimization problems at each update.

C. Actor-critic algorithms.

Actor-critic algorithms [18] represent a fusion of value-based and policy-based techniques within reinforcement learning. In this methodology, the ‘actor’, represented by a policy network, suggests actions for a specific state. In contrast, the ‘critic’, embodied by a value network, assesses these actions within the context of state-action pairs. By employing the Bellman equation, the critic learns the Q-function, and the actor is updated based on the Q-function to train the policy. Through this dual mechanism, the actor-critic approach harnesses the respective advantages of both value-based and policy-based methods. Up to date, the actor-critic based algorithms are mostly used and the classical algorithms include:

Advantage Actor-Critic (A2C). A2C [19] is an actor-critic algorithm that combines the advantages of both policy-based and value-based methods. It maintains a policy (actor) and a value function (critic) to estimate the advantage of each action. The policy is updated using the advantage as a baseline.

Asynchronous Advantage Actor-Critic (A3C). A3C [19] is an extension of A2C that parallelizes the learning process by using multiple agents that interact with different environment instances. Each agent updates the shared policy and value function asynchronously, improving sample efficiency.

Deep Deterministic Policy Gradient (DDPG). DDPG [20] is a reinforcement learning technique combining both deep Q-learning and deterministic policy gradients (DPG) [21]. DDPG has two main components: the actor-network and the critic network. The actor-network learns the policy, mapping states to actions, while the critic network estimates the Q-value of state-action pairs. These networks are usually implemented as deep neural networks, allowing them to approximate complex functions.

Twin Delayed Deep Deterministic Policy Gradient (TD3). TD3 [22] is an advanced variant of the Deep Deterministic Policy Gradient (DDPG) algorithm. TD3 was proposed to improve the stability and sample efficiency of DDPG in continuous action space reinforcement learning problems.

Soft Actor-Critic (SAC). SAC [23] is a state-of-the-art deep reinforcement learning algorithm designed for environments with continuous action spaces. It is based on the actor-critic architecture and has been proven to be highly effective and stable in various continuous control tasks. SAC has demonstrated impressive performance on various challenging continuous control tasks, such as robotic control, locomotion, and dexterous manipulation.

2.5 Network Architecture

Neural networks are effective function approximators in deep reinforcement learning, especially when the state or action space is too large to be fully known. Among the neural network architectures commonly employed in deep RL, the multilayer perceptron (MLP) is prevalent. An MLP comprises an input layer, a hidden layer, and an output layer. Unlike the input nodes, neurons in the hidden and output layers employ nonlinear activation functions. MLPs are trained through backpropagation, a supervised learning technique. However, MLPs possess a drawback in that they are fully connected, establishing connections between every perceptron. Consequently, this can result in many parameters and redundant information in high-dimensional spaces, making them inefficient.

A. Convolutional Neural Network.

Convolutional neural network (CNN) [24] is a type of artificial neural network specifically designed for processing and analyzing visual data, such as images and videos.

CNNs have proven to be highly effective in tasks related to computer vision, including image classification, object detection, and image segmentation. The key feature of CNNs is their ability to automatically learn and extract hierarchical features from input data through a series of convolutional layers. These layers use convolution operations to scan and filter the input data, gradually capturing patterns and features of increasing complexity. CNNs typically consist of multiple layers, including convolutional layers, pooling layers for down-sampling, and fully connected layers for classification or regression tasks. In a reinforcement learning algorithm, the output of the CNN can be used as the input for a value or policy function.

B. Recurrent Neural Network.
A recurrent neural network (RNN) [25] is a type of artificial neural network designed to process sequential data by maintaining internal memory. Unlike traditional feedforward neural networks, which process data in a one-way direction, RNNs have connections that loop back on themselves, allowing them to capture patterns and dependencies in sequences. The key feature of an RNN is its ability to retain information from previous time steps and use it to influence the processing of the current input. This makes RNNs particularly well-suited for sequences involving natural language processing, speech recognition, time series analysis, and more.

C. Graph Neural Network.
Graph Neural Network (GNN) [26] is a type of neural network designed to process and make predictions on graph-structured data. Graphs are mathematical structures that consist of nodes (vertices) connected by edges (links), and they are used to model relationships or interactions between different entities. GNN can be used for various tasks, including node classification, link prediction, graph classification, and recommendation systems, among others. Some popular GNN architectures include Graph Convolutional Networks (GCN) [27], GraphSAGE [28], and Graph Attention Networks (GATs) [29]. These architectures have been successfully applied to a wide range of domains, such as social networks, bioinformatics, chemistry, and knowledge graphs.

3 Deep Reinforcement Learning for Robotic Manipulation

This section explores how Deep Reinforcement Learning (Deep RL) techniques enhance robotic manipulation. From grasping techniques to gripper designs and theoretical simulation, these advancements empower robots with improved precision, dexterity, and adaptability in manipulation.

3.1 Robotic Grasping

Robotic grasping has improved Through various tactics meant to increase the effectiveness and adaptability of object manipulation. These include sole-grasping rules, in which specific algorithms precisely guide robots to acquire objects in challenging circumstances. Additionally, suction-based grasping methods have increased the variety of things that robots may handle by enabling secure grabbing by establishing a vacuum seal.

Thanks to multifunctional grippers, robots are now adaptable and can change their grip to suit different object shapes and sizes. In addition to these methods, “two-action synergy” has been developed, which entails the coordinated performance of several actions, such as grasping and pushing, for improved grasping performance.

Combining these robotic grasping approaches with machine learning can potentially have transformational effects on various automation and robotics-related businesses, especially when combined with the development of machine learning techniques, particularly reinforcement learning. For a comprehensive description, Table 1 depicts the literature on Deep RL in robotic grasping action.

A. Sole-grasping.

This policy is a technique for training a robot to only utilize the grasping action to catch objects when no other action (such as pushing, moving, and poking) are involved. Numerous research studies have focused on applying reinforcement learning to grasping individual or cluttered objects. For instance, there are studies [30, 31] in the case of learning to grasp individual objects. Investigators in [30] introduced a reinforcement learning model that utilizes a strategy search algorithm, demonstrating remarkable robustness in the generalization from simple to complex object shapes. However, the current form of guided policy search (GPS) faces limitations in its applicability to sequential multitask learning scenarios due to its batch-style training requirement. In another related study, developers in [31] concentrates on addressing the problem of chin-grasping poses in 3D space using 3D point clouds as inputs for the model. The research results showed promising simulation outcomes, and the simulation data can be further utilized for real-world applications.

Additionally, accurate grasping is a challenging problem with significant potential for applications in manufacturing. Researchers in [32] developed an innovative approach for training a robot to perform pick-and-place tasks using self-supervised learning, without relying on an object model. They combined two techniques, namely, robot learning of primitives estimated by fully convolutional networks (FCNs) and one-shot imitation learning (IL). To achieve precise pick-and-place actions without an object model, they formulated the place reward as a contrastive loss between real-world measurements and a task-specific noise distribution. The results showed great promise in terms of accuracy. However, the process required a substantial amount of time for exploring behaviors, which limited its efficiency for industrial applications. In general, this study represents a significant step forward in developing robotic capabilities for pick-and-place tasks without the need for explicit object models. A highly innovative grasping strategy is proposed in a study [33]. TossingBot, a robot system capable of accurately tossing various objects to designated target positions, showcases the potential of this approach. The authors introduce an end-to-end approach that simultaneously learns to infer control parameters for grasping and throwing by iteratively testing and adjusting based on images of objects within a container. This self-supervised learning process enables the system to identify optimal grasping positions that result in consistent and predictable throws. To simplify the throwing task, the system focuses on predicting the release velocity alone. The release velocity is determined using a physics-based controller and further refined based on the residual estimate obtained from the neural network. The integration of image-based learning and physics-based control in TossingBot demonstrates promising results in enhancing the ability of a robot to handle a wide array of objects and execute precise throwing actions.

Table 1. Literature of Deep Reinforcement Learning in Robotic Grasping

Author(s)	Publication year	Action	Gripper type	Sim package	Sim/Real-world	Method
Beltran-Hernandez et al. [30]	2019	Grasping	Parallel-jaw	Gazebo	Sim	CNN + Guided Policy Search
Mousavian et al. [31]	2019	Grasping	Parallel-jaw	FleX	Sim/Real	Point-Net + +
Berscheid et al. [32]	2020	Grasping and placing	Parallel-jaw	N/A	Real	FCNNs + Q-learning + one-shot imitation learning
Zeng et al. [33]	2020	Grasping and throwing	Two fingers	Bullet	Sim/Real	ResNet-FCNs + Q-learning
Shao et al. [34]	2019	Grasping	Suction	V-REF	Sim	Resnet with U-net (CNN) + Q-learning
Zakka et al. [35]	2020	Grasping and placing	Suction	N/A	Real	FCN ResNet + Q-learning
Cao et al. [36]	2022	Grasping	Suction	V-REF	Sim	A3C
Zeng et al. [37]	2022	Grasping	Two fingers, suction	N/A	Real	ResNet-FCNs + Q-learning
Zeng et al. [38]	2018	Grasping and pushing	Two fingers	V-REF	Sim/Real	DenseNet-FCNs + Q-learning
Ren et al. [39]	2021	Grasping and moving	Two fingers	V-REF	Sim/Real	Duelling DDQN
Tang et al. [40]	2021	Grasping and pushing	Three fingers	V-REF	Sim/Real	DenseNet-FCNs + Q-learning
Zhang et al. [41]	2023	Grasping and pushing	Two fingers	Isaac Gym	Sim/Real	PPO
Berscheid et al. [42]	2019	Shifting and grasping	Parallel-jaw	N/A	Real	Deep Q-learning

(*continued*)

Table 1. (*continued*)

Author(s)	Publication year	Action	Gripper type	Sim package	Sim/Real-world	Method
Hundt et al. [43]	2020	Pushing, Grasping and Placing	Two fingers	V-REF	Sim/Real	DenseNet-FCNs + Q-learning
Yang et al. [44]	2020	Grasping and pushing	Parallel-jaw	V-REF	Sim/Real	FCNs + Q-learning
Xu et al. [45]	2021	Grasping and pushing	Two fingers	V-REF	Sim/Real	DenseNet-FCNs + Q-learning
Huang et al. [46]	2021	Grasping and pushing	Two fingers	Bullet	Sim/Real	DQN + MCTS + DIPN
Chebotar et al. [47]	2021	Grasping and placing	Two fingers	N/A	Sim/Real	Q-learning
Ren et al. [48]	2022	Grasping and pushing	Two fingers	V-REF	Sim/Real	DenseNet-FCNs + Q-learning
Novkovic et al. [49]	2020	Grasping and pushing	Parallel-jaw	Bullet	Sim/Real	PPO
Chen et al. [50]	2020	Grasping and pushing	Parallel-jaw	MuJoCo	Sim	TD3

B. Suction-Based Grasping

Suction grasping is another mechanism strategy that has been increasingly utilized for performing object manipulation in dense cluttered environments. This approach involves using suction-based grippers or end-effectors that generate a vacuum force to firmly attach to the surface of an object. By creating a secure suction grip, robots can pick up and manipulate objects even in scenarios where traditional grasping mechanisms might struggle due to clutter or irregular shapes. Suction grasping proves to be particularly effective in industries such as logistics, warehousing, and agriculture, where objects may be randomly arranged or piled up in confined spaces. The ability to handle objects in dense clutter allows robots to operate efficiently, improving automation capabilities and expanding their scope of applications in real-world scenarios.

In [34], researchers introduced the concept of suction grasp as a viable alternative for object manipulation in cluttered environments, aiming to mitigate potential failure situations resulting from the combination of pushing and grasping actions. Their approach involved utilizing deep reinforcement learning, using techniques such as Q-learning with ResNet and the U-net structure. Their method faced a limitation in that the suction grasp points were randomly predicted, leading to difficulties in accurately identifying grasp points, especially in cluttered environments. To enhance the effectiveness of such frameworks, there is a need to incorporate more diversity in the shapes of objects during both training and testing phases. Moreover, their reported results were solely based on successful outcomes in simulated environments using CoppeliaSim (V-REP), raising the need for additional real-world validation To demonstrate a method for pick-and-place, using matching network [35] computes dense visual descriptors to associate picking actions to placing actions. Their framework can easily be generalized to new objects and kits. Yet, due to the only processing 2D rotations and some assumptions that objects are face-down, it would be interesting to explore a more complex action representation for 3D assembly. In another related study [36], researchers proposed using an Actor-Critic algorithm, A3C, as the reinforcement learning method to train the picking policy network for executing grasping tasks in cluttered areas. This opens exciting possibilities for improving robotic manipulation in complex scenarios.

C. Multifunctional Gripper-Based Grasping

Recently, another intriguing mechanism gaining attention in research studies is the training of reinforcement learning algorithms to coordinate the execution of grip and suction grasps. This approach involves equipping the robotic arm with a gripper capable of both finger gripping and suction-cup functions. By adopting this multifunctional gripper design, researchers can leverage the advantages of both gripping techniques in various scenarios. For instance, the finger-gripper excels at grasping objects in cluttered environments, overcoming the limitations of the suction-cup grasp, and vice versa. In one such study [37], the researchers proposed a method for robotic object manipulation, specifically pick-and-place tasks, by predicting both grip and suction affordances using the multifunctional gripper. Their approach employed a fully convolutional residual network to predict suction affordance for multi-view RGB-D images. A category-agnostic affordance prediction technique was then utilized to choose and execute one of four potential grasping primitive behaviors. However, it’s worth noting that the inclusion of planar grasps in their learning approach might pose challenges due to arm movement restrictions. Furthermore, the strategy of ‘Pick first, ask questions later’ employed in their study may not be suitable for tasks requiring pre-determining the target object. This ongoing research in RL coordination of grip and suction grasps holds great promise for advancing robotic manipulation capabilities and overcoming complex real-world challenges.

D. Synergy of Two Primitive Actions

The synergy of two primitive actions refers to the powerful combination of two fundamental movements or behaviors that, when integrated, create a more complex and efficient action. In the context of robotics and artificial intelligence, primitive actions represent basic building blocks of behavior, such as grasping, pushing, reaching, or turning. By combining these individual actions in a coordinated manner, robots can perform more sophisticated tasks and adapt to dynamic environments effectively. This synergy enables robots to handle complex real-world scenarios that require a sequence of actions, making them more versatile and capable problem-solvers. The concept of combining primitive actions is essential for developing advanced robotic systems that can accomplish a wide range of tasks autonomously and adaptively, bringing us closer to the realization of intelligent and adaptable robotic agents in various fields. As technology continues to evolve, harnessing the synergy of primitive actions opens up new frontiers for robotics research and applications, propelling us towards a future where robots seamlessly integrate into our daily lives, assisting us in a multitude of tasks and enriching our experiences.

In [38], scholars delved into exploring the synergistic relationship between pushing and grasping in the context of robotics. They aimed to achieve more stable and efficient results in densely cluttered environments by training deep end-to-end policies. Their approach was centered on the visual pushing-grasping (VPG) framework, where Q-learning was employed with the DenseNet pre-training model, specifically DenseNet-FCN, a fully connected network. However, it was noted that the VPG framework was primarily designed for target-agnostic tasks, necessitating re-projection before inputting the prediction network. One limitation observed was that the intrinsic pushing reward did not explicitly indicate whether a push would facilitate future grasping. Consequently, there were instances where the robot inadvertently pushed objects out of the workspace, leading to unnecessary actions and prolonged task execution times. Similar studies exploring the combination of pushing and grasping in the context of robotics have been of interest [39–41]. Developers [39] concentrate on the rapid acquisition of grasping and pre-grasping skills in goal-agnostic tasks. Their study aims to enhance performance and learning efficiency by introducing a mask function. This function plays a crucial role in guiding the robot's grasping behavior, enabling it to identify and focus on relevant object features during the learning process. By incorporating the mask function, the robot becomes more adept at grasping objects in diverse and unpredictable environments, ultimately leading to improved overall performance in goal-agnostic tasks. In [40] presented a novel method for collaborative pushing actions to aid in the process of grasping objects. Their approach employed Q-learning to learn a deterministic policy for both pushing and grasping. Interestingly, they did not assign any specific reward to pushing actions; instead, the agent received a reward only when the robot successfully completed the grasping task. Investigators in [41] develop a model-free Deep Reinforcement Learning framework to synergize pushing and grasping actions. The paper proposes an approach to tackle the challenge of handling objects in difficult positions. However, the push-to-wall method cannot effectively handle objects with bevels or hard flat sides.

Another approach to enhance object grasping in cluttered environments involves the technique of "shifting" objects, as discussed in [42], which involves putting a finger on top of the target object to increase grasp probabilities. The results of this approach demonstrate a high success rate and a capacity for generalization. Notably, the training process was conducted online without reliance on a simulation model. However, a critical aspect in improving the robotic bin grasping process is its ability to adapt when depth information is missing. This becomes particularly important because the availability of depth data from stereo cameras is often limited by factors such as shadows or reflective surfaces. Grasping objects near the edges or corners of tote boxes presents a notable challenge, and in scenarios where objects are densely packed together, the robot may encounter situations where suitable grasping options are not readily available.

Developers in [43] introduced a schedule outlining the Positive Task Framework (SPOT) and elaborated on the SPOT-Q RL algorithm. The SPOT framework serves the dual purpose of quantifying an agent's progress in multi-step tasks and offering crucial guidance with zero rewards, a masked action space, and situation removal. This framework can rapidly acquire policies that can generalize effectively from simulated environments to real-world scenarios. Nevertheless, it is worth noting that this approach places substantial demands on data resources and necessitates several iterative refinements to enhance its efficiency when applied to such tasks. Consequently, it is advisable to incorporate mechanisms for reactivity and failure recovery to counterbalance the precision loss that can occur due to policies trained in both simulated and real-world contexts.

In [44], researchers introduced a deep Q-learning approach for the purpose of grasping invisible objects, involving two distinct stages. The first stage revolves around determining the visibility of the target objects. If the thing is visible, the robot proceeds to the second stage and either performs a push or a grasp action. However, in cases where the target object remains unseen, the robot initiates an exploration process by continuously pushing until the thing is located. Following the discovery of the target object, the robot employs coordination of grasp-push actions to successfully grasp it. It is worth noting that this approach assumes prior knowledge and relies on the target object having a specific color. Additionally, the construction of the entire pushing reward function is done manually, potentially necessitating numerous tuning iterations and lacking adaptability to novel scenarios.

Target object-grasping tasks in cluttered environments have been the focus of interest in various research studies [45–50]. In [45], a goal-conditioned hierarchical reinforcement learning approach was proposed, demonstrating high sample efficiency in training a push-to-grasp technique for a particular object amidst clutter. Meanwhile, researchers in [46] focused on extracting a target object from a densely packed environment using quasi-static push and overhand grasp movements. To achieve this, they introduced the visual foresight tree (VFT) method, which identified the shortest sequence of actions. The VFT method combined a deep interactive prediction network (DIPN) to estimate push action outcomes and the Monte Carlo tree search (MCTS) to select the best actions. Despite its effectiveness, the VFT method had some limitations. The computation time was considerably long due to the size of the MCTS tree. In [48], developers introduced a bifunctional network that processes visual observations and produces comprehensive pixel-wise Q value maps for both pushing and grasping primitive actions. This innovation was aimed at augmenting the available data samples within the action space. The system is feasible for practical deployment as the pre-trained model in the simulation achieved a considerably high success rate in the real world without fine-tuning. Investigators in [49] employed Proximal Policy Optimization (PPO) to enable the agent to make predictions about its future actions based on its past experiences and its current state. To facilitate the grasping of occluded objects, a robotic arm equipped with an RGB-D camera was utilized, allowing the agent to observe the scene from various angles. For object tracking using RL, the scene state was encoded using a discretized truncated signed distance field (TSDF) volumetric representation. The study in [50] discusses the difficulty in gripping objects when they are near to one another and there isn't enough room for gripper fingers to do so. They use distinct fully convolutional networks (FCNs) to anticipate the grab point and push direction in their method. Then, using Q-learning, the corresponding action is carried out using the highest Q-value that was chosen. However, rule-based systems are less efficient since the robot could continue pushing after the push operations are completed without changing the robot's workspace, which has an impact on the robot's performance. A common limitation in these studies is the assumption of prior knowledge and reliance on target objects having specific colours, which poses significant constraints in real-world applications.

3.2 Assembly and Disassembly Tasks

Assembly plays a pivotal role in modern manufacturing and production processes. Researchers have delved into robot learning to augment the efficiency and effectiveness of assembly tasks. Several studies on this subject are listed in Table 2.

Table 2. Literature of Deep Reinforcement Learning in Robotic Assembly and Disassembly

Author(s)	Publication year	Task	Method	Observation	Action
Luo et al. [51]	2018	Inserting a peg in a hole	GPS	Force, torque and robot state	Reference to an admittance controller
De Winter et al. [52]	2019	Cranfield assembly benchmark	Q-learning	Labelled state of objects	Hierarchy of sub-procedure
Li et al. [53]	2019	Circuit breaker assembly	DQN	Force and torque	Rotate along three axis
Kristensen et al. [54]	2019	E-waste unscrewing disassembly	Q-learning	Force, joint angle, and position of the end-effector	Seven types of movements
Kim et al. [55]	2020	Inserting a peg in a hole	Imitation learning + DDPG	Force and position	Desired position and velocity
Ota et al. [56]	2019	Computer assembly	TD3	Joint angles and angular velocities	Angular velocities

In [51], researchers were pioneers in applying reinforcement learning to tackle an industrial problem involving deformable objects. Their primary emphasis was on achieving precision. They developed a policy search framework for robotic assembly tasks involving rigid and deformable components. Consequently, they successfully guided a position and velocity-controlled robot with haptic feedback to insert a wooden peg into a non-linear deformable part with a hole. However, it is worth noting that this system did not incorporate a visual perception system, and its performance suffered when the peg was not close to the hole. In subsequent studies, researchers in [52] utilized Q-learning along with a hierarchical task graph for robotic assembly, while investigators in [53] employed DQN and ROS Gazebo simulators to train a KUKA iiwa robot in completing a circuit breaker assembly task. Developers in [54] explored a robot simulation framework for e-waste disassembly using Q-learning, and the study in [55] focused on robot learning for square peg-in-hole assembly, ingeniously combining imitation learning and DDPG. Additionally, investigators in [56] investigated motion trajectory in unknown environments using TD3 with RRT as the reference and validated their approach through a computer assembly task. These studies collectively contribute to the advancement of robotic assembly techniques, showcasing the versatility and efficacy of RL in tackling various challenges across diverse tasks and environments. Their findings pave the way for further innovations and advancements in the field of robotics.

4 Human-Robot Collaboration and Deep Reinforcement Learning

The objective of this review is to offer a comprehensive understanding of the application of reinforcement learning in research related to human-robot collaboration.

4.1 Collaborative Robotics

Human-Robot Collaboration (HRC) is a synergistic approach that combines the strengths of two entities: the power, durability, repeatability, and precision of robots with the intuition, flexibility, problem-solving abilities, and sensory perception of humans. Depending on the scope of application, humans and robots can collaborate in various ways. In this article the meaning of collaboration level is adapted from [57], where the collaboration levels are illustrated in Fig. 3.

Currently, the prevailing approach in industrial production centers around human-robot coexistence, which prioritizes safety. The risk of injuries is notably low because distinct workspaces are designated for humans and robots. In this setup, humans and robots carry out their tasks in separate, dedicated areas.

Within human-robot cooperation, these workspaces are aligned, yet the operations are time-separated and executed sequentially. However, when it comes to collaboration, a different dynamic emerges, as both spatial and temporal separations are diminished. Here, direct interaction between humans and robots becomes common.

In this evolving collaborative landscape, the importance of safety requirements for human-robot cooperation and collaboration is rising. It is crucial to ensure that as humans and robots work more closely together, robust safety measures are in place to safeguard all parties involved.

The combination of robot and human capabilities brings about several significant benefits, including:

- Enhanced Efficiency: Robots often possess faster processing abilities and do not experience fatigue like humans. Humans can leverage the high efficiency of robots to perform tasks quickly and effectively.

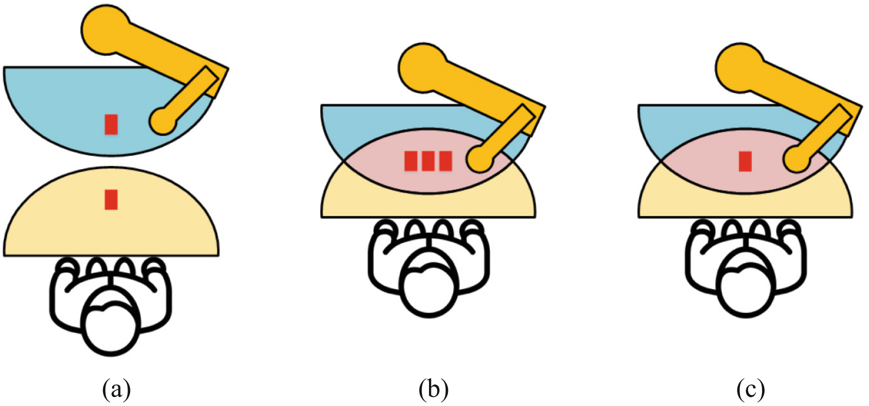
(a) (b) (c)

Fig. 3. Collaboration levels adapted for this study: (a) Coexistence; (b) Cooperation; (c) Collaboration.

- Safety in Hazardous Tasks: Robots can be deployed to carry out dangerous, difficult, or repetitive tasks that could pose risks to human safety. When combined with human supervision and control, robots help reduce hazards and enhance safety in hazardous work environments.
- Synergy of Computation and Creativity: The combination of machine computation and human creativity can lead to novel, breakthrough, and innovative solutions in work. Robots can be programmed to handle basic tasks, allowing humans to focus on analyzing and generating more complex solutions.
- Social Interaction with Humans: Robots can be programmed to interact socially with humans, creating a comfortable and easy communication environment in the workplace. This fosters better collaboration and information exchange between humans and robots.

4.2 Deep Reinforcement Learning in Human-Robot Collaboration

In the journey towards smart manufacturing, seamless interaction between humans and industrial robots holds significant importance. Automated robots enhance efficiency and precision, while human involvement introduces a crucial element of adaptability. A prevailing trend in this landscape involves incorporating human factors into interaction strategies to optimize human-robot collaboration. However, the challenge lies in addressing the unpredictability of human behaviors, which can complicate robot task planning and decision-making processes. To tackle this issue, deep reinforcement learning emerges as a vital tool in developing control approaches. Deep RL helps manage the uncertainty introduced by human actions and enhances the robot's ability to adapt to dynamic scenarios (Table 3).

Safety is of paramount importance in manufacturing operations. In human-robot interaction (HRI), one crucial aspect of safety control involves utilizing Deep Reinforcement Learning (DRL) to develop collision avoidance motion plans and navigation control strategies tailored for industrial robot arms. An illustrative example of this approach can be found in the work of researchers highlighted in reference [58]. They addressed safety

Table 3. Literature of Deep Reinforcement Learning in Robotic Grasping

Author(s)	Publication year	Application	Interactive interface	Method	Collaborative level	Observation	Action
Liu et al. [58]	2021	Safe interaction	Vision	DDPG	Cooperation	3D obstacle	Joint angles
Haage et al. [59]	2017	Assembly	Vision	Imitation learning	Coexistence	Human actions visually	Pre-defined optional actions
Zanchettin et al. [60]	2018	Assembly	Vision	Markov Chains	Cooperation	Human actions visually	Accept or not
Wang et al. [61]	2018	Assembly	Audio recognition Vision AR system	Imitation learning	Collaboration	12 assembly task states	Pre-defined movements
Akkaladevi et al. [62]	2019	Assembly	Vision	Q-learning	Collaboration	Human actions visually and object	Sub-procedure
Wu et al. [63]	2020	Object handling	Human reference	Q-learning	Collaboration	Human reference	Object motion
Zhang et al. [64]	2020	Human assistant/Assembly	Vision	RNN	Collaboration	Location of the body joints	Target trajectory of robot
Yu et al. [65]	2021	Assembly	Human reference	DQN	Coexistence	Task chessboard	Sub-procedure
Wang et al. [66]	2021	Path generation	Vision	Imitation learning and Q-learning	Cooperation/ Collaboration	Search parameters	Pre-defined movements
Wang et al. [67]	2021	Hand-over collaboration	Vision	Imitation learning	Collaboration	10 subclasses of hand-over intensions	Accept or not
Zhang et al. [68]	2022	Assembly	Vision	DDPG	Cooperation/ Collaboration	Current state of the robot	Task selection
Deng et al. [69]	2017	Object handling	Vision	Q-learning	Cooperation/ Collaboration	Human actions visually	Joint action
Ghadirzadeh et al. [70]	2016	Object handling	Vision	Q-learning	Cooperation/ Collaboration	Sensory states and actions	Velocities in the end-effector

concerns within industrial human-robot collaboration scenarios by implementing Deep Deterministic Policy Gradients (DDPG). Their innovative approach involved using data from the human arm's position as an observation input. With this information, DDPG generated precise joint angle commands for an ABB IRB1200 robot in real-time, facilitating dynamic trajectory planning to ensure safety and seamless collaboration between humans and the robot.

Additionally, the most talked-about application in production activities is assembly. Deep Reinforcement Learning (DRL) is widely employed as a tailored learning methodology to bolster robot capabilities in providing assembly assistance. A significant proportion of current approaches places their emphasis on aspects like human feedback, human demonstrations, or human-guided training methods. The study in [59] introduced the teach-by-demonstration framework for smartphones, which aims to reduce the time and expertise needed to set up a robotized assembly station. This method allows for changes in roles and tasks, enabling workers to decrease their physical or cognitive workload.

In reference [60], researchers proposed a cutting-edge technique to forecast human activity patterns. By anticipating when a human will request a specific collaborative task, this capacity enables the robot to engage in alternate autonomous tasks proactively. The prediction algorithm at the heart of this strategy is based on higher-order Markov Chains and was extensively tested by experiments carried out in an actual environment. In a broader context, a comprehensive approach is detailed in [61], wherein a three-pronged, integrated strategy encompassing teaching, learning, and operation is adopted. In this approach, humans initially instruct the robot using natural language commands, and subsequently, the robot learns from human assembly demonstrations via a Reinforcement Learning algorithm. Following the teaching and learning phases, the acquired knowledge is actively applied during collaborative assembly tasks to provide valuable assistance.

Academic researchers in [62] employed tabular Q-learning with the reward signal from the human collaborator. In this method, the robot actively scans the work area, understands the assembly procedure, and decides which activities to carry out. The robot system includes 3D sensors to keep an eye on the user and their surroundings and a dynamic graphical user interface (GUI) for user interaction. Additionally, this framework allows for different user types, enabling them to command the robot in various assembly processes. In an analogous situation, as explained in [63], the main goal was to limit item movement to a specific surface. In such cases, people control the robot's end effector in a predetermined area of interest. The robot then modifies its end effector while still being held by the operator, ensuring accurate positioning and orientation alignment for the requested task.

In [64], they utilized sequences of motion frames from the user for experimental validation. They implemented multiple cascaded RNNs, a common approach to using them. The robot observes the human, predicts their next pose, and proactively moves to pick up a screwdriver and hand it to the human based on the prediction. In reference [65], researchers introduced an innovative approach using Deep Q-Networks (DQN) to schedule collaborative tasks between humans and robots efficiently. This method aims to optimize task completion times within a simulated manufacturing environment, specifically in an assembly chessboard scenario. Remarkably, the agent autonomously learns the optimal scheduling policy without any reliance on human intervention or expert knowledge, leveraging a Markov game model to achieve this level of automation. Researchers in [66] proposed a complex method embodying imitation learning, Q-learning, and a simulated annealing algorithm. Gaussian noises were designed in demonstrations to overcome trembling and abrupt changes during a human's demonstrations to avoid jerky regression paths.

Developers in [67] first used multimodal processing to estimate ten subclasses of hand-over intentions and then took them as observation. The robot was trained to decide whether to accept the delivery from the human. Scientists in [68] proposed an approach that utilizes the DDPG method to generate a suitable action sequence for humans and robots during collaborative assembly tasks. The real-time behavior of the agent-human interaction is displayed to the operator. This makes it possible for the operator to complete the assembly work according to the planned assembly behavior, using the globally effective method for the anticipated performance.

In [69], the researchers introduce a hierarchical robot learning approach that consists of two learning hierarchies for HRC tasks. During the mission, the human collaboratively lifts an object with the robot and repeatedly moves it from point A to point B. The robot detects the intention of human and follows its movement, maintaining the thing at a specific orientation throughout the task. In [70], researchers introduced a sensorimotor reinforcement learning framework. This framework is designed to empower robots with the ability to acquire the skills needed for effective collaboration with human partners. The algorithm relies on inputs from vision and force/torque sensors to make informed decisions regarding motor commands. To account for the inherent unpredictability in human actions, a Gaussian process model is employed to model uncertainty. Bayesian optimization is employed to select the most optimal actions at each time step.

5 Open Problems and Research Directions

This section gives a general overview of the difficulties that reinforcement learning currently faces and investigates potential research trajectories for improving teaching strategies in intelligent robotic manufacturing.

A. Sim to Real

Simulation to Reality (Sim2Real) technology aims to transfer knowledge learned in simulators to the real world [71]. This involves setting control tasks, training an agent in a virtual environment provided by a simulator physics engine, and then deploying the acquired policies to control a physical agent in the real world. While training robots in simulators is often straightforward and effective, challenges arise if the simulator fails to accurately represent the complexities of real-world robotic tasks, leading to distribution shifts and failures when deploying the trained robot in the physical environment, despite good performance in the simulated environment. This discrepancy is known as the “sim-to-real gap” in robot learning. Luckily, manufacturing research is actively embracing digital twin technology [72], which holds promise in minimizing the sim-to-real gap in robot learning. The use of digital twins presents an open area of investigation in this context.

B. Multi-agent Reinforcement Learning

When considering robotic manufacturing systems from a higher perspective, the concept of multi-agent multi-task robot learning emerges in the field of artificial intelligence. This method has demonstrated its potential in esports, where multiple agents were trained to function as a virtual army with individual tasks [73]. Extending this approach to robotic manufacturing factories presents an open and challenging opportunity.

A class of techniques known as multi-agent reinforcement learning (MARL) uses reinforcement learning algorithms for individual agents in multi-intelligent systems. In such arrangements, each agent contains fundamental learning, thinking, and planning abilities. By utilizing MARL, an intelligent agent can collaborate with numerous entities that have simpler intelligence to attain complex intelligence, improving system robustness, dependability, and flexibility. Applications for MARL have been found in many fields, including distributed sensing networks, scheduling of transportation, power system optimization, and robot navigation, demonstrating its effectiveness in managing many intelligent systems.

C. Transfer Learning and Generalization

In the realm of manufacturing, environments often exhibit significant variations due to changes in products, production lines, and external conditions. To address this complexity, it is essential to develop reinforcement learning agents capable of transferring knowledge across diverse scenarios and generalizing learned policies to adapt to new settings. However, establishing a knowledge sharing architecture, managing knowledge from different robots, and determining the suitability of specific knowledge pose substantial challenges for researchers in [74]. Additionally, effectively utilizing knowledge gained from previous tasks in new ones remains an open question, especially in the field of smart robotic manufacturing [75]. Overcoming these challenges is crucial for improving the efficiency and adaptability of RL-based systems in manufacturing, unlocking new possibilities for intelligent automation in diverse and dynamic manufacturing environments.

6 Conclusion

Deep Reinforcement Learning (DRL) is a critical and promising technology that holds significant potential across various stages of the smart manufacturing lifecycle. Its adaptability and flexibility make it an appealing solution for enhancing smart manufacturing systems, ushering in a more cognitive and personalized manufacturing approach. To shed light on its essence, this study conducted a systematic literature review of 40 selected works from the past decade, examining the applications of DRL in the engineering product lifecycle. The review emphasizes the challenges faced and proposes future research directions for integrating DRL into smart manufacturing. By providing this comprehensive review, the aim is to inspire further in-depth research and discussions on DRL and its broader implementation in smart manufacturing practices.

Acknowledgement. This research is funded by Vietnam National University-Ho Chi Minh City (VNU-HCM) under grant number: **DS2023-20-02**.

References

1. Nguyen, H., La, H.: Review of deep reinforcement learning for robot manipulation. In: 2019 Third IEEE International Conference on Robotic Computing (IRC), pp. 590–595. IEEE (2019)

2. Ghadirzadeh, A., Chen, X., Yin, W., Yi, Z., Björkman, M., Kragic, D.: Human-centered collaborative robots with deep reinforcement learning. IEEE Robot. Autom. Lett. **6**(2), 566–571 (2020)
3. Fan, L., Zhu, Y., Zhu, J., Liu, Z., Zeng, O., Gupta, A., Fei-Fei, L.: Surreal: Open-source reinforcement learning framework and robot manipulation benchmark. In: Conference on Robot Learning, pp. 767–782. PMLR (2018)
4. Ribeiro, E.G., de Queiroz Mendes, R., Grassi, V., Jr.: Real-time deep learning approach to visual servo control and grasp detection for autonomous robotic manipulation. Robot. Auton. Syst. **139**, 103757 (2021)
5. Ngo, H.Q.T., Bui, T.T.: Application of the image processing technique for powerline robot. In: Phan, C.V., Nguyen, T.D. (eds.) Context-Aware Systems and Applications: 11th EAI International Conference, ICCASA 2022, Vinh Long, Vietnam, October 27-28, 2022, Proceedings, pp. 178–189. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-28816-6_14
6. Wang, B., Tao, F., Fang, X., Liu, C., Liu, Y., Freiheit, T.: Smart manufacturing and intelligent manufacturing: a comparative review. Engineering **7**(6), 738–757 (2021)
7. Nguyen, T.T., Nguyen, T.H., Ngo, H.Q.T.: Using real-time operating system to control the recycling waste system in beverage industry for circular economy: mechanical approach. Results Eng. **18**, 101083 (2023)
8. National Research Council: Frontiers in Massive Data Analysis. National Academies Press (2013)
9. Peters, J., Lee, D.D., Kober, J., Nguyen-Tuong, D., Bagnell, J.A., Schaal, S.: Robot learning, pp. 357–398. Springer, Springer handbook of Robotics (2016)
10. Elguea-Aguinaco, Í., Serrano-Muñoz, A., Chrysostomou, D., Inziarte-Hidalgo, I., Bøgh, S., Arana-Arexolaleiba, N.: A review on reinforcement learning for contact-rich robotic manipulation tasks. Robot. Comput. Integr. Manuf. **81**, 102517 (2023)
11. Watkins, C.J., Dayan, P.: Q-learning. Mach. Learn. **8**, 279–292 (1992)
12. Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., Riedmiller, M.: Playing Atari with deep reinforcement learning. arXiv 2013, arXiv:1312.5602
13. Van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, the Phoenix Convention Center, vol, 30, Phoenix, AZ, USA (2016)
14. Wang, Z., Schaul, T., Hessel, M., Hasselt, H., Lanctot, M., Freitas, N.: Dueling network architectures for deep reinforcementlearning. In: Proceedings of the International Conference on Machine Learning, pp. 1995–2003, New York City, NY, USA (2016)
15. Sutton, R.S., McAllester, D., Singh, S., Mansour, Y.: Policy gradient methods for reinforcement learning with function approximation. Adv. Neural Inform. Process. Syst. **12** (1999)
16. Schulman, J., Levine, S., Abbeel, P., Jordan, M., Moritz, P.: Trust region policy optimization. In: Proceedings of the International Conference on Machine Learning, Lille, France, pp. 1889–1897 (2015)
17. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O. In: Proximal policy optimization algorithms. arXiv 2017 arXiv:1707.06347
18. Konda, V.R., Tsitsiklis, J.N.: Actor-critic algorithms. In: Proceedings of the Advances in Neural Information Processing Systems, Denver, CO, USA, 28–30, pp. 1008–1014 (2000)
19. Mnih, V. et al: Asynchronous methods for deep reinforcement learning. In: Proceedings of the International Conference on Machine Learning, New York City, NY, USA 19–24 1928 1937 (2016)
20. Lillicrap, T.P., et al.: Continuous control with deep reinforcement learning. arXiv 2015 arXiv:1509.02971

21. Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., Riedmiller, M.: Deterministic policy gradient algorithms. In International conference on machine learning (pp. 387–395). PMLR (2014)
22. Fujimoto, S., Hoof, H., Meger, D.: Addressing function approximation error in actor-critic methods. In: Proceedings of theInternational Conference on Machine Learning, Stockholm, Sweden, 10–15, pp. 1587–1596 (2018)
23. Haarnoja, T., et al. Soft actor-critic algorithms and applications. arXiv 2018 arXiv:1812.05905
24. Ngo, H.Q.T.: Design of automated system for online inspection using the convolutional neural network (CNN) technique in the image processing approach. Results Eng., 101346 (2023)
25. Phan, L.A.D., Ngo, H.Q.T.: Application of the artificial intelligence technique to recognize and analyze from the image data. In: Deep Learning and Other Soft Computing Techniques: Biomedical and Related Applications (pp. 77–89). Springer Nature Switzerland, Cham. https://doi.org/10.1007/978-3-031-29447-1_8
26. Sperduti, A., Starita, A. Supervised neural networks for the classification of structures. IEEE Trans. Neural Netw. **8**, 714–735 (1997)
27. Morris, C., et al.: Weisfeiler and Leman go neural: higher-order graph neural networks. In: Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Conference and Artificial Intelligence, Honolulu, HI, USA, vol. 33, pp. 4602–4609 (2019)
28. Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, pp. 1025–1035 (2017)
29. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. Graph attention networks. arXiv 2017 arXiv:1710.10903
30. Beltran-Hernandez, C.C., Petit, D., Ramirez-Alpizar, I.G., Harada, K.: Learning to grasp with primitive shaped object policies. In: 2019 IEEE/SICE International Symposium on System Integration (SII) (pp. 468–473). IEEE (2019)
31. Mousavian, A., Eppner, C., Fox, D.: 6-dof graspnet: variational grasp generation for object manipulation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2901–2910 (2019)
32. Berscheid, L., Meißner, P., Kröger, T.: Self-supervised learning for precise pick-and-place without object model. IEEE Robot. Automation Lett. **5**(3), 4828–4835 (2020)
33. Zeng, A., Song, S., Lee, J., Rodriguez, A., Funkhouser, T.: Tossingbot: learning to throw arbitrary objects with residual physics. IEEE Trans. Rob. **36**(4), 1307–1319 (2020)
34. Shao, Q., et al.: Suction grasp region prediction using self-supervised learning for object picking in dense clutter. In 2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR), pp. 7–12. IEEE (2019)
35. Zakka, K., Zeng, A., Lee, J., Song, S.: Form2fit: learning shape priors for generalizable assembly from disassembly. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 9404–9410. IEEE (2020)
36. Cao, H.G., Zeng, W., Wu, I.C.: Reinforcement learning for picking cluttered general objects with dense object descriptors. In: 2022 International Conference on Robotics and Automation (ICRA), pp. 6358–6364. IEEE (2022)
37. Zeng, A., et al.: Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. Int. J. Robot. Res. **41**(7), 690–705 (2022)
38. Zeng, A., Song, S., Welker, S., Lee, J., Rodriguez, A., Funkhouser, T.: Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4238–4245). IEEE (2018)
39. Ren, D., Ren, X., Wang, X., Digumarti, S.T., Shi, G.: Fast-Learning Grasping and Pre-Grasping via Clutter Quantization and Q-map Masking. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3611–3618. IEEE (2021)

40. Tang, B., Corsaro, M., Konidaris, G., Nikolaidis, S., Tellex, S.: Learning collaborative pushing and grasping policies in dense clutter. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 6177–6184. IEEE (2021)
41. Zhang, H., et al.: Reinforcement learning based pushing and grasping objects from ungraspable poses. arXiv preprint arXiv:2302.13328 (2023)
42. Berscheid, L., Meißner, P., Kröger, T.: Robot learning of shifting objects for grasping in cluttered environments. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 612–618. IEEE (2019)
43. Hundt, A., et al.: Good robot!: efficient reinforcement learning for multi-step visual tasks with sim to real transfer. IEEE Robot. Autom. Letters **5**(4), 6724–6731 (2020)
44. Yang, Y., Liang, H., Choi, C.: A deep learning approach to grasping the invisible. IEEE Robot. Autom. Lett. **5**(2), 2232–2239 (2020)
45. Xu, K., Yu, H., Lai, Q., Wang, Y., Xiong, R.: Efficient learning of goal-oriented push-grasping synergy in clutter. IEEE Robot. Autom. Lett. **6**(4), 6337–6344 (2021)
46. Huang, B., Han, S.D., Yu, J., Boularias, A.: Visual foresight trees for object retrieval from clutter with nonprehensile rearrangement. IEEE Robot. Autom. Lett. **7**(1), 231–238 (2021)
47. Chebotar, Y., et al.: Actionable models: unsupervised offline reinforcement learning of robotic skills. arXiv preprint arXiv:2104.07749 (2021)
48. Ren, D., Wu, S., Wang, X., Peng, Y., Ren, X.: Learning bifunctional push-grasping synergistic strategy for goal-agnostic and goal-oriented tasks. arXiv preprint arXiv:2212.01763 (2022)
49. Novkovic, T., Pautrat, R., Furrer, F., Breyer, M., Siegwart, R., Nieto, J.: Object finding in cluttered scenes using interactive perception. In 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 8338–8344. IEEE (2020)
50. Chen, Y., Ju, Z., Yang, C.: Combining reinforcement learning and rule-based method to manipulate objects in clutter. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–6. IEEE (2020)
51. Luo, J., Solowjow, E., Wen, C., Ojea, J.A., Agogino, A.M.: Deep reinforcement learning for robotic assembly of mixed deformable and rigid objects. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2062–2069. IEEE (2018)
52. De Winter, J., De Beir, A., El Makrini, I., Van de Perre, G., Nowé, A., Vanderborght, B.: Accelerating interactive reinforcement learning by human advice for an assembly task by a cobot. Robotics **8**(4), 104 (2019)
53. Li, F., Jiang, Q., Zhang, S., Wei, M., Song, R.: Robot skill acquisition in assembly process using deep reinforcement learning. Neurocomputing **345**, 92–102 (2019)
54. Kristensen, C.B., Sørensen, F.A., Nielsen, H.B., Andersen, M.S., Bendtsen, S.P., Bøgh, S.: Towards a robot simulation framework for e-waste disassembly using reinforcement learning. Procedia Manuf. **38**, 225–232 (2019)
55. Kim, Y.L., Ahn, K.H., Song, J.B.: Reinforcement learning based on movement primitives for contact tasks. Robot. Comput. Integrated Manuf. **62**, 101863 (2020)
56. Ota, K., Jha, D. K., Oiki, T., Miura, M., Nammoto, T., Nikovski, D., & Mariyama, T. (2019, November). Trajectory optimization for unknown constrained systems using reinforcement learning. In 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS) (pp. 3487–3494). IEEE
57. Müller, R., Vette, M., Geenen, A.: Skill-based dynamic task allocation in human-robot-cooperation with the example of welding application. Procedia Manufacturing **11**, 13–21 (2017)
58. Liu, Q., Liu, Z., Xiong, B., Xu, W., Liu, Y.: Deep reinforcement learning-based safe interaction for industrial human-robot collaboration using intrinsic reward function. Adv. Eng. Inform. **49**, 101360 (2021)

59. Haage, M., et al.: Teaching assembly by demonstration using advanced human robot interaction and a knowledge integration framework. Procedia Manufacturing **11**, 164–173 (2017)
60. Zanchettin, A.M., Casalino, A., Piroddi, L., Rocco, P.: Prediction of human activity patterns for human–robot collaborative assembly tasks. IEEE Trans. Industr. Inf. **15**(7), 3934–3942 (2018)
61. Wang, W., Li, R., Chen, Y., Diekel, Z.M., Jia, Y.: Facilitating human–robot collaborative tasks by teaching-learning-collaboration from human demonstrations. IEEE Trans. Autom. Sci. Eng. **16**(2), 640–653 (2018)
62. Akkaladevi, S.C., Plasch, M., Pichler, A., Ikeda, M.: Towards reinforcement based learning of an assembly process for human robot collaboration. Procedia Manufacturing **38**, 1491–1498 (2019)
63. Wu, M., He, Y., Liu, S.: Adaptive impedance control based on reinforcement learning in a human-robot collaboration task with human reference estimation. Int J Mech Control **21**(1), 21–31 (2020)
64. Zhang, J., Liu, H., Chang, Q., Wang, L., Gao, R.X.: Recurrent neural network for motion trajectory prediction in human-robot collaborative assembly. CIRP Ann. **69**(1), 9–12 (2020)
65. Yu, T., Huang, J., Chang, Q.: Optimizing task scheduling in human-robot collaboration with deep multi-agent reinforcement learning. J. Manuf. Syst. **60**, 487–499 (2021)
66. Wang, Y.Q., Hu, Y.D., El Zaatari, S., Li, W.D., Zhou, Y.: Optimised learning from demonstrations for collaborative robots. Robotics and Computer-Integrated Manufacturing **71**, 102169 (2021)
67. Wang, W., Li, R., Chen, Y., Sun, Y., Jia, Y.: Predicting human intentions in human–robot handover tasks through multimodal learning. IEEE Trans. Autom. Sci. Eng. **19**(3), 2339–2353 (2021)
68. Zhang, R., Lv, Q., Li, J., Bao, J., Liu, T., Liu, S.: A reinforcement learning method for human-robot collaboration in assembly tasks. Robotics and Computer-Integrated Manufacturing **73**, 102227 (2022)
69. Deng, Z., Mi, J., Han, D., Huang, R., Xiong, X., Zhang, J.: Hierarchical robot learning for physical collaboration between humans and robots. In 2017 IEEE international conference on robotics and biomimetics (robio), pp. 750–755. IEEE (2017)
70. Ghadirzadeh, A., Bütepage, J., Maki, A., Kragic, D., Björkman, M.: A sensorimotor reinforcement learning framework for physical human-robot interaction. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2682–2688. IEEE (2016)
71. Zhao, W., Queralta, J.P., Westerlund, T.: Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In 2020 IEEE symposium series on computational intelligence (SSCI), pp. 737–744. IEEE (2020)
72. Tao, F., Qi, Q.: Make more digital twins. Nature **573**(7775), 490–491 (2019)
73. Vinyals, O., et al.: Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature **575**(7782), 350–354 (2019)
74. De Bruin, T., Kober, J., Tuyls, K., Babuska, R.: Experience selection in deep reinforcement learning for control. J. Mach. Learn. Res. **19** (2018).
75. Zhao, T. Z., Luo, J., Sushkov, O., Pevceviciute, R., Heess, N., Scholz, J., Levine, S.: Offline meta-reinforcement learning for industrial insertion. In: 2022 International Conference on Robotics and Automation (ICRA) (pp. 6386–6393). IEEE (2022)