

Research Article

A Visual Grasping Strategy for Improving Assembly Efficiency Based on Deep Reinforcement Learning

Yongzhi Wang¹, Sicheng Zhu,² Qian Zhang,^{1,3} Ran Zhou,¹ Rutong Dou,¹ Haonan Sun,⁴ Qingfeng Yao,⁵ Mingwei Xu,¹ and Yu Zhang¹

¹Department of Mechanical Engineering, Shenyang University of Technology, Shenyang 110000, China

²College of Information Science and Engineering, Zhejiang University, Hangzhou 310058, China

³Department of Robotics, Ritsumeikan University, Shiga 525-8577, Japan

⁴Department of Digital Factory, Chinese Academy of Sciences, Shenyang 110000, China

⁵College of Science, Shenyang University of Chemical Technology, Shenyang 110000, China

Correspondence should be addressed to Yu Zhang; zhangyu@sut.edu.cn

Received 12 May 2021; Revised 29 June 2021; Accepted 10 July 2021; Published 24 July 2021

Academic Editor: Kelvin Wong

Copyright © 2021 Yongzhi Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The adjustment times of the attitude alignment are fluctuated due to the fluctuation of the contact force signal caused by the disturbing moments in the compliant peg-in-hole assembly. However, these fluctuations are difficult to accurately measure or definition as a result of many uncertain factors in the working environment. It is worth noting that gravitational disturbing moments and inertia moments significantly impact these fluctuations, in which the changes of the peg concerning the mass and the length have a crucial influence on them. In this paper, a visual grasping strategy based on deep reinforcement learning is proposed for peg-in-hole assembly. Firstly, the disturbing moments of assembly are analyzed to investigate the factors for the fluctuation of assembly time. Then, this research designs a visual grasping strategy, which establishes a mapping relationship between the grasping position and the assembly time to improve the assembly efficiency. Finally, a robotic system for the assembly was built in V-REP to verify the effectiveness of the proposed method, and the robot can complete the training independently without human intervention and manual labeling in the grasping training process. The simulated results show that this method can improve assembly efficiency by 13.83%. And, when the mass and the length of the peg change, the proposed method is still effective for the improvement of assembly efficiency.

1. Introduction

The application and development of robots in the industrial field have been developed by leaps and bounds in the past two decades [1–9]; thus, they are gradually integrated into people's daily live. Robots are used to replace humans for completing work in many scenes. The traditional control methods of the robot are the patterns of hard coding in a structural environment. These methods limit robotics' adaptability and manufacturing flexibility, which increases the labor cost and reduces the suitable range of robotic application situations. Moreover, the traditional robot control methods have a huge gap with human intelligence when sensing the environment or learning some skills. Therefore, it becomes the main development direction in the field of

robot control that robots were trained to learn skills. The ways of perceiving the environment of humans are gradually implemented on robots through bionic means, such as visual sensors and force sensors [10–15]. The motion accuracy of robotic control has surpassed humans, such as speed, distance, and angle. Nevertheless, robotic intelligence has not yet met the demands of humans in the work of learning skills. In order to enable robots to acquire new knowledge or skills autonomously, researchers use the methods of machine learning to continuously improve robotic performance by training. Ultimately, robots can imitate or realize the learning behavior of humans [16–21].

The grasping function is the most basic manipulating function of robots in industry and daily life, and it is also the foundation of many complex manipulating actions [22–

26]. The assembly task is a complex robotic task, which often requires the grasping function. Therefore, researchers have been conducted much research in the field of grasping in recent years. The feature algorithm of deep learning based on multimodal group regularization has been able to do not rely on the hand design of features in the task of RGB-D image detection for robot grasping [27]. And it got better performance than the previous hand design of the features. The eye-hand coordination system with deep learning can perform real-time servo compensation, which does not depend on camera calibration and robot posture [28]. The deep learning method can solve the problem of grasping prediction well and has been able to be designed without relying on artificial features, which has greatly reduced the cost of learning. However, deep learning methods often require a large number of data sets to complete the analysis, and the results after training are very dependent on the quality of the data sets. This limits its range of application to a certain extent. Deep learning has good analysis and perception skills, but it lacks decision-making skills. This also limits the usage scenarios of grasping strategies based on deep learning. Deep learning needs an expensive cost to build a large number of data sets in practice, and even difficult to achieve it. Therefore, it is a good solution to use the trial and error method based on reinforcement learning to make the robot collect data sets autonomously. Therefore, the grasping method based on reinforcement learning has been widely studied. For example, the viewpoint optimization strategy based on reinforcement learning uses active vision to optimize the visual sensor viewpoint [29], which can make the grasping decisions with some information missing through the trial and error of reinforcement learning without relying on multi-angle image acquisition. And this method relaxes the assumption of the sensor viewpoint and improves the grasping success rate. In addition, the hierarchical strategy of reinforcement learning can automatically learn multiple grasping strategies to solve the limitation of a single grasping type for the robot system [30]. Low-level strategies learn how to grasp specific locations with specific grasping types, and high-level strategies learn how to choose grasping types and locations. This strategy can generate a grasping strategy from a given grasping position. Although reinforcement learning has good decision-making ability, it is limited to discrete action space due to the limitation of computing power, which limits its application range and makes it difficult to deal with the problems related to continuous action space. But many practical problems are working in continuous action space. Hence, scholars have carried out numerous studies with regard to deep reinforcement learning, which combines the perception ability of deep learning and the decision-making ability of reinforcement learning [31, 32]. And it achieves direct control from the original input to output through the end-to-end learning method. Subsequently, researchers proposed many grasping methods based on deep reinforcement learning. The visual grasping method based on deep reinforcement learning can output the predicted reward of all possible actions in the current state just by inputting the observation image and, then, choose the optimal action [33, 34]. The robot is entirely self-supervised to improve

the success rate for grasps by trial and error. Besides, in the real environment, the visual grasping method based on deep reinforcement learning does not need fine-tuning to successfully grasp previously seen objects, and even it can successfully grasp previously unseen semicompliant objects [35]. Therefore, deep reinforcement learning is more suitable to deal with the grasping problem for assembly in continuous action space.

The task of peg-in-hole is a classic assembly task. It is one of the basics for many complex assembly tasks [36]. In recent years, the research of the peg-in-hole assembly has also made many novel methods. For example, the automatic alignment method based on force/torque establishes a three-point contact model, which analyzes the autonomous correction before insertion through force analysis and geometric analysis [37]. In addition, the screw insertion method in the peg-in-hole assembly reduces the axial friction force by rotating shaft compensation and improves the collision contact of the peg and the hole during assembly [38]. Moreover, the compliance control method without force feedback can analyze the current contact state between the hole and the peg, which overcomes the unavoidable positional uncertainty in the identification process [39]. And the peg-in-hole assembly can be completed without relying on expensive force sensors or remote compliance machinery. Additionally, the assembly strategy of the variable compliance center has designed an elastic displacement device [40]. This method combines the advantages of active compliance and passive compliance without force/torque sensors, which simplifies the control system. This method can well solve alignment errors. The traditional control method in the peg-in-hole assembly has obtained many research results, but the traditional control method is limited to the specific working environment. Traditional assembly robots require a great number of parameters to be deployed before work. Therefore, the research of peg-in-hole assembly in a nonstructural environment is still a challenge. However, the method of intelligent assembly robots based on deep reinforcement learning can greatly reduce the work of related manual parameters deployment [41]. It uses the robot's sensor to perceive the environment and then analyzes the system state. This method can obtain better control accuracy and robustness. Furthermore, an assembly training method with deep reinforcement learning has been designed to dispose of the uncertainty in the complicated assembly process of circuit breakers [42]. It enables the robot to autonomously learn the skill of orientation and pose adjustments in the assembly training. This method has obtained a high assembly success rate.

The core work of the peg-in-hole is to align the peg with the hole, namely, adjusting the attitude and position of the peg. The assembly alignment efficiencies are affected by many uncertain factors in the real environment during the alignment adjustment. In order to solve this problem, this article has conducted the following research:

- (1) To analyze the relationship between the grasping position and the adjustment time of the assembly alignment

- (2) To investigate the impact of assembly efficiency when the length and mass of the peg have changed
- (3) To verify the effectiveness of the proposed method in improving assembly efficiency

The remainder of this paper is organized as follows: in the Section 2, the working principle of the device is introduced, and the relationship between grasping position and assembly efficiency is analyzed. The Section 3 puts forward the visual grasping strategy and explains the details of the assembly. The simulation results and analysis results are presented in Section 4. The last section introduces the conclusion of this paper and the future work.

2. Working Principles and Analyses

2.1. Working Principles and Analyses of Assembly. Peg-in-hole is divided into search phase, alignment phase, and insertion phase. Firstly, the job of the search phase is to find the location of the hole. And then, the alignment phase is to adjust the assembly attitude of the peg to align with the hole. Finally, the insertion phase is to insert the peg into the hole to complete the assembly tasks. It is often assumed, at the early research for peg-in-hole assembly, that the peg and the hole can be well aligned before insertion. In fact, the peg may not be well aligned with the hole, which needs to adjust the position and attitude of the peg to complete the alignment with the hole. The assembly time is also prolonged as the number of adjustments increases to reduce the assembly efficiency.

There is often an inclination angle between the peg and the hole during the assembly alignment phase early. This inclination angle is a key parameter of the assembly alignment, as shown in Figure 1. It is still possible to complete the assembly when there is an inclination angle between the peg and the hole if the assembly work has an assembled clearance. This inclination angle is the maximum inclination angle δ allowed by the assembly. The formula of the maximum inclination angle δ for assembly is described as follows:

$$\delta = \arctan \frac{\zeta}{K}, \quad (1)$$

where ζ is the peg-in-hole assembled clearance and K is the assembly insertion distance.

There are three contact states during the assembly of peg-in-hole, as illustrated in Figure 2. The robot moves the peg near the plane of the hole so that the bottom of the peg is in contact with the top of the hole. This contact state is defined as plane contact, as shown in Figure 2(a). The robot uses a spiral force to sweep the surface of the part to searching the hole. The peg will incline if the center of the peg is close enough to the center of the hole. The contact state becomes a two-point contact at this time, as illustrated in Figure 2(b). The peg slides along the edge of the hole while maintaining two-point contact. When the center of the peg approaches the center of the hole to a certain range, the contact state changes to a three-point contact, as shown in

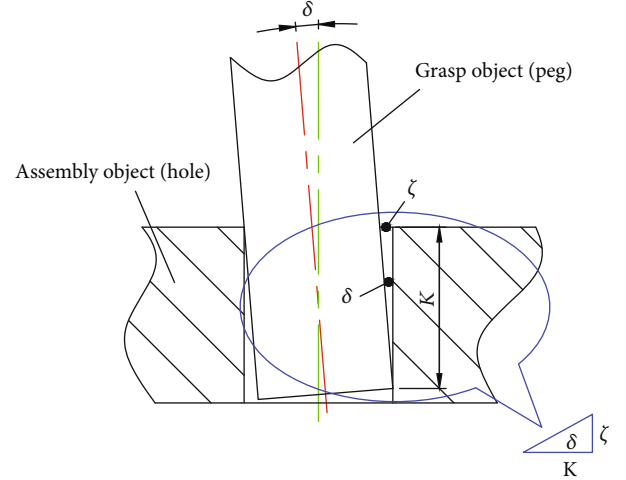


FIGURE 1: The schematic diagram of the assembly attitude angle. The red central line is the central line of the peg and the green central line is the hole's central line.

Figure 2(c). The peg needs to adjust attitude for alignment by this time.

The insertion action cannot be completed if the inclination angle is greater than the maximum inclination angle δ . That is to say, the insertion distance K is zero by this time. The training target of the robots is that the inclined angle can be faster adjusted to zero or less than the maximum inclination angle δ to insert the peg into the hole. This reduces the difficulty of assembly alignment but increases the difficulty of precise definition of the assembly model. The peg completes the alignment of adjusting attitude when the inclination angle ψ of the current alignment is adjusted to be lower than the maximum inclination angle δ or zero. The peg is then inserted into the hole to complete the assembly task. Thus, the alignment adjustment time is one of the important indicators that affect assembly efficiency.

The downward assembly force will be generated when the robot tries to insert the peg into the hole in the two-point contact state, which will generate a corresponding reaction force at the contact point. The direction of the sum of reaction forces F_{sum} always points to the center of the hole, as shown in Figure 3(a). The current inclination angle ψ_y is the angle between the line connecting the two contact points A and B and the y -axis. The peg will spontaneously slide toward the center of the hole under the action of the sum of the reaction force if the friction at the contact point is ignored. This spontaneous sliding is due to natural attraction, which is also the core control principle of the compliance-based robotic peg-in-hole. The robot with this ability can deal with the uncertainty of the hole position. F_r and F_z are the projections of the assembly force on the xy -plane and the z -axis, and F_z is consistent with the direction of gravity as illustrated in Figure 3(b). The assembly force of the peg is greater than the static friction force at the contact point if F_r and the resultant of reaction force F_{sum} are in the same direction. Thus, the peg and the hole generate relative sliding to each other, and then, the peg slides into

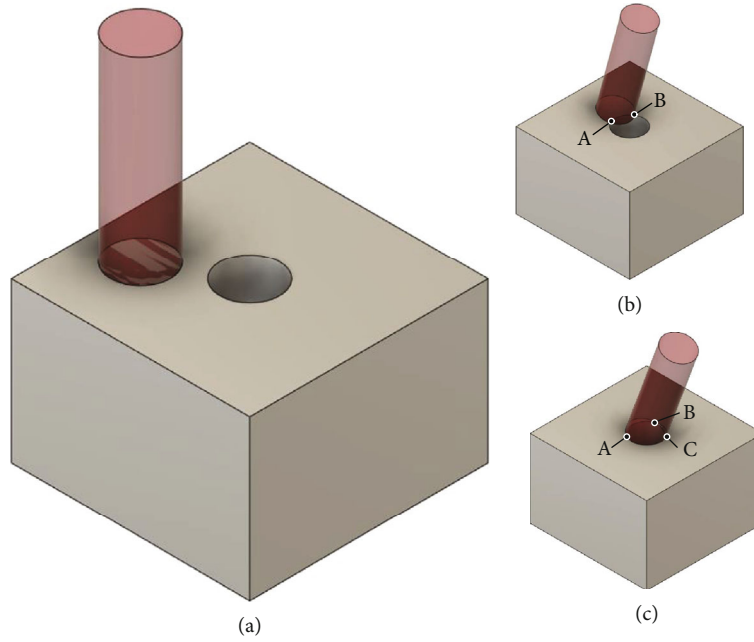


FIGURE 2: Schematic diagram of contact state. (a) Plane contact. (b) Two-point contact. (c) Three-point contact.

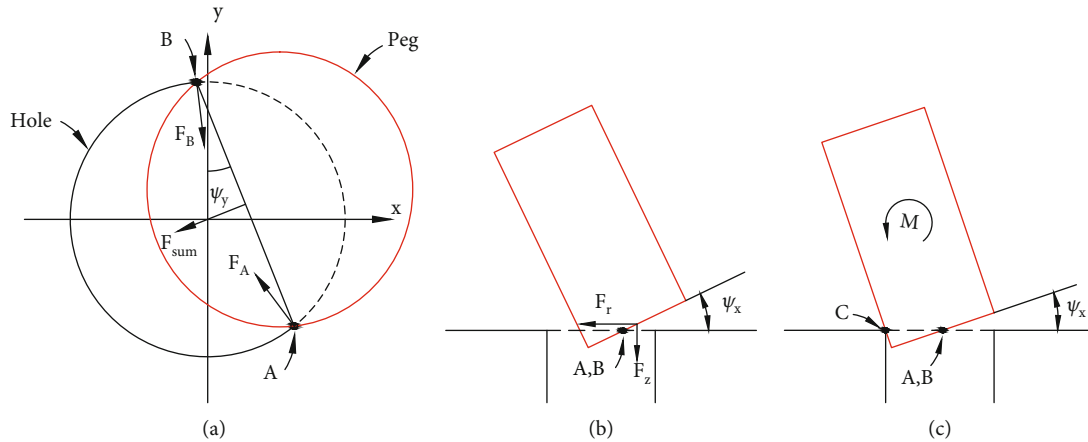


FIGURE 3: Schematic diagram of contact force. (a) Top view of two-point contact. (b) Side view of two-point contact. (c) Side view of three-point contact.

the hole. They cancel each other out if the direction of F_r and F_{sum} are opposite. The assembly force is less than the static friction force of the contact point at this time, and the peg and the hole will not slide relative to each other. It may cause the peg to miss the alignment position or slip out of the hole. The fluctuation of the disturbance moment will cause the assembly force to fluctuate suddenly, which makes the peg unable to complete the alignment. The robot then needs to readjust the alignment, which increases assembly time and reduces assembly efficiency. When the peg is attracted into the center of the hole, it can be inserted into the hole if the current inclination angle ψ_x is zero. Otherwise, the contact state changes to the state of three-point contact. The turning moment M is required to adjust the current inclination angle ψ_x to insert the peg into the hole, as shown in Figure 3(c). The peg is inserted

into the hole to complete the assembly after fulfilling the alignment adjustment.

2.2. Analyses of Disturbance for Assembly. This research is focused on the grasping position to impact the efficiency of assembly alignment, thereby improving assembly efficiency. Therefore, the search phase and the insertion phase are not researched and discussed deeply. Different grasping positions produce different alignment times for the same current inclination angle ψ , as shown in Figure 4. Different motion trajectories and disturbing moments will be produced by the different grasping positions, which produce the difference in assembly time. Among these disturbing moments, the effects of gravitational disturbing moments and inertia moments are particularly significant for assembly, which will also emerge the fluctuation when the robot adjusts the peg to

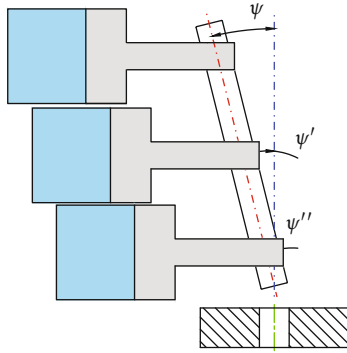


FIGURE 4: The schematic diagram of different gripping positions for adjustment alignment.

align the hole. In addition, it will be affected by the fluctuation of the disturbing moments that the robot adjusts the position and attitude of the peg. The signal fluctuations caused by disturbing moments will raise the difficulty of assembly alignment, which increases the adjustment time and reduces the efficiency of assembly work. Two special grasping positions are worth noting to reduce the influence of disturbing moments:

- (1) The grasping position is the point where the center of mass and the center of rotation coincide, which can produce the smallest disturbing moments of gravity during the adjusting alignment. But there are still the inertia moments
- (2) The grasping position is not only to coincide with the center of mass and the center of rotation but also to coincide with the inertial axis and the rotating axis, which produces minimum gravitational disturbing moments and inertia moments

The formula of gravitational disturbing moment M_G is shown as follows:

$$M_G = mg|l| \sin \beta, \quad (2)$$

where m is the mass of the peg, g is the gravitational acceleration, l is the distance from the rotation axis to the force functional point, and β is the angle between the gravity moment and the vector (i.e., $\beta = 90^\circ - \delta$).

The formula of the inertia couples M_I generated by the moment of inertia J is described as follows:

$$M_I = J \frac{d^2 \beta}{dt^2} = \frac{1}{2} mr^2 \frac{d^2 \beta}{dt^2}, \quad (3)$$

where r is the vertical distance between the center of mass and the rotating axis and $d^2 \beta / dt^2$ is the angular acceleration.

It can be seen from the above formula that the mass of the peg and the operating distance have an important influence on the gravitational disturbing moments and inertia moments. The position of the center of mass becomes uncertain due to the manufacturing error of the peg in the same

manufacturing batch. In addition, there are different coincidence degrees between the inertial axis and the rotating axis because of the different grasping positions. And the alignment process of adjusting attitude will move disparate distances and motion trajectories even when the peg has the same inclination angle. These factors have aggravated the uncertainty of the adjustment time. Therefore, selecting a suitable grasping position in the process of assembly can effectively reduce the fluctuation of the disturbing moments on the alignment adjustment, which is mainly caused by changes in mass, volume, and operating distance. The traditional control method for the robot cannot handle these complex and changeable assembly tasks. Hence, we hope to train the robot through the training method of deep reinforcement learning so that the robot can autonomously deal with these assembly tasks in an unstructured environment. Robots often need multiple times to adjust attitude in the alignment stage. The compliant control needs to constantly judge the current attitude based on the contact force. Some uncertain factors cause the fluctuation of the signal of contact force, which increases the difficulty of alignment. In particular, the gravitational disturbing moments and the inertia moments have a prominent influence on this fluctuation, which will lead to a prolonged time for alignment adjustment and ultimately reduce assembly efficiency. Therefore, the cost of time on the alignment stage can be reduced if the robot can reduce the fluctuation of the disturbing moments. Finally, the improvement of assembly efficiency is realized. Traditional control methods cannot handle these uncertain fluctuations of disturbing moments. However, artificial calibration of mechanical parameters or grasping positions is not only cumbersome, but also has certain errors, or even impossible to achieve. This difficulty can be avoided through trial and error learning based on deep reinforcement learning, which does not require artificial labels and prior knowledge of mechanical parameters. When the grasping position is restricted to a certain area with the proposed method, which is considered to improve assembly efficiency if the trained robot expends less assembly time than the untrained robot.

3. Assembly System with Deep Reinforcement Learning

The assembly task of peg-in-hole is divided into two branch tasks: grasping task and assembly task. Therefore, the robot is equipped with the grasping module and assembly module. The grasping task refers to the robot grasping the assembly peg before performing assembly. The assembly task is divided into three stages: searching, alignment, and insertion. This research proposes a visual grasping strategy to boost assembly efficiency by improving its grasping strategy based on the analysis in Section 2.2.

3.1. Grasping Module with Visual Grasping Strategy. The decision-making process of the grasping module is regarded as a Markov decision-making process. The grasping workflow is transformed into an interactive process that can be expressed in probabilistic form through the Markov decision process. Firstly, the robot observes the environmental state s_t

at time t and selects the performing action a_t from the available action set $A(s)$ through the strategy $\pi(s_t)$. And then the environmental state changes from s_t to s_{t+1} , meanwhile, the reward $R(s_t, s_{t+1})$ is obtained. The state-action-reward chain in the grasping decision-making process can be expressed as follows:

$$\{s_0, a_0, R_1, s_1, a_1, \dots, R_{t-1}, s_{t-1}, a_{t-1}, R_t, s_t\}. \quad (4)$$

The reward $R(s_t, s_{t+1})$ is composed of the grasping reward r_{t+1}^G and the assembly reward r_{t+1}^{AM} . The robot obtains a grasping reward $r_{t+1}^G = 0.3$ after successfully grasping the peg each time. The grasping network will also obtain an assembly reward $r_{t+1}^{AM} = 0.7$ if the assembly time is less than the threshold. The reward $R(s_t, s_{t+1})$ is described as follows:

$$R(s_t, a_t) = R_t = r_{t+1}^G + r_{t+1}^{AM}. \quad (5)$$

The training purpose of deep reinforcement learning is to obtain the optimal strategy π^* , which can maximize the total reward G_t :

$$G_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \gamma^3 R_{t+3} + \dots, \quad (6)$$

$$G_t = \sum_{t=0}^{\infty} \gamma^t R_t = \sum_{k=0}^{\infty} \gamma^t r_{t+k+1}.$$

The target of improving assembly efficiency is fulfilled if the robot can maximize the total reward G_t by establishing the mapping relationship between the grasping position and the assembly time. Therefore, the robot trained a greedy deterministic policy $\pi(s_t)$ using off-policy Q-learning, which chooses action a_t by maximizing the action-value function $Q_\pi(s_t, a_t)$:

$$Q_\pi(s_t, a_t) = \mathbb{E}_\pi [R_t | s_t = s, a_t = a]. \quad (7)$$

The optimal action-value function is expressed as follows:

$$Q^*(s_t, a_t) = \max_{\pi} Q_\pi(s_t, a_t) = R(s_t, a_t) + \gamma \max_{a \in A(s)} Q_\pi(s_{t+1}, a_{t+1}), \quad (8)$$

where γ is the future discount, which is set to a constant $\gamma = 0.5$.

The optimal strategy π^* , which was obtained by training, can select the optimal action a_t^* with the highest Q value from the set of available action $A(s)$ in the current state s_t . The formula of optimal strategy π^* is as follows:

$$\pi^*(s_t) = a_t^* = \operatorname{argmax}_{a \in A(s)} Q^*(s_t, a_t). \quad (9)$$

The fully convolutional networks based on DQN and DenseNet are used to build the network of grasping decision-making in this paper. The networks take the heightmap describing the observing environmental state s_t as input, which outputs a dense pixel-wise map of Q values with the

same size and resolution as the input. Any pixel point in the image has a Q value, which predicts the future reward of performing the grasping action a_t at the spatial position. To begin with, the agent observes the information of the environment to get the visual data, and then, it is reprojected onto the orthographic RGB-D heightmap. Whereafter, the color channel (RGB) and the clone depth channel (DDD) of the heightmap are input to two parallel 121-layer DenseNets to process the image features. And then, the image after channel-wise concatenation is sent to 3 additional 1×1 convolutional layers interleaved with ReLU activation functions and BatchNorm. Finally, the pixel-level probability map with Q value is obtained after bilinearly upsampled processing, and it is the same as the input image resolution by 224×224 . The robot will choose the performing action with the highest Q value based on this probability map. The grasping strategy has two fully convolutional neural networks with the same structure: target network and evaluation network. They have the same network architecture and initial network parameters. Firstly, the target network selects the action a_t with the highest Q value according to the strategy $\pi(s_t)$. Afterward, the evaluation network will evaluate this action. And two networks output Q_{tar} and Q_{eva} , respectively. The evaluation network updates the network parameters θ_i in real-time through the backpropagation operation according to the reward $R(s_t, a_t)$. But the target network only performs forward propagation operations, and it updates the network parameters θ_i' of the target network by copying the parameters θ_i of the evaluation network after completing a batch of iterative training, that is, $\theta_i' \leftarrow \theta_i$. The robot is considered to have completed training when the difference ΔQ in the predicted Q value between the target network and the evaluation network is less than the threshold through continuous iteration. ΔQ is described as follows:

$$\Delta Q = |Q_{\text{tar}} - Q_{\text{eva}}|. \quad (10)$$

The evaluation network uses the Huber loss function \mathcal{L}_i as follows:

$$\mathcal{L}_i = \begin{cases} 0.5 \times \left(Q_{\text{tar}}^{\theta_i'} - Q_{\text{eva}}^{\theta_i} \right)^2, & \text{for } |Q_{\text{tar}} - Q_{\text{eva}}| < 1 \\ \left| \left(Q_{\text{tar}}^{\theta_i'} - Q_{\text{eva}}^{\theta_i} \right) - 0.5 \right|, & \text{otherwise.} \end{cases} \quad (11)$$

3.2. Compliance-Based Assembly Module. The assembly module is based on compliant behavior control, which completes the peg-in-hole assembly by analyzing the contact state between the peg and hole to generate compliance behavior. The assembly module divides the assembly work into three stages: hole-searching stage, alignment stage, and insertion stage. The robot moves the peg to the surface of the hole after successfully grasping the peg. And the contact between the peg and the hole results in a plane contact state of the peg. The robot enters the hole-finding stage at the time. In order to simulate the uncertainty of the hole position during work, the initial position of the hole is randomly placed within a

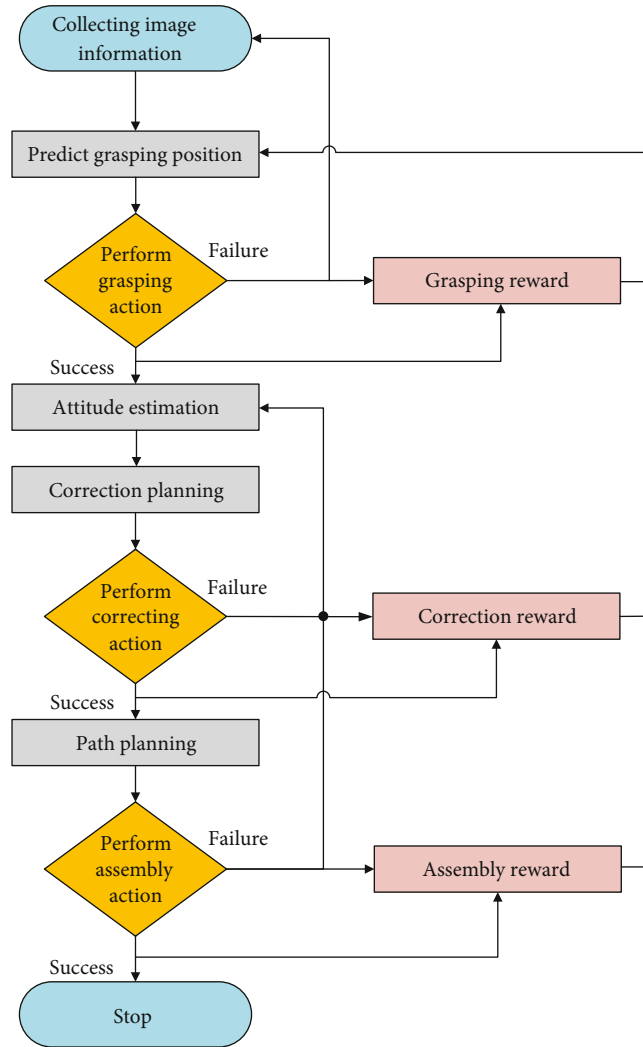


FIGURE 5: Flowchart of peg-in-hole procedure.

small range, which is equal to the area of the hole. Afterward, the robot searches holes on the surface through rubbing motion with the trajectory of Archimedes spiral. The peg will be inclined if the peg is close enough to the hole. And then, the contact state will change to two-point contact. Subsequently, the peg slides along the edge of the hole. There will be a three-point contact state when the peg is close enough to the center of the hole but still has an inclination angle. The alignment of the peg and the hole is completed by using the wiggling motion to adjust the error of the inclining angle. Finally, the assembly task is finished after performing the insertion action. The proposed method and baseline both use the same assembly module to ensure the fairness of the comparison in the assembly efficiency test. However, the grasp module of the baseline method is not equipped with an assembly reward r^{AM} . The effectiveness of the proposed method was proved if adding the assembly reward r^{AM} for alignment in robot training can improve assembly efficiency. The process of peg-in-hole assembly is shown in Figure 5.

4. Simulation Results and Analyses

4.1. Training of Visual Grasping Strategy. The assembly system established in the simulation software V-REP, which uses a UR5 robotic arm with an RG2 gripper, as shown in Figure 6. And it is also equipped with the RGB-D vision sensor, force sensor, and position sensor. The length of the peg is 100 mm, and its weight is 0.55 kg. The diameter is $\phi 30$ mm, and the assembly clearance is 1 mm. The peg and the hole have not chamfered. The CPU of the simulation workstation is Intel(R) Xeon(R) Gold 5222 at 3.80 GHz, the GPU is NVIDIA GeForce RTX 3090, and it is equipped with 128 GB of RAM. The robot uses trial and error in training to explore the law of the difference in assembly alignment efficiency caused by different grasping positions. The method of stochastic gradient descent with momentum is used for the training of the grasping networks. The learning rate is a constant at 10^{-4} , and the momentum is set as 0.9. The exploration strategy is a deterministic ϵ -greedy, and its initial value is set as 0.5 and then annealed overtraining to 0.1.

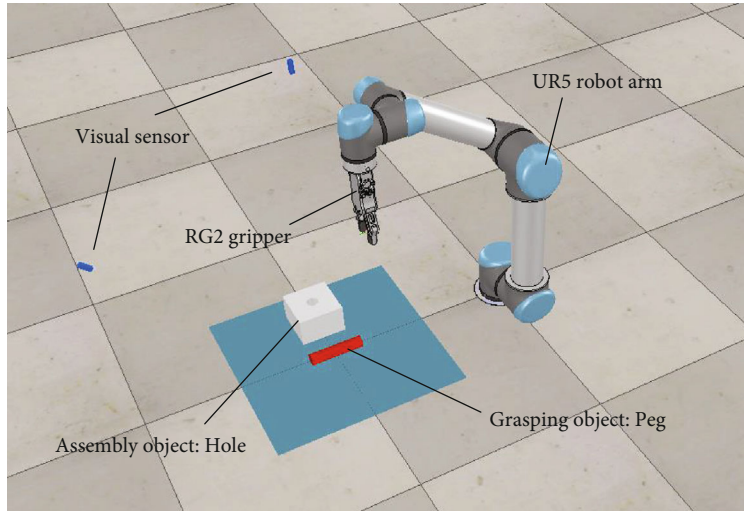


FIGURE 6: Schematic diagram of simulation system.

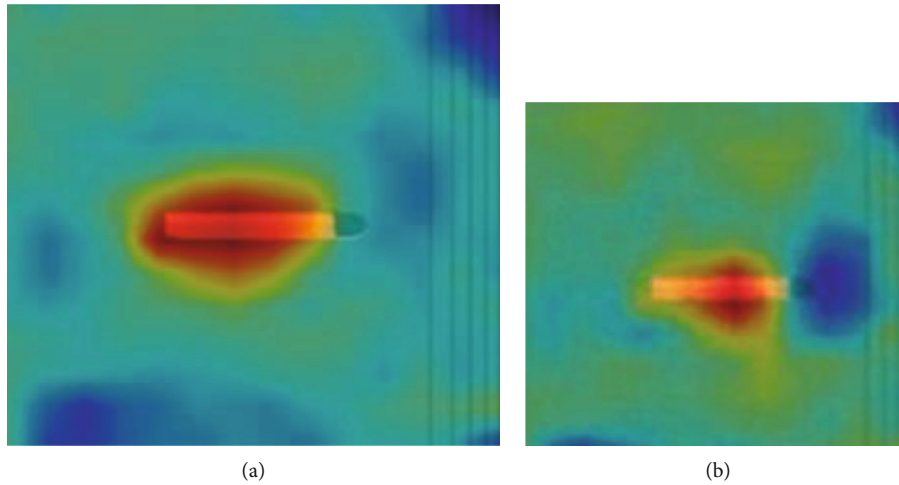


FIGURE 7: Heat maps of grasping decision-making. (a) Selection area of untrained grasping position. (b) Selection area of trained grasping position.

The first 1000 times of grasping training are to randomly select the grasping position. The purpose of random selection is to allow the robot to explore the impacts on the assembly efficiency for different grasping positions. The robot will choose the grasping position with the highest Q value in the remaining times of 4000 training. The heat maps of the grasping decision-making are shown in Figure 7. The red area represents the grasping position with a higher predicted Q value. The area where the untrained robot chooses the grasping position is spread over the whole peg, as illustrated in Figure 7(a). The robot will obtain the assembly reward r^{AM} when the assembly time is less than the threshold. The selection area of the grasping position will gradually shrink as the number of training increases, as shown in Figure 7(b). The grasping position is restricted to a specific area smaller than the previously selected area by establishing the mapping relationship between the grasping position and the adjusting time of alignment.

4.2. Simulation Test for Assembly Efficiency. The simulation tests have the purposes to prove the following two problems:

- (1) To verify whether the proposed grasping strategy can help robots improve assembly efficiency
- (2) To test whether this strategy is still effective when qualities, lengths, and mechanical parameters of the peg have changed

The proposed method is a visual grasping strategy (VGS) for the peg-in-hole task. The robot used baseline and VGS to conduct 1000 peg-in-hole assembly simulation tests, respectively, to compare the difference in assembly efficiency. The total assembly time of the baseline method is about 38.46 hours, while the total assembly time of VGS is only about 33.14 hours, which improves the assembly efficiency by 13.83%. VGS compares the distribution of the assembly time with baseline, as shown in Figure 8. In the test, the shortest

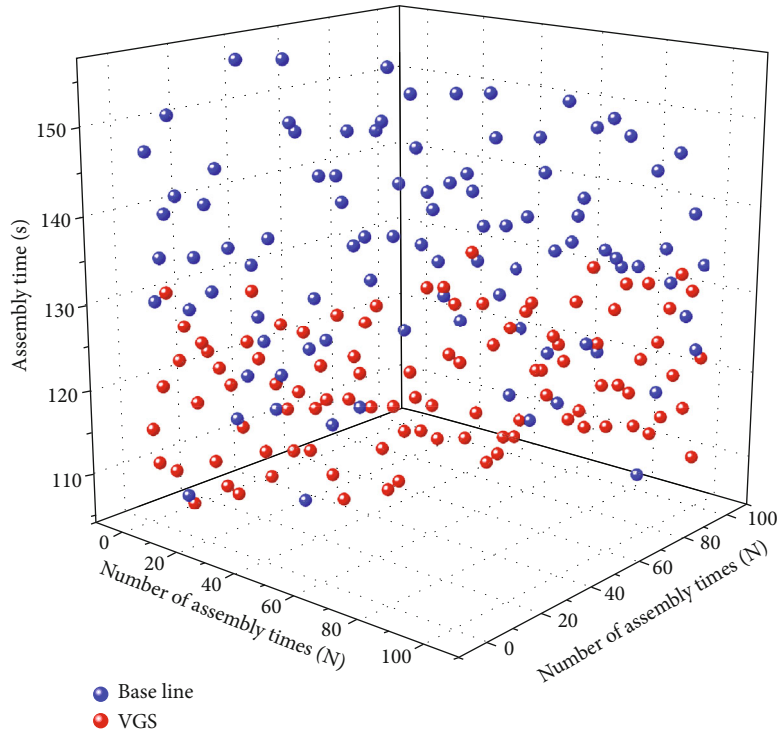


FIGURE 8: Comparison chart of assembly time.

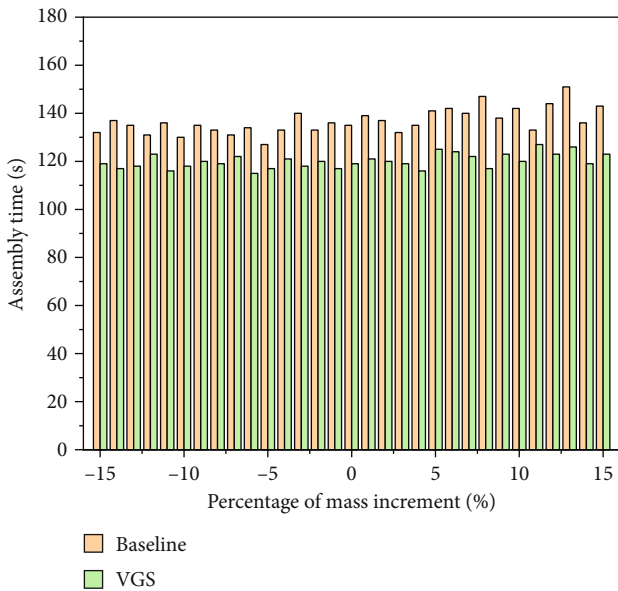


FIGURE 9: Comparing assembly time of different mass.

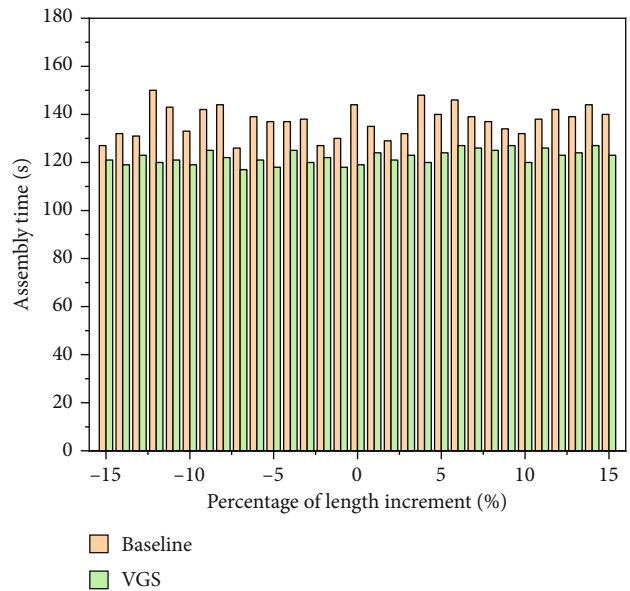


FIGURE 10: Comparing assembly time of different length and mass.

assembly time for baseline and VGS methods is both 106 seconds. But the longest time of baseline is 157 seconds, and VGS is 135 seconds. Baseline takes 19 seconds of the average assembly time more than VGS. It can be seen that the robot using VGS has a relatively shorter assembly time. The results of the simulation prove that our method can effectively improve assembly efficiency.

The change in the diameter of the peg causes a change in its mass, and these changes have certain effects on the assembly alignment. It is very cumbersome to manually calculate

and mark changes in mechanical parameters caused by these changes. The mass of the peg is additionally reduced or increased by 15% based on initial mass to imitate the random changes of the mass in the actual conditions. The robot, respectively, using baseline and VGS conducts 1000 assembly tests for the different mass of the peg, as shown in Figure 9. It can be shown that VGS can still improve the assembly efficiency by 12.13% when the diameter and mass of the peg make some changes. In this simulation result, the standard deviation of the baseline is 5.1756, and the standard deviation

of VGS is 3.0133. It can be seen that VGS has better stability for assembly efficiency.

Subsequently, the length of the peg is changed between 85% and 115% based on the original length of the peg. Not only has its mass been changed but also its mechanical parameters have been changed when the length of the peg changes. The robot, respectively, using baseline and VGS conducts 1000 assembly tests for the different length and mass of the peg, as shown in Figure 10. The result shows that VGS can also improve the assembly efficiency by 10.92%, even if the length and mass of the peg have certain changes. When the length and mass have changed, the standard deviations of baseline and VGS are 6.2242 and 3.8508, respectively. Obviously, VGS has a smaller fluctuation of the assembly time, and the assembly efficiency is more stable.

5. Conclusions and Future Work

In this paper, a visual grasping strategy based on deep reinforcement learning is proposed, which can improve assembly efficiency. The fluctuation of the contact force signal caused by the disturbing moments in compliance-based assembly is analyzed, and the visual grasping strategy introduces the assembly reward to reduce the fluctuation. In V-REP, the simulations of peg-in-hole are carried out, it can be obtained from simulation results that the grasping area is restricted at a special area less than the previous area, and the trained robot spends less assembly time than the untrained robot. Furthermore, the proposed method improves the assembly efficiency by 13.83% compared to baseline.

The proposed visual grasping strategy can also effectively improve the assembly efficiency when the size and mechanical parameters of the peg have changed, which provides some guidance in peg-in-hole assembly. Future research work will focus on extending the proposed strategy to different assembly parts to complete more complex tasks. At the same time, finding effective ways to improve the efficiency of training samples, assisting agents to obtain better assembly capabilities, and realizing multiagent collaborative assembly are also future research work.

Data Availability

The data used to support this study are available at <https://github.com/Bensonwyz/A-Grasping-Strategy-for-Improving-Assembly-Efficiency-based-on-Deep-Reinforcement-Learning>.

Conflicts of Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work; there is no professional or other personal interest of any nature or kind in any product, service, and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

References

- [1] T. Kamegawa, T. Akiyama, S. Sakai et al., “Development of a separable search-and-rescue robot composed of a mobile robot and a snake robot,” *Advanced Robotics*, vol. 34, no. 2, pp. 132–139, 2020.
- [2] L. el Hafi, S. Isobe, Y. Tabuchi et al., “System for augmented human–robot interaction through mixed reality and robot training by non-experts in customer service environments,” *Advanced Robotics*, vol. 34, no. 3–4, pp. 157–172, 2020.
- [3] D. Ferone, P. Festa, and M. G. C. Resende, “Hybridizations of GRASP with path relinking for the far from most string problem,” *International Transactions in Operational Research*, vol. 23, no. 3, pp. 481–506, 2016.
- [4] J. Dufek and R. Murphy, “Visual pose estimation of rescue unmanned surface vehicle from unmanned aerial system,” *Frontiers in Robotics and AI*, vol. 6, p. 42, 2019.
- [5] N. Vahrenkamp, E. Koch, M. Wachter, and T. Asfour, “Planning high-quality grasps using mean curvature object skeletons,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 911–918, 2018.
- [6] D. Watkins-Valls, J. Varley, and P. Allen, “Multi-modal geometric learning for grasping and manipulation,” in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7339–7345, Montreal, QC, Canada, 2019.
- [7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: a large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, FL, USA, 2009.
- [8] Y. Lin, S. Ren, M. Clevenger, and Y. Sun, “Learning grasping force from demonstration,” in *2012 IEEE International Conference on Robotics and Automation*, pp. 1526–1531, Saint Paul, MN, USA, 2012.
- [9] S. Ma, L. Du, E. Tsuchiya, and M. Fuchimi, “Paper-made grippers for soft food grasping,” in *2020 17th International Conference on Ubiquitous Robots (UR)*, pp. 362–367, Kyoto, Japan, 2020.
- [10] A. J. Choi, H. H. Yang, and J. H. Han, “Study on robust aerial docking mechanism with deep learning based drogue detection and docking,” *Mechanical Systems and Signal Processing*, vol. 154, p. 107579, 2021.
- [11] L. Gerez, G. Gao, and M. Liarokapis, “Employing magnets to improve the force exertion capabilities of adaptive robot hands in precision grasps,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7630–7635, Macau, China, 2019.
- [12] A. M. Schmidts, D. Lee, and A. Peer, “Imitation learning of human grasping skills from motion and force data,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1002–1007, San Francisco, CA, USA, 2011.
- [13] M. Abdeetdal and M. R. Kermani, “Grasp synthesis for purposeful fracturing of object,” *Robotics and Autonomous Systems*, vol. 105, pp. 47–58, 2018.
- [14] M. Bianchi, G. Averta, E. Battaglia et al., “Touch-based grasp primitives for soft hands: applications to human-to-robot handover tasks and beyond,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7794–7801, Brisbane, QLD, Australia, 2018.
- [15] H. Merzić, M. Bogdanović, D. Kappler, L. Righetti, and J. Bohg, “Leveraging contact forces for learning to grasp,” in *2019 international conference on robotics and automation (ICRA)*, pp. 3615–3621, Montreal, QC, Canada, 2019.

- [16] G. A. Garcia Ricardez, N. Koganti, P. C. Yang et al., "Adaptive motion generation using imitation learning and highly compliant end effector for autonomous cleaning," *Advanced Robotics*, vol. 34, no. 3-4, pp. 189–201, 2020.
- [17] X. Li, H. Cheng, H. Chen, and J. Chen, "Learning quasi-periodic robot motions from demonstration," *Autonomous Robots*, vol. 44, no. 2, pp. 251–266, 2020.
- [18] X. Xie, C. Li, C. Zhang, Y. Zhu, and S.-C. Zhu, "Learning virtual grasp with failed demonstrations via bayesian inverse reinforcement learning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019.
- [19] J. Zimmer, T. Hellebrekers, T. Asfour, C. Majidi, and O. Kroemer, "Predicting grasp success with a soft sensing skin and shape-memory actuated gripper," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 7120–7127, Macau, China, 2019.
- [20] X. Tan, Y. Lee, C. B. Chng, K. B. Lim, and C. K. Chui, "Robot-assisted flexible needle insertion using universal distributional deep reinforcement learning," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 2, pp. 341–349, 2020.
- [21] D. Wang, Y. Hu, and T. Ma, "Mobile robot navigation with the combination of supervised learning in cerebellum and reward-based learning in basal ganglia," *Cognitive Systems Research*, vol. 59, pp. 1–14, 2020.
- [22] J. Mahler, R. Platt, A. Rodriguez et al., "Guest editorial open discussion of robot grasping benchmarks, protocols, and metrics," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 4, pp. 1440–1442, 2018.
- [23] A. K. Tanwani, N. Mor, J. Kubiawicz, J. E. Gonzalez, and K. Goldberg, "A fog robotics approach to deep robot learning: application to object recognition and grasp planning in surface decluttering," in *2019 international conference on robotics and automation (ICRA)*, pp. 4559–4566, Montreal, QC, Canada, 2019.
- [24] B. Wu, I. Akinola, and P. K. Allen, "Pixel-attentive policy gradient for multi-fingered grasping in cluttered scenes," 2019, <https://arxiv.org/abs/1903.03227>.
- [25] J. Mahler, M. Matl, V. Satish et al., "Learning ambidextrous robot grasping policies," *Science Robotics*, vol. 4, no. 26, article eaau4984, 2019.
- [26] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 3, pp. 2460–2465, Seoul, Korea (South), 2001.
- [27] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.
- [28] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.
- [29] B. Calli, W. Caarls, M. Wisse, and P. Jonker, "Viewpoint optimization for aiding grasp synthesis algorithms using reinforcement learning," *Advanced Robotics*, vol. 32, no. 20, pp. 1077–1089, 2018.
- [30] T. Osa, J. Peters, and G. Neumann, "Hierarchical reinforcement learning of multiple grasping strategies with human instructions," *Advanced Robotics*, vol. 32, no. 18, pp. 955–968, 2018.
- [31] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Playing atari with deep reinforcement learning," 2013, <https://arxiv.org/pdf/1312.5602/>.
- [32] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [33] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "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, Madrid, Spain, 2018.
- [34] Z. Yang and H. Shang, "Robotic Pushing and Grasping Knowledge Learning via Attention Deep Q-learning Network," in *Knowledge Science, Engineering and Management. KSEM 2020*, G. Li, H. Shen, Y. Yuan, X. Wang, H. Liu, and X. Zhao, Eds., vol. 12274 of Lecture Notes in Computer Science, pp. 223–234, Springer, Cham, 2020.
- [35] O. M. Pedersen, E. Misimi, and F. Chaumette, "Grasping unknown objects by coupling deep reinforcement learning, generative adversarial networks, and visual servoing," in *2020 IEEE international conference on robotics and automation (ICRA)*, pp. 5655–5662, Paris, France, 2020.
- [36] K. Zhang, J. Xu, H. Chen, J. Zhao, and K. Chen, "Jamming analysis and force control for flexible dual peg-in-hole assembly," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 3, pp. 1930–1939, 2018.
- [37] T. Tang, H. C. Lin, Y. Zhao, W. Chen, and M. Tomizuka, "Autonomous alignment of peg and hole by force/torque measurement for robotic assembly," in *2016 IEEE international conference on automation science and engineering (CASE)*, pp. 162–167, Fort Worth, TX, USA, 2016.
- [38] Z. Liu, L. Song, Z. Hou, K. Chen, S. Liu, and J. Xu, "Screw insertion method in peg-in-hole assembly for axial friction reduction," *IEEE Access*, vol. 7, pp. 148313–148325, 2019.
- [39] H. Park, J. Park, D. H. Lee, J. H. Park, M. H. Baeg, and J. H. Bae, "Compliance-based robotic peg-in-hole assembly strategy without force feedback," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6299–6309, 2017.
- [40] S. Wang, G. Chen, H. Xu, and Z. Wang, "A robotic peg-in-hole assembly strategy based on variable compliance center," *IEEE Access*, vol. 7, pp. 167534–167546, 2019.
- [41] T. Inoue, G. De Magistris, A. Munawar, T. Yokoya, and R. Tachibana, "Deep reinforcement learning for high precision assembly tasks," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 819–825, Vancouver, BC, Canada, 2017.
- [42] F. Li, Q. Jiang, S. Zhang, M. Wei, and R. Song, "Robot skill acquisition in assembly process using deep reinforcement learning," *Neurocomputing*, vol. 345, pp. 92–102, 2019.