Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Slabstone Installation Skill Acquisition for Dual-Arm Robot based on Reinforcement Learning

Dong Liu
Autocontrol Institute
Xi'an Jiaotong University
Xi'an, Shaanxi, 710049, P.R. China
doctorsrn@stu.xjtu.edu.cn

Jianfu Cao and Xiaokang Lei
Autocontrol Institute
Xi'an Jiaotong University
Xi'an, Shaanxi, 710049, P.R. China
cjf@mail.xjtu.edu.cn

Abstract **- Slabstone installation is so widely used in the construction industry that it is very significant to use the robot to finish this process, making the process more automatic, effective and safer. Because of the uncertain factors in environments and filling material, it is difficult to finish the final step of slabstone installation that pressing the slabstone into good contact with filling material. There is currently no mature theory and method to achieve control of the process. In this paper, we use a dual-arm robot to accomplish the final step. First, we use the spring-damping elements to model and simulate the filling material. Then, based on deep reinforcement learning, a skill acquisition method is proposed to solve the problems of uncertain factors in the final step of slabstone installation process. Finally, we use a simulation-platform to train and evaluate our algorithm. The results have shown that our method is capable and effective in accomplishing the installation process.**

Index Terms - slabstone installation, reinforcement learning, dual-arm robot, cooperation motion control

I. INTRODUCTION

A. Motivation

Labor shortages, defect rates, safety problems, cost increases, and other factors cause conventional construction almost to reach its possible technological performance limit [1]. On the other hand, using widely in the traditional manufacturing industry, automation and robotics are considered as the key technology to increase construction productivity [2], even if they are still not widely used in the construction industry because of the complex site environment. The installation of slab building materials, including slabstone installation, tile installation, ceiling installation and curtain wall installation, is an extremely important part of the construction process while it is one of the most repetitive and labor-intensive parts [3]. As having the uniform and standard size specification, slab installation process is very suitable for using robots to complete.

Compared to the traditional manual installation method, slab installation robot has several important significance and advantages to complete slab installation process. Initially, the slab installation robot is an inevitable choice to ensure the safety of construction workers and improve the quality of work. Then, the robots are much more efficient than workers and can greatly reduce the construction period. What's more, the engineering cost can be cut down by using robots to solve the problem of labor shortages.

In the past few decades, some slab installation robots, most of them consist of one manipulator and a moving base [4,5,7,8], have been developed. The control method of single-arm robot is simple as well as its mechanical structure. However, as the slab gets heavier and bigger with the development of the construction industry, working ability, environmental adaptability and construction productivity of a slab installation robot with single manipulator is limited. In this paper, compared to previous work, we use two cooperative manipulators and a mobile base to form a dual-arm slab installing robot to enhance the working capacity of the robot.

The process of slab installing can be divided into two steps that the first step is to move the slab to the mounting point from the placement point and the second step is to adjust the pose of slab to fix it to the wall firmly. Generally, using filling materials, such as concrete and mortar, or screws to fix the slab are the two most common ways, corresponding to the wet sticking type and the dry mounting type. And the installation methods of different slab building materials have certain universality. In this paper, we focus on the wet sticking type way of slabstone installation by using slabstone installation robot with two cooperative manipulators and a mobile base. After analyzing and determining the process of slabstone installation, we propose a slabstone installing skill acquisition method for dual-arm robot based on reinforcement learning to solve the uncertainty in the complex installing process. Additionally, we present a model method that using the spring-damping elements to model and simulate the filling material to build the interactive environment for reinforcement learning.

B. Related Work

Over the past years, several slab installing robots have been invented while most of them consist of one manipulator and a mobile base. Santos et al. proposed an assisting robot for plaster panels installing, which consists of a four-wheeled platform and a ceiling-grasping manipulator with three degrees of freedom (DOF) [4,5]. By using a guidance system based on one or two joysticks, the active joints of the manipulator can move directly [6]. Yu and Lee presented a curtain wall installation robot combined with a commercial excavator and a 3-DOF manipulator under the control of human-robot cooperation [7,8]. The manipulator of this robot is driven by workers through a master system. The system has a 6-DOF

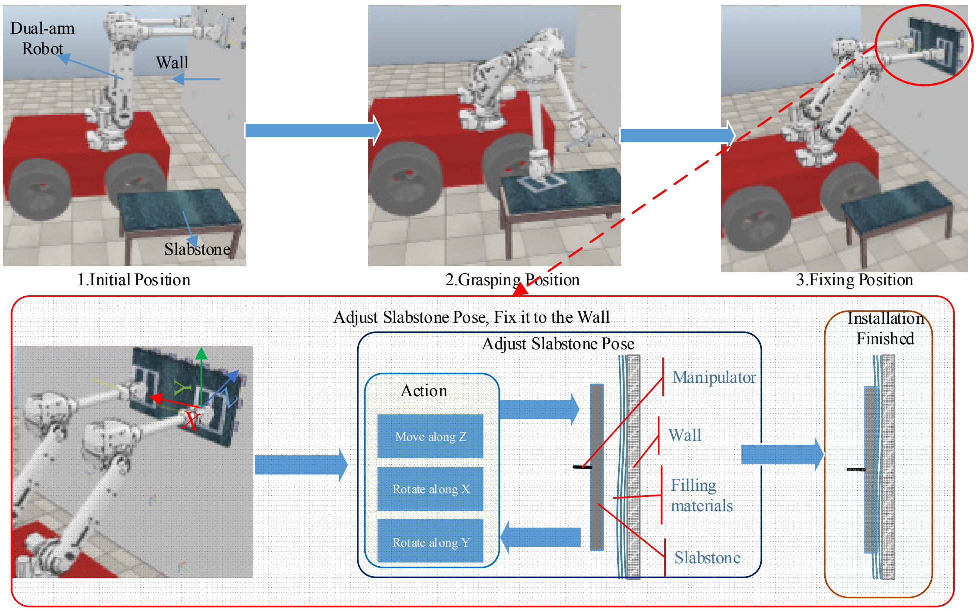

Fig. 1. Process of slabstone installation

force/torque sensor and can send command signal to the servo controller board based on the information from the sensor [9]. Lee et al. also proposed a ceiling glass installation robot consisted of an aerial lift and a multi-DOF manipulator [10,11]. Same as the human-robot cooperation control method in [9], based on two 6-axis force/torque sensors, workers can control the manipulator by supplying a force on a handler. A tile installation robot with a serial and parallel part was proposed [12].

The main ideas of those slab installing robots are combining a mobile base and a multi-degree of freedom manipulator with vacuum suction end-effector to form the body of the robot. Then, these robots are semi-automated assist system that works and coexist with workers—so called human-robot cooperation control. The whole process is not fully automated that it still needs people to complete some process manually. Additionally, due to the weight and deformation of the slab, the working ability of a single manipulator is limited. Because of the uncertain factors in site environments, traditional control methods are difficult to complete the slab installing automatically. With the breakthrough development of artificial intelligence technology, deep reinforcement learning(DRL) is considered to be a suitable method to solve complex high-dimensional control problems. Li et al. solved the posed problems of uncertainty in a complex assembly process based on DRL [13]. By using DRL, a dual-arm robot succeeded in finishing cloth manipulation [14]. In this paper, we use two manipulators and a mobile base to form a dual-arm slabstone installing robot to enhance the working capacity of the robot. Deep reinforcement learning method was used to control the robot to complete the final step of slabstone installing that adjusts the slabstone's pose to make it have good contact with filling materials. We propose a slabstone installing skill acquisition method for dual-arm robot based on reinforcement learning to solve the uncertainty in the complex installing process.

The rest of the paper is organized as follows: Section II introduces the final step of the complex slabstone installing process and formulates the problem to be solved. Section III presents the proposed method in detail. The configuration of the simulation environment, simulation results and analysis are provided in Section IV. Section V presents conclusions and future work.

II. Problem Formulation and Preliminaries

A. *Slabstone Installation Process*

Fig. 1 shows the steps to perform the slabstone installation process, which comprises three phases: moving to grasp position from the initial position and grasping the slabstone, moving the slabstone to the fixing position, and adjusting the pose of slabstone. In the first and second phases, motion planning and control techniques for dual-arm cooperation are generally used to realize the motion and grasping so that the slabstone can be transported to the fixed point smoothly and quickly. In the last phase, the pose of slabstone is adjusted (move along z-axis, rotate along x-axis and y-axis) by dual-arm robot according to the contact condition between the slabstone and filling materials to make slabstone have good contact with the filling materials. When the slabstone is fixed to the wall by filling materials, some features can be observed from the force/torque sensor data at the end effector of the manipulator. Herein, we focus on the last phase, i.e., adjusting the pose of slabstone to have good contact with filling materials.

B. *Formulation of the Problem*

On the one hand, with the development of control theory and other robotics techniques, industrial robots perform well in repetitive and high-precision tasks in a structured environment. On the other hand, uncertain factors from unstructured environment restrict the working scene and capacity of robot heavily. In the last few years, breakthroughs have been made in artificial intelligence technology that makes it possible for industrial robots to have the ability to learn and solve uncertain problems after learned and got skills by the interaction with the environment, in an unstructured environment.

This paper aims to realize the ability of industrial robots to learn the third phase of slabstone installation process (or TSI process, for short) through the sensing information inputs. The aim can be defined as below:

- The dual-arm robot can acquire the pose adjustment skill and make slabstone into good contact with filling materials by itself.
- The dual-arm robot's capacity in slabstone installation can be improved by learning more rounds.

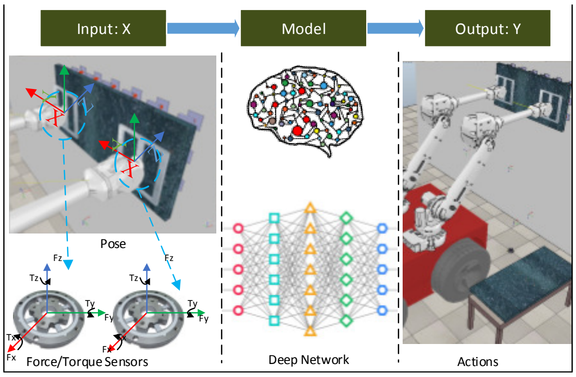

Fig. 2. Problems definition of TSI process

We assume that the main uncertain factors come from the filling materials and the slabstone is a rigid body without deformation. The learning model to solve the TSI process is shown in fig. 2, and it can be written as an unknown mathematical model: $Y=F(X)$. Here, X is the multi-dimension input vector of the two force/torque (F/T) sensors from end-effectors of a dual-arm robot, for representation to contact state, and Y is the multi-dimension output variable, for robotic actions to control the pose of slabstone. The problem of TSI process can be described as how to find the mapping function F between the contact states X and the robotic actions Y by intelligence algorithm. In this study, the function F is to be found by deep reinforcement learning, and the dual-arm robot can realize the TSI process by itself through learning. For the deep reinforcement learning algorithm, how to build the interactive training environment, how to choose actions and how to set a reward function must be solved. Using the spring-damping elements to model and simulate the filling material to build the interactive environment for reinforcement learning, and a reasonable reward function is constructed to guide dual-arm to complete the TSI process.

III. Proposed Method

A. Markov Decision Process of Slabstone Installation

The Markov decision process (MDP) is usually used to model the sequential decision problems which are solved by the reinforcement learning algorithm with agent-environment interaction. There are four key elements of the MDP that are interactive environment, state space, action space, and reward function design. Fig.3 shows the agent-environment of TSI process. The environment can be seen as consisting of dual-arm robot, a slabstone, the filling material and a wall. agent is a controller for the robot and can learn the mapping relationships between state and action. The state of the TSI process representing the contact state between a slabstone and the filling materials, which is the input of the agent, is by two F/T sensors. The action is the pose adjustment of the slabstone, which is decomposed into two arms to complete. The movement of the two arms is controlled by tight constraints that there is no relative displacement and rotation between two arms. The reward function is designed to guide the pose adjustment, making the slabstone have good contact with the filling materials.

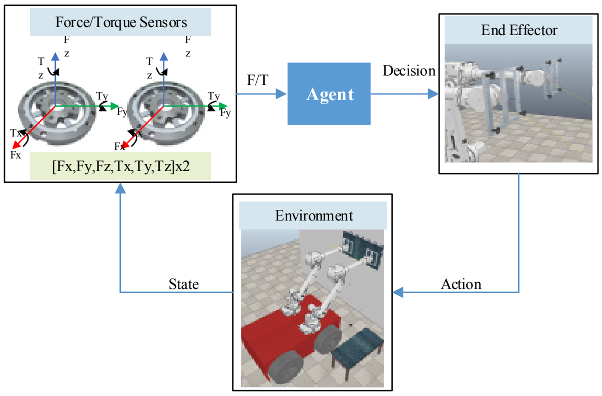

Fig. 3. The agent-environment interaction of TSI process

B. Interactive Environment

The agent of reinforcement learning improves its action and learns the optimal policies by the interaction with the environment. In addition, training the agent by interacting with the real world and robot is usually expensive and low effective. Modeling and simulation of the environment, therefore, are very important for the reinforcement learning algorithm.

For the filling material such a flexible and deformable object, a mass-spring model [15] and a finite element model [16] are widely used. However, these two modeling methods have high computational cost and poor real-time performance while common robot simulation software has limited support for flexible object modeling. On the other hand, the main effect between the filling materials and the slabstone is the axial force, while the extrusion force between the filling materials has less influence. And the displacement-load curve of the cement compression experiment shows that the force-displacement relationship is linear at the beginning stage [17]. Hence, considering feasibility and authenticity, we proposed using spring-damping elements to model the filling materials to simplify the setup of DRL environments, and adding some constraints according to reality, making it more realistic and reasonable. Using spring-damping elements to form an array to simulate the filling materials as shown in Fig.4, and some conditions and assumptions are introduced below:

- Filling materials meet Hooke's Law: $F=kx$, in the case of small displacement in one direction.
- After having good contact with filling materials, the slabstone can only move in one direction toward the wall because of the atmospheric pressure.
- Slabstone deformation and force among springs are not considered.

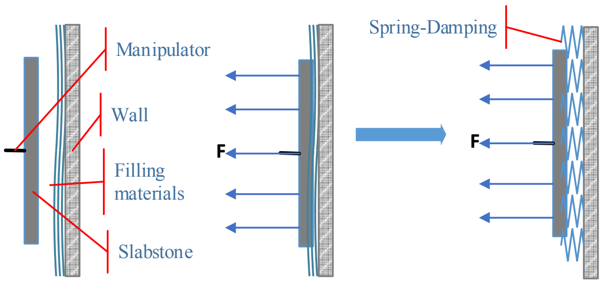

Fig. 4. The model of the filling materials

C. *State Space, Action Space and Reward Function Design*

The state space is defined as S, which represents all the contact state between the slabstone and filling materials in the TSI process, and is composed of two 6 dimension force/torque vector collected from two F/T sensors. There is a mapping relationship between the contact state and the force/torque data from F/T sensors, shown in [18, 19], while the change of contact state will lead to the obvious change of F/T data. Therefore, the state space is the set of possible contact state described as:

$$S=\{f_1,\cdots,f_6,t_1,\cdots,t_6 \mid f_i\in[-lf_i,rf_i],t_i\in[-lt_i,rt_i],i=1,\cdots,6\} \tag{1}$$

Here, $f_1,\cdots,f_3$ and $t_1,\cdots,t_3$ are corresponding to the force and torque in x, y, and z direction for the first arm, $f_4,\cdots,f_6$ and $t_4,\cdots,t_6$ are corresponding to the same F/T data for the other arm. And, lf_i and rf_i are corresponding to the minimum and maximum threshold value for the i^{th} force. And, lt_i and rt_i are corresponding to the minimum and maximum threshold value for the i^{th} torque.

As shown in Table I, there are seven discrete actions constitute the action space. The figure in the bottom left corner of the Fig.1 shows the definition of coordinate axis which is an absolute coordinate. Considering that the TSI process is mainly fine-tuning the pose of the slabstone, these actions are summarized and produced. Every action is to be decomposed into two motion by dual-arm cooperative constraint and be executed by two seven DOF arms. The value δ, α, and θ are the small increments for moving or rotating at every adjustment step. Therefore, the action space can be represented as (2):

$$A=[a_0,a_1,a_2,a_3,a_4,a_5,a_6] \tag{2}$$

TABLE I
DISCRETE ACTIONS

No.	Action
a_1	Move +δ along the Z axis
a_2	Move –δ along the Z axis
a_3	Rotate +α along the X axis
a_4	Rotate –α along the X axis
a_5	Rotate +θ along the Y axis
a_6	Rotate –θ along the Y axis

The reward function is designed to guide the slabstone has good contact with the filling materials, using some punishment and motivation terms for different stages of the movement. Same as the definition of the coordinate axis shown in Fig1, the reward function is designed as below:

- If the count of the pose adjustment N reaches the maximum count N_{th} or if the cumulative move distance d in Z axis exceeds the threshold d_{th} or if the cumulative rotate angle α along X axis exceeds the threshold α_{th} or if the cumulative rotate angle θ along Y axis exceeds the threshold θ_{th}, it is considered to fail the process and returns a negative reward r_1 = -1. Define the above failure condition as A.
- If both the forces along Z axis are approximately equal to the threshold and both torques in X axis are approximately equal to the zero and the torques in Y axis are equal and opposite, it is considered to succeed in TSI process and returns a positive reward r_2 = 1. Define the above success condition as B.
- If not meet condition A and B, it is considered to continue to explore and adjust. Then, a little negative value r_3 = -0.01 for punishment will be returned. Define the above condition as C.
- If meet the condition C and current action is a_1 and the cumulative move distance d in Z axis less than the thirty percent of the threshold d_{th}, it returns a little positive reward r_4 = 0.05 to accelerate the initial adjustment. Define the above condition as D.
- If meet the condition C and current action is a_2 and the cumulative move distance d in Z axis more than the half of the threshold d_{th}, it returns a negative reward r_5 = -0.5 to limit the movement along the negative Z axis. Define the above condition as E.

Therefore, the reward function $R(s, a)$ can be summarized as below:

$$R(s,a)=\begin{cases} -1 & A \\ 1 & B \\ -0.01 & C \\ 0.05 & D \\ -0.5 & E \end{cases} \tag{3}$$

Here, A, B, C, D, and E are the judgment conditions defined above, s and a correspond to a state and an action in their space.

D. Deep Reinforcement Learning for Skill Acquisition

The method of deep reinforcement learning for slabstone installation skill acquisition is shown in fig.5, which shows one episode to train the network. The state space S, action space A and reward function $R(s, a)$ of TSI process are already defined above. Policy $\pi(a|s)$ represents the probability of action a being taken under state s. Considering a fixed policy π, the state-action value function $Q_{\pi}(s,a)$ can be defined as:

$$Q_{\pi}(s,a) = R(s,a) + \sum_{t=0}^{\infty} \gamma^t R(s_t,a_t) \mid s_0 = s, a_0 = a \tag{4}$$

Here, $\gamma \in (0,1)$ is the discount parameter. The state-action value function $Q_{\pi}(s,a)$ is used to evaluate what is good or bad in a long-term process of a state-action pair (s, a) when following a fixed policy π, and can be used as an objective function to get the optimal policy of MDP by optimization. Therefore, the optimal value function $Q^*(s,a)$ can be written as:

$$Q^*(s,a) = \max_{\pi}(s,a) \tag{5}$$

Therefore, the optimal action is:

$$a = \arg\max_{\pi} Q^*(s,a) \tag{6}$$

Equation (4) can be rewritten:

$$\begin{aligned} Q^{\pi}(s,a) &= R(s,a) + \sum_{t=1}^{\infty} \gamma^t R(s_t,a_t) \\ &= R(s,a) + \max_{a' \in A} Q^*(s',a') \end{aligned} \tag{7}$$

This is the Bellman optimality equation, and s' is the state at the next time step. Based on deep reinforcement learning, the optimal state-action value function $Q^*(s,a)$ is estimated by a neural network $Q(s,a|\theta)$, where θ is the parameters of the deep neural network. In deep reinforcement learning, the deep neural network has strong feature extraction ability from state input.

As shown in fig.5, the input is a 12 dimension vector from state space, which is force/torque filtered data from two F/T sensors used to represent the contact state between a slabstone and the filling materials. The hidden layer is the multi-layer perceptron (MLP) consisted of full connective layers and uses Rectified Linear Unit (ReLu) [20] activation function in every layer, having 64, 64, and 32 nodes in it. The output layer is a single layer with sigmoid activation function, which output is a 7 dimension vector representing the valid actions. Priority experience replay technology [21] is used to accelerate the training process by changing the distribution of generated training data. Set up a data cache to store a series of learning

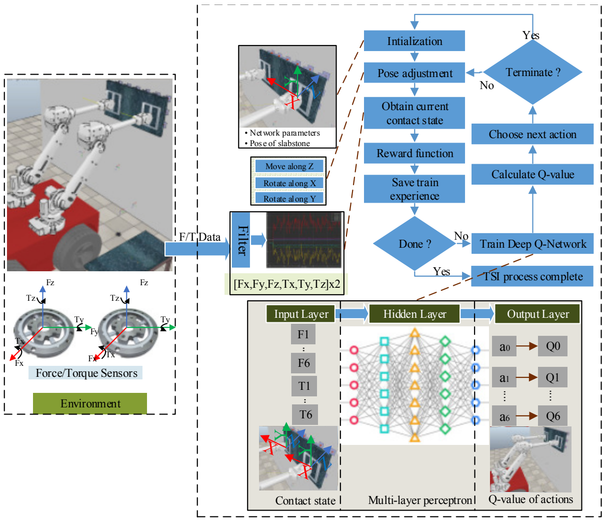

Fig. 5. Deep reinforcement learning for skill acquisition

process data $\{s_t, s_{t+1}, a_t, r_t\}$. Double Q-learning [22] technology is used as well.

The mean square error loss function of deep q-learning is used:

$$L(\theta) = E[(R(s,a) + \gamma * Q^{\pi}(s',a';\theta) - Q(s,a;\theta))^2] \tag{8}$$

The full algorithm is shown in TABLE II.

TABLE II
ALGORITHM 1

Algorithm 1 Skill Acquisition by Deep Q-learning in TSI process

Input:
State space S, action space A, discount rate γ, learning rate ς, action increment parameters, training episode M, maximum update times U_{max}, replay buffer maximum size N_r, training batch size N_b, target network replacement frequency N^-
Initialize: empty replay buffer D, initial network parameters θ, copy θ to θ^-
for *episode* $e \in \{1,2,\ldots, M\}$ **do**
  for update *times* $u \in \{1,2,\ldots, U_{max}\}$ **do**
    Sample next action a_t from A by greedy strategy
    Execute action a_t and receive reward r_t and new state s_{t+1}
    Store $\{s_t, s_{t+1}, a_t, r_t\}$ to D, and replacing the oldest if $|D| \geq N_r$
    Sample a minibatch of N_b tuples $\{s_t, s_{t+1}, a_t, r_t\} \sim \mathrm{Unif}(D)$
    Construct target values for every tuple in N_b tuples
    Define $a_{max}(s';\theta) = \arg\max_{a'} Q(s',a';\theta)$
    $$y_i = \begin{cases} r & \text{if } s' \text{ is terminal} \\ r + \gamma Q(s', a_{max}(s';\theta); \theta^-), & \text{otherwise} \end{cases}$$
    Perform a gradient descent step with loss $\|y_i - Q(s,a;\theta)\|^2$
    Replace target parameters $\theta^- \leftarrow \theta$ every N^- steps
  end
end

IV. SIMULATION AND EXPERIMENTS

A. Simulation System

The proposed method of slabstone installation skill acquisition was evaluated in V-rep [23] simulation platform by

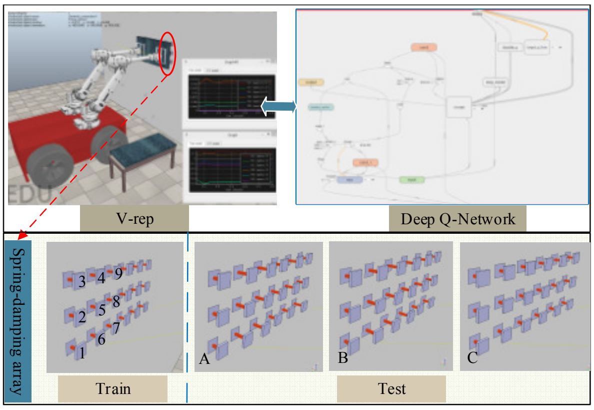

Fig. 6. Simulation system

using two 6-DOF manipulators to form a dual-arm robot. Using several spring-damping elements to model and simulate the filling materials for training as shown in fig.6, and the same initial length was set to these spring-damping elements. In addition, another three different distribution spring-damping arrays for testing were shown. The motion planning of dual-arm was through the open motion planning library (OMPL) [24] in V-rep, and cooperation was achieved through the position control. Deep Q-learning of the skill acquisition algorithm was implemented based on an open-source RL algorithm library named Stable-Baselines [25] and trained in a desktop with Intel® Core™ i5-8400 Processor and NVIDIA GTX 1060 6Gbps GPU. The simulation system is showed in fig.6. The source code of training and testing for the deep reinforcement learning model in this paper can be available here: https://github.com/doctorsrn/gym_vrep.

The setting of hyper-parameter and some object property parameters were shown in TABLE III.

TABLE III
PARAMETER SETTING

Parameter	Value
Learning rate ς	0.0001
Discount rate γ	0.95
training episode M	60000
training batch size N_b	64
maximum update times $U_{\max}$	30
replay buffer maximum size N_r	50000
target network replacement frequency N^{-}	300
Action increment move along Z-axis	0.005 *m*
Action increment rotate along X-axis	0.5 *deg*
Action increment rotate along Y-axis	0.5 *deg*
Spring-damping coefficient	250 *N/m*
Target force in Z-axis direction	-30 *N*

B. *Training and Test Results*

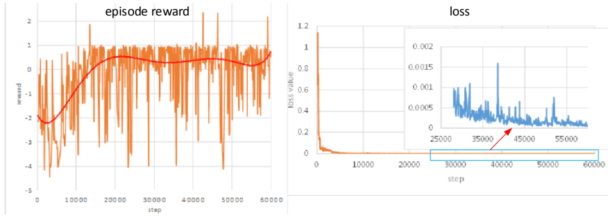

Fig. 7. Training reward and loss value

The network architecture of the skill acquisition method was described in fig.5 from the third section. The training process took 2399 minutes with 60000 episodes. As shown in fig.7, with the increment of training steps, the loss value of the deep Q-networks decreased from 1.14 to 5x10^{-5} and converged, the reward of the TSI process increased and stabilized to the range (0,1) finally. The result showed that the model can guide the installation process. The training reward curve had a large range of fluctuates due to the sparse success samples.

To evaluate the capability of the training process models, the success rate, average steps and reward were tested after every 3000 episodes of learning. The three parameters were tested with 100 episodes to get the average values. As shown in TABLE IV, the success rate and reward increased with the learning circle. Meanwhile, the average steps decreased. The learning had not much improvement after 33000 episodes and was overfitting probably. So, the model in 33000 episodes was chosen as the final trained model.

TABLE IV
TEST RESULTS OF THE TRAINING MODEL

Episode	**3000**	**6000**	**9000**	**12000**	**15000**	**18000**
success rate	0.29	0.47	0.27	0.91	0.67	0.85
average steps	24.53	21.58	26.75	11.91	17.84	11.75
reward	-1.902	-1.288	-1.335	0.373	0.294	0.635
Episode	**21000**	**24000**	**27000**	**30000**	**33000**	**36000**
success rate	0.88	0.78	0.65	1.0	0.96	0.78
average steps	12.36	13.7	16.15	8.52	9.59	12.46
reward	0.272	0.336	0.392	0.703	0.534	0.451
Episode	**39000**	**42000**	**45000**	**48000**	**51000**	**54000**
success rate	0.89	0.87	0.99	1.0	0.89	0.90
average steps	10.89	11.59	9.2	7.29	11.26	10.01
reward	0.457	0.282	0.67	0.78	0.593	0.61

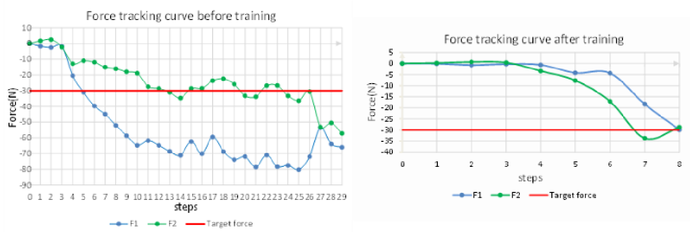

Fig. 8. Force response curve before and after the training

Fig. 8 shows the response curves of the contact force in the z-axis direction before and after the model training, where *F1* and *F2* correspond to the force of the left and right manipulator. As shown in the figure, the two forces are random before the model training but can track the target force quickly after training. That is to say, the robot can complete the TSI process after training.

To evaluate the generalization capacity of the model, three other different distributions of filling materials constructed by changing spring-damping elements length were used to test the average success rate, average success step and average reward for 100 times installation. The testing result shown in TABLE V shows that the model can complete the TSI process in the three other different conditions with a general success rate but below the success rate of the training result.

TABLE V
TEST RESULTS OF TESTING CASES

Parameter	Training	Test1	Test2	Test3
Average success rate	0.96	0.82	0.88	0.95
Average success step	9.59	14.08	11.14	11.04
Average reward	0.534	0.378	0.577	0.432

In short, the training and testing results in TABLE IV and TABLE V showed that the dual-arm robot can finish TSI process by the skill acquisition method in the simulation environment. But, there is still large room for improvement to the generalization capacity of the model.

C. Comparison with Installation Method by Construction Worker

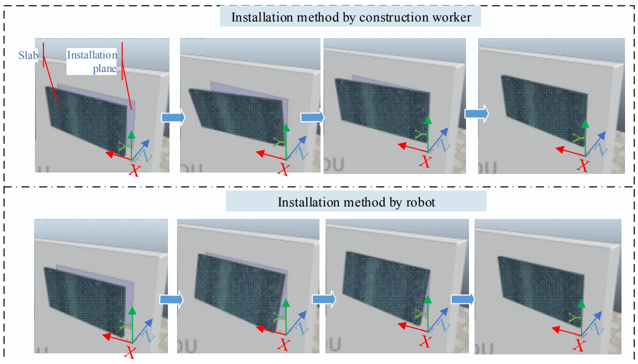

Fig. 9. Comparison of installation methods between the worker and robot

Fig. 9 shows the difference in the installation method between our robot based on DRL model and construction worker. The worker first aligns one side of the slab with the wall surface, then slowly lowers the other side and adjusts the contact state until the installation is completed [26]. As for robot, the simulation results show that it also aligns one side of the slab with wall and then fine-tunes the slab position until the installation is completed. But, most beginning actions are a_2 showing that the robot tends to align the top edge first. For example, the sequence of actions in the tracking curve of Fig.8 is $\{a_2,a_2,a_2,a_2,a_2,a_4,a_0,a_5\}$.The cause of this difference is probably because of the setup of reward function for DRL or some random factors. In a word, for the installation method, the robot can get the slabstone installing skill by DRL and the installation method is similar to a construction worker.

V. CONCLUSIONS AND FUTURE WORKS

It is difficult to finish slabstone installation process automatically by a robot because of the uncertainty in the complex process. To solve the slabstone installation problem, a slabstone installation skill acquisition method based on deep reinforcement learning for the dual-arm robot is proposed. Initially, with some constraints according to reality, this paper built reinforcement learning environment by using spring-damping elements to model the filling materials. Then, this paper defined the action space, state space, reward function and other elements for the skill acquisition algorithm based on deep reinforcement learning, and made a simulation experiment. The simulation results showed that the slabstone installation skill acquisition method can finish the slabstone installation process making slabstone have good contact with filling materials.

Future studies will focus on increasing the success rate and improving planarity of slabstone installing. On the one hand, the success rate of test results was not good enough. The reward function and success condition of states need be well designed in future work. On the other hand, not only good contact but the good planarity of slabstone installing must be considered in reality. Vision state feedback need be introduced to evaluate the planarity in future work. Additionally, to speed up the training process is also a necessary task. Finally, preparation for dual-arm robots and related equipment is underway, and real-world experiments are to be made in future work.

REFERENCES

[1] Bock T, Linner T. Robot oriented design[M]. Cambridge University Press, 2015.

[2] Chen Q, de Soto B G, Adey B T. Construction automation: Research areas, industry concerns and suggestions for advancement[J]. Automation in Construction, 2018, 94: 22-38.

[3] Gambao E, Balaguer C, Gebhart F. Robot assembly system for computer-integrated construction[J]. Automation in Construction, 2000, 9(5-6): 479-487.

[4] Gonzalez de Santos P, Estremera J, Jimenez M A, et al. Manipulators help out with plaster panels in construction[J]. Industrial robot: an international journal, 2003, 30(6): 508-514.

[5] De Santos P G, Estremera J, Garcia E, et al. A service robot for construction industry[C]//Proceedings World Automation Congress, 2004. IEEE, 2004, 15: 441-446.

[6] de Santos P G, Estremera J, Garcia E, et al. Power assist devices for installing plaster panels in construction[J]. Automation in Construction, 2008, 17(4): 459-466.

[7] Yu S N, Lee S Y, Han C S, et al. Development of the curtain wall installation robot: Performance and efficiency tests at a construction site[J]. Autonomous Robots, 2007, 22(3): 281-291.

[8] Lee S Y, Lee K Y, Lee S H, et al. Human-robot cooperation control for installing heavy construction materials[J]. Autonomous Robots, 2007, 22(3): 305.

[9] Han C S, Lee S Y, Lee K Y, et al. A multidegree- of- freedom manipulator for curtain- wall installation[J]. Journal of Field Robotics, 2006, 23(5): 347-360.
[10] Lee S, Gil M, Lee K, et al. Design of a ceiling glass installation robot[C]//Proceedings of the 24th International Symposium on Automation and Robotics in Construction. 2007: 247-252.
[11] Lee S, Lee J, Han C, et al. Human robot cooperative control and task planning for a glass ceiling installation robot[J]. 2008.
[12] Choi H S, Han C S, Lee K, et al. Development of hybrid robot for construction works with pneumatic actuator[J]. Automation in Construction, 2005, 14(4): 452-459.
[13] Li F, Jiang Q, Zhang S, et al. Robot Skill Acquisition in Assembly Process using Deep Reinforcement Learning[J]. Neurocomputing, 2019.
[14] Tsurumine Y, Cui Y, Uchibe E, et al. Deep reinforcement learning with smooth policy update: Application to robotic cloth manipulation[J]. Robotics and Autonomous Systems, 2019, 112: 72-83.
[15] Tokumoto S, Fujita Y, Hirai S. Deformation modeling of viscoelastic objects for their shape control[C]//Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C). IEEE, 1999, 1: 767-772.
[16] Cotin S, Delingette H, Ayache N. Real-time elastic deformations of soft tissues for surgery simulation[J]. IEEE transactions on Visualization and Computer Graphics, 1999, 5(1): 62-73.
[17] Ma S, Qian Y, Kawashima S. Experimental and modeling study on the non-linear structural build-up of fresh cement pastes incorporating viscosity modifying admixtures[J]. Cement and Concrete Research, 2018, 108: 1-9.
[18] Zhang S, Jiang Q, Li Y, et al. Contact State Classification in Industrial Robotic Assembly Tasks Based on Extreme Learning Machine[C]//2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2018: 617-622.
[19] Li F, Jiang Q, Li Y, et al. Modeling Contact State of Industrial Robotic Assembly Using Support Vector Regression[C]//2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2018: 646-651.
[20] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 807-814.
[21] Schaul T, Quan J, Antonoglou I, et al. Prioritized experience replay[J]. arXiv preprint arXiv:1511.05952, 2015.
[22] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
[23] Rohmer E, Singh S P N, Freese M. V-REP: A versatile and scalable robot simulation framework[C]//2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013: 1321-1326.
[24] Sucan I A, Moll M, Kavraki L E. The open motion planning library[J]. IEEE Robotics & Automation Magazine, 2012, 19(4): 72-82.
[25] Hill, Ashley and Raffin, et al. Stable Baselines[EB/OL].https://github.com/hill-a/stable-baselines,2018.
[26] Gil M S, Kang M S, Lee S, et al. Installation of heavy duty glass using an intuitive manipulation device[J]. Automation in Construction, 2013, 35: 579-586.