# Stable Robotic Grasping of Target Object Using Deep Reinforcement Learning

1st In Jun Park
*Digital Convergence Research Laboratory*
*Electornics and Telecommunications Research Institute*
Daejeon, South Korea
ijpark@etri.re.kr

2nd Hyonyoung Han
*Digital Convergence Research Laboratory*
*Electornics and Telecommunications Research Institute*
Daejeon, South Korea
hyonyoung.han@etri.re.kr

***Abstract*—Robotic grasping within a cluttered environment has shown its potential for improvement through the integration of pushing. Nevertheless, achieving an effective pushing strategy remains a persistent challenge. Furthermore, the complexity is heightened in highly cluttered environments, particularly when dealing with stacked objects, posing significant difficulties for successful grasping. To address these issues, we propose a method aimed at enhancing pushing mechanisms by incorporating an information-theoretic measure of entropy. This integration facilitates the efficient clearance of obstacle objects surrounding the target object. Additionally, we leverage depth information as a foundational action prior, seamlessly integrating it into Q-value calculations. This integration results in an enhanced exploration strategy, consequently improving the grasping of stacked objects. Furthermore, we ensure the stability of grasping through reward shaping, achieved by accounting for changes in the object's pose. We conducted an evaluation of our approach using three distinct object types within a simulation environment, which revealed a notably higher rate of successful grasping in comparison to the baseline method. To substantiate our findings, we proceeded to test the model within a real-world setting, where the proposed approach showcased a substantial enhancement, reinforcing the efficacy of our method.**

***Keywords—Robotic Grasping and Pushing, Reinforcement Learning, Deep Q-learning***

## I. Introduction

Effectively performing object grasping in a cluttered robotic environment remains a notable challenge, primarily due to the close presence of surrounding objects in relation to the target objects. This proximity heightens the potential for collisions between the gripper and nearby obstacles during the grasping process, which, in turn, elevates the likelihood of task failures. In such scenarios, employing pushing maneuvers can serve to strategically rearrange cluttered objects, creating sufficient space for the fingers to facilitate successful grasping.

Numerous research endeavors have explored the realm of manipulation techniques, encompassing both grasping and pushing. Some studies aimed at developing effective grasping policies have concentrated on optimizing affordance or analytic metrics [1, 2, 3], yielding promising results. Similarly, a significant body of research has been devoted to pushing techniques, which focus on altering an object's pose [4, 5, 6]. Nevertheless, the challenge remains in seamlessly integrating these distinct approaches, given their independent developmental paths and the absence of a shared underlying objective. Another noteworthy limitation in prior research is its emphasis on relatively straightforward tasks, often overlooking the incorporation of precise target-related specifics [7, 8, 9]. Even when target-centric grasping has been explored, the objects tend to be distinctly separable [10, 11]. One particular study that addresses both target-driven grasping and complex, cluttered environments has shown promising results. However, achieving such outcomes necessitates the inclusion of supplementary information, such as a target object mask, facilitated through perception mechanisms [12].

In this work, we introduce an approach that enhances both pushing and grasping models using vision-based reinforcement learning (RL). Our primary goal for pushing is to increase the efficiency of removing obstructing objects around the target object, which is accomplished by integrating an information-theoretic measure. Furthermore, we utilize depth information as a fundamental action guideline, seamlessly incorporating it into the computation of Q-values. This fusion leads to an enhanced exploration strategy, resulting in improved grasping of stacked objects. Moreover, we bolster the stability of the grasping procedure by tailoring rewards to account for object pose variations. Our method encompasses comprehensive enhancements to both grasping and pushing models, culminating in superior performance with respect to grasping success rates. This superiority is substantiated through testing in simulated as well as real-world scenarios, showcasing notable advancements over the baseline methodology.

## II. Methods

### *A. System Overview*

Our system consists of three modules, as depicted in Fig. 1: a perception module, a task/motion plan module, and an action prediction module. The perception module extracts the identification and center position of all objects from an RGB scene image captured by a side camera using OpenCV library. The task planning module then establishes the sequence for

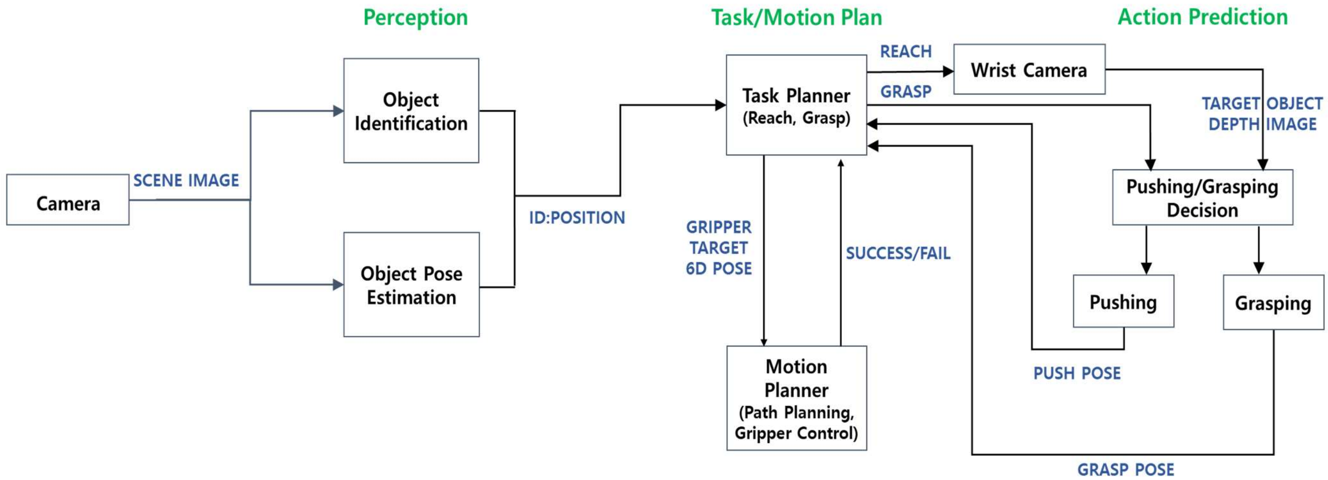


Fig. 1. Overview of the system: The system consists of three distinct sub-tasks - perception, task/motion planning, and action prediction.

grasping a target object and assigns primitive actions, such as reaching and grasping, based on the object's information obtained from the perception module. The motion planning module, using the Moveit library [13], is used to move the robot arm based on the target actions received from the task planning module. The grasping module is trained using reinforcement learning in a simulation to determine the gripper's pose based on the depth image of the target object. This results in achieving the most stable grasp, assisted by pushing actions.

### B. *Action Prediction Model*

The objective of the action prediction model is to achieve stable grasping of the target object in a cluttered environment, aided by pushing actions. Our model is essentially built upon the foundation of VPG in [9], with certain modifications implemented. These modifications involve incorporating an action prior to enhance sample efficiency and employing an information-theoretic measure for making decisions regarding pushing. The overall model architecture for the action prediction model is depicted in Fig. 2.

The task planner initially guides the robot arm to the center of the target object in order to capture a top-view depth image using a wrist camera. Subsequently, the depth image is transformed into a 3D point cloud in camera coordinates, utilizing the camera's intrinsic parameters. This 3D point cloud is then converted from camera coordinates to world coordinates, employing the camera's extrinsic parameters. Ultimately, this sequence of transformations culminates in the generation of a heightmap, which is defined as a state within the context of reinforcement learning.

$$H(X) = -\sum_{i=1}^{n} P(x_i) log P(x_i) \tag{1}$$

The entropy, defined in (1) and serving as a measure of image complexity, is calculated to determine the necessity of the pushing action. As illustrated in Fig. 3, the entropy increases with an increasing number of objects in an image. Pushing is carried out when the image's entropy surpasses a predetermined threshold, which ultimately leads to the removal of objects. This process subsequently aids grasping by eliminating obstacle objects from the target object's vicinity. We assign the reward function for pushing, as described in (2), to maximize the change in entropy. This change is defined as the difference between the entropies of the previous image and the current image, aiming to reduce the complexity around the target object as effectively as possible.

$$R_{\text{PUSH}}(\boldsymbol{s_t}, \boldsymbol{a_t}) = H_{t-1}(X) - H_t(X) \tag{2}$$

This differs from the one in VPG in the sense that a constant reward is provided when a detectable heightmap change occurs. However, this approach does not always yield the optimal action, as the reward is given based on the evident heightmap change, even in cases where nearby objects might actually move closer to the target object, potentially worsening the grasping outcome. Our method demonstrates an improvement in pushing actions by elevating grasping performance. The action space of the pushing model consists of the 3D gripper's pose $(x, y, \varphi)$, where only planar motion is assumed to take place.

The reward for the grasping model is devised to be inversely proportional to the change in the object's pose, as depicted in (3), where $q$ represents the quaternion of the grasped object.

$$R_{\text{GRASP}}(\boldsymbol{s_t}, \boldsymbol{a_t}) = \alpha / |1 + \sqrt{\Delta x^2 + \Delta y^2}| + \beta / |1 + \sqrt{\Delta q^2}| \tag{3}$$

This signifies that minimal change yields heightened grasp stability. The heightmap, containing 3D scene information, offers insights into the spatial distribution of objects. Integrating this data into the Q-value computation could potentially enhance the grasping model and reduce exploration time. The incorporation of this action prior, as depicted in (4), along with the heightmap, adjusts the Q-value as described in (5).

$$AP(s_t, a_t) = \begin{cases} 1, & \text{if } s_t > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $s_t$ is the heightmap.

$$a_t = \max_{a_t} Q(s_t, a_t)\, AP(s_t, a_t) + s_t \tag{5}$$

Adding the heightmap to the Q-value prompts the policy to acquire the ability to grasp objects from above. This proves advantageous in scenarios involving densely stacked objects,

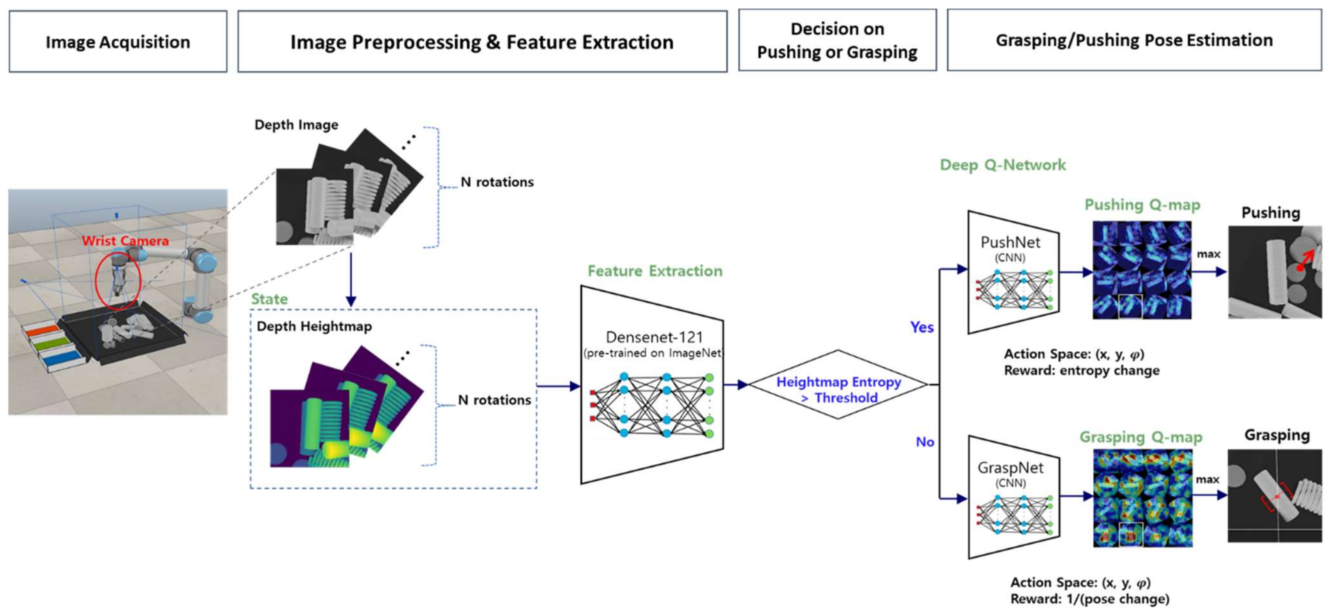


Fig. 2. Architecture of the action prediction model: This model's architecture involves image acquisition, image feature extraction, decision-making for pushing/grasping, and estimation of grasping/pushing poses.

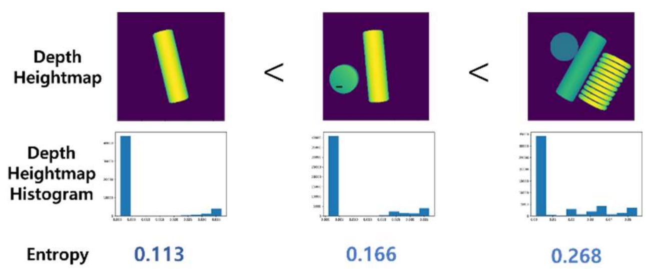


Fig. 3. Image complexity metric: entropy

such as bin picking, where the grasping sequence needs to occur from the top down.

## C. Training Policy

The training of the action predictive model is conducted in two phases, as illustrated in Fig. 4. In the first phase, only the grasping model is trained in a simple environment where the target object is positioned solely at the center of the image. Once the grasping model has converged, additional objects are introduced into the environment, leading to the presence of cluttered objects. The second phase of training is subsequently resumed to further learn the pushing model. Both the grasping and pushing models undergo updates while dealing with complex object configurations. The significance of the initial training lies in the grasping model's ability to retain the knowledge of grasping the target object located at the center, even without explicit cues such as a segmented mask indicating the target object's region.

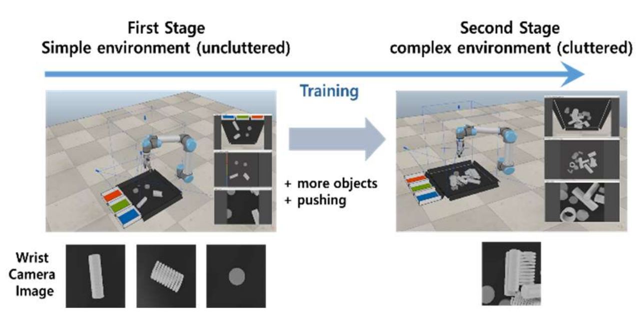


Fig. 4. Training procedure for grasping and pushing model

# III. Experiments

## A. Simulation Experiments

The action predictive model underwent training within a simulated robotic environment utilizing CoppeliaSim [14]. Our testing encompassed three distinct object types: a vehicle suspension model, a teeth model, and the YCB model [15], as shown in Fig. 5. In each episodes, the objects are introduced into the environments, resulting in a cluttered workspace. The

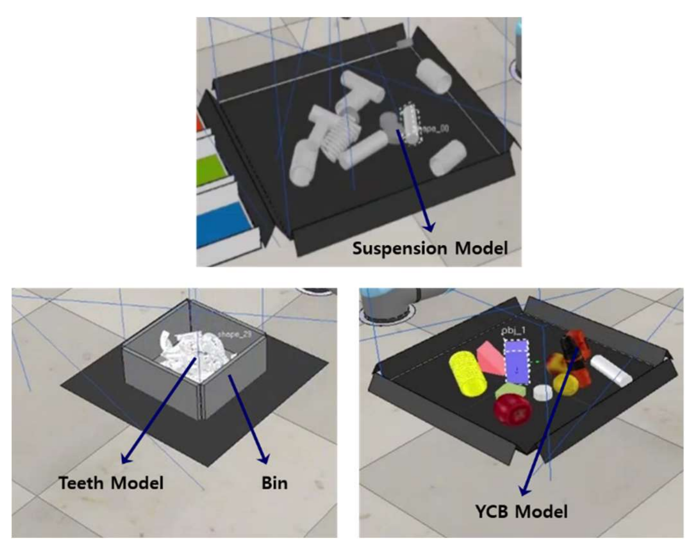


Fig. 5. Tested object types in a simulated environment.

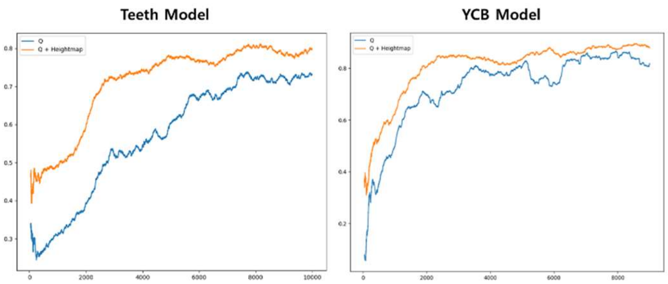


Fig. 6. Learning curves for teeth model (left) and YCB model (right).

learning curves for both the teeth model and the YCB model in Fig. 6 demonstrate that our modified Q-value-based approach exhibits superior performance in comparison to the non-modified version. For the teeth model, there is a notable enhancement in both sample efficiency and performance, with improvements of 37% and 10%, respectively. Similarly, for the YCB model, we observe advancements of 50% in sample efficiency and 5% in performance. A comprehensive overview of the evaluation results can be found in Table 1. It is evident that our model exhibits superior performance across all three types, even in challenging scenarios such as teeth bin picking, where objects are smaller, more densely stacked, and present in a cluttered arrangement.

TABLE I. GRASPING SUCCESS RATE (%)

| Object Types | Methods | |
|---|---|---|
| | Our model | Baseline (VPG) |
| **Suspension** | 97.3 % | 88.7 % |
| **Teeth** | 87.2 % | 77.6 % |
| **YCB** | 92.3 % | 83.8 % |

## B. *Real-World Experiments*

We conducted testing of our model using a 3D-printed suspension model within a real-world setting, as depicted in Fig. 7. The experimental setup comprises a UR5 robot arm, a RealSense L515 camera mounted on the robot's wrist, and a Robotiq 2F-85 gripper. The suspension model itself consists of three distinct part types: the cap, shaft, and spring. We conducted tests on 50 different scenarios by altering the position and orientation of objects. In each scenario, seven objects were closely located to each other. The evaluation results presented in Table 2 highlight that our model surpasses the VPG, even within the real-world environment. The observed performance decline, when contrasted with simulation results, can be attributed to the presence of noisy depth images captured by the L515 camera.

TABLE II. GRASPING SUCCESS RATE (%)

| Object Types | Methods | |
|---|---|---|
| | Our model | Baseline (VPG) |
| **Cap** | 90.3 % | 80.7 % |
| **Shaft** | 93.2 % | 83.6 % |
| **Spring** | 92.1 % | 84.8 % |

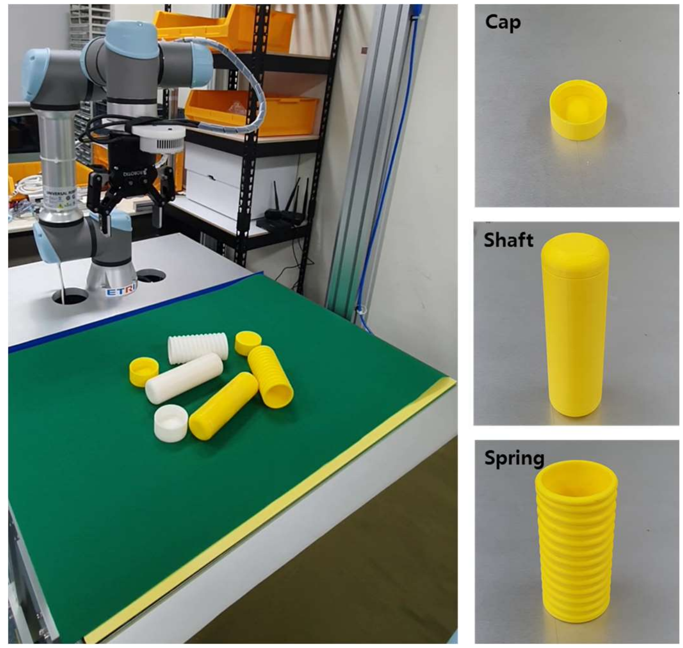


Fig. 7. Real-world experiment setup (left) and test objects (right).

# IV. DISCUSSION AND FUTURE WORK

We presented an approach rooted in vision-based reinforcement learning, addressing the obstacles posed by cluttered environments to attain proficient grasping outcomes. Our method encompasses enhancements to both the grasping and pushing models, leading to superior performance in terms of grasping success rates. This superiority was demonstrated across both simulated and real-world settings, showcasing substantial advancements compared to the baseline methodology. In our future endeavors, our focus will extend towards incorporating a broader range of objects encompassing diverse shapes, sizes, and textures for training and method evaluation. This comprehensive approach seeks to thoroughly assess the efficacy of our methodology across a wider spectrum of real-world environments. Furthermore, our upcoming research will delve into addressing the challenges of sim-to-real transfer, aiming to mitigate any decline in real-world performance that might arise during the deployment of our approach.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government (23ZR1130, A Study of Hyper-Connected Thinking Internet Technology by autonomous connecting, controlling and evolving ways), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No.2022-0-00187, Development of an edge brain framework to make manufacturing equipment and robots intelligent and No.2022-0-01049).

# REFERENCES

[1] L. Pinto and A. Gupta, “Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours,” in ICRA, 2016.

[2] A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, et al., “Robotic pick-and-place of novel objects in clutter with multiaffordance grasping and cross-domain image matching,” in ICRA, 2018.

[3] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, “Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics,” in RSS, 2017.

[4] M. Bauza and A. Rodriguez, “A probabilistic data-driven model for planar pushing,” in ICRA, 2017.

[5] C. Finn and S. Levine, “Deep visual foresight for planning robot motion,” in ICRA, 2017.

[6] C. Ignasi, D. Held, and P. Abbeel, “Policy transfer via modularity,” in IROS, 2017.

[7] Eitel, N. Hauff, and W. Burgard, “Learning to singulate objects using a push proposal network,” in Proc. of the International Symposium on Robotics Research (ISRR), 2017.

[8] C. Choi, W. Schwarting, J. DelPreto, and D. Rus, “Learning object grasping for soft robot hands,” IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 2370–2377, 2018.

[9] 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 IROS, 2018.

[10] E. Jang, C. Devin, V. Vanhoucke, and S. Levine, “Grasp2vec: Learning object representations from self-supervised grasping,” in Conference on Robot Learning (CoRL), 2018.

[11] K. Fang, Y. Bai, S. Hinterstoisser, S. Savarese, and M. Kalakrishnan, “Multi-task domain adaptation for deep learning of instance grasping from simulation,” in ICRA, 2018.

[12] Y. Yang, H. Liang, and C. Choi, ‘‘A deep learning approach to grasping the invisible,’’ *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2232–2239, Apr. 2020.

[13] D. Coleman, I. Sucan, S. Chitta, N. Correll. “Reducing the barrier to entry of complex robotic software: a MoveIt! case study”, Journal of Software Engineering for Robotics, vol. 5, no. 1, pp. 3-16, May 2014.

[14] E. Rohmer, SPN. Singh, and M. Freese, “CoppeliaSim (formerly V-REP): a Versatile and Scalable Robot Simulation Framework (PDF),” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 1321–1326.

[15] Berk Calli, Aaron Walsman, Arjun Singh, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar, Benchmarking in Manipulation Research: The YCB Object and Model Set and Benchmarking Protocols, IEEE Robotics and Automation Magazine, pp. 36 – 52, Sept. 2015.