Physics-Guided Hierarchical Reward Mechanism for Learning-Based Multi-Finger Object Grasping

Yunsik Jung*, Lingfeng Tao*, Michael Bowman*, Jiucui Zhang^, and Xiaoli Zhang*, Member, IEEE

Abstract—Autonomous grasping is challenging due to the high computational cost caused by multi-fingered robotic hands and their interactions with objects. Various analytical methods have been developed yet their high computational cost limits the adoption in real-world applications. Learning-based grasping can afford real-time motion planning thanks to its high computational efficiency. However, it needs to explore large search spaces during its learning process. The search space causes low learning efficiency, which has been the main barrier to its practical adoption. In this work, we develop a novel Physics-Guided Deep Reinforcement Learning with a Hierarchical Reward Mechanism, which combines the benefits of both analytical methods and learning-based methods for autonomous grasping. Different from conventional observation-based grasp learning, physics-informed metrics are utilized to convey correlations between features associated with hand structures and objects to improve learning efficiency and learning outcomes. Further, a hierarchical reward mechanism is developed to enable the robot to learn the grasping task in a prioritized way. It is validated in a grasping task with a MICO robot arm in simulation and physical experiments. The results show that our method outperformed the baseline in task performance by 48% and learning efficiency by 40%.

I. INTRODUCTION

Multi-finger robotic grasping is necessary to accomplish object manipulation that can replace human activities in various environments, such as manufacturing industry, space, and deep-sea maintenance. Despite the potential, it is still challenging in several aspects. As high dimensional robotic hands that can accomplish complex tasks have been developed, it significantly increases computational demands, which harms real-time performances in real-world applications. In addition, interactions between robotic hands and objects with various contours are demanding to accomplish stable performances. Therefore, the current performance of robotic grasping is limited.

Although analytical methods have been widely adopted to solve autonomous grasping tasks, the high computational cost makes them challenging to optimize a solution in the large search space for high dimensional multi-fingered robotic hands to support real-time manipulation. To deal with it, approaches that simplify kinematic structures and/or reduce the degree of freedom (DOF) of robotic arms/hands have been proposed. However, the simplification may cause model inaccuracy and thus reduce grasp performance in control and optimization.

Compared with the analytical methods, learning-based methods have improved the computational efficiency for grasping tasks with various objects or environments. Unlike the analytical methods depending on a kinematic model of robot structures, learning-based methods can solve the problems without a kinematic model. Rather, the process learns a control policy that maximizes an objective function/reward. Specifically, the Deep Reinforcement Learning (DRL) method has made significant progress in improving performance by handling high dimensional problems and enabling real-time autonomous grasping.

However, a common issue of DRL approaches and other observation-based robot learning in general is that training a robot is a time-consuming process to reach sufficient stabilities and performances. Training requires exploring a broad search space because of complex configurations of robotic hands and interactions with objects, which results in low learning efficiency. Further, the generalizability of the trained policy to grasp similar objects is limited unless they are identical to the trained objects. One of the critical reasons for the learning efficiency and the generalizability issues is that current DRL methods mainly use task related dense reward components as the only criterion to define the reward function. However, a grasp normally has multiple quality-related criteria components with different priorities such as grasp pose, contact points/regions on the object, and grasp stability. These quality-related criteria are commonly used in physics-based grasping methods but have rarely been considered in learning-based grasping. Thus, it is not easy for current learning-based robots to fundamentally understand how to achieve and improve grasp quality other than task completion. Ideally, considering both task completion and these quality-related criteria as the reward can help RL-based robots to efficiently explore the environment and generalize the learned policy.

Physics-informed learning methods have been proven in many other domains to handle the complexity of learning and dynamic temporal aspects. However, it has been rarely reported in robotic grasp learning. Further, most physics-based methods have used simplified assumptions that result in a lack of generalizability. In this paper, we introduce the physics-guided DRL with a hierarchical reward mechanism (PG-H-RL) for autonomous grasping.
The rationale of this work is that physics-informed learning leverages both the positives of learning and physics to facilitate computationally efficient yet high-quality grasping solutions by enabling the robot to fundamentally understand the problem at hand. The contributions of this work are:

- Developed a physics-guided learning strategy for autonomous grasping, which integrates physics-based metrics as rewards so they can guide the robot to understand the grasping task resulting in the improvements of learning efficiency and yielding physically consistent performances.
- Developed a hierarchical reward mechanism to learn the physics-based rewards in a prioritized logical way to help the robot further understand the priority of different metrics and improve learning efficiency.

II. RELATED WORK

A. Analytical Methods for Autonomous Grasping

Analytical approaches consider physics, kinematics, and dynamics of objects and hands to get the correct grasp, which is a vital aspect to accomplish grasping tasks. In [10], they proposed an interactive grasping simulator with the embedded dynamics engine to compute robot and object motions under the influence of external forces and contacts. Form closure and force closure properties of grasps as basic grasp quality criteria were utilized in these approaches to find the correct grasp [11]. Further, grasp quality measures for grasping have been developed to interpret the quality of robotic grasping. The measures associated with contact points on the objects and the configurations of the robotic hands were developed to provide informative measures [12, 13]. In [2], they introduced an approach combining empirical and analytical methods by imitating humans to reduce the computation time of calculating force-closure grasps. Finding the optimal solution to meet these criteria requires high computational power and limits real-world applications.

B. Learning-based Methods for Autonomous Grasping

DRL for robotic grasping has been actively studied in recent years. In earlier work, it aimed to acquire strategies of robotic grasps using DRL with images [4]. Recently, DRL methods have been proposed to accomplish autonomous grasping with vision-based observations [5]. However, these learning methods require a massive amount of training data and time to explore. To overcome this, human preferences, demonstration data, and potential contact regions were utilized [14, 15, 16]. In [17], they estimated the probability of a successful grasp using the contact region database collected from human demonstrations. The probability was considered as a partial reward to increase learning speed. Although these pure learning approaches could improve learning efficiency by reducing the search space, they do not fundamentally learn physics as a physics-informed approach would.

C. Physics-guided Learning Strategies

Despite the validated effectiveness of physics-guided learning in many applications [18], few studies have been reported in the robotic grasping field. In [19], they used the physics-guided target poses as the input for the learning process to improve performance for manipulation tasks on a physics simulator. In [16], they proposed an RL method that utilized a grasp quality metric as the reward for a good grasping configuration by using the potential grasp locations estimated with the database of the contact information of successful grasps on the objects. In addition, [20] defined the reward function that summed the force-closure quality index [21]. However, all these methods treated the grasp quality as binary bonus rewards and used a linear summation which can be easily biased or lose the information of grasping.

D. Hierarchical Reward in RL

The reward formulation of DRL is usually a linear summation of the reward components [8, 16, 17], which is implicit and inefficient to learn the multi-objective priorities and causes poor learning performance for multi-objective tasks (i.e., takes a long time to learn or even fail to learn a correct policy). Hierarchical reward methods have been proposed to enable a robot to learn multi-objective tasks such as achieving autonomy or human-like merging actions for driving [22] and performing home service activities [23]. The formulation of the reward hierarchies contains logical or weighted connections. Logical connections are strict constraints, where the higher-level hierarchy must be learned before the lower-level hierarchy. Weighted connections are soft constraints, where the higher-level hierarchy and lower-level hierarchy are learned together with a weighted summation. In [24], an RL agent for swarm robot control is trained with a logically connected hierarchical reward function. Inspired by these studies, this paper introduces the hieratical reward in physics-guided grasping learning to learn multiple physics metrics and their correlations explicitly and efficiently.

III. METHODOLOGY

PG-H-RL is developed to enable a robot to learn to stably grasp and lift objects to the target height from a table. It is assumed that the position of the object and contact information between the robotic hand and the object can be detected to calculate the grasp quality. Further, the joints of the robotic arm and hand can be controlled.

A. Reinforcement Learning Formulation

The grasping task is formulated as an RL problem that follows the Markov Decision Process (MDP). The MDP is defined as a tuple \( \{S, A, R, \gamma\} \), where \( S \) is the state of the environment and \( A \) is the set of actions. \( R(s')|s, a\) is the reward function to give the reward after the transition from state \( s \) to state \( s' \) with action \( a \) and \( \gamma \) is a discount factor. Since the task is required to consider interactions with objects and more accurate controls, the continuous control domain is considered. To solve the problem, PG-H-RL adopts the Twin Delayed Deep Deterministic policy gradient (TD3) algorithm, which is a model-free reinforcement learning [25].

B. Physics Metrics and Constraints

1) Object perspective

Grasp quality is an important factor for the agent to achieve a stable grasp. To evaluate grasp quality, contacts between the robotic hand and the object are important and necessary. The grasp matrix \( G \) is defined by the relevant velocity kinematics and force transmission properties of the contacts on the object.
in three-dimensional space [13]. In this work, there are two measures of grasp quality computed with \( G \) to evaluate the grasp quality: the measure for being graspable \( (r_{\text{graspable}}) \) and the normalized volume of the ellipsoid \( (r_{\text{rvew}}) \). Inspired by previous studies with the traditional criteria of physics metrics as binary evaluations, the null space of the grasping matrix \( \mathbb{N}(G) \) is considered as a reward to indicate whether a grasp is graspable or ungraspable based on internal object forces [13].

\[
  r_{\text{graspable}} = \begin{cases} 
    0 & \mathbb{N}(G)=0 \\
    0.1 & \mathbb{N}(G)\neq0
  \end{cases}
\]  
(1)

where \( \mathbb{N}(G)\neq0 \) reveals being graspable. It judges the grasp quality by providing an initial guide before further evaluation, which can reduce the search space. The binary value is empirically determined considering its importance level relative to other components in the entire reward. Using \( G \), the contribution of all contact forces on the object can refer to the continuous grasp quality measure using the volume of the ellipsoid in the wrench space [7] as:

\[
  Q_{\text{rvew}} = \sqrt{\det(GG^T)} = \sigma_1\sigma_2\sigma_3 \cdots \sigma_m
\]  
(2)

where, \( \sigma_1, \sigma_2, \ldots, \sigma_m \) denotes the singular values of \( G \) and \( m \) is the number of contact points on the object. This value is continuous and must be maximized to obtain the optimum grasp. In addition, the maximum value of \( Q_{\text{rvew}} \) affected by the number of contacts is used to normalize the reward, \( r_{\text{rvew}} \) as:

\[
  r_{\text{rvew}} = \text{norm}(Q_{\text{rvew}})
\]  
(3)

Fig. 1 (a) illustrates the examples of optimal grasps of \( Q_{\text{rvew}} \). Further, there are comparisons of the example robotic grasps with their \( r_{\text{rvew}} \)s in Fig. 1 (b).

2) Robotic hand perspective

Constraints were placed on the fingers to prevent closing the fingers and avoid contact between the fingers. These constraints act as penalties to the reward to re-shape the robot hand for the following grasping tasks. The former provides a penalty when the fingers attempt to close the fingers before the hand approaches close enough to grasp the object, and the latter provides a penalty if there are any contacts between the fingers.

C. Hierarchical Reward Mechanism with Physics Metrics

Using the physics metrics, multiple components in the reward function are prioritized logically to learn autonomous grasping progressively. An autonomous grasp task can be broken down into three sequential stages: 1) approaching the object, 2) grasping the object, and 3) lifting the object from the table. In the approaching stage, the robot perspective constraints were included in the reward function before touching and after grasping the object. The grasping stage is designed to include grasp quality physics metrics with a hierarchical structure considering their priorities. The hierarchical structure reflects that a measure and/or constraint with a lower priority is not considered when a condition of one with a higher priority is not satisfied. It improves the learning efficiency because the agent can explore action/state spaces efficiently depending on the satisfactions of the higher level of hierarchies. Fig. 2 illustrates the hierarchical physics-guided reward mechanism, including the three sequential stages for each training episode.

The reward function consists of multiple reward components for each stage. The approaching stage includes \textit{get close to the target} (a penalty to the reward for the distance between the robotic hand and the object), \textit{prevent closing the fingers}, and \textit{avoid contact between the fingers}. get close to the target is \( p_{\text{target}} \) and can be defined as:

\[
  p_{\text{target}} = -\left( \epsilon \times \text{dist}_{\text{obj\_hand}} \right)
\]  
(4)

where \( \epsilon \) is a weighted coefficient and task dependent. It is determined as 10 to balance with other reward components. In the grasping stage, pre-grasp preparation is a condition to determine the agent gets close enough to the object. The reward determined by the normalized exponential value of the distance between the robotic hand and the object, \( r_{\text{dist}} \), is added to the reward.

\[
  r_{\text{dist}} = \text{norm} \left( e^{0.1 \times \text{dist}_{\text{obj\_hand}}} \right)
\]  
(5)

where \( \text{dist}_{\text{obj\_hand}} \) is the distance between the hand and the object. Being graspable is a condition to decide whether a grasp can grasp or not based on (1). If it reveals being graspable then an additional reward, \( \text{weight} \), is added to the reward, which is determined by prudent considerations to balance with other reward components. Contact forces to grasp is a condition to show how much contact forces are applied to grasp the object based on \( r_{\text{rvew}} \) in (3).

In the lifting stage, a further reward is added to the reward function as a binary guidance.

![Figure 1](image1.png)

Figure 1 The examples of the volume of ellipsoid in wrench space of \( G \). (a) illustrates the optimal grasps with symmetric locations of contact points on the 2-D object. (b) shows the example grasps of the robotic hand. The grasp quality comparisons of these 3 \( r_{\text{rvew}} \)s is \( r_{\text{rvew}}(1) < r_{\text{rvew}}(2) < r_{\text{rvew}}(3) \). \( r_{\text{rvew}}(2) \) has only two contact points with the object, and \( r_{\text{rvew}}(1) \) is one of unstable grasps with contact points.

![Figure 2](image2.png)

Figure 2 The autonomous robotic grasp task is decomposed into three stages: approaching, grasping, and lifting. In the grasping stage, the hierarchical physics-guided mechanism is implemented. Blue lines and green lines represent robot and object perspectives, respectively.
stage, \( r_{\text{obj\_height}} \) is calculated and added to the total reward, which is a measure related to the error between the height of the object and the target height:

\[
r_{\text{obj\_height}} = a \times \left| \beta - \text{error}_{\text{obj\_height}} \right|
\]

where \( a \) is a weighted coefficient for the reward and \( \beta \) is the maximum error of \( \text{error}_{\text{obj\_height}} \). Fig. 3 illustrates the total reward function, \( \lambda \), \( \mu \), and \( \nu \) are binary coefficients that are determined as 1 when both conditions of higher hierarchies and the corresponding condition are satisfied, otherwise they are 0. They can be described as:

\[
\begin{align*}
\lambda &= \begin{cases} 0 & \text{if } 1^{\text{st}} \text{ condition is not satisfied} \\ 1 & \text{if } 1^{\text{st}} \text{ condition is satisfied} \end{cases} \\
\mu &= \begin{cases} 0 & \text{if } 2^{\text{nd}} \text{ condition is not satisfied} \\ 1 & \text{if } 2^{\text{nd}} \text{ condition is satisfied} \end{cases} \\
\nu &= \begin{cases} 0 & \text{if } 3^{\text{rd}} \text{ condition is not satisfied} \\ \lambda \times \mu \times 1 & \text{if } 3^{\text{rd}} \text{ condition is satisfied} \end{cases}
\end{align*}
\]

IV. EXPERIMENTS

A. Experimental Setup

1) MICO arm: to accomplish the task described above, a Kinova’s MICO arm is used [26], which has 6 rotational joints for the arm and the three fingered gripper. The MICO arm can be controlled to open and close the fingers to grasp the object.

2) Simulator: CoppeliaSim (V-REP) is used as the simulator [27]. It provides a precise physics engine for interactions between the robot, the object, and the environment. Using an API of V-REP, scripts were programmed in the scene to remotely connect for the kinematics and the sensing details. The maximum number of steps for each training episode is 300.

For the TD3 agent, we define the state observation, including the positions of the object, the joint angles for the fingers, the position of the robotic hand, and the joint angles of the arm. Table I shows the hyper-parameters for the TD3 agent. \( \alpha \) and \( \beta \) in (6) are empirically determined as 30 and 0.05. In the training, the cube with a side length 0.065m was used to train the policy. The cylinder and polyhedron (Fig. 4) and the various sizes as \( \pm 10\% \) of the original size of objects were used to evaluate the generalizability to different objects since they are similar to the trained object but with different contours and require different contact points and grasping shapes. The target lifting height was 0.05m above the table.

B. Evaluation Methods and Metrics

To evaluate the influence of the physics metrics and the hierarchical reward mechanism of the PG-H-RL method, two different reward functions are considered as baselines. Except the reward functions, the baselines are in the same conditions and setups with PG-H-RL. A baseline, Task Only, has a reward function with task related dense reward components:

\[
r_{\text{Task Only}} = p_{\text{target}} + r_{\text{obj\_height}}
\]

It uses a linear summation instead of a hierarchical reward mechanism. The second baseline, Linear Summed, uses all reward components that are considered in PG-H-RL, but a linear summation is used instead of a hierarchical reward mechanism. The PG-H-RL method and the two baselines were trained with the same grasping task to evaluate the learning efficiencies and the learning outcomes. Trends of the total reward at different training episodes are used to compare the learning efficiency. The trends are evaluated by comparing how fast the total reward increases and maintaining this increase. To evaluate the learning outcome, the success rate and the height error are used for the task completion. The success rate is the percentage of object lifting to the target height. The height error indicates the difference between the object actual height and the target. Further, the distance to the object center and \( Q_{\text{approx}} \) are used to evaluate the grasp quality. \( Q_{\text{approx}} \) is considered to assess the stable and firm grasp quality. The distance to the object center is the distance between the desired location of the object center and the actual object center to evaluate the stability of a grasp pose (shown in Fig. 7 (a)).

V. RESULTS AND DISCUSSION

A. Hierarchical Reward Mechanism

Fig. 5 shows an episode reward using PG-H-RL in late
training. The duration from 0 to 21 steps indicates the robotic hand is approaching the object. Then, the reward entered the next stage for grasping with three physics-guided reward components associated with the object perspective. The condition of pre-grasp preparation as the first hierarchy was satisfied after 21 steps. The condition of being graspable as the second hierarchy was satisfied after 25 steps. The condition of contact forces to grasp as the third hierarchy was satisfied after 29 steps. It reveals that the hierarchical reward mechanism allows the agent to precede learning of the higher level of a constraint than lower ones.

B. Learning Efficiency

The experiments executed 5 cases which means that agents learned the task 5 times for each method to generate statistical results, including the means and variances during training. The total reward indicates the earned rewards during 300 steps for an episode, and Fig. 6 shows the means of the total rewards for each training episode. The results reveal that PG-H-RL reaches 56.40% of its maximum total reward with 200 training episodes, while Task Only reaches 34.04% and Linear Summed reaches 47.25% of their maximum rewards. Further, PG-H-RL shows a steadier trend with the mean of the total rewards than the other two methods. Even further training proceeds, Task Only and Linear Summed show reducing the total rewards due to overfitting or overwriting the agent’s experiment by the new experience. The above results validate that considering the physics metrics and constraints as the reward are effective in guiding agents to learn the task faster. However, introducing more objectives without considering their relative priorities in the reward function confuses the agent to balance among different reward components and makes the learning unstable. Involving the physics and performance constraints with the appropriate hierarchical mechanism in the reward function is effective in improving the learning efficiency.

C. Learning Outcome

Although the success rates for the methods indicate there are still failure cases due to the difficulties to predict the interactions between the robotic hand and the object or the environment with the high-dimension robotic system, PG-H-RL outperforms Task Only, and Linear Summed with various numbers of training episodes, 200, 600, and 1000. Table II shows the success rates for different shapes and sizes of the object that are performed with the learned policies with 1000 training episodes. Since the policies are trained with the cube shape of the object, the results with the cube are relatively higher than other shapes. The results with the polyhedron show the worst success rates due to the bigger differences of the contact positions with the cube than the cylinder. The size reduction of 10% resulted in reducing the success rates for all methods and shapes since the policies are trained with the larger object. With respect to the methods, PG-H-RL always outperforms Task Only and Linear Summed for both shapes and sizes since PG-H-RL considers the physics metrics, which makes it more generalizable to different objects. The results with 200 and 600 numbers of training episodes also have consistent inclinations in the performances. It is difficult to reach a 100% rate since the training environment is a stochastic environment that contains uncertainty and noise that may cause task failure. Further, it can be caused by the nature of reinforcement learning that brings stochastic behavior and leads to different grasping behaviors in each testing. Especially for the grasp task, the failures occur by several reasons such as wrong approaching direction, finger closing timing, and interactions with object. It indicates that learning still cannot fully cover all possible testing cases. These are the same influences on all the methods, but PG-H-RL outperforms the baseline methods due to utilizing the physics metrics and the hierarchical reward mechanism. To confirm statistically significant differences for the success rates between PG-H-RL and the baselines, N-1 Chi-Square test [28] that compares two binary variables for two independent groups is used to calculate p-values in Table III. A p-value less than 0.05 means a statistically significant difference. Only the p-value of the comparison between PG-H-RL and Task Only with 1000 training episodes do not show statistically significant difference, while all other p-values

![Figure 6](image-url)

Figure 6 The mean and variance of the total rewards that are normalized for each learning method. The shaded areas indicate their standard deviations for the 5 cases.

| Shape  | Size    | PG-H-RL [%] | Task Only [%] | Linear Summed [%] |
|--------|---------|-------------|---------------|-------------------|
| Cube   | -10%    | 72          | 66            | 44                |
|        | Original| 86          | 72            | 56                |
| Cylinder | -10%  | 50          | 32            | 40                |
|        | Original| 78          | 56            | 44                |
| Polyhedron | Original | 42          | 16            | 28                |
|        | +10%    | 64          | 36            | 30                |

| Table III. P-VALUES OF N-1 CHI-SQUARE TEST FOR SUCCESS RATES |
| Training episodes | PG-H-RL vs Task Only | PG-H-RL vs Linear Summed |
|-------------------|----------------------|--------------------------|
| 200               | 0.0074               | 1.3912e-5                |
| 600               | 0.0004               | 0.0014                   |
| 1000              | 0.2183               | 8.6378e-5                |

a. $\chi^2$ can be calculated using the binary result (pass and fail) and the total number of trials [28].
b. The N-1 Chi-squared distribution (p-value) can be obtained using $\chi^2$ and a table of Chi-square values or the Excel function CHIDIST.
show statistically significant differences.

Further, Fig. 7 shows the performance comparisons with 1000 training episodes based on the three criteria: the error of height, the distance to the object center, and the grasp quality. All parameters are illustrated in Fig. 7 (a). The error of height is defined as:

\[
err_{\text{obj.height}} = |T_c - \text{Obj}|
\]

The distance to the object center is defined as:

\[
dist_{\text{obj.hand}} = |H_{xyz} - \text{Obj}_c|
\]

PG-H-RL surpasses Task Only and Linear Summed in the distance to the object center and the grasp quality, which means it performs firmer and more stable grasps. The experiment results with various shapes and sizes of the object show that PG-H-RL outperforms Task Only and Linear Summed for the three criteria. To confirm statistically significant differences for the comparisons, one-way analysis of variance (ANOVA) is used to calculate p-values for the pair-wise distribution comparisons between PG-H-RL and the baselines with the least significant difference (LSD) correction factor [29], shown in Table IV. Using the threshold 0.05, only the p-value between PG-H-RL and Task Only in the error of height does not show a statistically significant difference. The comparisons for the distance to the object center and the grasp quality verify that PG-H-RL is effective in learning how to maintain appropriate distances between the robotic hand and the object to achieve a firm grasp. With respect to the grasp quality, the results verify a reward function—with multiple reward components and without considering their priorities—makes it difficult to learn. In other words, utilizing the hierarchical reward mechanism to learn multiple objectives can improve learning performance. PG-H-RL and Task Only show similar performances in the error of height. Given that the Task Only reward solely focuses on task related rewards with a specific lifting height, this similarity shows that the PG-H-RL can learn the additional grasp-quality-related rewards without sacrificing learning the conventional task related reward.

Fig. 8 illustrates the performance trained using PG-H-RL and the performance trained using Task Only for a specific grasp instance when the object reaches the target height. Although both cases succeed in the task, the former result shows a higher grasp quality with three contacts with the object to secure the object more stably. In contrast, the latter result shows a lower grasp quality with two contacts with the object. Since Task Only encourages task related rewards but cannot provide finer reward differences in terms of grasp quality, it is difficult to achieve secure grasps. In contrast, the PG-H-RL method can provide a finer evaluation of postures since it considers both task related rewards and grasp quality. Thus, PG-H-RL performed outperforms Task Only as the results of the overall success rates and the performance criteria are showing, even it is possible to find a secure grasp from the individual cases of Task Only.

### Table IV. P-values Using One-Way Analysis of Variance for Learning Performances

| Category           | PG-H-RL vs Task Only | PG-H-RL vs Linear Summed |
|--------------------|----------------------|--------------------------|
| Error of Height    | 0.4474               | 8.9922e-18               |
| Distance to the Object Center | 9.8952e-16             | 5.8884e-13               |
| Grasp Quality      | 1.5815e-15           | 4.0366e-19               |

a. The One-way Analysis of variance for the performance data can be obtained using the MATLAB function (ANOVA1). The p-value from the distributions can be derived by the function which performs the One-way Analysis of Variance.

### References

[1] F. Ficuciello, G. Palli, C. Melchiorri and B. Siciliano, ”A model-based strategy for mapping human grasps to robotic hands using synergies,” 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1737-1742, 2013

[2] El-Khoury, S., & Sabhani, A., “A new strategy combining empirical and analytical approaches for grasping unknown 3D objects,” *Robotics and Autonomous Systems*, 58(5), 497–507, 2010

[3] K. Cobbe, O. Klimov, C. Hesse, T. Kim, & J. Schulman, “Quantifying generalization in reinforcement learning,” 36th International Conference on Machine Learning, ICML 2019, 2019, 2280–2289, arXiv:1812.02341
[4] Moussa, M. A., & Kamel, M. S., “Connectionist model for learning robotic grasps using reinforcement learning,” IEEE International Conference on Neural Networks - Conference Proceedings, 3, 1771–1776, 1996

[5] Kalashnikov, D., Ir pan, A., Pastor, P., Ibarz, J., Herzog, A., Jang, E., Quillen, D., Holly, E., Kalakrishnan, M., Vanhoucke, V., & Levine, S., “Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation,” ArXiv CoRL, 1–23, 2018

[6] H. Sekkat, S. Tigan, R. Saadane, A. Cherihi, “Vision-Based Robotic Arm Control Algorithm Using Deep Reinforcement Learning for Autonomous Objects Grasping,” Appl. Sci, 11, 7917, 2021

[7] I. Popov, N. Hcess, T. Lillicrap, Roland Hafer, Gabriel Barth-Maron, Matej Vecerík, T. Lampe, Y. Tassa, T. Erez and Martin A. Riedmiller, “Data-efficient Deep Reinforcement Learning for Dexterous Manipulation,” ArXiv abs/1704.03073, 2017

[8] Rajeswaran, A., Kumar, V., Gupta, A., Vezzani, G., Schulman, J., Todorov, E., & Levine, S., “Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations,” arXiv:1709.10087, 2018.

[9] Kalakrishnan, M., Righetti, L., Pastor, P., & Schaal, S., “Learning force control policies for compliant robotic manipulation,” Proceedings of the 29th International Conference on Machine Learning, ICML, 1, 4639–4644, 2012

[10] A. T. Miller and P. K. Allen, 'Graspit! A versatile simulator for robotic grasping,' in IEEE Robotics & Automation Magazine, vol. 11, no. 4, pp. 110-122, Dec. 2004

[11] E. Rimon and J. Burdick, "On force and form closure for multiple finger grasps," Proc. IEEE ICRA, pp. 1795–1800, 1996

[12] Roa, M. A., & Suárez, R., “Grasp quality measures: review and performance,” Autonomous Robots, 38(1), pp. 65–88, 2014

[13] Prattichizzo, D., & Trín kle, J. C., “Springer Handbook of Robotics,” In Springer Handbook of Robotics, 2008

[14] Pinzler, R., Akour, R., Osa, T., Peters, J., & Neumann, G., “Sample and Feedback Efficient Hierarchical Reinforcement Learning from Human Preferences,” Proceedings - IEEE International Conference on Robotics and Automation, 596–601, 2018

[15] Mandikal, P., & Grauman, K., “Dexterous Robotic Grasping with Object-Centric Visual Affordances,” 1–11, 2020, http://arxiv.org/abs/2009.01439

[16] Osa, T., Peters, J., & Neumann, G., “Hierarchical reinforcement learning of multiple grasping strategies with human instructions,” Advanced Robotics, 32(18), 955–968, 2018

[17] E. Valarezo Alazco et al., “Natural object manipulation using anthropomorphic robotic hand through deep reinforcement learning and deep grasping probability network,” Appl. Intell., 2020

[18] Zhao, P., & Liu, Y., “Physics-Informed Deep Reinforcement Learning for Aircraft Conflict Resolution,” IEEE Transactions on Intelligent Transportation Systems, 1–14, 2021

[19] Garcia-Hernando, G., Johns, E., & Kim, T. K., “Physics-based dexterous manipulations with estimated hand poses and residual reinforcement learning,” IEEE International Conference on Intelligent Robots and Systems, 9561–9568, 2020

[20] Monforte, M., & Ficuciello, F., “A Reinforcement Learning Method Using Multifunctional Principal Component Analysis for Human-like Grasping,” IEEE Transactions on Cognitive and Developmental Systems, 8920(c), 1–1, 2020

[21] Bicchi A. “On the closure properties of robotic grasping. International Journal of Robotics Research,” 14:319–334, 1994

[22] Sun, L., “Intelligent and High-Performance Behavior Design of Autonomous Systems via Learning, Optimization and Control”, Ph. D Thesis, Mechanical Engineering, Univ. of California, Berkeley, California, 2019.

[23] Zhang, M., Tian, G., Zhang, Y., & Duan, P., “Service skill improvement for home robots: Autonomous generation of action sequence based on reinforcement learning,” Knowledge-Based Systems, 212, 106605, 2021

[24] Clayton, N. R., & Abbass, H., “Machine Teaching in Hierarchical Genetic Reinforcement Learning: Curriculum Design of Reward Functions for Swarm Shepherding.” 2019 IEEE Congress on Evolutionary Computation, CEC 2019 - Proceedings, 1259–1266, 2019

[25] Fujimoto, S., Van Hoof, H., & Meger, D., “Addressing Function Approximation Error in Actor-Critic Methods,” 35th International Conference on Machine Learning, I CML 2018, 4, 2587–2601, 2018

[26] Robotics company—Robotic assistive technology—Kinova. https://www.kinovarobotics.com/en. Accessed 18 Feb 2021

[27] E. Rohmer, S. P. Singh, and M. Friesic, “V-REP: A versatile and scalable robot simulation framework,” Proceedings of the International Conference on Intelligent Robots and Systems (IROS), pp. 321–326, 2013.

[28] Sauro, J., Lewis, J., “Quantifying the user Experience,” 2nd ed., Netherlands: Elevier Inc. of Publisher, ch. 5, pp. 61-102, 2016

[29] Wu, C. F. J., Hamada, M. S., “Experiments: Planning, Analysis, and Parameter Design Optimization,” 2nd ed., Wiley of Publisher, 2009