electronics

MDPI

Article

Research on Perception and Control Technology for Dexterous Robot Operation

Tengteng Zhang and Hongwei Mo *

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China; zttdouble@hrbeu.edu.cn

* Correspondence: mhonwei@163.com

Abstract: Robotic grasping in cluttered environments is a fundamental and challenging task in robotics research. The ability to autonomously grasp objects in cluttered scenes is crucial for robots to perform complex tasks in real-world scenarios. Conventional grasping is based on the known object model in a structured environment, but the adaptability of unknown objects and complicated situations is constrained. In this paper, we present a robotic grasp architecture of attention-based deep reinforcement learning. To prevent the loss of local information, the prominent characteristics of input images are automatically extracted using a full convolutional network. In contrast to previous model-based and data-driven methods, the reward is remodeled in an effort to address the sparse rewards. The experimental results show that our method can double the learning speed in grasping a series of randomly placed objects. In real-word experiments, the grasping success rate of the robot platform reaches 90.4%, which outperforms several baselines.

Keywords: deep reinforcement learning; attention mechanism; reward remodeling; learning and adaptation; dexterous manipulation

Citation: Zhang, T.; Mo, H. Research on Perception and Control Technology for Dexterous Robot Operation. *Electronics* **2023**, *12*, 3065. https://doi.org/10.3390/electronics12143065

Academic Editor: Dah-Jye Lee

Received: 3 June 2023
Revised: 9 July 2023
Accepted: 11 July 2023
Published: 13 July 2023

1. Introduction

Robot grasping, or the capacity for object manipulation, is a key skill required for robot contact with the physical world. While robotic grasping has come a long way, the challenges intensify when dealing with cluttered environments where multiple objects are present. In clutter, objects can be densely packed, occluded by other objects, or have varying orientations and positions. These variables add complexity, which conventional grasp planning techniques find difficult to manage. The difficulties presented by ambiguous object postures and probable collisions must be addressed in order to ensure robust grasp operation under clutter. Traditional grasp planning approaches often struggle in cluttered environments due to their reliance on explicit models or heuristics that may not generalize well. Robots can acquire grasping rules directly from interactions by using a deep reinforcement learning approach, which enables them to adapt and overcome the difficulties of cluttered environments.

Currently, grasping flexibly in the continuous motion space for robots in a complicated environment is difficult. The robot's performance will be hampered by the crowded objects in an unstructured environment. The following three factors are crucial to robot grasping operations: (1) How can an unstructured or complicated environment be perceived in an autonomous and high-precision manner? (2) How can the generalization performance be improved? (3) How may new operational abilities be acquired at a lower cost and with less training data? It has been demonstrated that deep reinforcement learning is capable of solving the complex control issues of the robot arm in an efficient manner [1,2]. Nevertheless, in reinforcement learning, the robot is not informed about which actions to take but instead needs to determine through iterative experimentation which actions will yield the maximum reward by trial and error. The majority of the time, these actions not

only affect immediate rewards but also impact the subsequent action, thus influencing all future rewards. When training a real robot arm with reinforcement learning algorithms, it will suffer from issues such as a low sample efficiency and long learning cycles [3,4]. Self-supervised learning methods could reduce manual inputs and improve data efficiency for many objects and scenarios [5]. As a result, its implementation on real robot platforms is quite challenging.

Due to the high sample complexity, the direct application of deep reinforcement learning for manipulator control is frequently limited to fairly straightforward tasks [6]. Additionally, reinforcement learning still struggles with sparse rewards, and when the task and action dimension increase, it becomes harder to learn non-zero rewards during learning and exploration. We propose a method called CBAM-Dueling DQN, which combines the attention mechanism with DQN for grasping invisible objects in order to overcome these difficulties. This method enables the virtual to real conversion and exhibits some adaptability in dynamic environments.

This paper presents a comprehensive framework that tackles the challenge of achieving robust grasp operations in cluttered environments for multi-objective robotic tasks using deep reinforcement learning. Our proposed approach aims to overcome the drawbacks of earlier studies by incorporating robustness, adaptability, and multi-objective considerations within the deep RL framework. By simultaneously optimizing grasp success rates, the efficiency, and other relevant objectives, our framework enables the robot to learn policies that exhibit an improved performance and generalize effectively to cluttered scenarios. Additionally, we conduct a comparative analysis against several baselines to evaluate the effectiveness of our approach. We run actual robot testing to confirm the effectiveness of our strategy and show how it generalizes to unknown objects (see Figure 1). The experimental results confirm the practicality and generalization capabilities of our approach in grasping objects in diverse and cluttered environments.

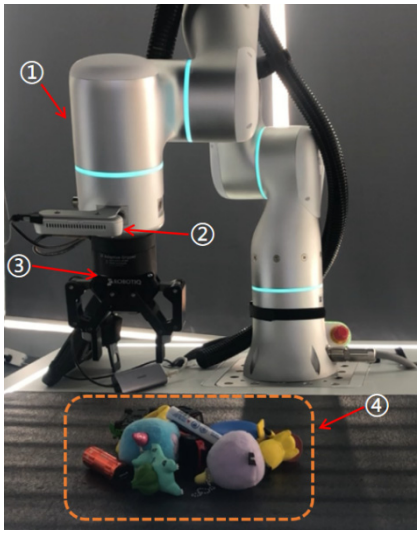

Figure 1. Our robot is able to grasp unknown objects with precise perception, perform decision making at a high rate of speed and reduced training time and receive remarkable results. Our experimental setup consists of (1) a cooperative robot, (2) a depth camera, (3) a manipulator with a two-finger gripper, and (4) a manipulation platform of various unknown objects.

In summary, the main contributions of this paper are as follows:

(1) To enable the robot to actively observe the surroundings, an attention-based active exploration dueling DQN method is presented.
(2) The experiment results demonstrate that the system has improved the grasping efficiency in a real chaotic environment because it fully incorporates the attention mechanism and reinforcement learning algorithm.

(3) We reduce the training time while significantly improving the results: after 10,000 iterations training for 25 h, the optimal grasping success rate is 90.4%.
(4) The low sample efficiency and sparse reward issues have been resolved, and the modified method can double the learning speed.

The rest of the paper is organized as follows. Briefly, the related work will be presented in Section 2. We will next go into great depth on the proposed approach in Section 3. Section 4 presents the experimental results, and finally, Section 5 concludes this work.

2. Related Work

Robotic grasping in cluttered environments is a challenging problem that has garnered significant attention in the robotics community. In order to tackle this issue, researchers have investigated a number of strategies using both analytical and learning-based methods [7–9]. These methods for graspable objects are time consuming and impractical for real-time implementation. Similar to iterative neural networks, RL is a process of running agents through a series of state-action pairs. It extracts information from data by sampling and combines the Markov decision process (MDP) with a large number of state–action pairs. The complex probability distribution model of the reward is associated with it.

Analytical methods often rely on geometric reasoning and physics-based models to plan grasps. These methods provide workable grasp configurations by taking into account elements like the object shape, contact locations, and hand kinematics. These approaches, however, have difficulty in cluttered contexts because of occlusions, ambiguous item positions, and the incapability to take into account dynamic environments. On the other hand, learning-based systems use data-driven methodologies to learn grasp policies. Supervised learning methods utilize labeled training data to predict grasp configurations from visual or tactile inputs.

Reinforcement learning (RL) has demonstrated remarkable success in training robots to perform complex tasks, including robotic grasping. Robots can learn policies by interacting with their surroundings and receiving feedback in the form of rewards or penalties based on task performance using RL-based techniques. Mokhtar et al. [10] presented a deep reinforcement learning approach to learn grasping and pushing policies for manipulating a goal object in highly cluttered environments to address the problem of present objects preventing the grasp action. A dual reinforcement learning model technique allowed the robot to handle complex scenarios with excellent resilience and grasping success rates. However, due to numerous grasping attempts, the method suffers from a relatively slow speed.

To address the challenges of cluttered environments, researchers have looked into various methods for reinforcement-learning-based grasping. Some studies have incorporated depth information into the state representation to enable better perception of objects in clutter. For instance, Duan et al. [11] proposed an end-to-end, multi-task semantic grasping convolutional neural network (MSG-ConvNet) which enables the robot to select an optimal grasping area in an active perception way through simply reasoning on the multi-modal information output by the proposed model. Self-supervised learning for suction grasping in a congested environment was examined in another work [12]. It made the robotic picking system learn picking skills from scratch, but this work only considered cylinders. It is exciting to combine Resnet with the U-net structure, a unique convolutional neural network (CNN) framework, to forecast the picking region without recognition or a pose estimate [13]. The authors trained the network end to end with online examples, which took a lot of time due to the variety of poses, types of stacks, and complex backgrounds in bin picking situations. However, the methods in [11–13] suffer from a very poor speed because of the repeated gripping attempts.

Additionally, researchers have investigated the use of reinforcement learning for multi-objective robotic tasks [14–16]. Agents can optimize several objectives by learning a set of trade-off policies using multi-objective reinforcement learning (MORL) algorithms. A method is to consider the states of all the targets so that the pushing action can expand

the grasping space of all targets as much as possible to achieve the minimum total number of pushing and grasping actions, which in turn increases the effectiveness of the entire system for multi-objective robotic tasks [17]. Many of these methods make the assumption that the scene is adequately scattered, with objects well isolated. However, depending exclusively on grasp motions is insufficient when dealing with dense clutter where objects are packed closely together.

Overall, robot grasping is mainly focused on model-based [18–20] and data-driven [21–23] methods. Many learning-based approaches have tried to overcome the problem of grasping objects from a cluster of multiple objects [24–26]. Traditional analytical methods rely on 3D models of known objects in order to identify stable force closures for grasping. Nonetheless, it might frequently be difficult to find accurate models for unique objects. In order to improve the efficiency, Azizzadenesheli et al. [27] presented a unique RL algorithm that blends model-free and model-based approaches. Related to our work, Stephen et al. enhanced the performance of exploration with efficient learning on environments with sparse rewards [28]. In this paper, we proposed a model-free self-supervised learning method for robot grasping in cluttered environments. Instead of employing ResNet and a feasible grasping set, for feature extraction, we adopt a different strategy and use a Densenet network [29]. Our work is based on a framework of an attention-based dueling DQN in combination with a full convolutional network. The self-supervised learning enables our method to achieve excellent grasp success rates even without an object model.

3. Approach

In order to identify the answer to the MDP, which is a policy that maps the present state (s_t) to an action (a_t) that maximizes the sum of expected rewards, we trained the robot using attention-based dueling deep reinforcement learning (Q-learning). We restrict our procedure to a discrete action space to increase the sampling efficiency because of the enormous action space and sparse rewards.

3.1. Problem Formulation

Our suggested methodology aims to provide a robot with robust grasp operations in cluttered environments while taking several objectives into account. Specifically, we will train a deep reinforcement learning (RL) agent to learn grasping policies that optimize the grasp success rates and efficiency. In order to accomplish this, we formulate the issue as a Markov decision process, in which the robot interacts with the environment and makes decisions based on the observed states in order to maximize its anticipated cumulative rewards.

3.2. Prioritized Experience Replay

Prioritizing the experience replay is combined with strengthening the experience replay mechanism. How to judge whether a sample is valuable, or to cause a greater TD error (temporal difference error), is the key to the prioritized experience replay mechanism. The sample value increases with the size of the estimate to the target value error. If the TD deviation at sample i is defined as σ_i, the following is true for the sampling probability:

$$C_i = \frac{C_i^j}{\sum_m C_m^j} \tag{1}$$

Each sample TD error is represented by C_i during calculations, and j is used to modify the significance of the TD error. The TD error value is utilized immediately when j is 1. When j is less than 1, the influence of samples with a high TD error can be reduced, while the influence of samples with a low TD error can be increased. There are often two different definition methods of C_i: the proportion of priority $C_i = |\theta_i| + \varepsilon$ and the priority-based sorting method $C_i = 1/rank(i)$, where $rank(i)$ is given by ordering $|\theta_i|$.

The sample is drawn with unequal probability when the probability distribution of the priority playback is used. Since the distributions of the sample and the action value function are not identical, the model update is biased. The important sampling weight w is used to correct this deviation, and the corresponding formula is:

$$w_\alpha = \left(\frac{1}{N \cdot P(\alpha)}\right)^\beta \tag{2}$$

where N represents the number of samples stored in the replay buffer and β represents the degree of correction. Add a weighted w_α in front of each learned sample to make the update unbiased.

3.3. Reward Remodeling

When the allowable error between the end-effector position of the manipulator and the target position reaches a certain value, the manipulator will get a higher reward value, such as 10. In this process, when the target is not reached, each step will only get a small reward, such as -0.01. According to Equation (3), the judgment of the reward is related to the adaptive size of the target. Its form can be expressed as:

$$r_t = \begin{cases} 10 & \|X_{\theta_t} - X^T\| \le \rho(e) \\ -0.01 & \|X_{\theta_t} - X^T\| > \rho(e) \end{cases} \tag{3}$$

Nevertheless, because there are so few target rewards, it is difficult to sufficiently train the learning policy. We changed the rewards and adjusted the intermediate rewards when the end-effector and target point were separated by a specific distance. The setup of reward is as follows:

$$r_{st} = \frac{\|X_{\theta_{t-1}} - X^T\| - \|X_{\theta_t} - X^T\|}{\|X_{\theta_t} - X^T\|} \tag{4}$$

where r_{st} must be stabilized in $[-0.08, 0.08]$ since it represents the reward determined by reward modification at step t in Formula (4). The stability of training will be impacted if the intermediate reward is too substantial.

3.4. Network

Drawing the attention mechanism into the visual network to build the attention network can elevate the object perception. Our attention architecture (CBAMNet) is a convolutional block attention module and is mainly based on the deep residual network (DenseNet-121). The network includes a convolutional layer and four attention blocks. The spatial attention (SA) mechanism and channel attention (CA) mechanism are used for residual concatenation in the attention blocks. One generates the channel attention map, which can effectively draw attention to the global information. The other focuses on the spatial feature maps of the attention space and target space, respectively. The CA and SA are independent of each other and are combined in sequence to enhance attention to the position and feature information of the objects in the workspace. The output features are merged and input into two action networks, and then the visualization maps of the shifting and grasping action are generated. The pixel-wise prediction value Q and probability of the action are obtained by using the greedy strategy. The purpose of self-training is to achieve a better target value:

$$Q_{i+1}(s_t, a_t) = R_{t+1}(s_t, s_{t+1}) + \gamma \max_a Q(s_{t+1}, a; \theta_{t+1}) \tag{5}$$

where Q_{t+1} is the predicted value of the executed action, $R_{t+1}(s_t, a_t)$ is the reward value obtained after executing action a_t, θ_{t+1} is the network parameter at time $t + 1$, and the maximum predicted value Q is derived from selecting the optimal action.

The value function shows how much the agent conducts behavior α in state s, whereas the Q function in the deep Q network shows how much the agent performs behavior in state

s. Prioritized experience replay is introduced to improve the decision-making process. The advantage function shows how much better a robot performs behavior a. The adversarial network DQN considers the Q network to be divided into two parts. The first part is only related to the state s and has nothing to do with the specific action a. This part is called the value function and is denoted as $V(s, w, \alpha)$. The second part is related to state s and action a. Its symbol is $A(s, a, w, \beta)$ and it is known as the advantage function. The final Q function can then be rewritten as follows:

$$Q(s, a, w, \alpha, \beta) = V(s, w, \alpha) + A(s, a, w, \beta) \tag{6}$$

where w is the network parameter, α is the network parameter of the unique part of the value function, and β is the network parameter of the unique part of the advantage function.

In the network architecture (see Figure 2), we propose an estimated Q value more precisely and use the advantage function to determine whether the currently selected activity receives a higher reward value than other behaviors. The efficiency and execution time can be increased by removing unimportant experience sequences using the priority sorting method.

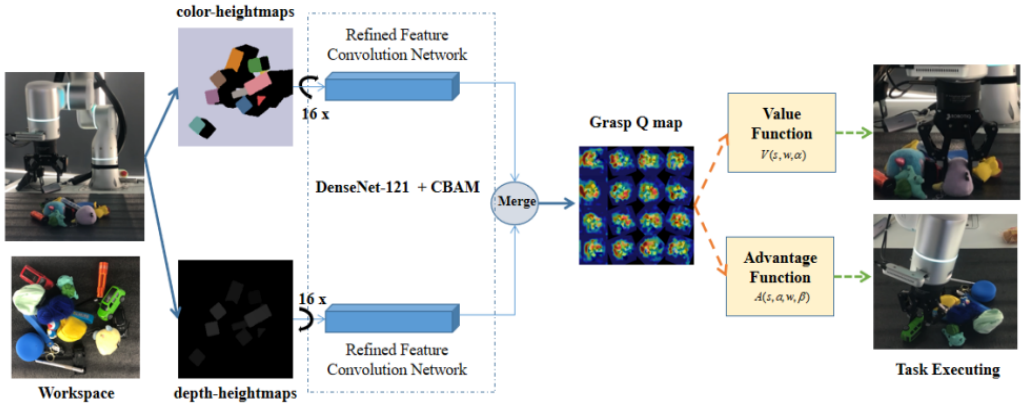

Figure 2. Overview. To construct height maps, the visual 3D data captured by a statically mounted RGB-D camera are orthographically projected. The height maps are then input into the FCN after being rotated by 16 orientations. The attention mechanism is utilized to raise the target's expressiveness, extract the workspace features, and produce the target affordance map after the action network. The dense pixel-level map predicted by the full convolutional network has several alternative locations where the grasp can be executed at a given angle.

4. Experiments

Usually, there are multiple objects presented in the actual grasping scene, which brings significant difficulties to object grasping. We carried out model training in the simulation environment to lessen the loss of the robot. The trained model would then be transferred to the actual robot arm. We verified how to improve the grasping success rate in real-word experiments.

4.1. Experimental Setup

We used a simulated six-DOF robot arm (UR5) and a two-finger parallel gripper (RobotIQ 2F-85) with an adjustable range of 0–85 mm to train the model. The performance of our proposed method was evaluated in various scenes in a simulation. The v-rep3.6 simulation platform, which is based on bullet 2.8, was used for all experiments. An Intel RealSense D435 depth camera was placed at the end effector.

4.2. Training

In the simulated environment, we conducted extensive experiments to evaluate the performance of the trained grasping policies. We varied the clutter densities, object config-

urations, and task requirements to assess the robustness and adaptability of the policies. In the simulation environment, multiple objects are randomly generated, and their states are uncertain per episode. There are adhesion and stacking cases of the objects. The shifting action is used as the grasping operation to create a better space and finally succeed in grasping.

The primary CBAM-Dueling DQN was trained with 10,000 grasp attempts. Multiple objects were placed in both random and challenging arrangements in a workspace (1 m $\times$ 1 m) in training. We executed 10 runs for each object per episode. The durations for grasp attempts were around 9 s, resulting in an overall training time of 25 h. The exploration rate discount was set as 0.99. The momentum coefficient in this experiment was set as 0.95 and the network parameters were updated using stochastic gradient descent. To avoid a bad algorithm training effect caused by insufficient sample data, the training was initiated when the number of sequence samples stored in the replay buffer reached 5000 and the maximum replay buffer capacity was set as 580,000. For verifying the model in the subsequent replay, a mini-batch sample (32) was randomly chosen from the replay buffer. Each layer was followed by ReLU, batch normalization, and dropout (between 0.2 and 0.4). The Adam optimizer was employed for training, and its learning rate was 10^{-4}.

Table 1 displays the detailed training parameters.

Table 1. Parameter settings.

Parameter	Value
Learning rate	0.0004
Momentum coefficient	0.95
Exploration rate (start)	0.01
Exploration rate discount	0.99
Discount factor	0.9
Replay buffer capacity	580,000
Mini batch	64
Max episode	10,000
Optimizer	Adam

We directly trained the model simulation on 10 objects and unknown objects were grasped during evaluation. For various object tests, we placed them randomly in the workspace. In addition, we created 30 different scenes for DQN [26], A3C [2], CNN [30], and our method to make 10 grasping attempts. The grasp success rates are presented in Figure 3. To our surprise, exhaustive experimental results indicate that our grasping method achieves higher success rates than DQN [28], A3C [31], and CNN [32] for a wide variety of objects in clutter. Interestingly, when we only trained robot grasping for 10,000 episodes (see Figure 3), it yielded the strongest overall performance.

We also evaluated the task completion time, which measures the time taken by the robot to complete a specific robotic task involving grasping objects in the cluttered scene. The trained grasp policies demonstrated faster task completion times compared to baseline methods. This improvement can be attributed to the efficient exploration and exploitation capabilities of the RL agent, which enabled it to identify and execute successful grasps more quickly. The integration of multi-objective optimization further contributed to task efficiency by finding a balance between grasp success rates and other relevant objectives.

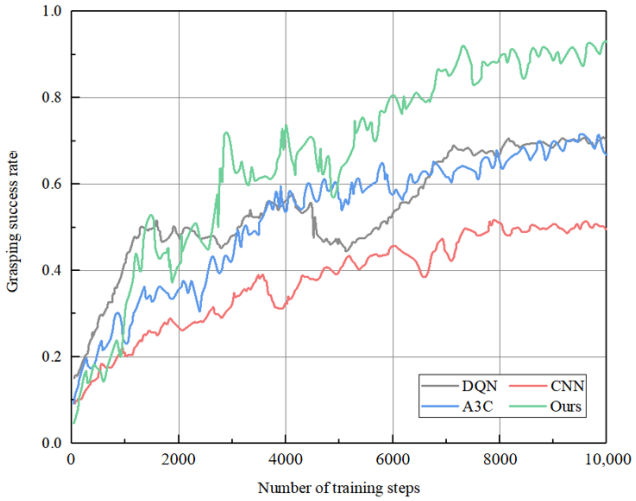

Figure 3. Training curve comparison for various approaches.

4.3. Real-World Experiments

Our real experiments consisted of a ROBOTIQ-85 gripper on a FLEXIV rizon4 robot arm. An Intel RealSense D435 camera captured RGB-D images at a resolution of 1280 × 720. The camera was attached to the end-effector, affording a good visual coverage of the graspable objects. In each grasp attempt, our network received the visual signals from the depth camera mounted on the robot end-effector (shown in Figure 4). We contrast our approaches with the approaches described in [31–33] for grasping the unknown objects, which performed better than other baselines in our simulation studies.

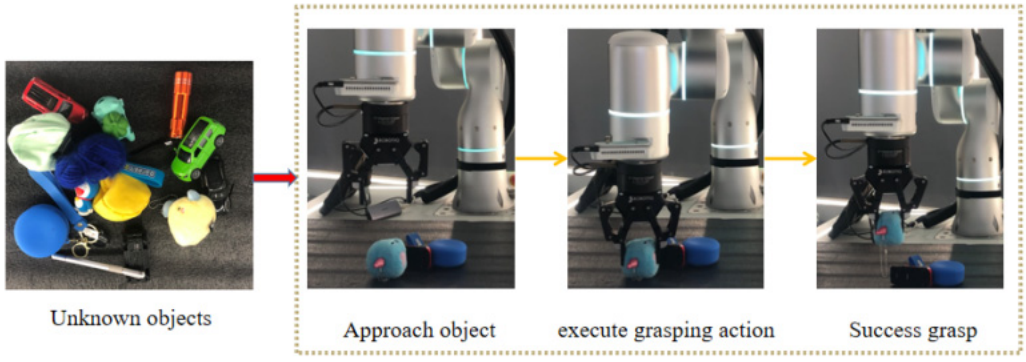

Figure 4. The invisible object in a cluttered environment is either a toy or a novel object never seen in training.

We executed three experiments (10 items, 20 items, and 30 items) to empty the objects in the workspace for further analysis. According to Table 2, we can see that our system achieves around 90.4% average grasping success with 511 grasp attempts, much higher than 3DCNN (85%). Our method is easily generalized to handle grasping manipulation for invisible objects in clutter. In particular, while our method has never been trained on these novel objects (30 items), it is able to achieve a grasping success rate of 87.2%. We address multi-object grasping tasks, attention mechanisms for exact feature extraction, and sparse rewards by remodeling the reward. This decreases the number of randomly predicted grasps in the background. Overall, the grasping success rate in real experiments is generally lower than that in simulation. We attribute this mostly to the fact that clutter and a variety of objects are present.

Table 2. Quantitative results considering the grasp success rate on the test object.

Authors	Total Grasps	Average Success Rate/Grasping Time per Item	Success Rate of Emptying Objects		
			10 Items	20 Items	30 Items
Coordinator [33]	509	85% (17.3 s)	94.5%	81%	79.5%
3DCNN [31]	471	87% (12.7 s)	92.5%	89.5%	79%
UCB [32]	523	82% (15.8 s)	89%	83%	75%
Ours	511	90.4% (8.9 s)	96%	88%	87.2%

4.4. Experimental Validation

To validate the effectiveness of the proposed methodology, we performed extensive experiments in both the simulated environment and with the physical robot platform. The experiments are designed to evaluate the grasp success rates, grasp success rates of object emptying in different scenes, and task completion times per item. The performance was evaluated with three metrics: (1) the average % grasp efficiency for all test runs, defined as $\frac{\sum_{i=1}^{runs} successful \quad numbers}{\sum_{i=1}^{runs} action \quad numbers}$, (2) the success rate of object emptying, defined as $\frac{\sum_{i=1}^{n} successful \quad numbers}{n(total \quad numbers)}$, and (3) the grasping time per item, defined as $\frac{total \quad time}{number \quad of \quad successful \quad objects}$. The environment will be reset for the next epoch of grasping if the robot grasps all the objects within the threshold of action numbers or not before exceeding the threshold. For all of these metrics, the higher the better.

In the simulated environment, we conducted several experiments with varying clutter densities, object configurations, and task requirements. These experiments allowed us to assess the robustness of the trained grasp policies under different conditions. We compared the performance of our methodology with baseline methods, such as traditional grasp planning algorithms or handcrafted heuristics, to demonstrate the superiority of the proposed approach. For the real-world experiments, we transferred the trained grasp policies from the simulation to the physical robot platform. This transfer involves adapting the policies to the specific characteristics and constraints of the physical robot. The evaluation metrics were computed, and the performance of the trained grasp policies was compared to that of the baseline methods in the real-world setting.

Overall, the experimental setup encompasses both the simulated environment and the physical robot platform, enabling a comprehensive evaluation of the proposed methodology's performance in both settings. It demonstrates the effectiveness, robustness, and generalization capabilities of the proposed methodology in real-world scenarios.

5. Discussion

The results obtained from the experiments highlight the effectiveness and robustness of the proposed methodology for robust grasp operation in cluttered environments using deep reinforcement learning. By integrating perception modules, the RL agent could leverage an accurate grasp ability assessment, leading to significantly improved grasp success rates compared to the baseline methods. The perception modules played a crucial role in providing the necessary information for object detection in both simulated and real-world experiments. The task completion times achieved by the trained grasp policies demonstrated efficiency and expedience. The RL agent's exploration and exploitation capabilities enabled it to quickly execute successful grasps, contributing to faster task completion. The multi-objective optimization framework further facilitated task efficiency by finding a suitable trade-off between grasp success rates and other objectives.

It is worth noting that while the trained grasp policies exhibited a strong performance, there are still limitations and challenges that need to be addressed. Many robotic grasping tasks require the use of perception information, but designing effective features can be difficult and time consuming, especially when working with RGB-D data. In future work,

our approach can be extended to a broader range of such problems. For instance, the perception modules may encounter difficulties in accurately perceiving highly occluded or visually ambiguous objects. Additionally, the grasp policies may be sensitive to variations in object properties, such as shape, texture, or weight. Further research and improvements in perception and grasp planning algorithms can address these challenges and enhance the overall performance of the methodology.

6. Conclusions

In this paper, we present a framework for a self-supervised robot grasping task in cluttered scenes. The results obtained from the extensive experiments in both simulated and real-world settings demonstrate the effectiveness, robustness, efficiency, and transferability of the proposed methodology for robust grasp operations in cluttered environments using deep reinforcement learning. The task success percentage for the real robot is 90.4%, which is significantly higher than that of the baselines. The attention-based dueling DQN approach can be accurately transferred to the real world and even generalized to unknown items. The sparse reward problem is successfully solved by the proposed method, which also increases the learning efficiency and grasping success rate. Future research will investigate how the proposed system model can be used to handle a robotic manipulator with more than two links and utilize simulation-based methodology to effectively learn and even generalize unidentified objects.

Author Contributions: Conceptualization, T.Z. and H.M.; methodology, T.Z.; software, T.Z.; validation, T.Z.; formal analysis, H.M.; investigation, T.Z.; resources, H.M.; data curation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
2. Mnih, V.; Badia, A.P.; Mirza, M.; Graves, A.; Lillicrap, T.P.; Harley, T.; Silver, D.; Kavukcuoglu, K. Asynchronous methods for deep reinforcement learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19 June 2016; pp. 1928–1937.
3. Kalashnikov, D.; Irpan, A.; Pastor, P.; Ibarz, J.; Herzog, A.; Jang, E.; Quillen, D.; Holly, E.; Kalakrishnan, M.; Vanhoucke, V.; et al. Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. *arXiv* **2018**, arXiv:1806.10293.
4. Joshi, S.; Kumra, S.; Sahin, F. Robotic grasping using deep reinforcement learning. In Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Hong Kong, China, 20 August 2020; pp. 1461–1466.
5. Berscheid, L.; Rühr, T.; Kröger, T. Improving data efficiency of self-supervised learning for robotic grasping. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 12 August 2019; pp. 2125–2131.
6. Hafner, R.; Riedmiller, M. Reinforcement learning in feedback control: Challenges and benchmarks from technical process control. *Mach. Learn.* **2011**, *84*, 137–169. [CrossRef]
7. Wei, Y. An Intelligent Human-like Motion Planner for Anthropomorphic Arms Based on Diversified Arm Motion Models. *Electronics* **2023**, *12*, 1316. [CrossRef]
8. Vitabile, S.; Franchini, S.; Vassallo, G. An Optimized Architecture for CGA Operations and Its Application to a Simulated Robotic Arm. *Electronics* **2022**, *11*, 3508. [CrossRef]
9. Zhang, J.; Dai, X. Adaptive Fuzzy Control for Flexible Robotic Manipulator with a Fixed Sampled Period. *Electronics* **2022**, *11*, 2270. [CrossRef]
10. Mokhtar, K.; Heemskerk, C.; Kasaei, H. Self-Supervised Learning for Joint Pushing and Grasping Policies in Highly Cluttered Environments. *arXiv* **2022**, arXiv:2203.02511.
11. Duan, S.; Tian, G.; Wang, Z.; Liu, S.; Feng, C. A semantic robotic grasping framework based on multi-task learning in stacking scenes. *Eng. Appl. Artif. Intell.* **2023**, *121*, 106059. [CrossRef]

12. Cao, H.-G.; Zeng, W.; Wu, I.-C. Reinforcement Learning for Picking Cluttered General Objects with Dense Object Descriptors. In Proceedings of the 2022 International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 12 July 2022; pp. 6358–6364.
13. Shao, Q.; Hu, J.; Wang, W.; Fang, Y.; Liu, W.; Qi, J.; Ma, J. Suction Grasp Region Prediction Using Self-supervised Learning for Object Picking in Dense Clutter. In Proceedings of the 2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR), Singapore, 16 September 2019; pp. 7–12.
14. Ye, J.; Wang, J.; Huang, B.; Qin, Y.; Wang, X. Learning Continuous Grasping Function With a Dexterous Hand From Human Demonstrations. *IEEE Robot. Autom. Lett.* **2023**, *8*, 2882–2889. [CrossRef]
15. Wang, C.; Zhang, Q.; Wang, X.; Xu, S.; Petillot, Y.; Wang, S. Multi-Task Reinforcement Learning based Mobile Manipulation Control for Dynamic Object Tracking and Grasping. In Proceedings of the 2022 7th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), Tianjin, China, 18 August 2022; pp. 34–40.
16. Mosbach, M.; Behnke, S. Efficient Representations of Object Geometry for Reinforcement Learning of Interactive Grasping Policies. *arXiv* **2022**, arXiv:2211.10957.
17. Wu, L.; Chen, Y.; Li, Z.; Liu, Z. Efficient push-grasping for multiple target objects in clutter environments. *Front. Neurorobot.* **2023**, *17*, 1188468. [CrossRef] [PubMed]
18. Starke, J.; Eichmann, C.; Ottenhaus, S.; Asfour, T. Synergy-Based, Data-Driven Generation of Object-Specific Grasps for Anthropomorphic Hands. In Proceedings of the 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), Beijing, China, 24 January 2019; pp. 327–333.
19. Danielczuk, M.; Mahler, J.; Correa, C.; Goldberg, K. Linear push policies to increase grasp access for robot bin picking. In Proceedings of the 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), Munich, Germany, 6 December 2018; pp. 1249–1256.
20. Berscheid, L.; Friedrich, C.; Kröger, T. Robot Learning of 6 DoF Grasping using Model-based Adaptive Primitives. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 18 October 2021; pp. 4474–4480.
21. Hegedus, M.; Gupta, K.; Mehrandezh, M. Towards an Integrated Autonomous Data-Driven Grasping System with a Mobile Manipulator. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 12 August 2019; pp. 1601–1607.
22. Santina, C.D.; Arapi, V.; Averta, G.; Damiani, F.; Fiore, G.; Settimi, A.; Catalano, M.G.; Bacciu, D.; Bicchi, A.; Bianchi, M. Learning From Humans How to Grasp: A Data-Driven Architecture for Autonomous Grasping With Anthropomorphic Soft Hands. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1533–1540. [CrossRef]
23. Marios, K.; Sotiris, M. Robust object grasping in clutter via singulation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 12 August 2019; pp. 1596–1600.
24. Levine, S.; Pastor, P.; Krizhevsky, A.; Ibarz, J.; Quillen, D. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *Int. J. Robot. Res.* **2018**, *37*, 421–436. [CrossRef]
25. Gualtieri, M.; Ten Pas, A.; Saenko, K.; Platt, R. High precision grasp pose detection in dense clutter. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 1 December 2016; pp. 598–605.
26. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.
27. Azizzadenesheli, K.; Yang, B.; Liu, W.; Lipton, Z.C.; Anandkumar, A. Sample-efficient deep RL with generative adversarial tree search. *arXiv* **2018**, arXiv:1806.05780.
28. Gou, S.Z.; Liu, Y. DQN with model-based exploration: Efficient learning on environments with sparse rewards. *arXiv* **2019**, arXiv:1903.09295.
29. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 9 November 2017; pp. 2261–2269.
30. Pinto, L.; Gupta, A. Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 9 June 2016; pp. 3406–3413.
31. Choi, C.; Schwarting, W.; DelPreto, J.; Rus, D. Learning object grasping for soft robot hands. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2370–2377. [CrossRef]
32. Boularias, A.; Bagnell, J.A.; Stentz, A. Learning to manipulate unknown objects in clutter by reinforcement. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 1336–1342.
33. Yang, Y.; Liang, H.; Choi, C. A deep learning approach to grasping the invisible. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2232–2239. [CrossRef]