DeepRob Final Project Report: Choosing how to Grasp the Ungraspable

Jason Brown*
Aerospace and Robotics Department
University of Michigan
Ann Arbor, USA
jaybrow@uUSAch.edu

Eli Fox*
Robotics Department
University of Michigan
Ann Arbor, USA
elijfox@umich.edu

Jacob Harrelson*
Robotics Department
University of Michigan
Ann Arbor, USA
jharrels@umich.edu

Srushti Hippargi*
Integrative Systems Design Department
University of Michigan
Ann Arbor, USA
shipparg@umich.edu

***Abstract*—In this paper, we aim to reproduce and extend the results from *Learning to Grasp the Ungraspable with Emergent Extrinsic Dexterity* by W. Zhou and D. Held [1]. The original paper demonstrates that a simple gripper using intuition about its environment can still perform complex manipulation tasks. That work studies the task of "Occluded Grasping" that aims to reach a grasp in configurations that are initially intersecting with the environment. While the original work only considered occlusions by the ground, our work extends their work by considering occlusions by side walls along with unoccluded configurations. Our system trains different policies for each occlusion type and selects between them at run-time. In simulation, our policy selector was 100% successful at choosing the correct policy for the occlusion type and the policy was then 100% successful at picking up the object. Videos can be found at https://deeprob.org/w24/reports/grasping-ungraspable/ and our code can be found at https://github.com/HarrelsonJ/DeepRob_Ungraspable/tree/main.**

I. Introduction

In the realm of robotics, achieving dexterous manipulation comparable to that of human hands has long been a challenge. Traditional multi-fingered robotic hands [2], while capable, are often expensive to produce and prone to fragility. However, recent advancements in robotics research have introduced a fascinating concept: extrinsic dexterity [3]. This approach proposes that rather than focusing solely on the capabilities of the gripper itself, manipulation tasks can be accomplished by leveraging the surrounding environment. This paradigm shift opens doors to new possibilities, enabling even simple grippers to perform intricate maneuvers by exploiting external resources such as contact surfaces and gravity.

The paper under review, "Learning to Grasp the Ungraspable with Emergent Extrinsic Dexterity" by Wenxuan Zhou and David Held [1] delves into the realm of extrinsic dexterity, specifically focusing on a task known as "occluded grasping." Unlike conventional grasping tasks that typically involve reaching for an object in unobstructed space, occluded grasping requires the robot to grasp objects in poses that are initially unreachable due to the target grasp intersecting with the environment. These grasps are shown in Figure 1.

Imagine a scenario where a cereal box lies on its side on a table, partially obscured by the table's surface. The only manipulator available is a narrow parallel gripper, too small to grasp the cereal box from the top. A traditional approach would find the desired grasp unattainable. However, by employing extrinsic dexterity, the robot can manipulate the object using the environment—perhaps by pushing the object against a vertical wall—to expose an achievable grasp. This behavior is shown in Figure 2.

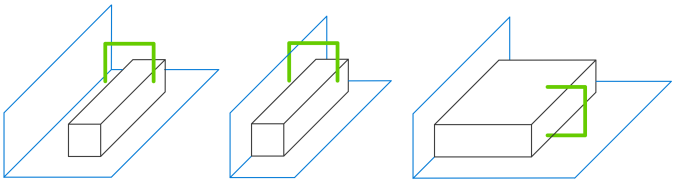

Fig. 1. Three types of grasps. The green C shape represents the gripper. From left to right: Unoccluded, side occlusion, ground occlusion.

Wenxuan Zhou and David Held achieved this advanced level of extrinsic manipulation by utilizing goal conditioned reinforcement learning (RL) to develop a closed loop policy for performing ocluded grasps. Previous research in this domain has shown promising results but has been limited by factors such as reliance on hand-designed motions, specific gripper designs, or the inability to generalize across different objects and environments [4]–[6]. The original paper showed that it was possible with a single 1 degree of freedom parallel gripper (Figure 3) to achieve a very high level of dexterity.

Their work, however, was limited to ground occlusions where the base of the object is in contact with the ground, and was not tested against other occlusion types. In this paper, we extend their work to show that using similar methods of goal-conditioned reinforcement learning, it is possible to achieve similar extrinsic dexterity as in the basic ground occlusion case.

II. Related Work

The field of robotic manipulation has seen significant advancements in addressing the complexities associated with in-hand manipulation utilizing external influences. Prior research has delved into various methodologies aimed at overcoming challenges such as continuous contact dynamics and discrete contact switch-overs. These challenges arise from the intricate interactions between the robot's fingers and the manipulated

*Equal contribution, listed alphabetically by last name

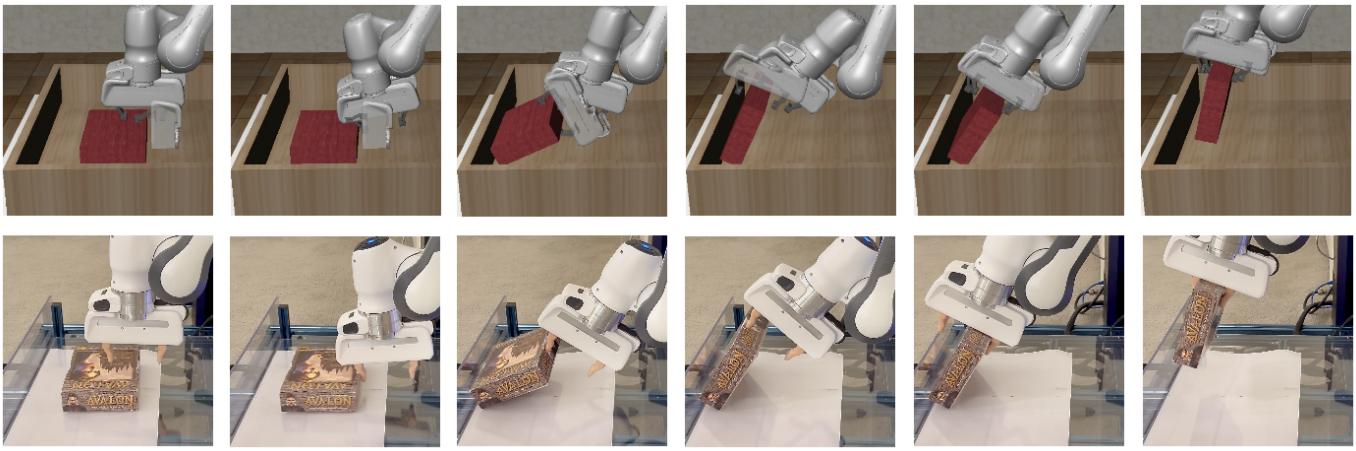

Fig. 2. Example of a robot using extrinsic dexterity to grasp an occluded object. Figure reused from [1].

Fig. 3. Simple 1 DoF parallel gripper used in trials

objects, requiring precise planning and control to achieve desired manipulation tasks effectively. The notable approach explored involves the integration of low-level optimization-based inverse dynamics with high-level sampling-based planning techniques [4]. By combining these methods, researchers have sought to generate push sequences that facilitate the reconfiguration of objects between different grasps while minimizing contact switch-overs. These efforts have contributed to a better understanding of the dynamics involved in in-hand manipulation with external pushes, laying the groundwork for further advancements in planning frameworks tailored to this context. There are several external resources usable in extrinsic dexterity tasks, such as gravity, environment surfaces, and arm motions used to manipulate objects within the hand. This approach offers advantages in scenarios requiring significant adjustments of object, potentially leading to more versatile and adaptable manipulation tasks [3]. The long term goal of developing a repertoire of pre-grasp actions to navigate grasps represents a promising direction for future research in this area. Efforts have also been made to extend traditional grasping techniques through the concept of shared grasping. By leveraging external contacts to maintain force closure, researchers have proposed frameworks that emphasize robustness and versatility in object manipulation [5].

III. Original Methodology [1]

The proposed system aims to overcome these limitations by employing RL to learn a closed-loop policy $\pi(s_t, \eta)$ that guides the robot's interactions with both the object and the environment, considering both planning and control aspects. The policy's inputs are the state of the system s_t and the goal η, as the paper uses goal-conditioned RL.

A. System Definition

One of the key innovations of the paper lies in its use of model-free RL to optimize pre-grasp and grasping motions without the need for separate stages as seen in previous work. To do so, the original paper crafted a single, overall reward function:

$$r = \alpha D(g, E) + \beta \Sigma_i P(m_i)$$

Where α and β represent argument weights, g represents the object-frame target grasp configuration, E represents the world-frame end effector position, and m_i represents prospective grasp positions on the target gripper as seen in Figure 4. This reward function combines rewards for the box's pose $D(g, E)$ and the box's non-occluded grasps $\Sigma_i P(m_i)$. The box position reward is defined as:

$$\alpha D(g, E) = \alpha_1 \Delta T(g, E) + \alpha_2 \Delta\theta(g, E)$$

Where α_1 and α_2 are weight arguments. This function includes both the translational (ΔT) and rotational ($\Delta\theta$) difference of the box position to its desired position.

The available non-occluded points on the target gripper are defined as $\Sigma_i P(m_i)$. This function penalizes points on the target grasp that are intersecting with the environment. In doing so, we reward shifting the box to positions where the target grasp becomes unoccluded.

B. System Simulation

The system is then trained using a Soft Actor Critic [7] trainer using the MuJoCo robot simulation suite [8]. During training, the system incorporates Automatic Domain Randomization (ADR) [9] to enhance the policy's robustness. This allows the model to generalize across environmental conditions (including friction), object sizes, and object locations. Through iterative refinement and expansion of environmental parameters via ADR, the system learns to perform occluded grasping tasks with a high degree of success.

C. Robot Low-Level Motion Controller

To translate desired end-effector positions into actual motion, the paper controls the robot using an Operational Space Controller (OSC). This controller results in tunable end-effector stiffnesses and speeds. Tuned correctly, the end-effector is able to move relatively slowly but with enough force to move the boxes. This compliant style allows for quick low-level robot control loops (100 Hz) that do not result in adverse effects from the comparatively slow RL policy loops (2 Hz).

D. Translation to the Physical System

After training in simulation, the RL model is then transferred to a physical version of the robot for validation tests. These tests include a variety of item positions and types, introducing different-shaped containers such as cylindrical tubs (whereas the simulator only used boxes).

IV. Reproduction

We were able to replicate the results of the paper in simulation. After training with the provided hyperparameters and methods, we successfully trained a policy to pick up a block using extrinsic dexterity in simulation. Our replicated policy was without using grasp selection ADR and without using physical ADR. This took 13.5 hours to train on one of our local workstations. Due to time constraints, we did not train the ground occlusion policy any more times. Additionally, as we do not have access to a Franka Emika Panda robot, we did not attempt to replicate the real-robot experiments.

A video of our replicated ground occlusion policy is shown on our project page website.

V. Algorithmic Extension

A. Multiple Occlusion Types

While the original paper only considered grasps occluded by the ground, we extend the possible occlusion types to side occlusions and no occlusions. While ground occlusions are defined as grasp positions where the end effector would need to intersect with the ground to reach the grasp in the initial configuration of the object, side occlusions are defined as grasp positions where the end effector would need to intersect with one of the side walls. We additionally consider unoccluded grasps where the initial position is reachable from the top in a basic pick-and-place maneuver. Figure 1 shows the types of occlusions considered.

The two new types of grasp would not necessarily need extrinsic dexterity, although there are ways to complete the side occluded grasp using extrinsic dexterity. The side occluded grasp is reached without extrinsic dexterity by first scraping the block away from the wall, then moving into position to reach the grasp when it is no longer occluded. The side occluded grasp is reached using extrinsic dexterity by bumping the block into the wall, slightly rotating it towards the wall and causing it to rebound away from the wall. That leaves a gap for the gripper, which can then move into position and pick the block up.

B. Policy Switching

In principal, a policy could be trained to handle multiple cases, but we were unsuccessful in creating a single policy that could handle all cases. Instead, we trained two policies: One for ground occlusions, and the other for side occlusions and no occlusions. We found that the policy trained on side occlusions was able to pick up unoccluded configurations, despite those configurations not being in its training data. During testing, we have two possible ways to select a policy. We evaluate with both methods.

1) Simple Size Selector: During testing, the simple size selector decides which policy to use at the start of an episode. This decision is made based on the dimensions of the block and uses a simple algorithm rather than a neural network. The basic logic of the algorithm is as follows: If the block could be picked up from the top, we use the side occlusion policy. If the block could be picked up from the side, we use the ground occlusion policy. Otherwise, we determine that the block is too large for the gripper and we do not attempt to pick it up. A diagram showing the cases is in Figure 1 and the formal algorithm is shown in Algorithm 1.

Algorithm 1 Simple Size Policy Selection

```
x ← length of the block
y ← width of the block
z ← height of the block
g ← width of the gripper
if x < g or y < g then
    SetPolicy(side)
else if z < g then
    SetPolicy(ground)
else
    SetPolicy(none)
end if
```

2) Q-Function Maximizer: The Q-function maximizer picks the action that results in the highest Q-function value. For each time step, we pass the current observations s_t and goal η into both policies $\pi_g(s_t, \eta)$ and $\pi_s(s_t, \eta)$ to see what action the robot would take given each policy. We then pass each action along with the current observations into the Q-function of the corresponding policy, yielding $Q_g^\pi(s_t, \pi_g(s_t, \eta), \eta)$ and $Q_s^\pi(s_t, \pi_s(s_t, \eta), \eta)$. We perform the action that would yield a higher Q-function value. If the ground Q-function yields a

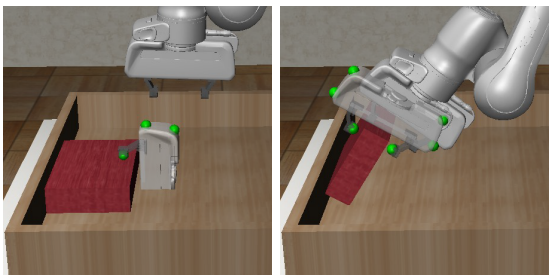

Fig. 4. Marker points to calculate occlusion penalty. Figure reused from [1].

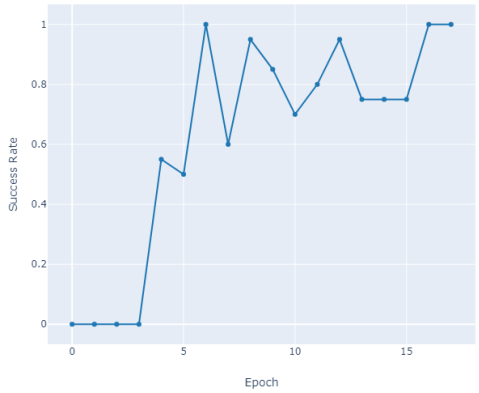

Fig. 5. Training Curve of Side Policy with Grasp Selection ADR

higher value than the side Q-function, we choose the action output by the ground policy, and vice versa.

VI. Experiments and Results

A. Training

We trained the side occlusion policy similar to the ground occlusion policy given to us in the paper's provided code. We changed a few things to make training work with the side occlusion. First, we changed the grasp selection method to choose grasps from the top rather than the side. We also added an additional penalty of how far outside the box the target grasp was. Similar to the ground occlusion reward function including a term to penalize the target grasp being inside the ground, this extra penalty encouraged the arm to move the box away from the side to where it was graspable. This penalty was computed based on seven marker points on the target gripper, as shown in Figure 4. If a marker is to the left of the left wall of the bin, the distance to the wall is used as the penalty.

The side occlusion policy was trained with a 0.06 m $\times$ 0.20 m $\times$ 0.06 m block. This size is small enough to be picked up from the top. The side occlusion policy was trained with grasp selection ADR, starting with a single grasp in the middle of the block and expanding outwards. Due to not having a real robot to experiment on, we did not train with physics ADR. A plot of the training curve is shown in Figure 5. As visible, after 14 epochs the success rate rose to 100%.

An example of the side occlusion behavior is shown in Figure 6. This policy uses extrinsic dexterity by first nudging the block against the wall to create a gap, then scraping it away from the wall and picking it up. To see a video of this policy, please see our project webpage.

Occlusion Type	Box Size [m]	Distance to Wall [m]
Ground	0.15 $\times$ 0.20 $\times$ 0.05	0.00
Side	0.06 $\times$ 0.20 $\times$ 0.06	0.00
None	0.06 $\times$ 0.20 $\times$ 0.06	0.05-0.10

TABLE I
Evaluation Experiment Parameters

Occlusion Type	Success Rate	Mean Final Reward (less negative is better)
Ground	100%	-0.71
Side	100%	-0.75
None	100%	-0.71

TABLE II
Results of Simulated Grasps using Hard-coded Selector

B. Evaluation Experimental Setup

During evaluation, we considered three possible cases: ground occlusions, side occlusions, and no occlusions. The box size and distance from the wall are shown in Table I. All other parameters were set to the same between cases.

To evaluate, we simulated each occlusion type for ten episodes and record the success percentage and average reward per occlusion type. We simulated using both policy selectors to compare them. Although the occlusion type was passed into the environment setup, this information was not made directly available to the robot.

C. Results

Table II shows the results of simulating each occlusion type for ten episodes using the simple policy selector. Table III shows the results for the Q-maximizer policy selector. In both cases, the policies were 100% successful at grasping and lifting the boxes in simulation. Additionally, both the simple selection algorithm and Q-maximizer were able to find the correct policy to use 100% of the time. In the test cases, the Q-maximizer had a slightly better mean final reward, but that is likely random variation. If more occlusion types were considered, the Q-maximizer may be a better idea, as it eliminates the need of a hand-crafted selector.

These results are hardly surprising given the limited cases and controlled environment we tested in. It would be more interesting to see if the policy selector could perform as well if it was given bounding boxes of other, non-rectangular objects.

VII. Conclusions

This paper successfully replicated and extended the findings of Zhou and Held's work on extrinsic dexterity in robotic

Occlusion Type	Success Rate	Mean Final Reward (less negative is better)
Ground	100%	-0.70
Side	100%	-0.70
None	100%	-0.71

TABLE III
Results of Simulated Grasps using Q-Maximizer Selector

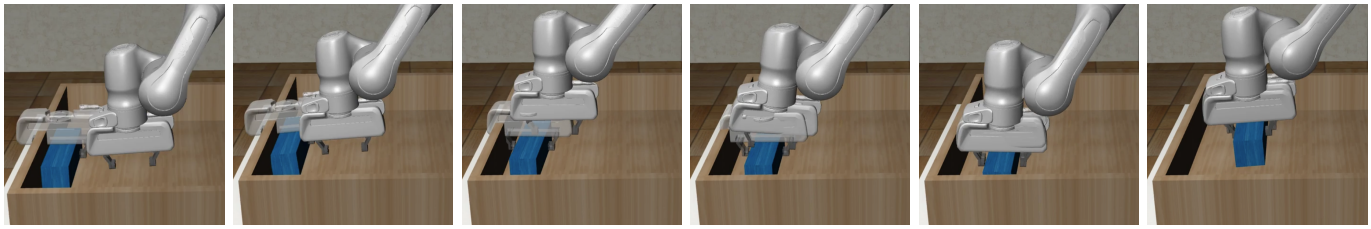

Fig. 6. An example of the side occlusion policy successfully picking up a block.

manipulation [1]. Through the utilization of reinforcement learning (RL), the study demonstrated the feasibility of achieving complex manipulation tasks with simple grippers by leveraging the environment. The original work focused on ground occlusions, where objects were partially obstructed by the ground, and showed promising results. Our paper extends their research to consider the additional occlusion types of side occlusions and unoccluded grasps. By training separate policies for different occlusion scenarios and implementing a policy-switching algorithm, the study achieved successful manipulation outcomes across various occlusion conditions. The experiments conducted in simulation showcased high success rates in grasping objects under different occlusion types, indicating the effectiveness of the proposed approach.

The future research could explore several directions building upon these findings. Firstly, further validation and experimentation in real-world environments with physical robots would be essential to assess the system's performance and generalization capabilities beyond simulation. It would also be interesting to consider even more possible occlusion types. We only considered side occlusions on one side of the bin, but there are many more combinations of occlusion types and object sizes that we did not train on. Additionally, extending the approach to handle more complex object shapes and types could enhance its practical utility in various applications. To handle more complex shapes, it would be interesting to consider using a richer representation of object shape, such as a point cloud or voxels. The pose of the object works for boxes but may not generalize to more complex objects.

Overall, our findings and extension contribute valuable insights into the potential of extrinsic dexterity and reinforcement learning in robotic manipulation, paving the way for advancements in cost-effective and versatile robotic systems capable of intricate manipulation tasks.

REFERENCES

[1] W. Zhou and D. Held, "Learning to grasp the ungraspable with emergent extrinsic dexterity," in *Conference on Robot Learning (CoRL)*, 2022.

[2] M. Pfanne, *n-Hand Object Localization and Control: Enabling Dexterous Manipulation with Robotic Hands*. Cham: Springer International Publishing, 2022, vol. 149.

[3] N. C. Dafle, A. Rodriguez, R. Paolini, B. Tang, S. S. Srinivasa, M. Erdmann, M. T. Mason, I. Lundberg, H. Staab, and T. Fuhlbrigge, "Extrinsic dexterity: In-hand manipulation with external forces," *IEEE transactions on cognitive and developmental systems*, pp. 1578–1585, 2014.

[4] N. C. Dafle and A. Rodriguez, "Sampling-based planning of in-hand manipulation with external pusges," 2017.

[5] Y. Hou, Z. Jia, and M. T. Mason, "Manipulation with shared grasping," 2020-06.

[6] X. Cheng, E. Huang, Y. Hou, and M. T. Mason, "Contact mode guided sampling-based planning for quasistatic dexterous manipulation in 2d," pp. 6520–6526, 2021.

[7] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018.

[8] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 5026–5033.

[9] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, "Solving rubik's cube with a robot hand," 2019.