Versatile and Generalizable Manipulation via Goal-Conditioned Reinforcement Learning with Grounded Object Detection

Cheryl Wang, Fahim Shahriar, Seyed Alireza Azimi, Gautham Vasan, Rupam Mahmood, Colin Bellinger

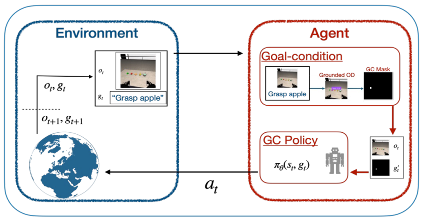
CoRL 2024 Workshop on Mastering Robot Manipulation in a World of Abundant Data

Videos

Paper

Introduction / Motivation

Goal conditioned reinforcement learning (GCRL) provides a framework to learn a single policy to achieve arbitrary reaching and grasping tasks.



Overview:

Agent converts text string goal to object mask goal-condition.
Object mask improves
- Generalization
- Sample efficiency, and
- Performance on unseen objects.

Objective:

- Develop robotic systems with versatile manipulation skills for dynamic environments, such as households and workplaces and human-centric labs.
- Enable robots to handle a variety of tasks autonomously, adapting to evolving objects and goals.

GCRL Challenges:

- Object Diversity: Struggles with generalization across diverse and unfamiliar objects, requiring extensive task-specific training.
  - Out-of-distribution objects requires significant additional training.
- Interaction Complexity: Learning efficient interaction with objects is resource-intensive and slow to converge, limiting adaptability in real-world scenarios.

Long training times and the assumption of sufficient experience on all target tasks limits the applicability of GCRL on robotics tasks.

Solution:

- Leverage language and perception capabilities of grounded object detectors to automatically produce a target object mask.
- Goal-conditioning mask simplifies GCRL agent's learning by enabling it exploit object detectors trained on large, generic datasets
  - Mask indicates the size and location of object in a general format
  - Mask can be automatically generated for any arbitrary new object.

Simulated UR10e Reach and Grasp Environment

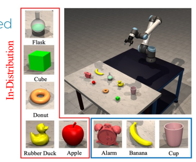
Setup:

Each episode:
- Environment randomly selects a goal from a set of in-distribution target objects
- Ends when agent grasps target object, or a max of 300 steps
- Object locations are randomized

Observations:
- UR10e proprioception + ego-centric RGB image
- Goal-condition

Actions
- Angular velocity + open / close gripper

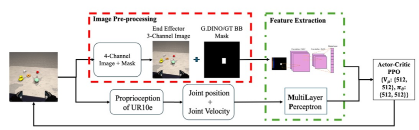
Reward
- Negative Distance to object
- Plus 2 for grasping



Out-Of-Distribution

Proposed Method

Mask-Based GCRL Framework



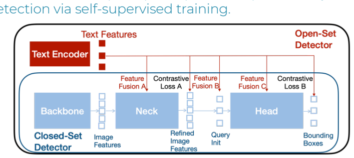
Ground Truth Masking

The GT mask is calculated based on the objects size, location and distance from the gripper. This can be produced based on prior knowledge or a depth sensor.

Inferred Masking with G.DINO

The inferred mask setting uses a pretrained object detector.
- Does not require prior knowledge nor a depth sensor.

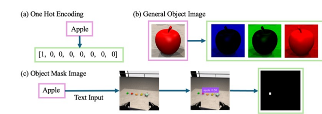
We utilize Grounded DINO. It combines an closed-set object detector with text prompts to produce open-set object detection via self-supervised training.



[1] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. 2017 Jul 20.
[2] Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, Li C, Yang J, Su H, Zhu J, Zhang L. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. arXiv preprint arXiv:2303.05499. 2023 Mar 9.

Evaluation

We evaluate 3 forms of goal-Conditioning on 10 random seeds in terms of total return, episode length and in- and out-of-distribution object grasp.

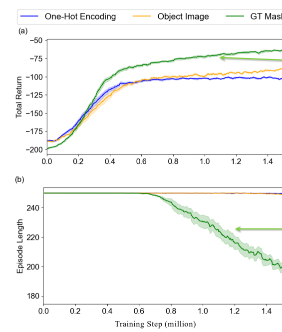


- 1-Channel mask-based goal-conditioning
  - Ground-truth mask (GT) using known object size, distance and location
  - Inferred mask using Grounded DINO (GD) [2]

with PPO [1] on the developed simulated UR10e

Results

GT Masking: PPO reach and grasp learning curves



Mask-based goal-conditioning converges faster to approx. 25% higher return

Mask-based goal-conditioning ends the episode early by successfully grasping the object, whereas the alternatives do not.

Grasp Success Rate:

GT Mask: Comparison of grasp success rates for in-distribution and out-of-distribution objects.

Goal-Conditioning	Grasp Success Rate	
	In-Distribution	Out-of-Distribution
One-Hot Encoding	0.13	0.2
3-Channel Image	0.62	0.28
GT Mask	0.89	0.9

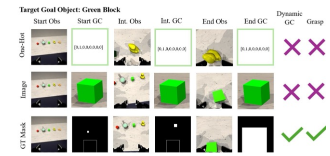
Goal-conditioning with GT mask significantly out-performs the alternatives.
- One-hot encoding GC fails to learn an acceptable policy
- Image-based GC learns a mediocre grasp policy that fails to generalize to the out-of-distribution objects
- GT Mask GC achieves excellent grasp performance that generalizes to the out-of-distribution objects

GD Mask: Masking with Grounded DINO (GD) compared to GT masking on grasping with 0, 1 and 2 distractor objects.

Masking	Grasp Success Rate			
	In-Distribution	Out-of-Distribution (number of distractors)		
		0	1	2
Train with GD, evaluate with GD	0.21	0.28	0.22	0.24
Train with GT, evaluate with GD	0.9	0.82	0.79	0.67

Training with GD masking suffers from false positives by the object detector. However, applying GD to the policy trained with GT masking achieves good performance on out-of-distribution objects with 0 and 1 distractors and moderate performance with 2 distractors.

Policy Demonstration



Unlike the alternatives, the dynamic goal-condition in our method provides real-time information to the agent on its progress towards the target object.
- This provides feedback to the agent on its progress to the goal.

Conclusion

Future Work

- Improved mask generation
  - Evaluate alternative grounded object detectors
  - Reduce cycle time for real-time control
  - Reduce the impact of false positives
- Extend to handle household and scientific glassware
- Deploy on UR10e and Franka Emika Panda robots.

Summary

- Propose object masking as a general and versatile goal-condition for object manipulation tasks
- Mask can be generated from:
  - Ground truth object location, distance and size, or
  - Automatically via pre-trained object detectors
- Mask-based goal conditioning out-performs the standard alternatives:
  - Converge faster and to a higher return
  - Grasp generalizes to out-of-distribution objects

Acknowledgements

This research is funded through the National Research Councils of Canada's Artificial Intelligence for Design Challenge program.