LM-RtA · Topic 6

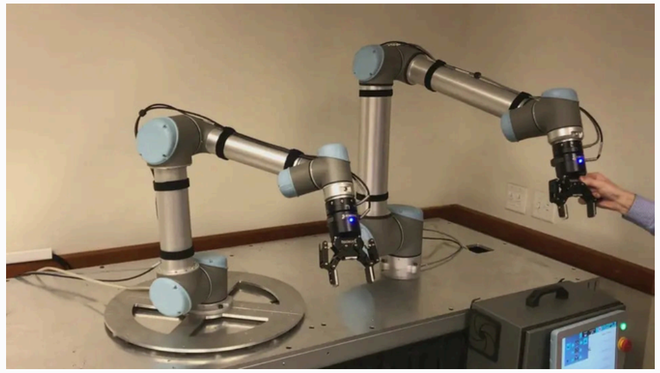
Learning to Grasp Diverse Objects

How does training variety affect generalization of a reinforcement learning policy?

Anastasya Zabolotna (LPNU) · Anna Danilowicz (SUT)

Supervisor: Jakub Musiał

The Problem



01 **Object variety**

An industrial robot must pick up objects that differ in shape, size, mass and surface friction

02 **Narrow training**

A policy trained on a narrow object set may fail on unseen objects; it is cheap but brittle

03 **Broad training**

Randomized training is robust, but raises cost and can slow or destabilize learning

FROM OUR RL LOCOMOTION WORK

The training setup itself (reward targets, contacts, evaluation) changed our conclusions more than we expected

The Reality Gap

01

Long training

RL needs millions of simulated attempts, and every new set of objects means more training time

02

Reachable only in theory

A pose that exists in simulation may be unreachable for the real UR10e because of joint, workspace, speed and collision limits

03

Physics mismatch

Mass, friction, inertia and contact are never modelled exactly, so a grasp that holds in simulation can slip or swing on a real object

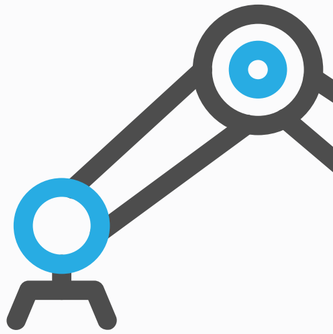
04

Safety risk

An unexpected reaction can drop the object or cause a fast, uncontrolled arm motion, dangerous for people and equipment

FROM OUR RL LOCOMOTION WORK

In our humanoid walking work, diagnosing why a policy fails took more effort than training it



CENTRAL RESEARCH QUESTION

How does the variety of objects and physical parameters seen during training affect the generalization of an RL grasping policy to unseen objects, and at what cost in training effort?

Hypotheses

H1

Smooth control

Different actions need different pose-generation frequencies, and penalties on pose changes reduce jerky, unstable motion

H2

Pose cloud

Generating a cloud of candidate poses before the action, then choosing at run time as the scene changes, is more robust than one fixed pose

H3

Randomization

Randomizing geometry and physical parameters improves held-out success and disturbance robustness

H4

Cost of variety

More variety needs more steps and time to a target success rate; a curriculum reduces this cost

FROM OUR RL LOCOMOTION WORK

Confounders to control: reward targets, evaluation settings, and contact/friction parameters, which silently changed results in our walking work

Environment and Scope

01

Platform

NVIDIA Isaac Sim + Isaac Lab, Python, PyTorch, RSL-RL or skrl

02

Robot and task

UR10e with two-finger gripper; grasp an object and lift it to a target height

03

Observations

Simulator state (object pose, joint states); vision is an optional extension

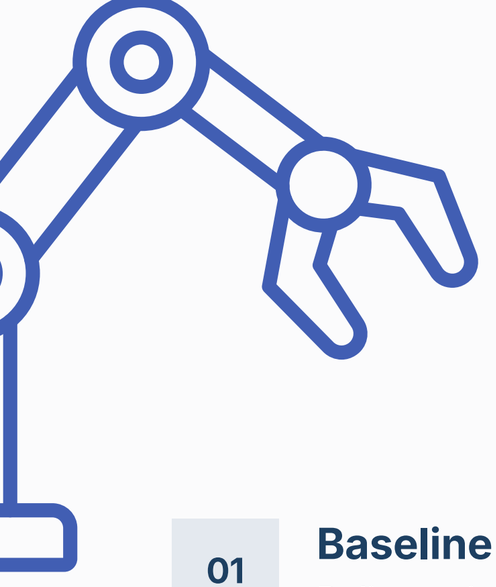
04

Constraints

Position targets within real joint, workspace and speed limits; no physical robot testing

FROM OUR RL LOCOMOTION WORK

Sanity gates before training: collision pairs, friction and joint limits checked in the simulator, not only in the CAD model



Experimental Design

01 **Baseline**
Policy trained on a single object

02 **Randomization**
Object geometry and physical parameters (mass, friction) randomized

03 **Curriculum**
Gradual increase of variety during training

04 **Objects and seeds**
Primitives and YCB models; strictly separate train and held-out sets; several seeds per condition

FROM OUR RL LOCOMOTION WORK

Every run logs its full configuration (targets, action scaling, randomization ranges); evaluation reuses the exact training settings

Evaluation Criteria

01

Generalization

Grasp-and-lift success rate on training and held-out objects

02

Cost

Environment steps and wall-clock time to a target success rate

03

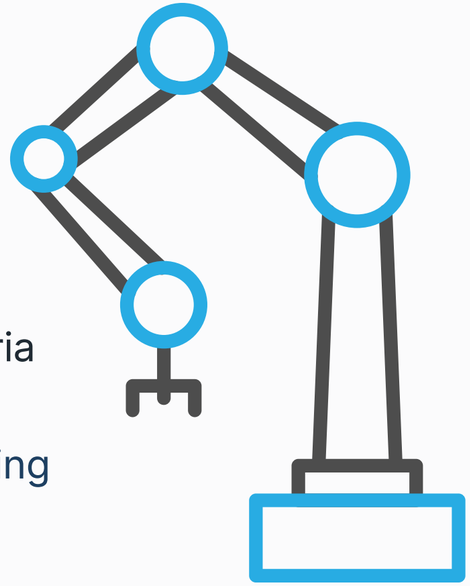
Robustness

Grasp stability under external disturbance; variance across seeds

FROM OUR RL LOCOMOTION WORK

Track per-minibatch statistics, not only averages: a few outlier updates can hide unstable learning

Preliminary Research Plan



01 Set up Isaac Lab scene, UR10e and gripper; define reward and success criteria

02 Run sanity gates and single-environment debugging before any parallel training

03 Select and split objects into training and held-out sets

04 Train the single-object baseline and verify the pipeline

05 Train randomization and curriculum variants with several seeds

06 Evaluate on seen and unseen objects, with disturbance tests

07 Analyse the variety, cost and generalization trade-off; report results

Questions for Discussion

01

Clarity

Is the problem sufficiently clear and challenging?

02

Hypotheses

Do H1–H3 relate directly to the problem, and would testing them give meaningful insight?

03

Plan and risks

Is the plan capable of testing the hypotheses? Which risks are we missing?

Thank you. We welcome your feedback.