LM-RTA · TOPIC 6 · 17 NOVEMBER 2026

Learning to Grasp Diverse Objects

State of the art and reusable resources for an RL grasping policy that must generalize to unseen objects

Anastasya Zabolotna (LPNU) · Anna Danilowicz (SUT)

Supervisor: Jakub Musiał

SOTA and resources

RECAP

Our question and what we looked for

> 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?

01

Evidence

What does prior work say about variety, generalization and training cost?

02

Resources

Which code, object sets and benchmarks can we adopt or adapt?

03

Gaps

What is still open for an industrial gripper with state observations?

STATE OF THE ART · 1 OF 3

More variety helps, with diminishing returns

Cobbe et al. · ICML 2019

Quantifying generalization in RL

Separate training and test levels in CoinRun. The gap between them shrinks as training levels grow, and good generalization needs far more training environments than earlier transfer studies used.

Tobin et al. · IROS 2018

Domain randomization for grasping

A grasp model trained only on procedurally generated, unrealistic objects reaches over 90% success on unseen realistic objects in simulation and about 80% on a real robot.

Lin et al. · ICLR 2025

Data scaling laws in manipulation

Generalization follows roughly a power law in the number of training objects and environments. Object diversity matters more than more demonstrations per object.

WHAT WE TAKE FROM THIS

Measure a curve, not two points: held-out success against the number of training objects on a log axis, for example 1, 4, 16 and 64.

STATE OF THE ART · 2 OF 3

Variety has a cost, and structure lowers it

OpenAI (Akkaya et al.) · 2019

Automatic domain randomization

Hand-set ranges are hard to choose: too much randomization makes learning difficult, too little hurts transfer. ADR starts narrow and widens the ranges as training progresses.

Wan et al. · ICCV 2023

UniDexGrasp++ curriculum

A geometry-aware curriculum grows the task set from 1 object to 300, 900 and finally 3000+. Success: 85.4% on training, 78.2% on test objects, about 11 points above UniDexGrasp.

Chen et al. · ICLR 2022

Why domain randomization works

Theory: transfer with domain randomization can succeed under mild conditions without real data, and history-dependent policies, that is policies with memory, matter.

WHAT WE TAKE FROM THIS

Our curriculum condition can follow either idea: grow the object set (UniDexGrasp++) or widen parameter ranges (ADR). Observation history is a cheap extra ablation.

STATE OF THE ART · 3 OF 3

How generalization is measured

WHAT THE FIELD RECOMMENDS

Kirk et al. · JAIR 2023 survey

Zero-shot generalization means training and test contexts are disjoint. The authors also argue that purely procedural content generation is not a sufficient basis for benchmarks.

Cobbe et al. · ICML 2019

The generalization gap is training performance minus test performance on unseen levels.

PROTOCOL WE ADOPT

- Disjoint object sets, fixed before any training run
- Held-out set includes real scanned objects, not only procedural shapes
- Report held-out success, generalization gap, steps and wall-clock to a target success rate, success under disturbance
- At least 5 seeds per condition; interquartile mean with bootstrap confidence intervals (Agarwal et al. 2021)

POSITIONING

Where our project fits

Work	Embodiment	Observations	Variety studied	Learning
Cobbe et al. 2019	Game agent	Pixels	Procedural levels	RL
Tobin et al. 2018	Robot gripper	Images	Procedural objects	Supervised
UniDexGrasp++ 2023	Dexterous hand	Point cloud	3000+ objects, curriculum	RL, distillation
DextrAH-G 2024	Arm with dexterous hand	Depth	Many objects	RL, teacher-student
Our project	UR10e, Robotiq 2F-85	State, finger forces	Geometry and physics, separately	RL (PPO), MuJoCo

THE GAP

In the work we reviewed we found no controlled study that separates geometry from physical parameters for an industrial parallel gripper with state observations, and reports training cost next to held-out success.

RESOURCES · SOFTWARE

Simulation and learning stack

MuJoCo

Adopted

Todorov et al. · IROS 2012

Fast and deterministic on CPU, accurate contact for grasping, one pip install. Chosen with the supervisor after Isaac Sim proved incompatible with Colab.

Stable-Baselines3

Adopted

Raffin et al. · JMLR 2021

The simplest reliable PPO on Colab. Its callback mechanism already records an evaluation video every 100 iterations.

MuJoCo Playground, MJX

Reference

Zakka et al. · 2025

GPU-parallel MuJoCo in JAX, with manipulation tasks transferred to real robots. Our route if CPU throughput limits the number of seeds.

Isaac Lab

Reference

Mittal et al. · 2025

The original target platform. Kept as an option for a later sim-to-sim cross-check of the main trend.

Decision: every condition trains with SB3 PPO and the same hyperparameters, so differences come from training variety and not from tuning.

RESOURCES ALREADY ADOPTED · NOTEBOOKS 00 TO 02

What is already in place

Robot model

UR10e with Robotiq 2F-85

Menagerie models joined with MjSpec under the prefix g_: 14 joints, 7 actuators, 2 ms step. Joint mapping checked by name with automatic tests; tracking errors about 0.01 rad.

Control

Final pose plus small corrections

Damped least-squares IK gives the final pose. PD with a rate-limited reference keeps motion within 50% of UR10e speed limits. PPO adds small Δq corrections and a gripper command.

Sensing

Forces, weight and slip

Normal and tangential finger forces, gripper effort, mass from arm torques (0.300 kg estimated for 0.300 kg), and slip margin 1 − max |Ft| / (μ Fn).

Scripted test

Grasp, lift and push

Lift by 20 cm, then an 8 N lateral push for 0.3 s: the object stayed in the gripper, slip margin about 0.70. One run, one object: a pipeline check, not a result.

RESOURCES · CODE

MuJoCo components we can reuse

Component	What it gives us	Verdict
MuJoCo Menagerie	Curated UR10e and Robotiq 2F-85 models	Adopted
MjSpec API (MuJoCo 3.2+)	Attach gripper to arm; build objects in code with set size, mass, friction	Adopted
mink	Differential IK with joint limits and collision avoidance	Adapt if needed
robosuite	Lift task, controllers, object and dynamics randomization	Reference
Gymnasium-Robotics (Fetch)	Goal-conditioned reach and pick-and-place environment structure	Reference
TorchRL MuJoCo tutorial	RL scene with Menagerie UR5e and the same 2F-85 gripper	Reference
rliable	Interquartile mean and bootstrap confidence intervals over seeds	Adopt

Plan: keep our own IK and PD layer, borrow environment structure from robosuite and Gymnasium-Robotics, and report statistics with rliable.

RESOURCES · OBJECT SETS

Object sets for training and held-out tests

Set	Content	Role for us	Caveat
Procedural primitives (MjSpec)	Boxes, cylinders, spheres, capsules, ellipsoids; any size, mass and friction	Baseline object, main training set	Too regular on their own
EGAD (Morrison et al. 2020)	2000+ generated shapes graded by shape complexity and grasp difficulty	Controlled variety axis	Convex decomposition, set mass
YCB (Calli et al. 2015)	About 77 everyday objects with scanned meshes and measured mass	Held-out real objects	Convex decomposition
Google Scanned Objects (Downs et al. 2022)	1000+ scanned household items, simulator-ready models	Larger held-out pool	Decompose, filter by size

SELECTION RULE FOR ALL SETS

Keep objects that fit the 2F-85 (about 92 mm between pads when open in our model). MuJoCo collides meshes as convex hulls, so non-convex objects need decomposition, for example with CoACD or V-HACD.

RESOURCES · BENCHMARKS AND PRIOR IMPLEMENTATIONS

Closest benchmarks and implementations

ManiSkill3

Reference

Tao et al. · RSS 2025 · PickSingleYCB-v1

GPU-parallel benchmark with a single-object YCB picking task and RL baselines. Reference for protocol and a rough check of success rates.

ManiSkill2

Reference

Gu et al. · ICLR 2023 · PickSingleEGAD

The same pick task on EGAD shapes, with training and evaluation model lists. A precedent for using EGAD in an RL pick task.

UR10e thesis in Isaac Lab

Reference

Politecnico di Torino · MSc 2025

UR10e with Robotiq 2F-140, PPO on Reach, Lift and OpenDrawer, with randomization and URSim checks. Other simulator, closest task: a source of reward and limit choices.

DextrAH-G

Reference

Lum et al. · CoRL 2024

RL teacher with privileged state distilled into a depth-based student that grasps diverse objects. A template for our optional vision extension.

Caution: ManiSkill (SAPIEN) and Isaac Lab (PhysX) model contact differently from MuJoCo, so we compare protocols and trends, not raw success numbers.

FROM REVIEW TO PLAN

From resources to experiments

Experiment	What we reuse	What we build
Baseline: one object	Menagerie models, IK, PD and sensors from notebook 02, SB3 PPO	Gymnasium environment and reward (notebooks 03, 05)
Randomized variety	MjSpec object builder (size, mass, friction), primitives and EGAD	Variety levels of 1, 4, 16 and 64 objects; geometry and physics as separate axes
Curriculum	ADR idea, UniDexGrasp++ object curriculum	Schedule driven by training success rate (notebook 06)
Evaluation	YCB and Scanned Objects held out, force-pulse test, video callback, rliable	Standardized push test, steps and wall-clock logs (notebook 07)

EFFORT ESTIMATE

Robot, control and sensing are done in notebooks 00 to 02. What remains is the environment, the object library and the experiments themselves.

WHAT CHANGES IN OUR PLAN

Five adjustments from the literature

01 **Variety as a scale, not a switch**
Train on 1, 4, 16 and 64 objects and fit the curve (Cobbe et al., Lin et al.).

02 **Two separate axes**
Vary geometry and physics separately, then together. MuJoCo takes the larger of pad and object friction, so low object friction alone changes nothing.

03 **Shape-aware state as an ablation**
Finger forces and slip margin say little about shape before contact. Test object dimensions or a small point cloud (Dexsuite uses 64 points).

04 **A fair cost axis**
Compare at equal environment steps and equal wall-clock time; mesh objects may cost more CPU time than primitives.

05 **Memory as an ablation**
Add a short observation history, following Chen et al. on domain randomization.

OPEN POINTS

Risks found in the resources, and our questions

TECHNICAL RISKS AND MITIGATION

- The floor is the table; the UR10e shoulder capsule can hit it: add pedestal and table
- Gripper glitches reported when attaching the 2F-85 to the UR10e: keep the gripper test
- Contact-based mass estimate is biased (0.67 vs 0.30 kg): use the torque-based one
- CPU throughput on Colab may cap the seeds: MJX as fallback

QUESTIONS FOR DISCUSSION

- Is a log-scale variety curve a better answer than a two-condition contrast?
- Should shape information enter the state, or is the shape-blind policy the more honest test?
- Equal environment steps or equal wall-clock time as the primary cost axis?

NEXT STEPS

From review to first results

NB 03

Gymnasium environment

Observations: joints, object pose, finger forces, slip margin. Actions: Δq_goal and gripper.

NB 04

Object library

Primitives, EGAD levels, YCB and Scanned Objects; fixed held-out splits.

NB 05

Baseline PPO

One object, at least 5 seeds; steps and wall-clock logged.

NB 06

Variety and curriculum

Levels 1 to 64, geometry and physics axes, object and range curricula.

NB 07

Evaluation

Held-out success, standardized push test, IQM with bootstrap intervals.

Thank you

We welcome pointers to work we may have missed.

SOURCES

References

- Agarwal et al., Deep RL at the Edge of the Statistical Precipice, NeurIPS 2021
- Akkaya et al. (OpenAI), Solving Rubik's Cube with a Robot Hand, arXiv:1910.07113, 2019
- Calli et al., The YCB Object and Model Set, IEEE RAM 22(3), 2015
- Chen et al., Understanding Domain Randomization for Sim-to-real Transfer, ICLR 2022
- Cobbe et al., Quantifying Generalization in Reinforcement Learning, ICML 2019
- Downs et al., Google Scanned Objects, ICRA 2022
- Gu et al., ManiSkill2, ICLR 2023
- Kirk et al., A Survey of Zero-shot Generalisation in Deep RL, JAIR 76, 2023
- Lin et al., Data Scaling Laws in Imitation Learning for Robotic Manipulation, ICLR 2025
- Lum et al., DextrAH-G, CoRL 2024
- Mittal et al. (NVIDIA), Isaac Lab, arXiv:2511.04831, 2025
- Morrison et al., EGAD!, arXiv:2003.01314, 2020
- Raffin et al., Stable-Baselines3, JMLR 22, 2021
- Tao et al., ManiSkill3, RSS 2025
- Tobin et al., Domain Randomization and Generative Models for Robotic Grasping, IROS 2018
- Todorov et al., MuJoCo: A Physics Engine for Model-Based Control, IROS 2012
- Wan et al., UniDexGrasp++, ICCV 2023
- Zakka et al., MuJoCo Playground, arXiv:2502.08844, 2025
- Zhu et al., robosuite, arXiv:2009.12293
- Politecnico di Torino, Deep RL for Robotic Manipulation on UR10e, MSc thesis, 2025
- Code: google-deepmind/mujoco_menagerie · kevinzakka/mink · google-research/rliable (GitHub)