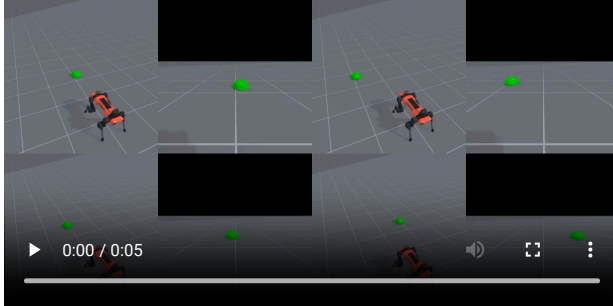


# Supplemental

This is the page documenting and showcasing various demonstration videos of the ManiSkill3 system. Best viewed by opening the supplemental.html file in a web browser such as Chrome, otherwise one can look through the videos folder for the individual videos and follow this document for descriptions of each video.

## Parallel Rendering



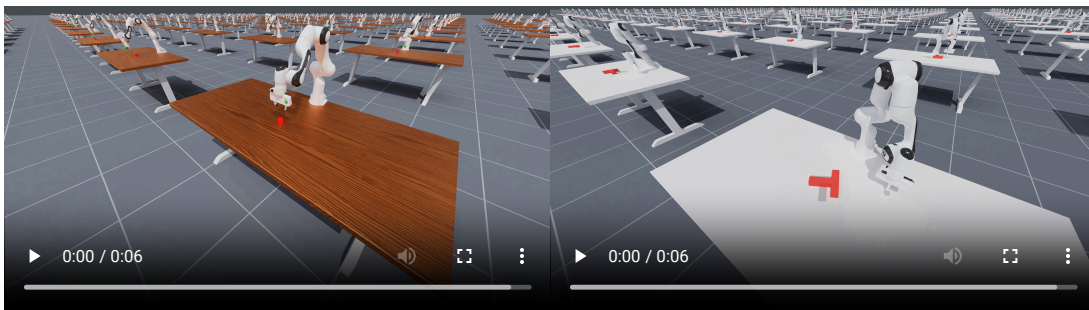
Parallel rendering of the AnymalC Quadruped robot controlled by a visual-based RL policy walking to a goal, showcasing a subset of the 1024 environments being rendered in parallel.

## Heterogeneous Simulation



Parallel heterogeneous simulation of the mobile manipulator Fetch robot opening different cabinets of different degrees of freedom, showcasing the ability to simulate different geometries and articulations in one GPU simulation. The robot is controlled by a state-based RL policy trained in 15 minutes on a single 4090 GPU.

## Fast Visual Training Speed



Fast training speed of a state and vision-based RL policy for the PickCube and PushT tasks. With state inputs PickCube is solved in about 1 minute, PushT is solved in about 5 minutes. With visual inputs PickCube is solved in about 10 minutes and PushT is solved in about 50 minutes. PPO is used for training with 4096 parallel environments for state-based experiments and 1024 parallel environments for vision-based experiments, running on a single 4090 GPU.

# Vision-Based Zero-shot Sim2Real Manipulation

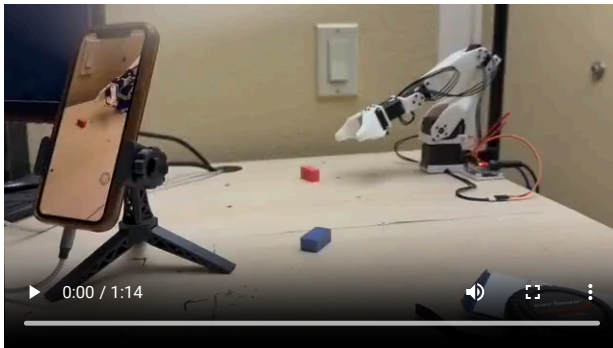
We demonstrate some zero-shot sim2real manipulation results using the low-cost \$300 Koch v1.1. robot arm and 🤖 [LeRobot](#) code for robot hardware interface/control. Policy is trained with PPO on RGB camera inputs and robot proprioceptive data for about an hour on a single 4090 GPU on a domain randomized simulation environment.

## Real World Uncut Evaluation



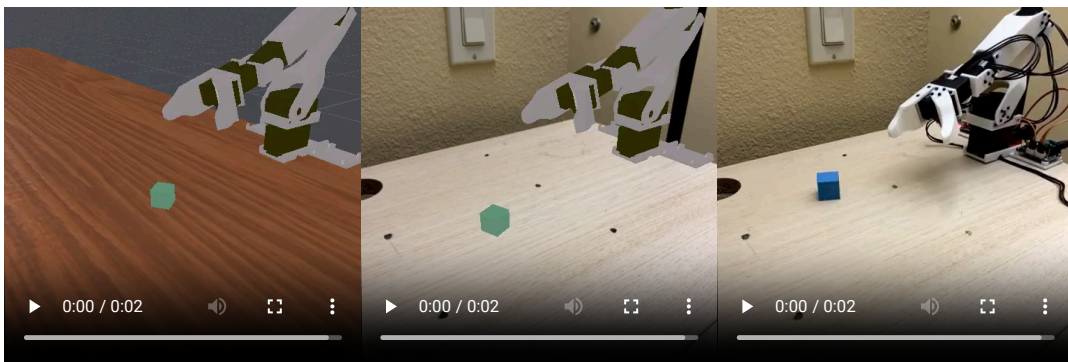
Real world evaluation of the PickCube task at 1x speed. 18/20 trials were successful where success is defined as the robot arm being able to pick up the cube and move it back to a rest position. In all 20/20 trials the robot arm was always able to grasp the cube. The camera observation fed to the policy is displayed on the phone screen. This demo was setup by picking a random table in a house and following our to be open-sourced tutorial to setup the robot and perform training in the new scene.

Interestingly there are some untrained behaviors such as being able to pick up non cube-shaped objects, although we do not claim this kind of generalization always works.



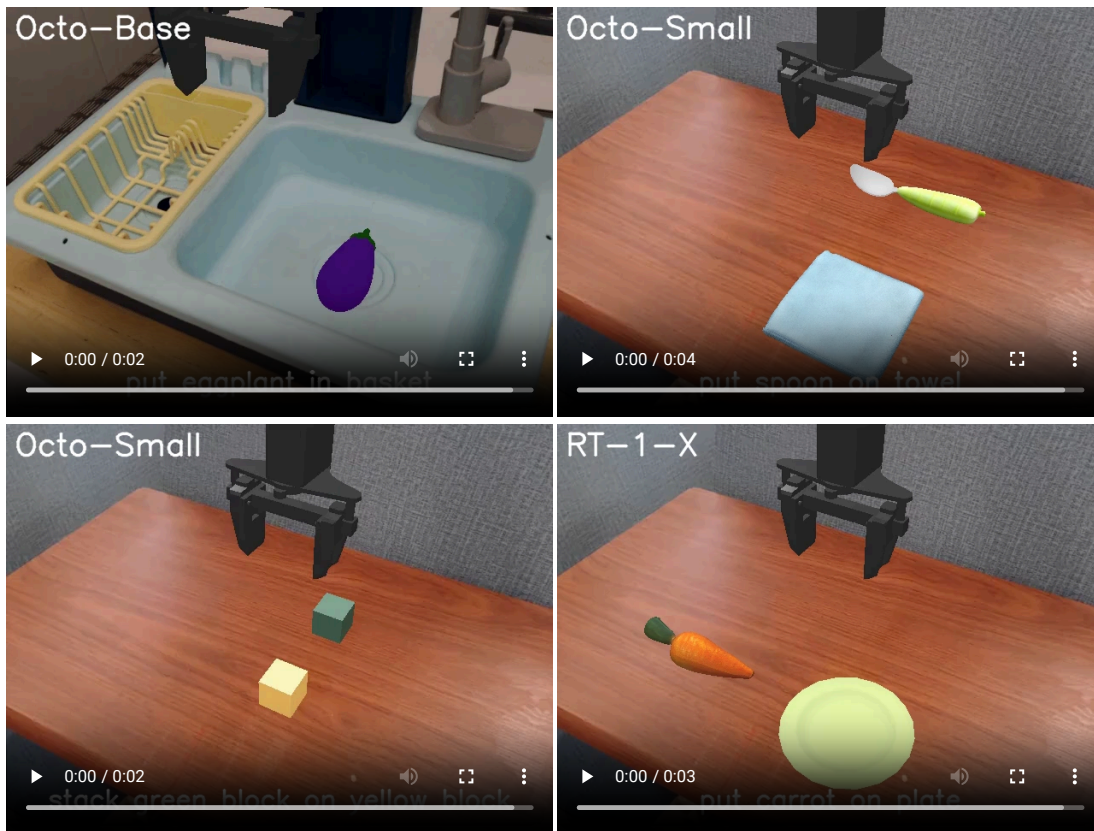
Real world evaluation of the PickCube task at 1x speed on unseen object shapes.

## Reset Distributions



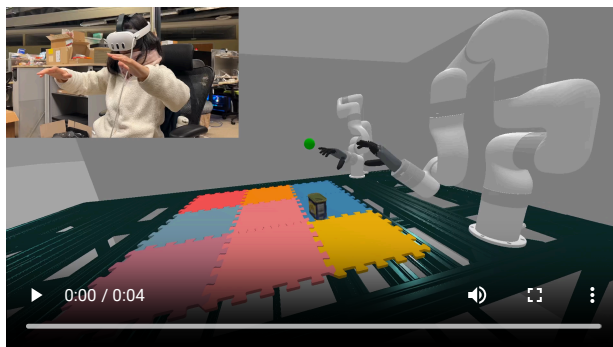
Reset distribution of the PickCube task with the low-cost Koch v1.1. robot arm from LeRobot. Left: Simulation without overlay. Middle: Simulation with overlay. Right: Real world. Reset distribution here shows the domain randomizations all applied together to the simulation environment as well as the robustness testing we perform in the real world by testing on different cube sizes, colors, and poses.

## Real2Sim Evaluation Environments



We port over some of the Real2Sim evaluation environments from the SIMPLER project. The videos below show 4 different vision language action (VLA) models being evaluated on 4 different tasks (videos are originally from SIMPLER). These videos are subsets of the 128 environments that are being simulated and rendered in parallel to evaluate VLAs.

## Teleoperation



We provide a VR based teleoperation system for various robot configurations that is integrated with ManiSkill3, enabling low-latency teleoperation with 4K stereo video streaming to the user at 60 Hz. The video shows a teleoperator controlling two dextrous 5-fingered hands to pick up an object from the YCB dataset in simulation via the Meta Quest 3.