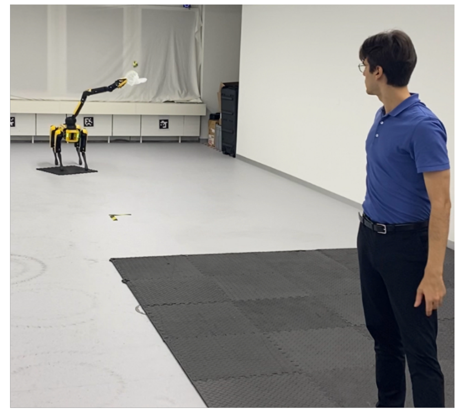
Active Perception for Ball Catching with a Quadrupedal Manipulator

Motivation:



Catching a moving object is a complex real-world challenge requiring precise coordination between perception, locomotion, and manipulation. For robots, this problem is made even harder by limited sensing fields of view and delayed visual feedback. In human ball-catching behavior, we instinctively move not only our hands but also our eyes and heads to keep the object in sight. This kind of active perception—adjusting one's viewpoint to maintain visibility—is largely missing in current mobile manipulation systems.

Traditional robotic systems rely on static or body-mounted cameras, which restrict the robot's ability to resolve occlusions or improve viewpoint during a task. To address this, we propose mounting a RealSense RGB-D camera [1] on the end-effector of a Unitree Z1 arm [2], integrated onto a Unitree B2 quadruped base [3]. This design allows the robot to actively control its own viewpoint by moving the camera in space—similar to how humans reposition their arms or heads to follow a moving object. This enhanced perceptual mobility is crucial for maintaining visibility of the ball, especially during fast, dynamic trajectories.

This project builds on insights from recent work in perception-aware control for robotic sports [4], emergent active visual behavior in humanoids [5], and athletic whole-body manipulation using reinforcement learning [6], aiming to enable intelligent and dynamic response behaviors from simulation-trained policies with no real-world fine-tuning.

Goal:

Our goal is to develop a unified, zero-shot-deployable visuomotor control policy that enables a quadrupedal mobile manipulator—equipped with an arm-mounted RGB-D camera—to perceive, track, and catch a thrown ball using only onboard sensing. Specifically, we aim to:

- Enable viewpoint-aware visual tracking of a moving target via active camera motion
- Coordinate whole-body loco-manipulation for real-time interception and grasping
- Achieve successful ball catching in the real world using policies trained entirely in simulation

Approach:

We train an end-to-end reinforcement learning policy in simulation to perform perception-driven ball catching. The robot receives RGB-D images from the end-effector-mounted camera, along with proprioceptive feedback, and learns to produce joint commands for both locomotion and arm control.

Key elements of our approach include:

- **Camera-on-Arm Active Perception** The RealSense camera is mounted on the robot's end-effector, enabling dynamic viewpoint adjustment. The policy learns to reorient the arm not just for catching, but to actively optimize visual tracking throughout the ball's flight [4].
- **Perception-Aware Simulation** Inspired the work by Ma et al. [4] and Luo et al. [5], We incorporate realistic noise models and randomized ball trajectories in simulation to promote generalization. This encourages the emergence of robust visual behaviors—such as tracking, anticipation, and smooth viewpoint control—that transfer to the real world.

- **Zero-Shot Policy Deployment** The control policy is trained entirely in simulation and deployed in the real world without fine-tuning. By using domain randomization and perception noise injection during training, we prepare the policy to handle real-world uncertainties at test time.

General Details:

The student should bring along the following attributes:

1. Proficiency in Python and machine learning (familiar with PyTorch).
2. Good knowledge of reinforcement learning and image processing.
3. Experience with hardware is preferred.

Interested?

Reach out to Jin Cheng (jicheng@ethz.ch) and Tianxu An (tianan@ethz.ch) with your CV and transcripts.

References

[1] Intel Corporation. Intel realsense technology. `https://www.intelrealsense.com/`. Accessed: 2025-05-30.

[2] Unitree Robotics. Unitree z1 robotic arm. `https://www.unitree.com/z1`, . Accessed: 2025-05-30.

[3] Unitree Robotics. Unitree b2 quadruped robot. `https://www.unitree.com/b2`, . Accessed: 2025-05-30.

[4] Yuntao Ma, Andrei Cramariuc, Farbod Farshidian, and Marco Hutter. Learning coordinated badminton skills for legged manipulators. *Science Robotics*, 10(102):eadu3922, 2025. doi: 10.1126/scirobotics.adu3922. URL `https://www.science.org/doi/abs/10.1126/scirobotics.adu3922`.

[5] Zhengyi Luo, Chen Tessler, Toru Lin, Ye Yuan, Tairan He, Wenli Xiao, Yunrong Guo, Gal Chechik, Kris Kitani, Linxi Fan, et al. Emergent active perception and dexterity of simulated humanoids from visual reinforcement learning. *arXiv preprint arXiv:2505.12278*, 2025.

[6] Nolan Fey, Gabriel B Margolis, Martin Peticco, and Pulkit Agrawal. Bridging the sim-to-real gap for athletic loco-manipulation. *arXiv preprint arXiv:2502.10894*, 2025.