1 Appendix

1.1 Detailed Implementation of Real Robot System

Bimanual Hands System. For our system, we have developed a ROS-based pipeline that operates at a control frequency of 20Hz. This pipeline serves as the foundation for controlling our setup, enabling efficient communication and coordination between the different components. In our configuration, the Arm-Hand subsystems are controlled by a single policy utilizing multiple agents. This unified policy governs the actions of both subsystems, promoting synchronized and collaborative behavior in our setup. To achieve this, we control the motion of the robotic arms through Modbus TCP (Transmission Control Protocol) using an AC/DC Control Box. The control boxes of the two robotic arms are connected to a router via Ethernet cables, and the router is then connected to the host computer. Additionally, the two robot hands are directly connected to the same computer using RS-485 serial communication.

Object Tracking. Real-time object tracking is performed with an Intel RealSense D435 stereo camera. Since the object has a high color contrast from its background, we first use a simple color detector on the RGB image to find the pixel location of the object. The color range for detecting a blue object is constrained between $[80, 200, 0]$ and $[120, 255, 0]$ in HSV color space. Next, the 3D position of that pixel is obtained from querying the corresponding depth value on the depth image, where post-processing filters including disparity, spatial and temporal, are applied to reduce depth noise. Finally, we get the 3D object position in robot frame from image frame with calibrated camera extrinsics parameters.

1.2 Sim2Real Transfer

System Identification. To achieve a successful sim-to-real transfer, we utilize system identification techniques to align the behavior of the PD (Proportional-Derivative) controller of the arm and hand in simulation with that in the real world. This involves tuning the PD coefficients of the controllers to ensure that their responses to impulse and sinusoidal inputs are aligned. This step is crucial in ensuring that the control actions generated in simulation can be effectively applied to the real robot setup, enabling a reliable sim-to-real transfer of our system.

Domain Randomization

Isaac Gym offers several domain randomization functions for reinforcement learning training. We apply randomization to the task, as indicated in Table. 1 for each environment. We generate new randomizations every 1000 simulation steps.

Table 1: Domain randomization parameters.

Parameter	Type	Distribution	Initial Range
Robot			
Mass	Scaling	uniform	[0.5, 1.5]
Friction	Scaling	uniform	[0.7, 1.3]
Joint Lower Limit	Scaling	loguniform	[0.0, 0.01]
Joint Upper Limit	Scaling	loguniform	[0.0, 0.01]
Joint Stiffness	Scaling	loguniform	[0.0, 0.01]
Joint Damping	Scaling	loguniform	[0.0, 0.01]
Object			
Mass	Scaling	uniform	[0.5, 1.5]
Friction	Scaling	uniform	[0.5, 1.5]
Scale	Scaling	uniform	[0.95, 1.05]
Observation			
Obs Correlated. Noise	Additive	gaussian	[0.0, 0.001]
Obs Uncorrelated. Noise	Additive	gaussian	[0.0, 0.002]
Action			
Action Correlated Noise	Additive	gaussian	[0.0, 0.015]
Action Uncorrelated Noise	Additive	gaussian	[0.0, 0.05]
Environment			
Gravity	Additive	normal	[0, 0.4]

1.3 Hyperparameters of the RL algorithms

Table 2: Hyperparameters of MAPPO.

Hyperparameters	Throw and Catch
Num mini-batches	1
Num opt-epochs	5
Num episode-length	8
Hidden size	[1024, 1024, 512]
Use popart	True
Use value norm	True
Use proper time limits	False
Use huber loss	True
Huber delta	10
Clip range	0.2
Max grad norm	10
Learning rate	5.e-4
Opt-eps	5.e-4
Discount (γ)	0.96
GAE lambda (λ)	0.95
Std x coef	1
Std y coef	0.5
Ent-coef	0

Table 3: Hyperparameters of PPO.

Hyperparameters	Throw and Catch
Num mini-batches	1
Num opt-epochs	5
Num episode-length	8
Hidden size	[1024, 1024, 512]
Use popart	True
Use value norm	True
Use proper time limits	False
Use huber loss	True
Huber delta	10
Clip range	0.2
Max grad norm	10
Learning rate	5.e-4
Opt-eps	5.e-4
Discount (γ)	0.96
GAE lambda (λ)	0.95
Std x coef	1
Std y coef	0.5
Ent-coef	0