

APPENDIX I

RANDOMIZATION DETAILS

A. *Demonstration Generation*

We summarize the initial-state randomization for the cabinet (simulation) and microwave (sim-to-real) opening tasks in Table VI. Each parameter is sampled independently per trajectory during demonstration generation. In addition, the robot’s initial arm joints are sampled such that the robot end-effector is within 0.8 m of the object. For both tasks, camera poses are sampled to cover a broad range of over-the-shoulder placements (Table VII). Each camera pose is sampled independently, and the camera pose is re-sampled if the object handle is not visible at the initial timestep.

B. *Visual Domain Randomization*

Table V lists the visual domain randomization applied during IsaacLab rendering for sim-to-real transfer on the microwave task, where we randomize the appearance of the object, robot, table, and background. Backgrounds are drawn uniformly from 485 HDR environment maps from PolyHaven, and object and table textures from 437 material sets from ambientCG. To narrow the sim-to-real gap, we measure the table at our robot workstation and instantiate an analogous table in simulation, randomizing its dimensions about these real-world measurements.

For robust sim-to-real transfer, we perturb the camera parameters and depth maps to make the policy robust against such noise in the real world. We perturb the wrist and external cameras by applying a small amount of jitter to the camera focal length, orientation, and translation such that there is a small mismatch between the camera configuration used to render the images and the recorded camera poses. Additionally, we follow [4], [38] and model edge artifacts in depth maps by introducing correlated depth noise using bilinear interpolation on a shifted depth map.

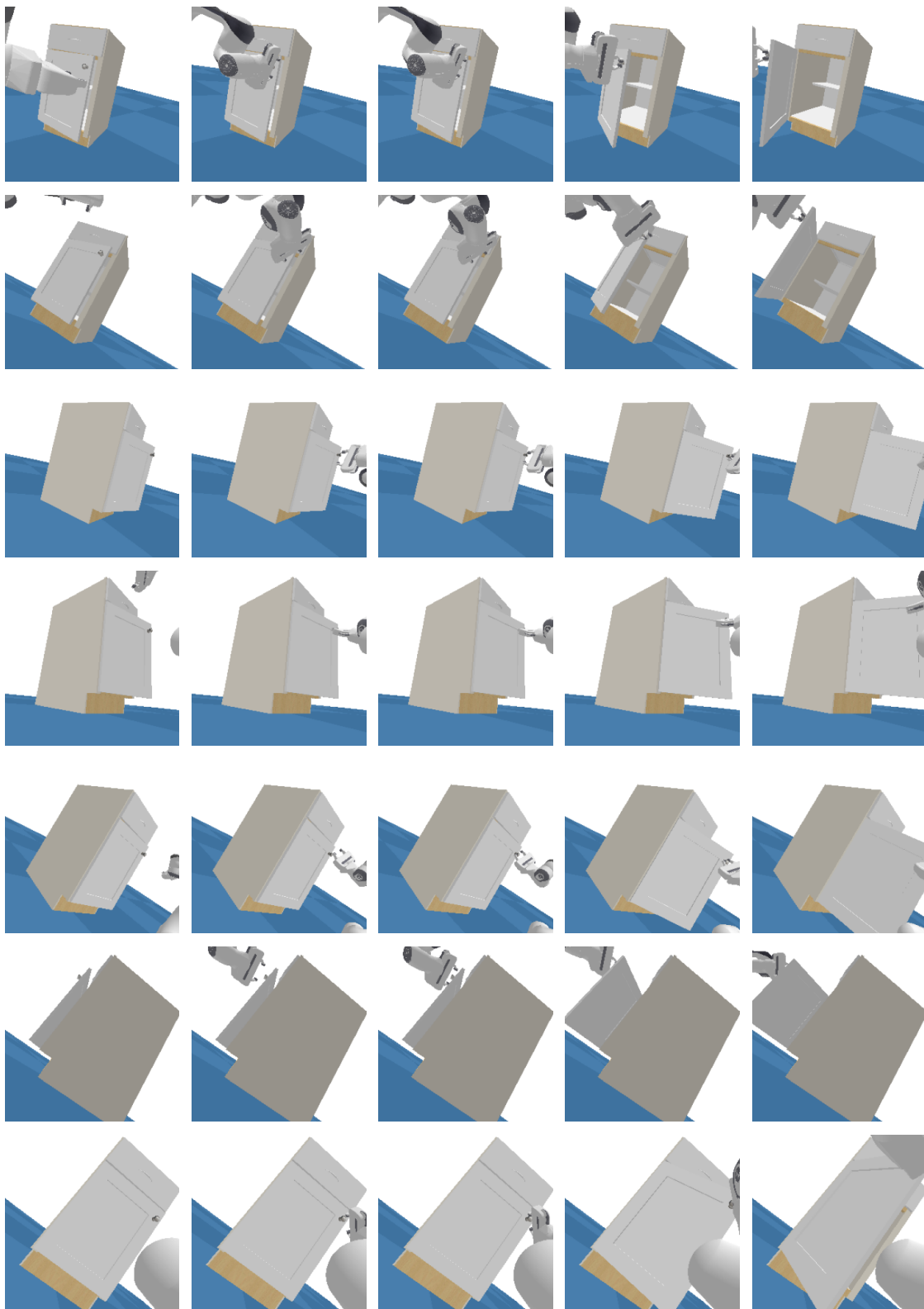


Fig. 7: Visualization of the Cabinet task dataset. Each trajectory is rendered from a random camera pose.

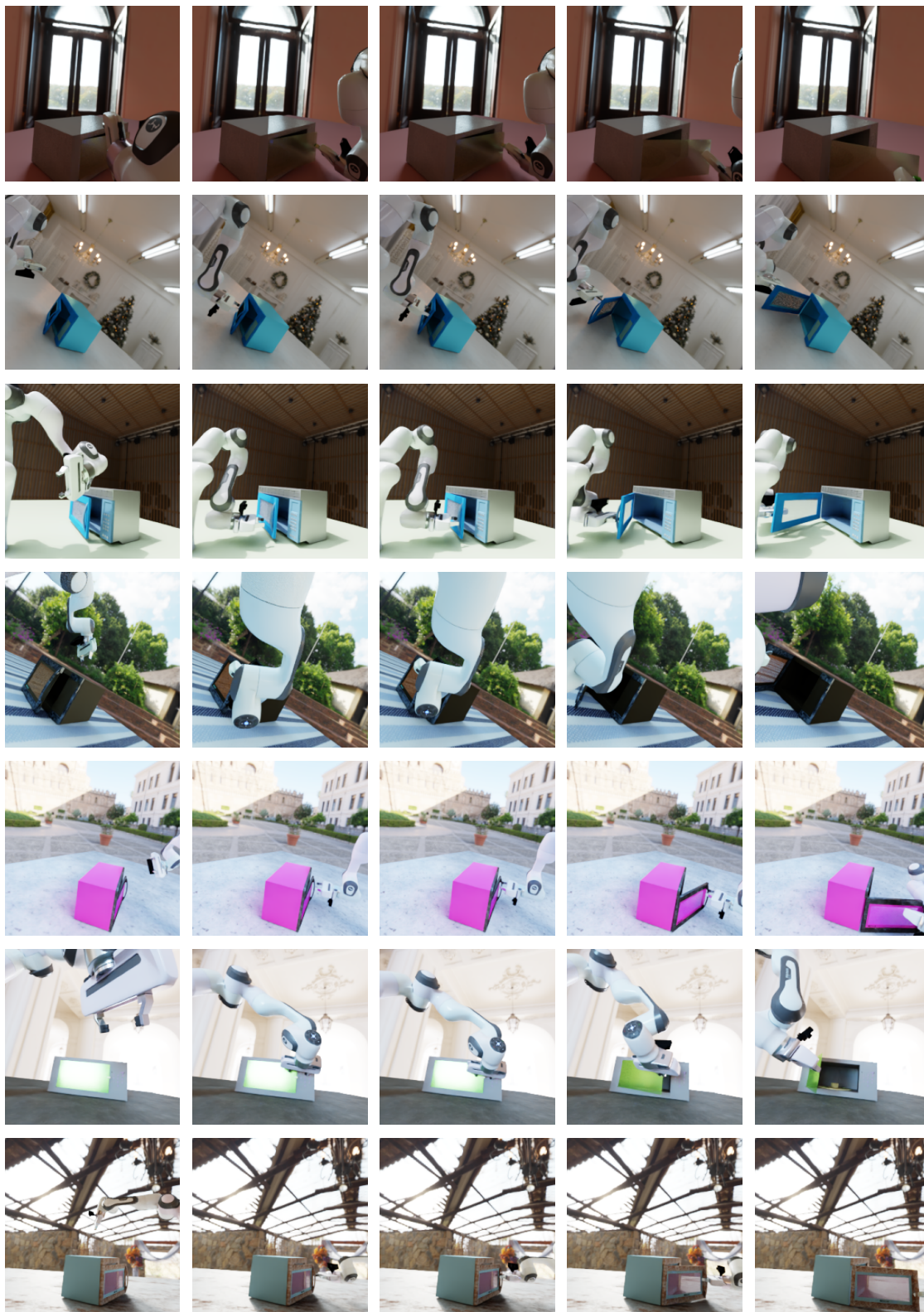


Fig. 8: Visualization of the microwave opening task dataset. Each trajectory is rendered from a random camera pose in IsaacLab.

TABLE V: Visual domain randomization applied in IsaacLab for sim-to-real transfer. Visual appearance, lighting, and camera intrinsics/extrinsics are randomized on top of the demonstration trajectories.

Parameter	Range / Distribution
<i>Object (microwave) appearance</i>	
Material type	50% PBR texture / 50% solid color
Solid color (per channel)	$\mathcal{U}(0.0, 1.0)$
Texture scale	$\mathcal{U}(0.7, 5.0)$
Roughness	$\mathcal{U}(0.0, 1.0)$
Metallic	$\mathcal{U}(0.0, 1.0)$
Specular level	$\mathcal{U}(0.0, 1.0)$
<i>Robot appearance</i>	
Arm material color (per channel)	$\mathcal{U}(0.8, 1.0)$
Gripper base color	$(0.78, 0.80, 0.82) + \mathcal{U}(-0.08, 0.08)$ per channel
Gripper metallic	$\mathcal{U}(0.85, 1.0)$
Gripper roughness	$\mathcal{U}(0.10, 0.30)$
<i>Table</i>	
Length (X) / Width (Y)	$\mathcal{U}(0.78, 0.92)$ m / $\mathcal{U}(1.10, 1.35)$ m
Thickness	$\mathcal{U}(0.02, 0.08)$ m
Center X offset / Center Y	$\mathcal{U}(-0.06, 0.06)$ m / $\mathcal{U}(-0.12, 0.12)$ m
Surface height jitter (Z)	$\mathcal{U}(-0.025, 0.025)$ m
Yaw	$\mathcal{U}(-5^\circ, 5^\circ)$
Surface gray level	$\mathcal{U}(0.85, 0.98)$
<i>Lighting and background</i>	
HDRI environment map	uniform over map set
Dome light intensity	$\mathcal{U}(100, 3000)$
Dome light rotation (Z)	$\mathcal{U}(0^\circ, 360^\circ)$
<i>Camera intrinsics and calibration noise</i>	
Shoulder / wrist focal length	$\times \mathcal{U}(0.98, 1.02) / \times \mathcal{U}(0.99, 1.01)$
Wrist mount pitch (down)	$\mathcal{U}(15^\circ, 25^\circ)$
Wrist mount yaw / roll	$\mathcal{U}(-3^\circ, 3^\circ)$ each
Wrist translation offset	$\mathcal{U}(0, 0.015)$ m, random direction
Shoulder extrinsic noise (trans / rot)	$\mathcal{N}(0, 0.005)$ m / $\mathcal{N}(0, 0.5^\circ)$
Wrist extrinsic noise (trans / rot)	$\mathcal{N}(0, 0.015)$ m / $\mathcal{N}(0, 2.0^\circ)$
Depth edge noise (shoulder / wrist, per frame)	$\sigma = 1.0$ px / 1.5 px, corr. length 18 px

TABLE VI: Randomization applied during demonstration generation for the Cabinet (sim) and microwave (real) opening tasks. Object pose, scale, and initial robot/object configurations are sampled per trajectory.

Parameter	Range / Distribution
Object size ratio	$\mathcal{U}(0.7, 0.9) \times$ base size
Object position (X, Y)	$\mathcal{U}(-0.1, 0.1)$ m about base, per axis
Object yaw	base + $\mathcal{U}(-30^\circ, 30^\circ)$
Door initial joint angle	$\mathcal{U}(0, 0.2) \times$ joint range
Robot arm initial joints	per-joint $\mathcal{U}$ within Franka limits, shrunk 20% each side

TABLE VII: Camera pose randomization for both cabinet and microwave tasks. All parameters are sampled per trajectory.

Parameter	Range / Distribution
Distance to target	$\mathcal{U}(0.8, 1.2)$ m + $\mathcal{N}(0, 0.05)$ m
Look-at target	$(0.7, 0.0, 0.4) + \mathcal{N}(0, 0.05)$ per axis
Shoulder-cam pitch	$\mathcal{U}(-20^\circ, 20^\circ)$
Shoulder-cam roll	$\mathcal{U}(-40^\circ, 0^\circ)$
Shoulder-cam yaw	$\mathcal{U}(-160^\circ, -110^\circ)$ or $\mathcal{U}(-70^\circ, -20^\circ)$