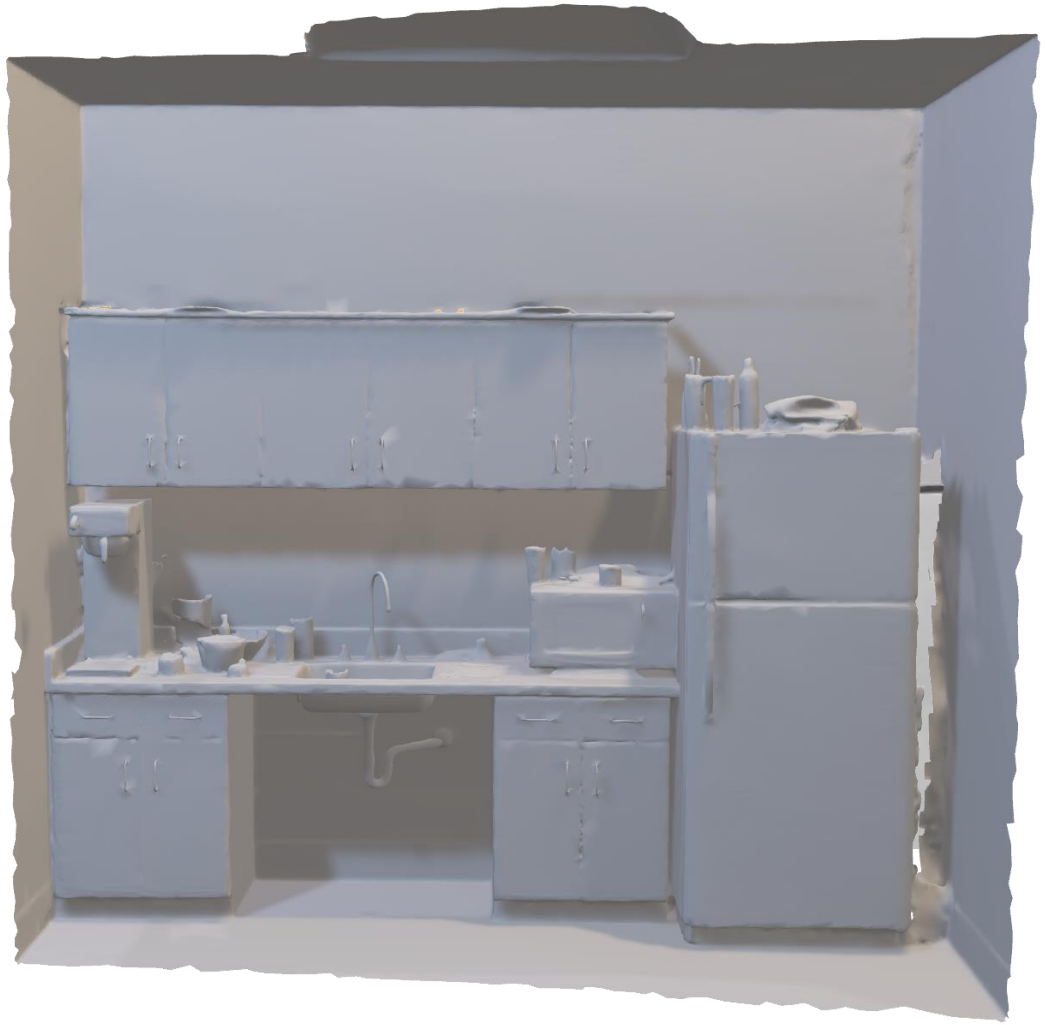
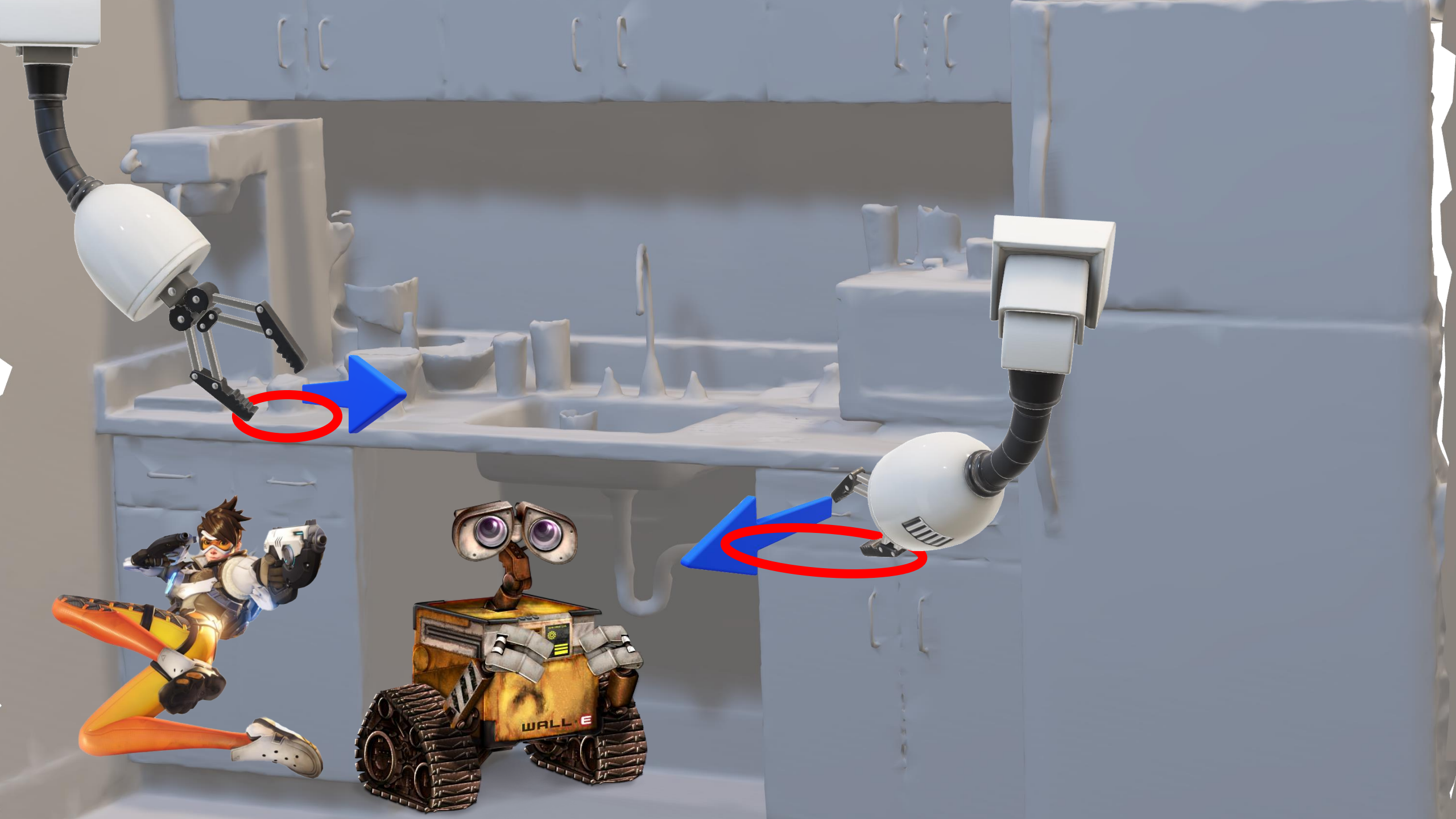
Towards Interactable and Actionable Digital Twins

Wei-Chiu Ma

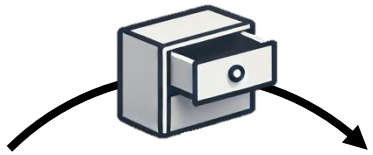
Cornell University

3D Reconstructions are cool!




WALL·E

Interactable and actionable digital twin



Real-to-Sim
(DRAWER)

Real World
(**w/o** any physical interactions)

Sim-to-Real
(Robotics, Gaming)

Digital Twin

Related work includes Torne et al. 2024, Chen et al. 2024, Liu et al. 2025, etc.
See our paper for the full list and a comprehensive discussion.

Project page: https://drawer-art.github.io/

Which 3D representation should we use?

Gaussian splatting

Which 3D representation should we use?

Gaussian splatting

🙂 Visual Fidelity

☹️ Geometry

Neural SDF

Which 3D representation should we use?

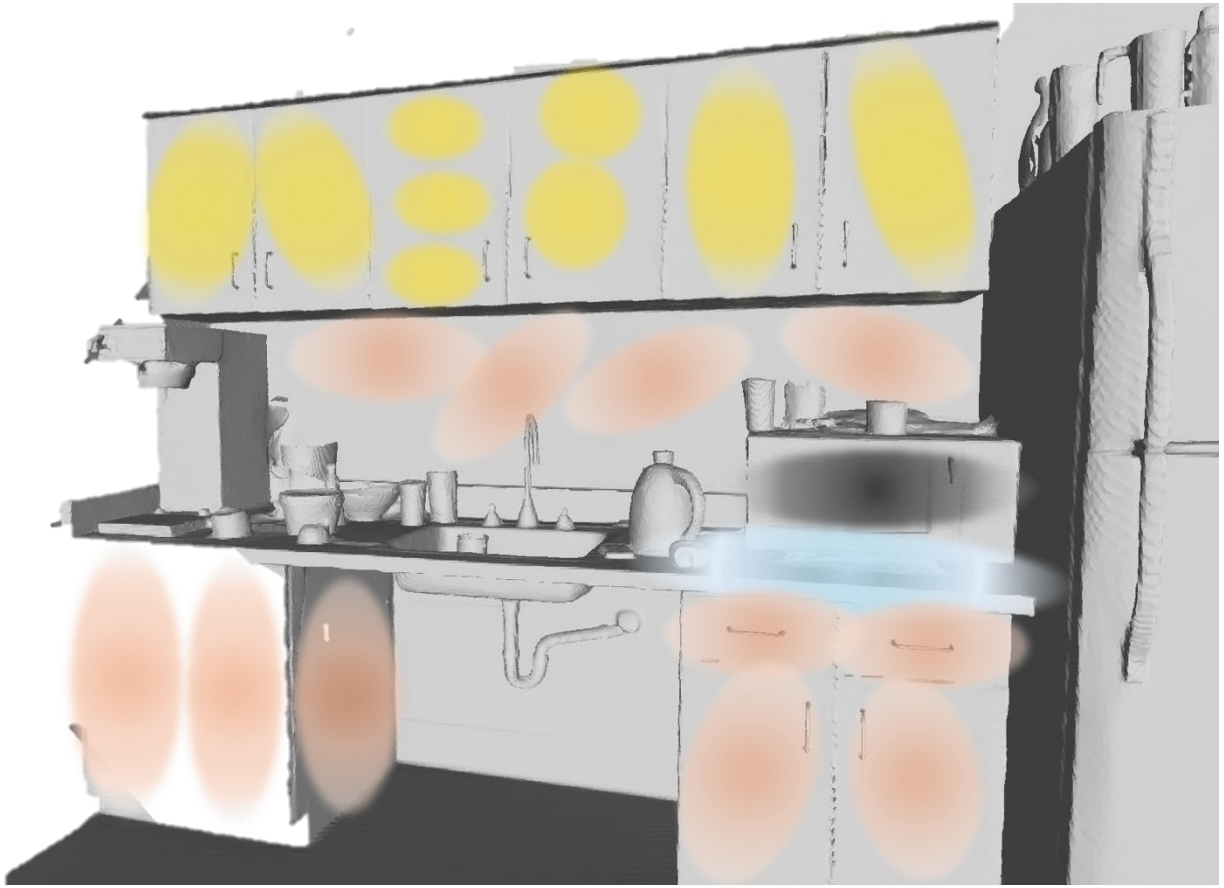
Gaussian splatting

🙂 Visual Fidelity
☹️ Geometry

Neural SDF

🙂 Geometry
☹️ Visual Fidelity

Dual scene representation



Reconstructs accurate surface w/ noise

Related work includes Gao et al. 2024, Qian et al. 2024, Wen et al. 2024, Paudel et al. 2024, etc.
See our paper for the full list and a comprehensive discussion.

Dual scene representation

Gaussian splatting

🙂 Visual Fidelity

☹ Geometry

Neural SDF

🙂 Geometry

☹ Visual Fidelity

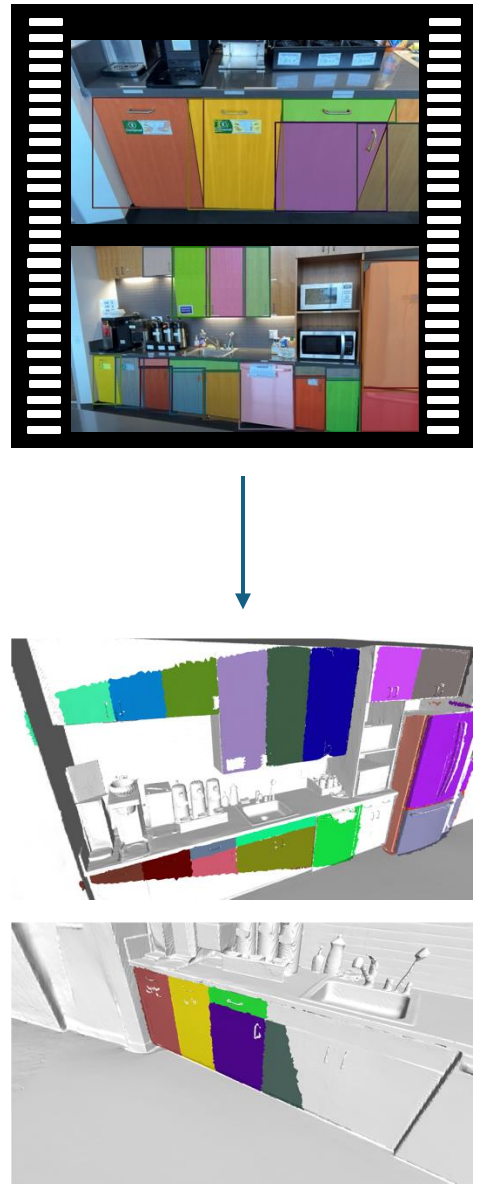
Ours

🙂 Geometry

🙂 Visual Fidelity

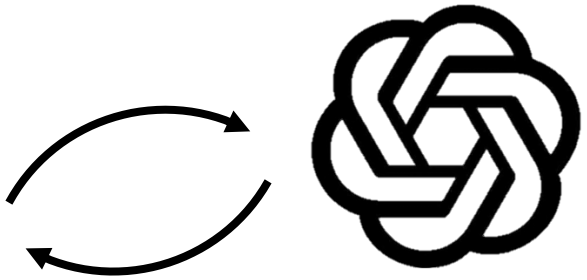
How do we interact with the scene?

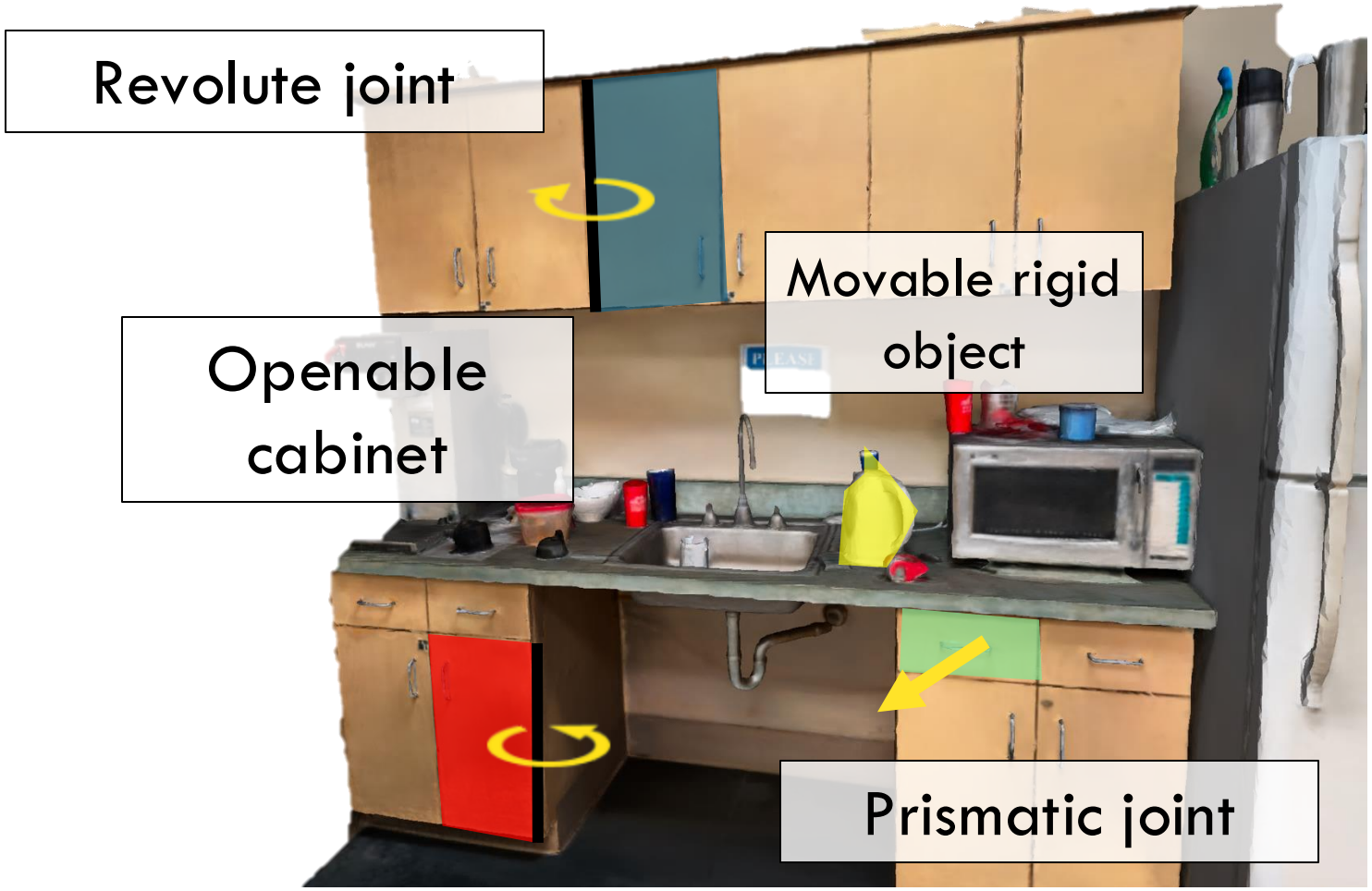
Scene decomposition



Identify <u>interactable</u> and <u>actionable</u> objects

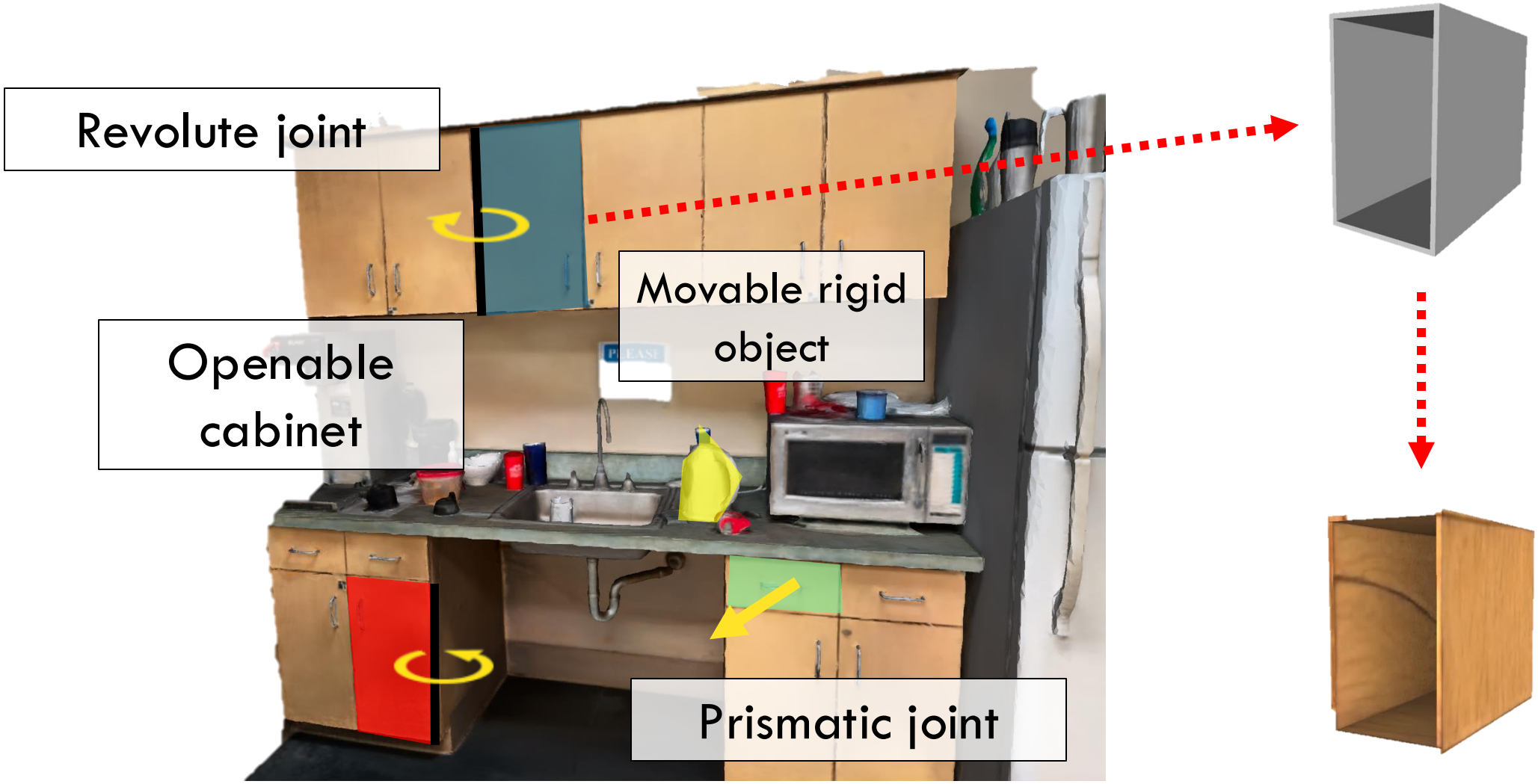
Physical reasoning





Reason about physical properties (e.g., articulation, mass)

Amodal shape estimation



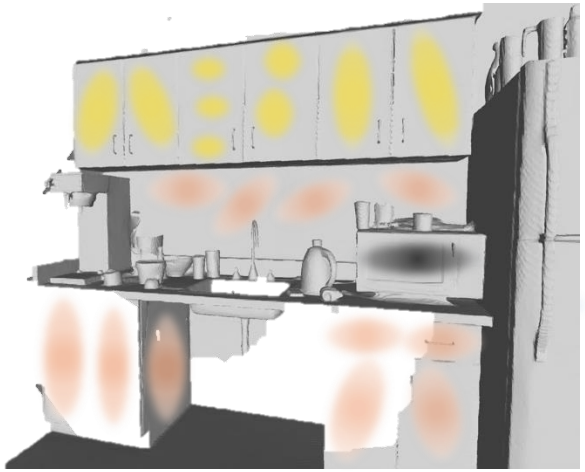
Recover invisible regions via differentiable rendering + generative modeling

DRAWER: Digital Reconstruction and Articulation With Environment Realism

Video

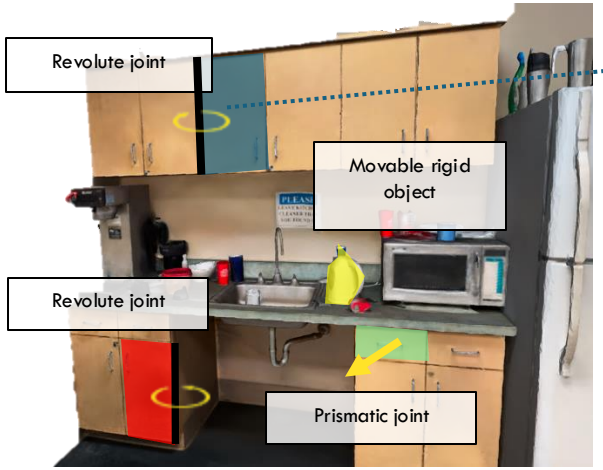
No Interactions

Dual Scene Representation

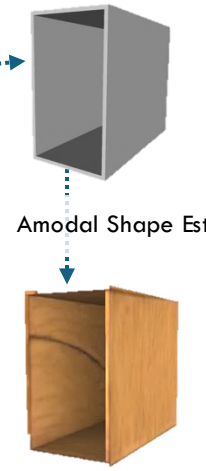


Anchoring Gaussian Splats with SDF

Animating the Scene



Physical Reasoning



Texturing Hidden Region

Interactable Digital Twin

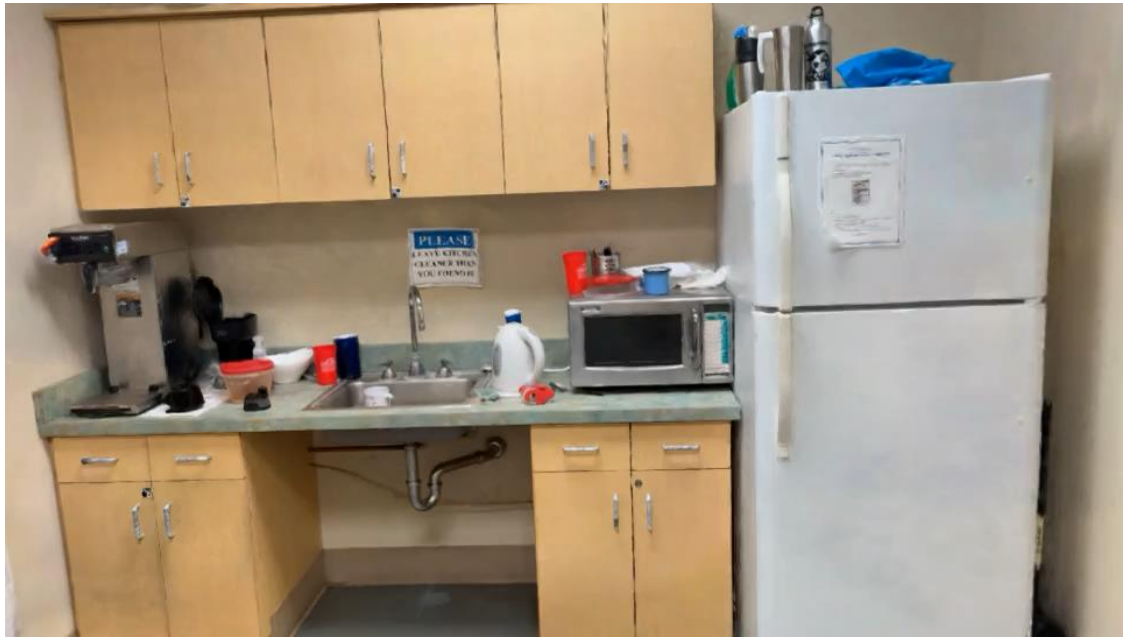
Open Drawers

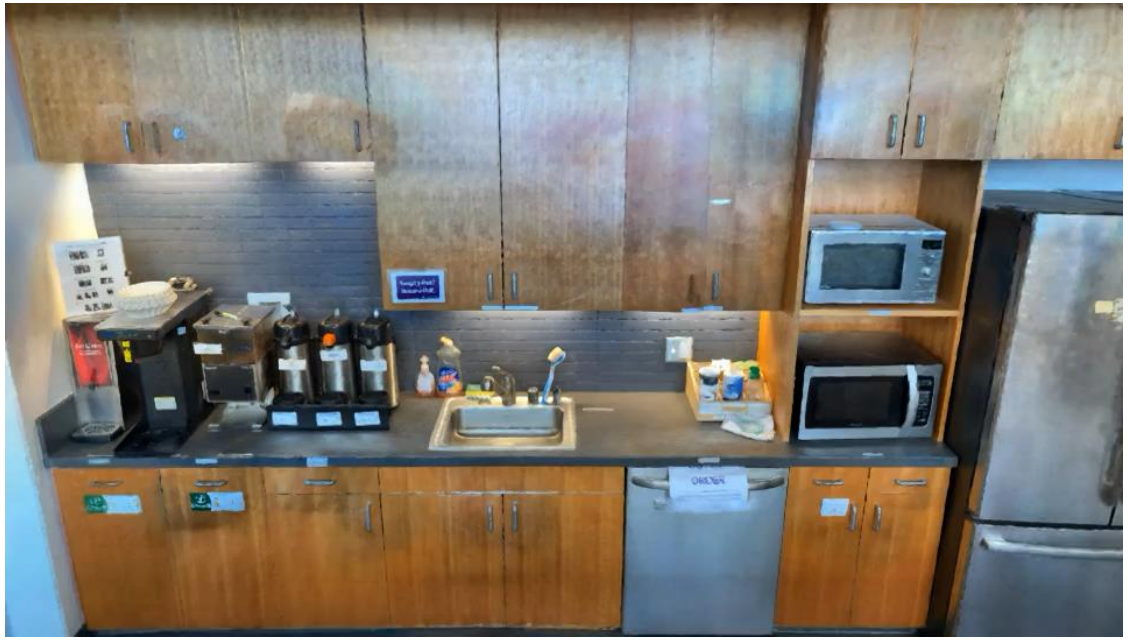
Open Cabinets

Moving Objects

Novel View Synthesis

Reconstructed twins





Gaming

First person shooting

Real-to-Sim-to-Real Transfer

Training in digital twins

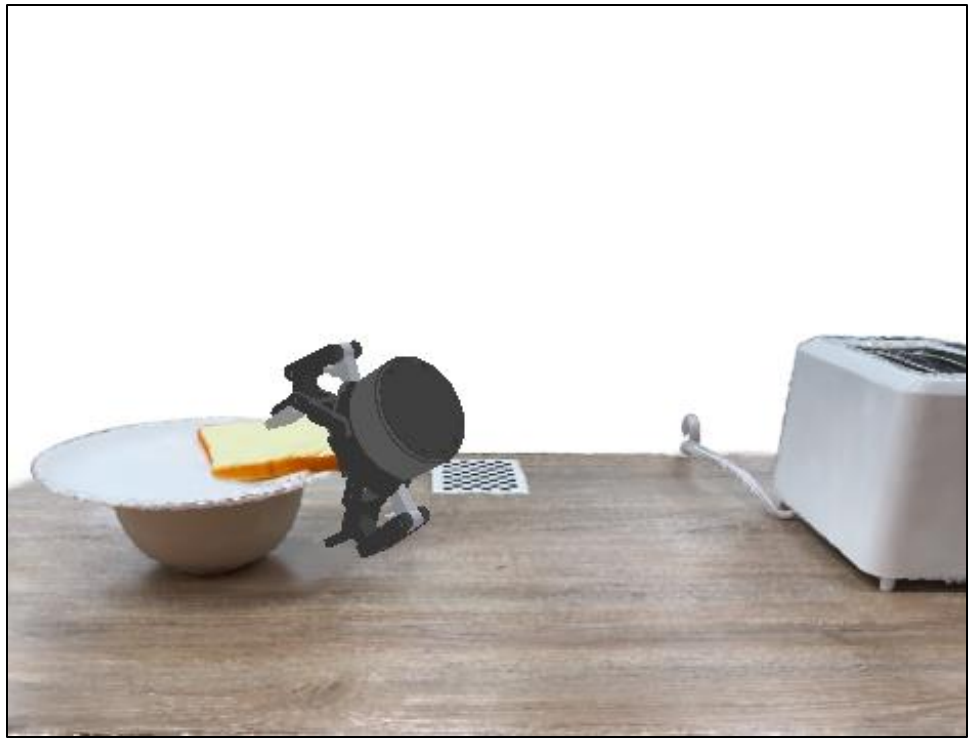
Deploy in the real world

What if we don’t want to train a policy?

Zero-shot Open World Manipulation

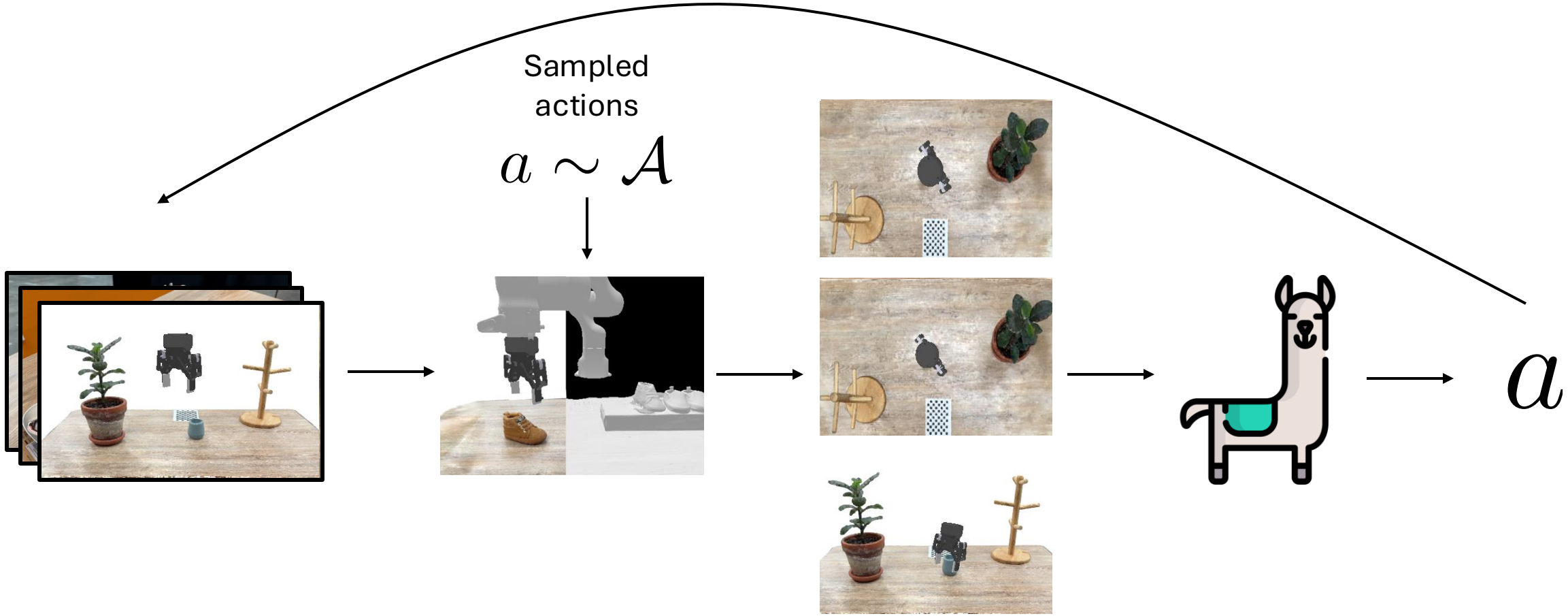
Vision Language Models (VLM) as a Critic

Toast the bread

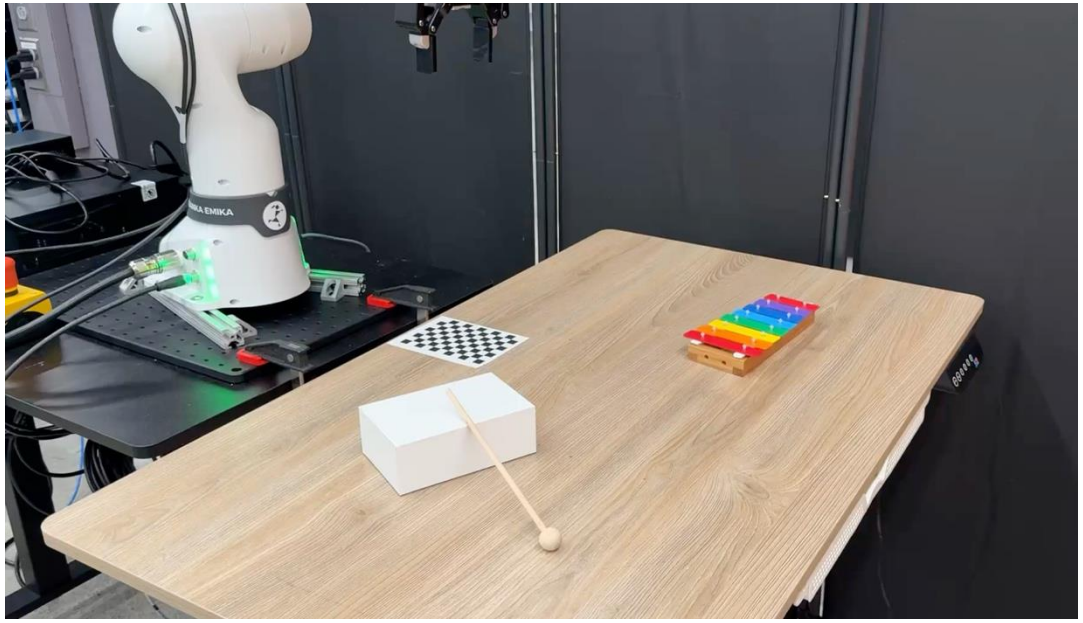


>

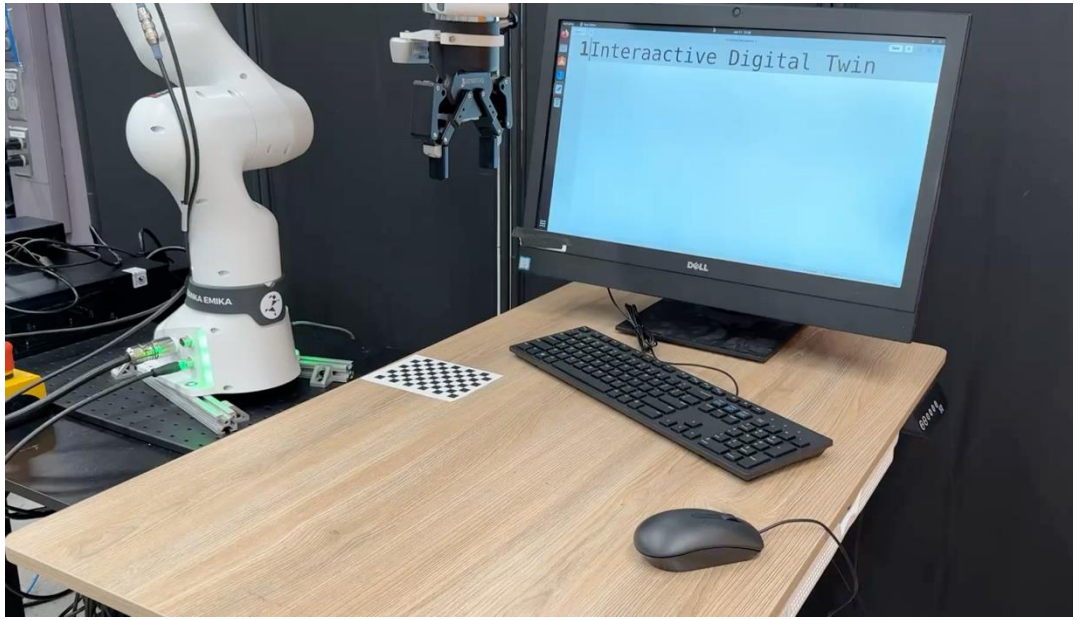
Twin-Assisted Reasoning for Planning (TARP)



Zero-Shot Open World Manipulation



“Play lowest tune”



“Press spacebar”

Acknowledgement

Hongchi Xia

Chuanruo Ning

Ai2