Scalable Trajectory Generation for Whole-Body Mobile Manipulation

Website Code

Yida Niu[1,3,4,5*] Xinhai Chang[1,3,4,5,6*] Xin Liu[4*] Ziyuan Jiao[2,4†] Yixin Zhu[3,4,5,7†]

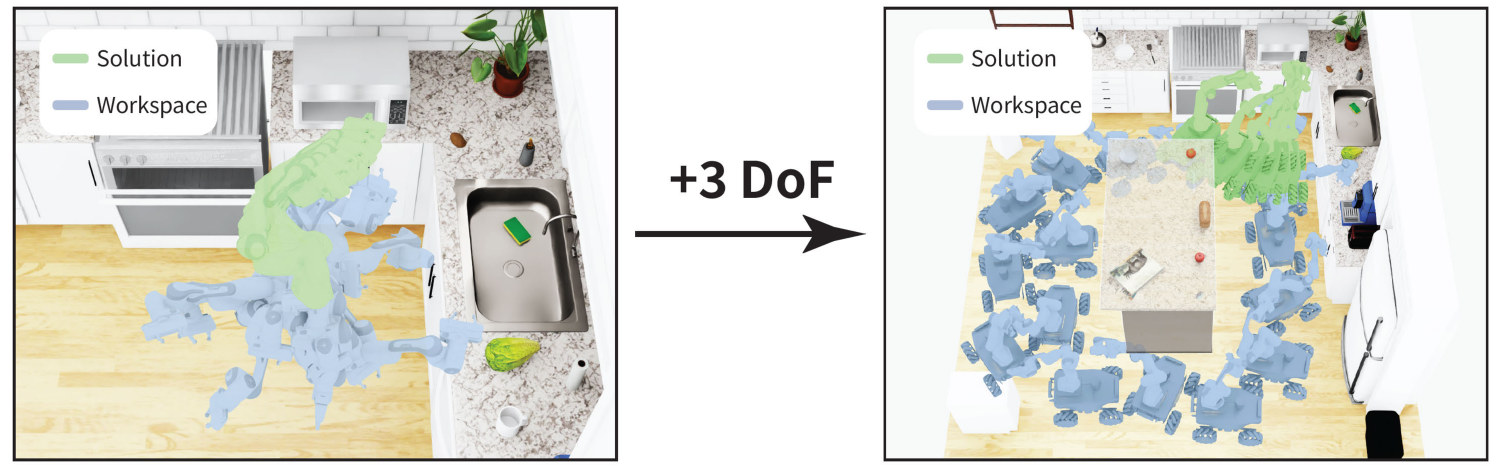
[1]Institute for AI, Peking University [2]Institute of Unmanned Systems, Beihang University [3]School of Psychological and Cognitive Sciences, Peking University [4]State Key Laboratory of General Artificial Intelligence
[5]Beijing Key Laboratory of Behavior and Mental Health, Peking University [6]Yuanpei College, Peking University [7]Embodied Intelligence Lab, PKU-Wuhan Institute for Artificial Intelligence
*Equal contribution †Corresponding authors

CVPR JUNE 3-7, 2026 DENVER COLORADO

A scalable pipeline for whole-body mobile manipulation data — generating 5,000 trajectories/GPU-hr, 80× faster than existing CPU-based methods.

PROBLEM

The mobile base explodes the search space



Fixed base
~98% success rate with 800 episodes
small space, dense solutions

Mobile manipulator
~20% success rate with 3200 episodes
vast space, sparse solutions

Adding a mobile base (+3 DoF: x, y, yaw) turns a small, densely-solved workspace into a vast one where valid solutions are sparse — so **the required training data increases by orders of magnitude.**

Why existing datasets fall short

Dataset	Space	# Episodes	Coord.*	# Scenes	Collection Method
Open X-Embodiment (OXE)	Mixed	1,000,000+	×	~300	Aggregated
DROID	Tabletop (Franka)	76,000	×	564	Human Teleoperation
AgiBot World	Bimanual Humanoid	1,000,000+	×	106	Human Teleoperation
Mobile ALOHA	Mobile (Bimanual)	276	✓	few	Human Puppeteering
M^{3}Bench	Mobile	30,000	✓	119	Automated Planning (CPU)
AutoMoMa (ours)	Whole-body Mobile	500,000+	✓	330	Automated Planning (GPU)

* Coord.: base-arm coordinated whole-body motion.

No existing dataset offers large scale, base-arm coordination, and scene/object diversity at once. Existing data is therefore **insufficient to support learning robust whole-body mobile manipulation.**

Why existing collection methods fall short

Every existing route to coordinated data hits a wall.

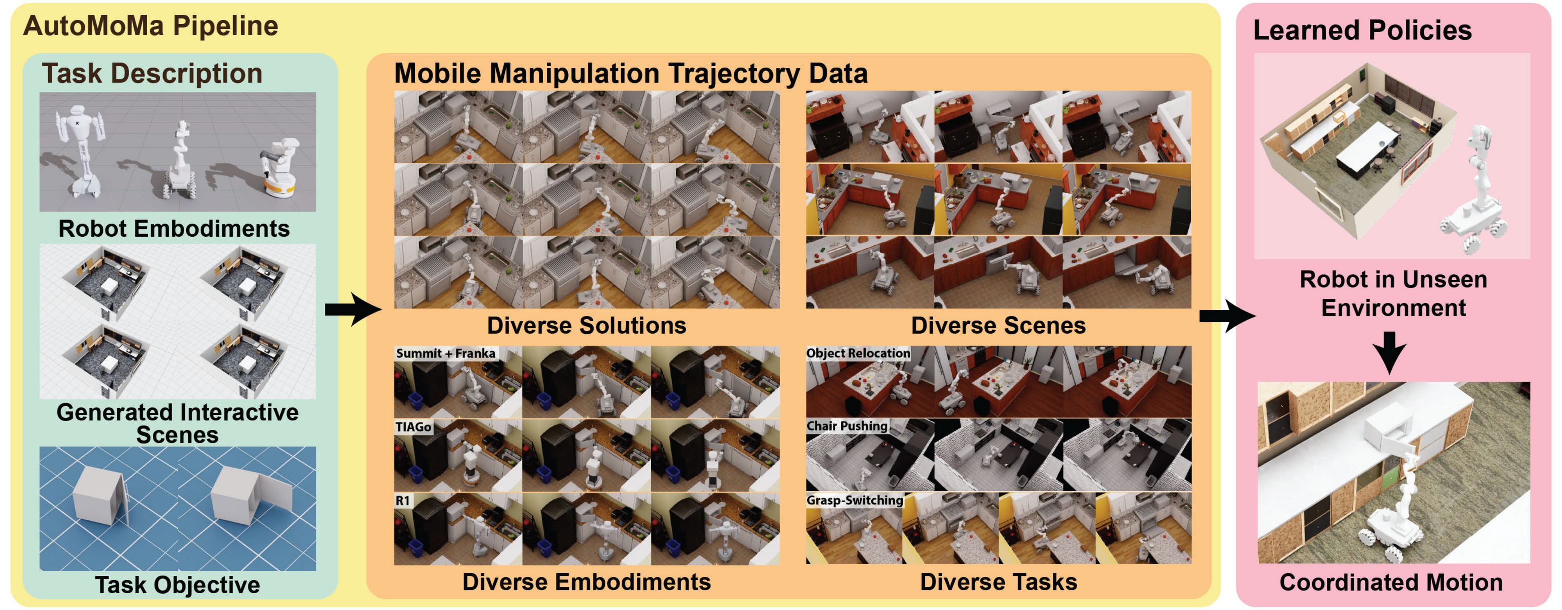
- **Teleoperation** — coordinated whole-body control overwhelms human operators.
- **Reinforcement learning** — brittle reward design and a persistent sim-to-real gap.
- **Motion planning** — physically valid, but limited by CPU at ~60 traj/hr.

Producing physically valid, whole-body coordinated data at scale requires **automatic generation, physical grounding, and high throughput** — a combination prior methods have not achieved.

Whole-body data is scarce — costly to collect and hard to scale.

METHOD

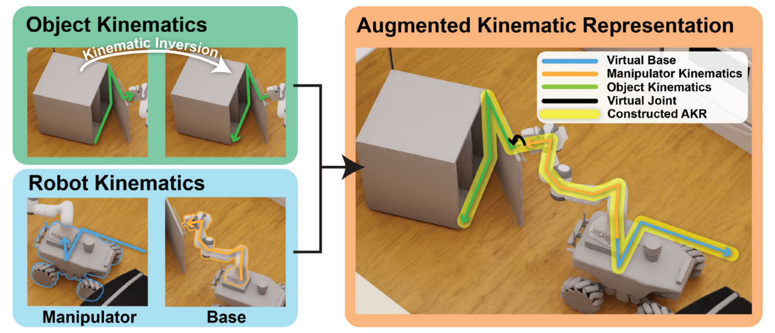
AutoMoMa: one pipeline, diverse coordinated data



Our pipeline generates **whole-body coordinated** mobile manipulation trajectories that offer **diverse solutions** across various robots, environments, and tasks.

AKR — base, arm, and object as one chain

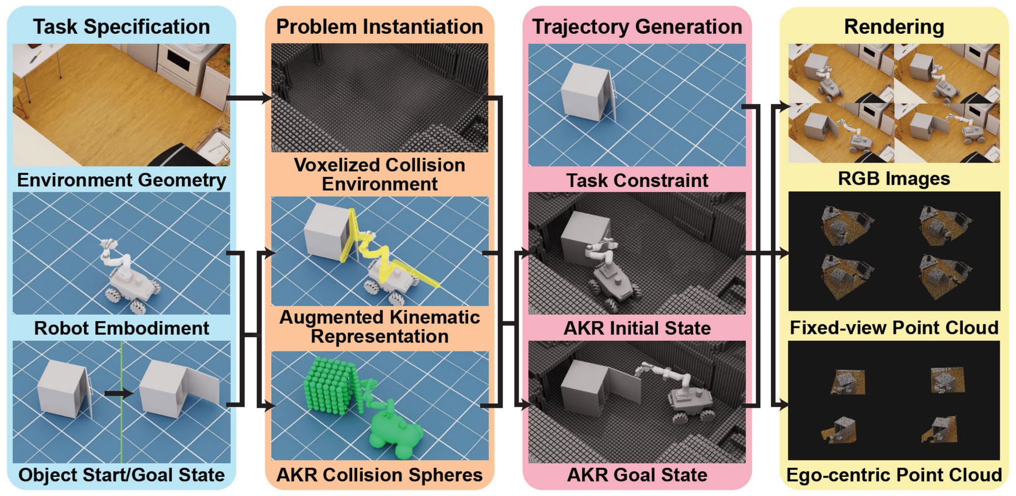
(Augmented Kinematic Representation)

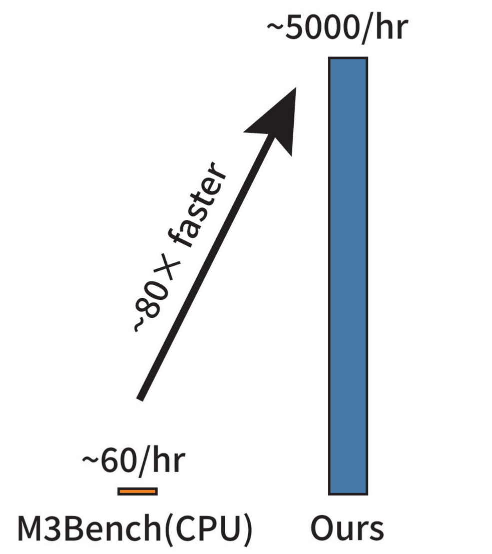


- **One unified kinematic chain:** base, arm & object in one representation
- **Single whole-body optimization:** no decoupled base / arm planning
- **Physically valid by construction:** satisfies grasp, collision & task constraints

A kinematic representation that unifies the mobile base, manipulator, and target object into one chain, **turning whole-body planning into a single optimization problem.**

Four-stage pipeline, 80× faster on GPU





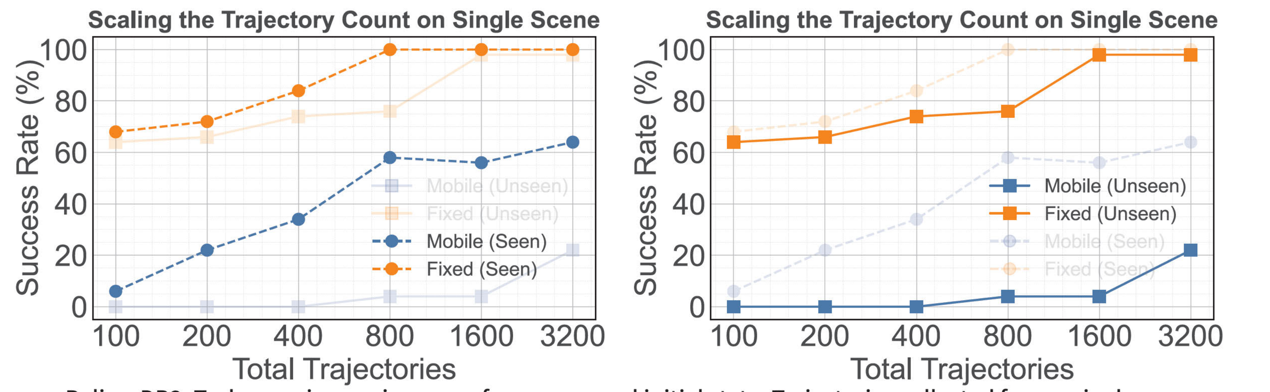
A four-stage pipeline that produces physically-valid whole-body trajectories from a task specification, **with GPU-accelerated planning.**

~80× faster at the same validity as its CPU baseline.

AKR + GPU together unlock whole-body data at the scale.

RESULTS

Mobile manipulation needs far more data than fixed-base

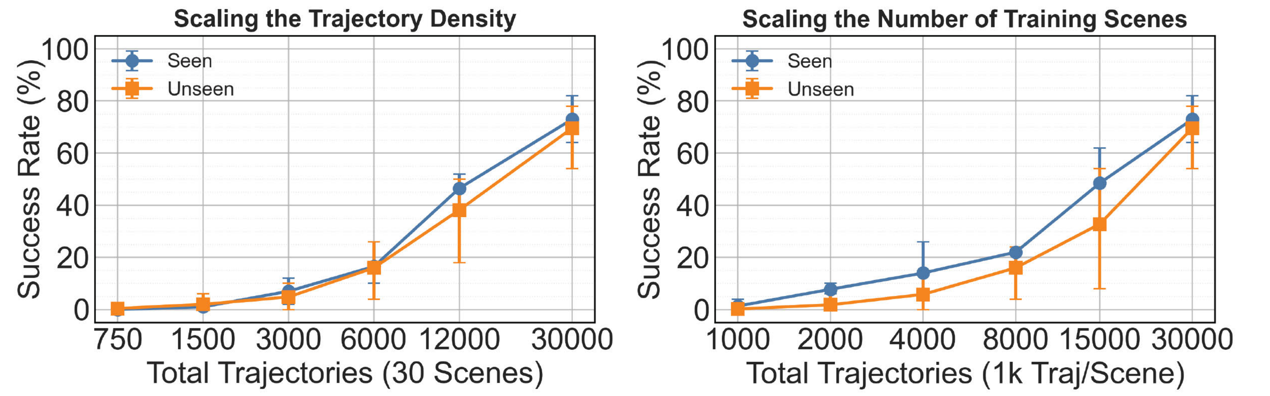


Policy: DP3. Task: opening a microwave from a grasped initial state. Trajectories collected from a single scene.

Fixed-base **saturates fast at 100%.** Mobile improves slowly and **reaches ~60%** — tested on **seen initial states.**

Fixed-base still **saturates at ~98%.** Mobile struggles to learn, **reaching only ~20%** — its larger search space **demands far more data** than 3,200.

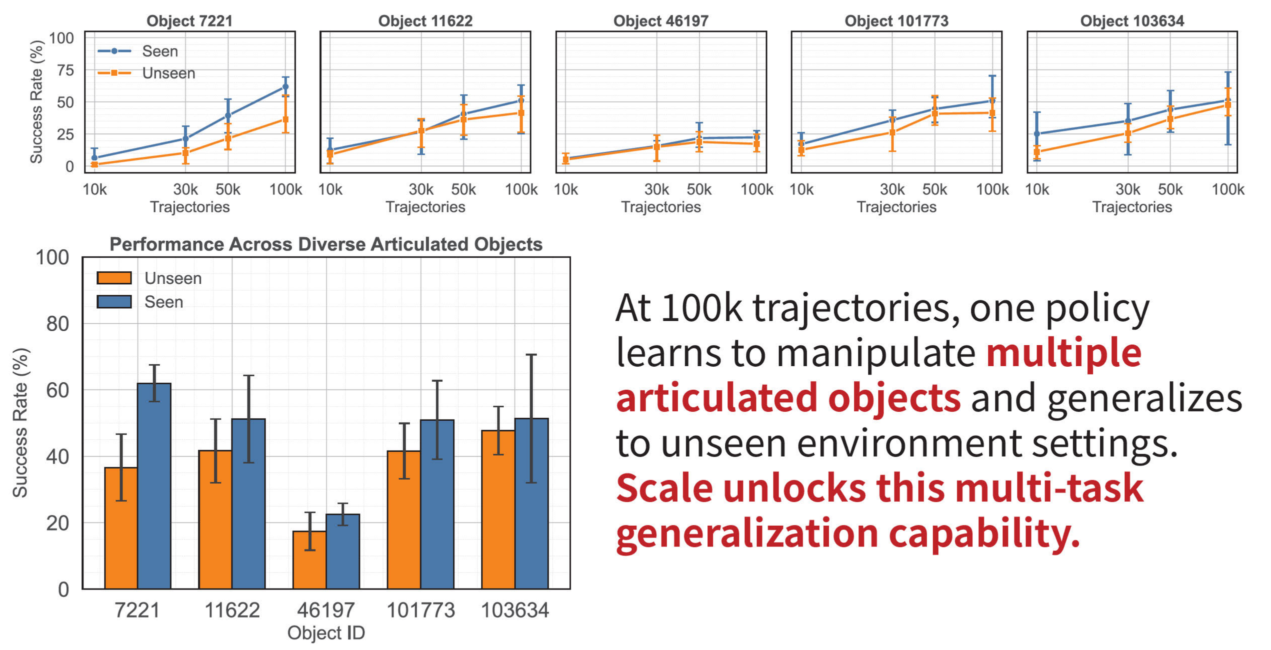
Scale and diversity close the gap



Policy: DP3. Task: opening a microwave from a grasped initial state.

At 30k trajectories with diverse scenes and solutions, the policy **generalizes robustly on unseen scenes.** Both trajectory density and scene diversity contribute.

At 100k data, IL handles multiple objects

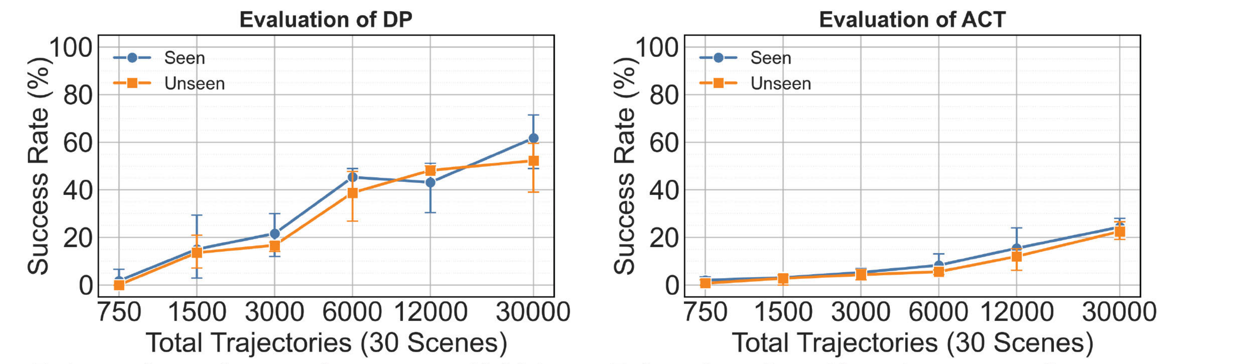


At 100k trajectories, one policy learns to manipulate **multiple articulated objects** and generalizes to unseen environment settings. **Scale unlocks this multi-task generalization capability.**

Scale and diversity make whole-body mobile manipulation generalizable.

ABLATIONS AND CONCLUSIONS

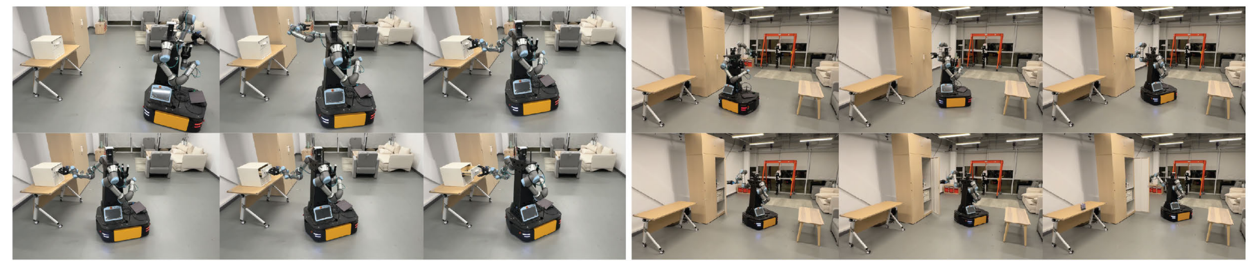
The scaling pattern is policy-agnostic



Task: opening a microwave from a grasped initial state. Trajectories collected from the same set of scenes.

Replacing DP3 with Diffusion Policy or ACT preserves the same scaling trend on both seen and unseen scenes. **Scale benefits whole-body IL across all diverse architectures.**

Trajectories execute on a real robot



Ridgeback-UR5 executing open drawer task. Ridgeback-UR5 executing open closet task.

AutoMoMa-generated trajectories execute as planned on a real Ridgeback + UR5 platform, confirming the data is **physically valid beyond simulation.**

Conclusion

500k mobile manipulation episodes 330 kitchen arrangements support objects from SAPIEN

AutoMoMa provides scalable, physically-valid whole-body data for the field.

- ✓ **Scalable trajectory generation** — 80× faster than CPU planning
- ✓ **Physically-valid whole-body data** — base + arm + object as one chain
- ✓ **Diverse coverage** — 500k+ episodes across 330 scenes and multiple objects
- ✓ **Real-world validation** — trajectories execute successfully on Ridgeback + UR5
- ✓ **Reusable pipeline** — for new robots, scenes, and tasks beyond this dataset

We introduce AutoMoMa, a GPU-accelerated **data generation pipeline for whole-body mobile manipulation** (80× CPU throughput), and a **500k+ trajectory dataset.** Our experiments reveal that effective generalization **requires data at a scale and diversity far beyond what tasks with fixed-base settings need.**

Limitations: known object geometry • sphere-approximated collisions • open-loop planning.

Data scale is the variable that matters most — and now it's available.