LeHome: A Simulation Environment for Deformable Object Manipulation in Household Scenarios

Zeyi Li[2], Yushi Yang[1], Shawn Xie[3], Kyle Xu[3], Tianxing Chen[4], Yuran Wang[1], Zhenhao Shen[1],
Yan Shen[1], Yue Chen[1], Wenjun Li[4], Yukun Zheng[4], Chaorui Zhang[3], Siyi Lin[3], Fei Teng[3],
Hongjun Yang[2], Ming Chen[3], Steve Xie[3] and Ruihai Wu[1,*]

Fig. 1: **LeHome** provides a high-fidelity simulation platform by integrating various household scenarios and various objects within the scenarios, especially **deformable** objects.

Abstract—Household environments present one of the most common, impactful yet challenging application domains for robotics. Within household scenarios, manipulating deformable objects is particularly difficult, both in simulation and real-world execution, due to varied categories and shapes, complex dynamics, and diverse material properties, as well as the lack of reliable deformable-object support in existing simulations. We introduce LeHome, a comprehensive simulation environment designed for deformable object manipulation in household scenarios. LeHome covers a wide spectrum of deformable objects, such as garments and food items, offering high-fidelity dynamics and realistic interactions that existing simulators struggle to simulate accurately. Moreover, LeHome supports multiple robotic embodiments and emphasizes low-cost robots as a core focus, enabling end-to-end evaluation of household tasks on resource-constrained hardware. By bridging the gap between realistic deformable object simulation and practical robotic platforms, LeHome provides a scalable testbed for advancing household robotics. Webpage: lehome-web.github.io/.

I. INTRODUCTION

For humans, the household scenario is the living space with the highest frequency of daily activities and the most critical needs. It encompasses a wide variety of essential daily tasks, such as organizing personal belongings, preparing food, and managing clothing. Unlike those in more structured scenes such as industries, these tasks feature distinct challenges: interactions with diverse, non-standardized objects and adaptation to unstructured, dynamic environments.

Amid these challenges, household scenarios have attracted increasing attention, ranging from the development of simulated household simulation platforms [15], [16], [28] to the design of robotic policies for household tasks [1], [14], [36].

However, a key mismatch remains between current benchmarks and real-world needs: most existing methods target rigid and articulated objects, while many household tasks inherently involve diverse deformable objects (*e.g.*, garments and food), for which current support remains limited. Such objects lack fixed shapes, deform nonlinearly under applied forces, and exhibit dynamic physical parameters that are

*Corresponding author: Ruihai Wu, `wuruihai@pku.edu.cn`
[1]Peking University.
[2]Institute of Automation, Chinese Academy of Sciences.
[3]Lightwheel.
[4]The University of Hong Kong.

TABLE I: **Features Comparisons** across different open-source robotics environment.

Feature	LeHome (ours)	RoboTwin2.0	DexGarmentLab	Behavior-1K	Libero	RLbench	Robocasa
Household Scenarios	✓	✗	✓	✓	✓	✗	✓
Photorealistic Rendering	✓	✓	✓	✓	✗	✗	✓
Food Deformation Simulation	✓	✗	✗	✗	✗	✗	✗
Flame and Particle Simulation	✓	✗	✗	✓	✗	✗	✗
Fluid Simulation	✓	✗	✓	✓	✗	✗	✗
Garment Manipulation	✓	✗	✓	✓	✗	✗	✗
Articulation and Rigid Manipulation	✓	✓	✗	✓	✓	✓	✓
Multi-material Manipulation	✓	✗	✓	✓	✗	✗	✗
Teleoperation	✓	✗	✗	✓	✗	✗	✓
Low-cost Robot	✓	✗	✗	✗	✗	✗	✗

difficult to characterize. Consequently, creating large-scale, realistic, and diverse deformable-object datasets for training robust policies remains a major challenge, due to two fundamental issues: (i) collecting real-world household data is prohibitively expensive and labor-intensive, as deformable objects' variable states and the inherent complexity of household environments make it difficult to obtain sufficient high-quality data; and (ii) achieving accurate and authentic modeling of deformation is intrinsically hard, since it requires simultaneously capturing complex material properties, nonlinear dynamics, and realistic interactions.

As shown in Fig. 1, we present **LeHome**, a household simulation environment designed for high-fidelity manipulation of diverse deformable objects. LeHome provides a household asset library spanning liquids, gaseous fluids, granular objects, linear objects, thin shells, and volumetric objects. To model these categories, LeHome combines appropriate simulation engines, including Position-Based Dynamics (PBD), Finite Element Method (FEM), and Eulerian fluid simulation, to improve physical realism while preserving broad task coverage. Furthermore, to enable complex manipulation mechanisms, such as a sausage being cut into multiple pieces with a knife. LeHome introduces a graphical logical modeling method called the **Action Graph**, representing causal relationships. This method enforces causal consistency between actions and outcomes, ensuring the simulation aligns with real-world dynamics. In summary, these designs allow LeHome to provide a reliable physical foundation for deformable object manipulation and establish a strong basis for generating high-quality training data.

In terms of supported robot platforms, LeHome spans multiple robotic embodiments, ranging from mainstream commercial manipulators (e.g., UR, Franka) to open-source low-cost platforms (e.g., LeRobot, XLeRobot). Compared with widely-used industrial robots, low-cost robots are more compact, easier to deploy, and cheaper to maintain. These properties make them better suited for scalable deployment in household environments. Therefore, LeHome places particular emphasis on supporting a diverse set of open-source, low-cost robot arms, which are easy to set up, modify, and maintain, while still providing sufficient manipulation capability for everyday household tasks. By lowering the hardware and deployment barriers, this emphasis enables scalable, repeatable validation of core household-robot functionalities in realistic home environments, and in turn supports broader adoption and large-scale deployment.

Finally, to further strengthen LeHome in both scalable data acquisition and user-friendliness, we integrate diverse teleoperation methods, offering multiple devices to fit different demands and allowing anyone to operate it intuitively. Moreover, it is compatible with both virtual and real scenarios, enabling simulation datasets to be seamlessly supplemented with real-world demonstrations under the same pipeline, which in turn supports efficient collection of manually annotated data. Meanwhile, we employ domain randomization to augment existing demonstrations by injecting randomized scene variations during trajectory replay, thereby automatically generating more diverse data with richer visual appearances and interaction conditions. Real-robot evaluations validate that LeHome supports effective sim-to-real transfer for diverse deformable object manipulation tasks, and that domain randomization mitigates distribution shift, resulting in more robust real-world deployment.

II. RELATED WORK

A. Simulation and Manipulation in Household Scenarios

Household scenarios are particularly valuable for robotic simulation because they closely reflect the complexities of real-world settings. Existing simulation platforms have established a solid foundation for related research: many [4], [8], [22], [20], [13] show practical value in building desktop-level or single-object scenes, effectively supporting targeted research. Some platforms [28], [15] further extend scene coverage to more diverse scenarios, yet their visual fidelity remains limited, falling short of near-photorealistic realism. Additionally, studies [16], [29], [23], [9] have actively addressed these limitations, however, high-fidelity interactions with deformable objects (*e.g.*, garments, food) remain challenging. These deformable objects are ubiquitous in daily life, and accurately simulating their manipulation is critical to realistic household scenarios. To address these existing gaps, we introduce the LeHome platform, which integrates both flexibility and scalability. As Tab. I, Lehome incorporates high-fidelity assets across various room types and enables complex interactions with deformable objects (*e.g.*, folding garments and cutting sausages), thereby effectively addressing a key limitation in current simulation platforms.
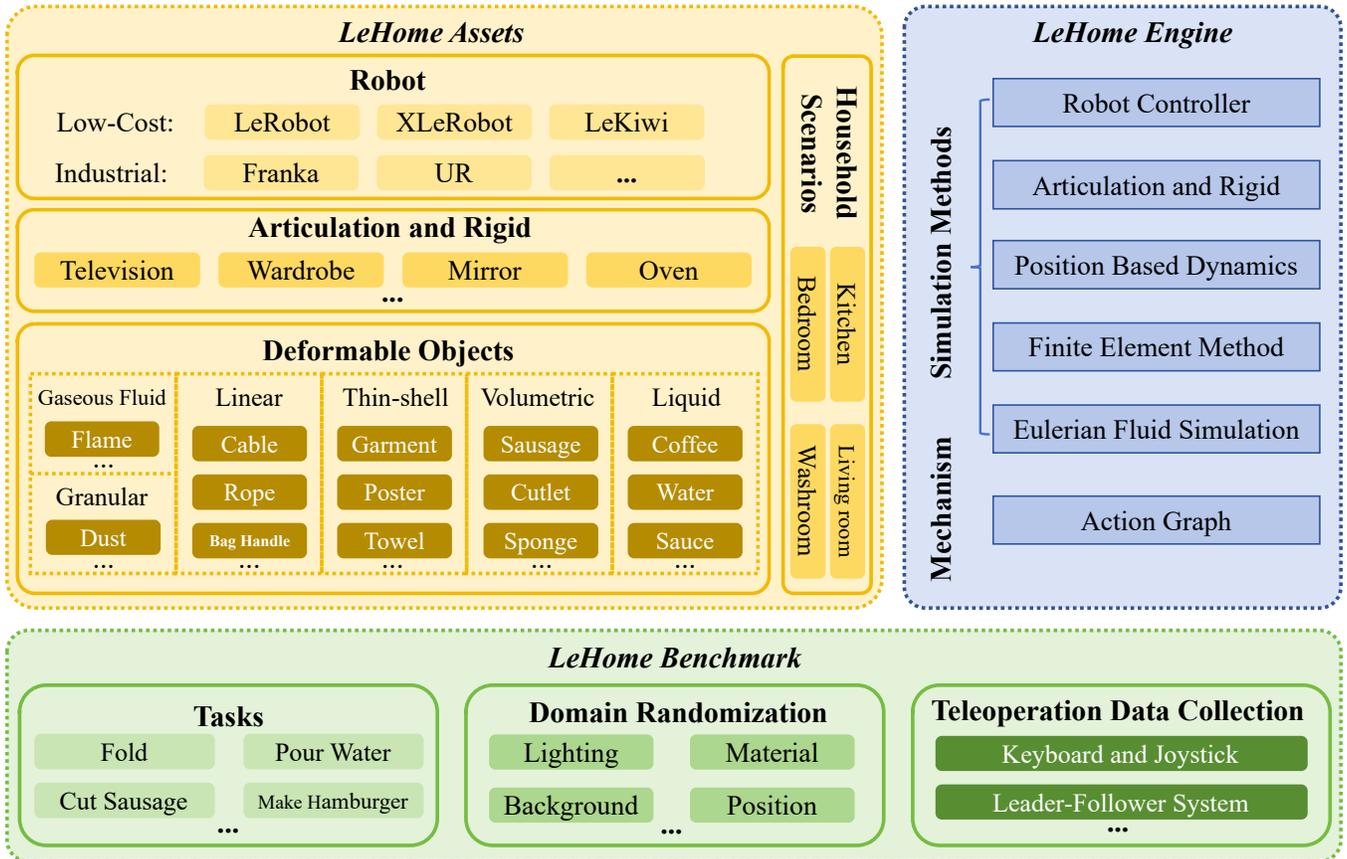
Fig. 2: **The Architecture of LeHome**. (Left) LeHome Assets can deliver realistic simulations of different robots, articulation/rigid objects, deformable objects, and cover multiple household scenarios. (Right) Leveraging diverse simulation methods and mechanisms, LeHome Engine enables various simulation capabilities. (Bottom) Subsequently, LeHome Benchmark utilizes these assets to construct tasks, conduct domain randomization, and facilitate teleoperation data collection.

B. Deformable Object Simulation and Manipulation

Robotic manipulation of deformable objects has recently gained attention, with research on distinct object types.

Some works [38], [21], [35], [10], [33], [19], [30] specialize in the simulation and manipulation of cloth-related tasks. For soft materials, fluids,, and food, dedicated efforts [17], [34], [12], [11], [18], [25] have also focused on their simulation and manipulation. Additionally, studies have addressed the manipulation of linear and thin-shell objects, supported by dedicated platforms [32], [24]. However, most simulators remain confined to a single category of deformable objects and cannot capture multi-modal interactions, *e.g.*, cloth with furniture or sausages with knives. In contrast, LeHome unifies high-quality deformable assets, including cloth, food, and fluids, within furnished household scenarios, enabling systematic study of multi-modal deformable manipulation in realistic environments.

C. Manipulation with Low-Cost Robot

Mainstream industrial robots such as Franka and UR demonstrate strong reliability while remaining costly and bulky, with potential obstacles to broad household deployment. A variety of low-cost and open-source platforms (*e.g.*,

Aloha [37], Mobile Aloha [7], OpenArm [6]) have emerged, providing modular robotic arms and mobile systems for household research. Among these platforms, the LeRobot series (LeRobot [3], Lekiwi [27], XLeRobot [31]) has further reduced costs by integrating insights from preceding products. Moreover, supported by an open-source community, it integrates the implementation of many manipulation algorithms for deployment. Consequently, LeHome chooses to integrate this LeRobot series across diverse household scenarios, thereby offering a potential path for future large-scale deployment of household robots.

III. ENVIRONMENT

A. Overview

LeHome is a household simulation environment designed for manipulation research, with a particular emphasis on deformable objects. As illustrated in Fig. 2, **LeHome** is organized into three tightly connected components. The first is **LeHome Assets**, which provides realistic household scenes, diverse object instances, and multiple robot embodiments as the content foundation of the environment. The second is **LeHome Engine**, which serves as the computational core by integrating physics simulation and interaction mechanisms
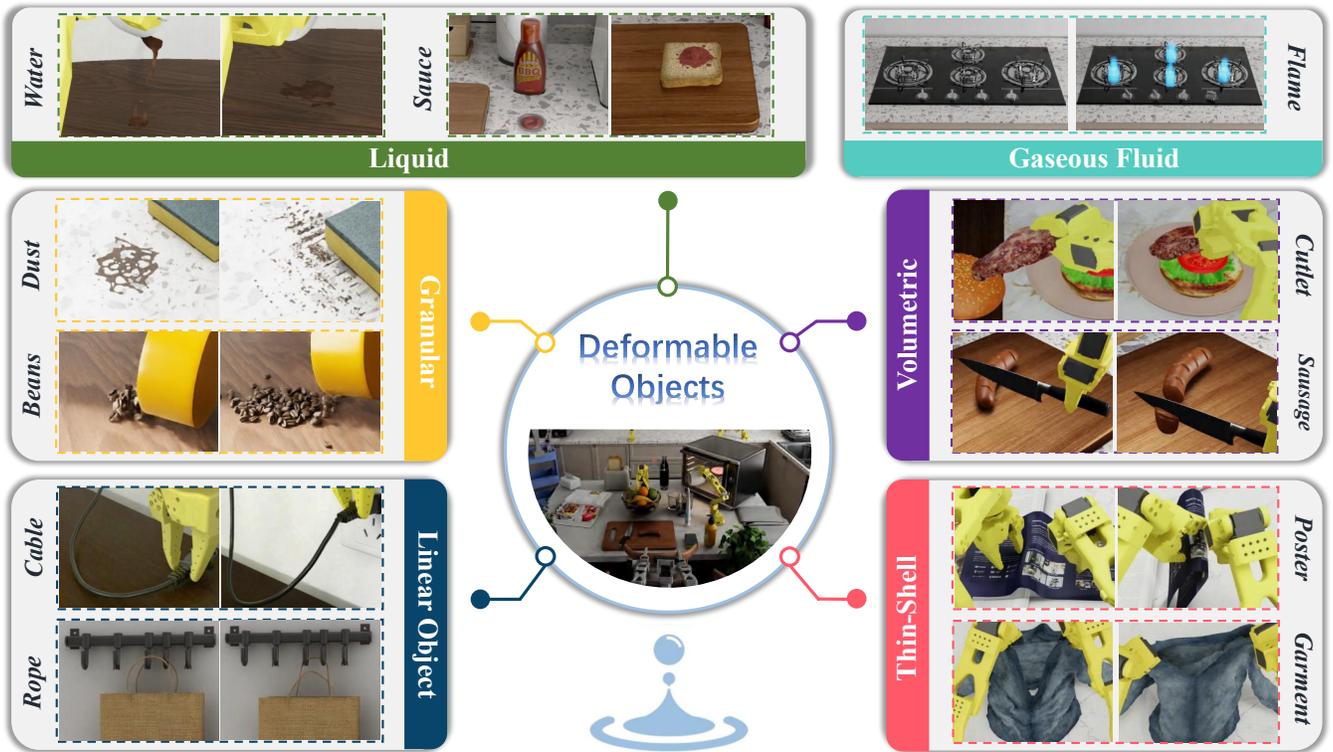
Fig. 3: **Simulated Deformble Objects** cover 6 categories with visually and physically high-fidelity assets for each category.

to model fine-grained manipulation dynamics. The third is **LeHome Benchmark**, which defines representative household tasks and supports teleoperation data collection for training and evaluation. Together, these components form a unified framework that combines realistic scene modeling, controllable simulation, and efficient data acquisition.

B. Assets and Physics

Beyond rigid and articulated objects, LeHome provides diverse deformable assets for household tasks. To model heterogeneous physical behavior in a structured manner, we categorize deformable objects into six classes according to their mechanical characteristics, as shown in Fig. 3:

1) **Liquid:** Materials that flow readily and thus have no fixed shape. While their volume is approximately conserved, they deform under shear and adapt to container shape. Examples: water, juice.
2) **Gaseous Fluid:** Here, fluid specifically refers to the gaseous phase, rather than liquid. It is compressible and has no fixed volume, with density varying with time and pressure. Example: flame.
3) **Granular Object:** Composed of discrete granular aggregates, with mechanical behaviors dominated by inter-particle collision, friction, and stacking. Example: Dust, Beans.
4) **Linear Object:** Characterized by a highly slender geometry. Under outer loads, deformation is dominated by bending. Examples: cable and rope.

5) **Thin Shell:** Characterized by thin planar geometry, deformation is dominated by shell bending and membrane stretching. Example: poster, garment, bag.
6) **Volumetric Object:** Governed by continuum mechanics, these are treated as continuous solids with stress/strain fields. Examples: cutlet, sausage, burger.

To achieve both realism and efficiency, we assign simulation methods to each category based on physical characteristics and task requirements:

1) **Liquid:** All Liquids are simulated using PBD, which efficiently captures liquid motion and interactions under frequent contact.
2) **Gaseous Fluid:** For flame simulation, we leverage Omniverse Flow with a sparse voxel grid representation to update key fields, enabling efficient Eulerian Fluid Simulation with realistic visual appearance.
3) **Granular Object:** Fine granules (e.g., dust) are modeled with PBD for dispersed motion, while coarse objects (e.g., coffee beans) are approximated as rigid bodies for stable contact interactions.
4) **Linear Object:** We support two simulation modes for linear objects: (1) a multi-rigid-body chain model, simple and efficient; and (2) an FEM-based deformable model capturing detailed bending and stretching at higher computational cost.
5) **Thin Shell:** For thin-shell objects with wrinkling (e.g., garments), we use PBD for efficient simulation of stretching and bending. When wrinkling is less critical (e.g., posters), FEM is used for stable elastic responses.
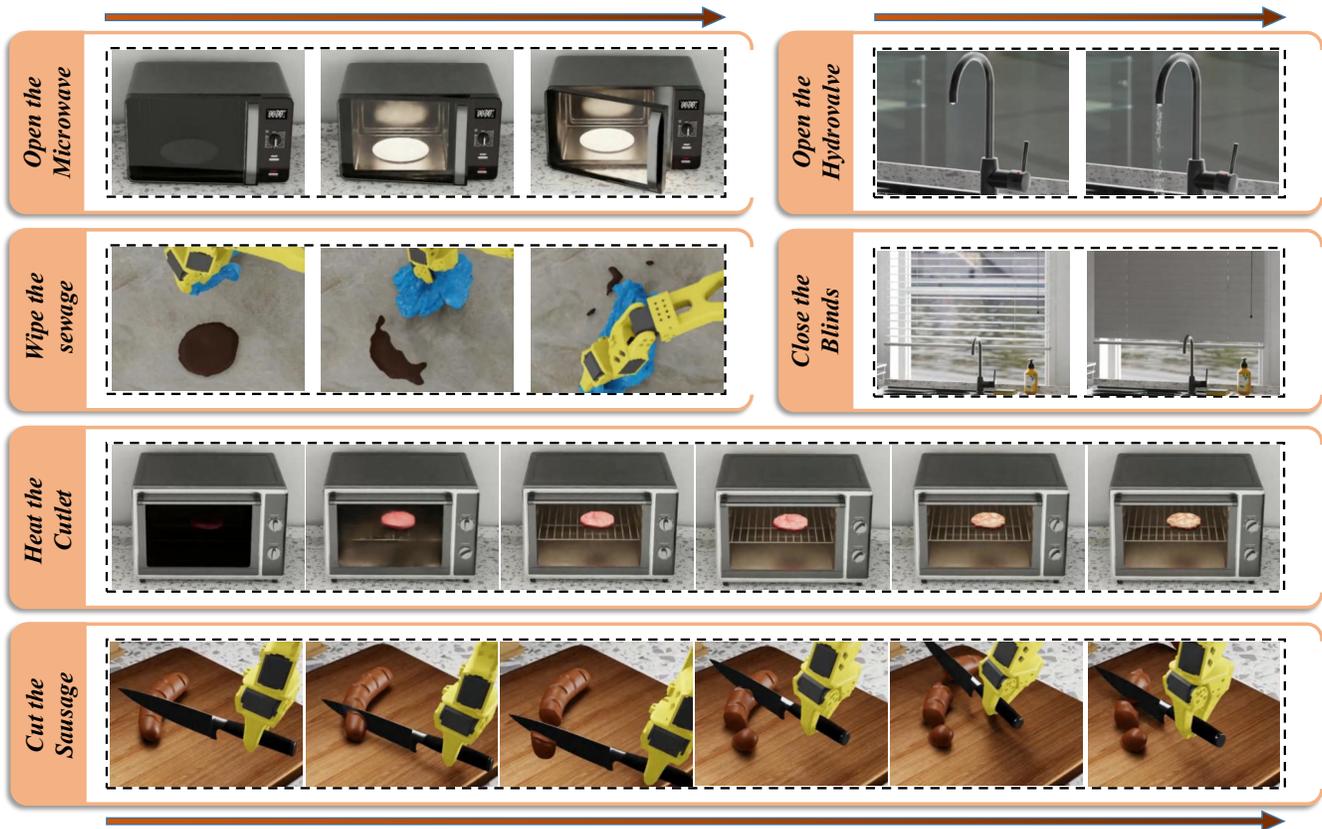
Fig. 4: **Diverse Manipulation Mechanisms.** LeHome models causal relationships of manipulation through the action graph, ensuring the simulation results align with real-world causal relationships and providing high-fidelity interactions.

6) **Volumetric Object:** All volumetric deformable objects (e.g., patty, sausage) are simulated using FEM with volumetric discretization, enabling elastic or elasto-plastic stress-strain modeling.

C. Mechanisms

Previous simulation systems struggle to capture mechanisms induced by physical interactions (e.g., object morphological splitting and state transitions), and suffer from weak coupling between action logic and physical changes, making it hard to replicate real-world cause–effect dynamics. Our LeHome achieves high-fidelity dynamic mechanisms, as shown in Fig. 4, by proposing the **Action Graph**.

Specifically, the core lies in decomposing key simulation links (action triggering, state transition, and physical computation) into controllable modular units, enabling physics-compliant interactive simulation while balancing efficiency and real-time response. Supported by an "event-response" model, Action Graph consists of three key components:

1) **Attributes:** carriers of a name, a data type, or a set of metadata (e.g., a constant or a coordinate).
2) **Nodes:** basic units of the action graph. There are many types based on function, such as on-trigger, computation, and state update nodes.
3) **Connections:** edges in the action graph connecting different nodes, representing the flow of data and logical relationships.

As shown in Fig. 5, the Action Graph achieves simulation of sausage cutting as follows: First, the On Trigger Node captures collision signals between the cutting tool and the sausage. When the cutting tool collides with the sausage, a "cut-trigger" signal is generated. Then, the Computation Node initiates geometric mesh segmentation for the sausage model based on the cutting plane. Finally, after the mesh segmentation by the cutting computation, the State Update Node performs new object creation and updates the physical properties as well as textures of the split halves, realizing the "Split" process.

Compared with previous works, Action Graph provides modularity (node-level reuse), controllability (explicit dependencies), and extensibility (new mechanisms can be added by composing nodes). It is designed to support diverse household interaction patterns while remaining compatible with real-time simulation loops.

D. Robots and Teleoperation

We provide a suite of low-cost LeRobot-family embodiments (LeRobot, LeKiwi and XLeRobot), covering single, bimanual, and mobile-manipulation configurations for household tasks.

As shown in Fig. 6, we provide multiple teleoperation methods for demonstration collection. We support keyboard and joystick inputs that issue joint-space commands, offering a low-cost and widely accessible control modality. We further
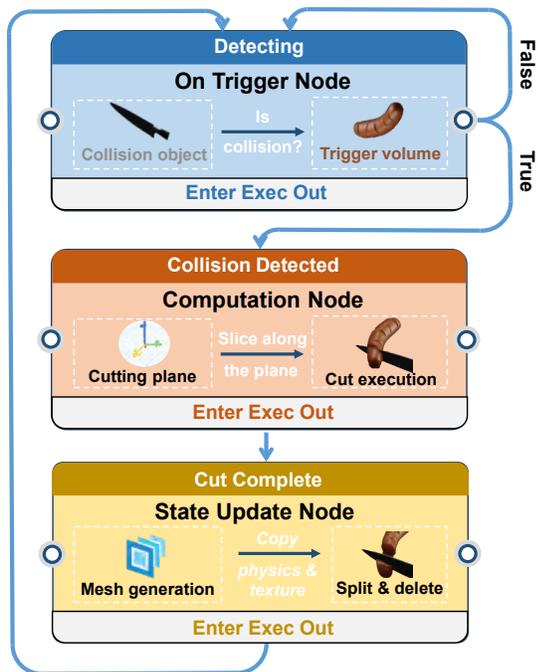
Fig. 5: **Action Graph Workflow** on sausage cutting, consisting of an On-Trigger Node, a Computation Node, and a State Update Node, connected in an ordered way.

integrate a leader-follower system that enables intuitive joint-synchronized teleoperation. For mobile robots, the system supports hybrid control by combining leader-follower arm synchronization with keyboard/joystick commands for base locomotion.

Overall, by emphasizing low-cost, compact form factor, and ease of operation for household settings, this robot lineup lowers the deployment barrier and facilitates the broader adoption of home-robot applications in everyday families.

E. Domain Randomization

To mitigate the sim-to-real gap in diverse household settings, LeHome applies domain randomization (DR) during teleoperation and further augments the dataset via trajectory replay. At the start of each episode, four factors are randomized to generate varied yet feasible configurations: (1) **Initialization position**: objects sampled within workspace bounds while avoiding invalid states; (2) **Lighting**: illumination intensity and color temperature vary; (3) **Background texture**: tabletop textures sampled from a curated set; and (4) **Material property**: visual materials (e.g., cloth/liquid appearance) randomized while preserving physical dynamics.

After data collection, we replay recorded trajectories under randomized appearance factors to further increase visual diversity. Object positions remain fixed to preserve task geometry and contacts. Trajectories are then filtered using task-specific success detectors (e.g., state validation or geometric constraints), retaining only successful ones for training. This process improves robustness with minimal additional demonstration cost.
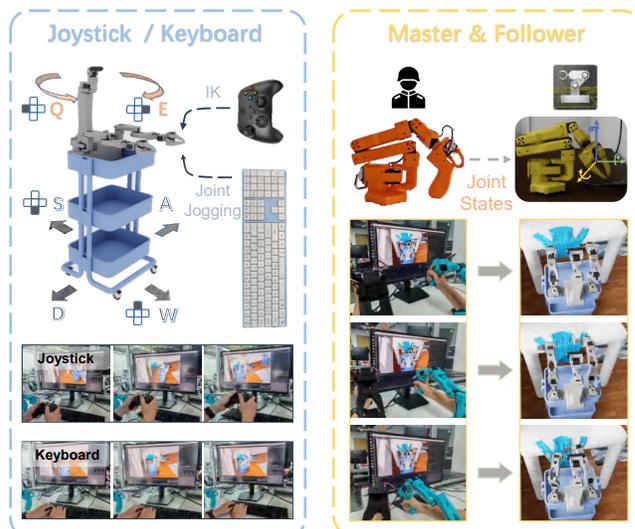

Fig. 6: **Teleoperation Methods.** (Left) We use Joystick and Keyboard to teleoperate XLeRobot, and (Right) Leader-Follower Teleoperation for LeRobot.

IV. EXPERIMENT

With the proposed LeHome environment, we conduct experiments to demonstrate that LeHome offers high-fidelity simulation (especially for deformable objects) to support policy learning across diverse household tasks, and that training in LeHome facilitates effective sim-to-real transfer and improves real-world performance.

A. Tasks and Settings

To demonstrate the capability to support different policy learning algorithms on diverse daily household tasks, we selected six representative tasks from LeHome, covering single-arm and bimanual manipulation, tool use, deformable food interaction, fluids, and rigid object manipulation, rigid-deformable and deformable-deformable interactions. The tasks span multiple rooms (bedroom, kitchen, living room, bathroom) to reflect realistic household diversity.

1) (bedroom) **Fold Garment** 7a requires the dual-arm robot to neatly fold a piece of clothing placed on the table.
2) (bedroom) **Fling Garment** 7b requires the robot to grasp and fling wrinkled clothes to fully unfold and flatten them on the table surface.
3) (kitchen) **Assemble Burger** 7c requires the robot to stack the patty on top of the cheese and bread slices to complete a burger assembly.
4) (kitchen) **Cut Sausage** 7d requires the robot to use a knife to cut the sausage into two separate pieces, demonstrating the cutting mechanism in simulation.
5) (living room) **Pour Coffee** 7e requires the robot to control the cup orientation to pour coffee into a bowl without spilling.
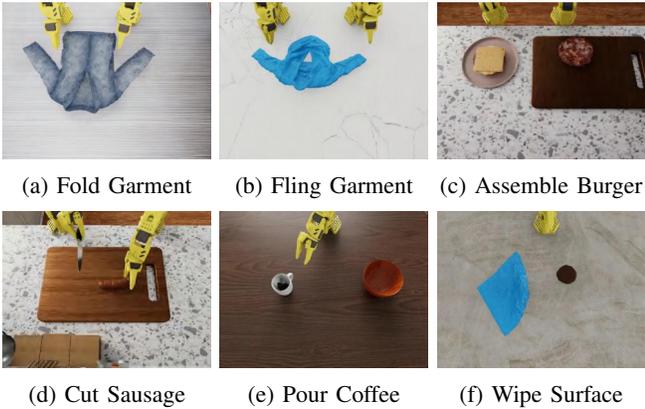6) (bathroom) **Wipe Surface** 7f requires the robot to use a cloth to wipe water off the floor surface.

(a) Fold Garment   (b) Fling Garment   (c) Assemble Burger

(d) Cut Sausage   (e) Pour Coffee   (f) Wipe Surface
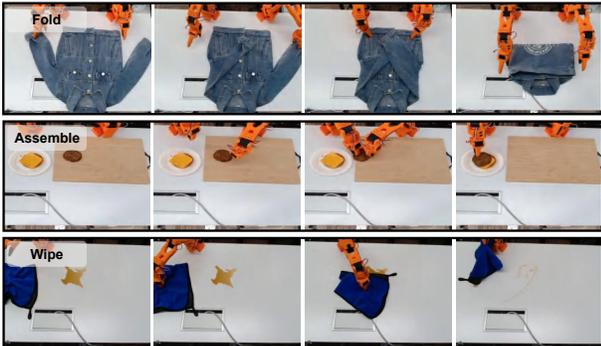
Fig. 7: Gallery of **Evaluated Tasks**.

Fig. 8: **Real-world execution on diverse household tasks.**

B. Baselines

We compare four representative policies under a unified imitation-learning setup. To analyze the effect of language conditioning, we include two RGB-only imitation baselines, **Diffusion Policy (DP)** [5] and **ACT** [37], and two language-conditioned VLA policies, **Pi0** [2] and **SmolVLA** [26].

1) **Diffusion Policy (DP)** [5], a generative policy leveraging diffusion models to propose action trajectories.
2) **ACT** [37], transformer-based policy architecture that predicts sequences of action chunks.
3) **Pi0** [2], a flow-matching VLA policy leveraging a pretrained vision-language backbone to generate robot action sequences.
4) **SmolVLA** [26], a lightweight VLA developed by the Hugging Face team.

C. Simulation Experiments

Table II reports simulation success rates on six tasks. For fair comparison, all policies are trained with the same data budget (50 teleoperated demonstrations per task) and evaluated over 100 test trials.

The results show clear task-dependent strengths: SmolVLA performs best on garment manipulation tasks, DP performs best on Cut Sausage, and ACT is competitive on Assemble Burger and Pour Coffee. Notably, Fling Garment is consistently difficult for all methods, with success rates remaining low (ACT: 25.0%, DP: 20.0%,

TABLE II: **Results of Simulation Evaluation.**

Task	ACT	DP	SmolVLA	Pi0
Fold Garment	45.0%	30.0%	70.0%	44.0%
Assemble Burger	78.0%	35.0%	40.0%	39.0%
Fling Garment	25.0%	20.0%	36.0%	14.0%
Cut Sausage	77.0%	93.0%	90.0%	75.0%
Pour Coffee	80.0%	30.0%	40.0%	40.0%
Wipe Surface	60.0%	30.0%	60.0%	60.0%

TABLE III: **Results of Real-World Experiment.**

Task	ACT		SmolVLA	
	Real	Co-Training	Real	Co-Training
Fold Garment	2/10	**5/10**	2/10	**4/10**
Assemble Burger	2/10	**4/10**	1/10	**4/10**
Wipe Surface	1/10	**7/10**	1/10	**6/10**

SmolVLA: 36.0%, Pi0: 14.0%). This indicates that large-deformation, long-horizon garment manipulation is still a major challenge. Overall, these results suggest that LeHome provides a competitive evaluation setting that can effectively distinguish the performance of different policies.

D. Real-World Experiments

To evaluate the visual and physical fidelity of LeHome and its contribution to real-world manipulation, we conducted experiments with two LeRobots to assess whether incorporating simulation data can improve policy performance.

Using dual-arm LeRobot, we compare two training settings: (1) **Real**, where policies are trained using 10 real-world demonstrations only; and (2) Sim+Real **Co-Training**, where policies are trained with both LeHome simulation demonstrations and the same 10 real-world demonstrations. We evaluate ACT and SmolVLA under both settings, and report real-world results in Table III, with qualitative examples shown in Fig. 8. Across all evaluated tasks, **Sim+Real Co-Training** consistently outperforms **Real-Only**. On average, the success rate improves from about 15% to about 50%, suggesting that LeHome pretraining improves data efficiency and real-world robustness in low-data settings.

V. CONCLUSIONS

In this paper, we present LeHome, a comprehensive household simulation environment and benchmark for manipulating deformable objects with everyday items. LeHome provides a diverse suite of physically grounded deformable assets embedded within complete household scenarios, enabling realistic and reproducible experimentation. To expand the application of embodied intelligence to price-sensitive households, we leverage the low-cost LeRobot platform to execute everyday tasks and establish an end-to-end tool chain for training and validating manipulation algorithms.

REFERENCES

[1] Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025.

[2] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. pi_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

[3] Remi Cadene, Simon Alibert, Alexander Soare, Quentin Gallouedec, Adil Zouitine, Steven Palma, Pepijn Kooijmans, Michel Aractingi, Mustafa Shukor, Dana Aubakirova, Martino Russi, Francesco Capuano, Caroline Pascal, Jade Choghari, Jess Moss, and Thomas Wolf. Lerobot: State-of-the-art machine learning for real-world robotics in pytorch. https://github.com/huggingface/lerobot, 2024.

[4] Tianxing Chen, Zanxin Chen, Baijun Chen, Zijian Cai, Yibin Liu, Qiwei Liang, Zixuan Li, Xianliang Lin, Yiheng Ge, Zhenyu Gu, et al. Robotwin 2.0: A scalable data generator and benchmark with strong domain randomization for robust bimanual robotic manipulation. *arXiv preprint arXiv:2506.18088*, 2025.

[5] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, page 02783649241273668, 2023.

[6] enactic. OpenArm. https://github.com/enactic/openarm, 2025.

[7] Zipeng Fu, Tony Z Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

[8] Haoran Geng, Feishi Wang, Songlin Wei, Yuyang Li, Bangjun Wang, Boshi An, Charlie Tianyue Cheng, Haozhe Lou, Peihao Li, Yen-Jen Wang, et al. Roboverse: Towards a unified platform, dataset and benchmark for scalable and generalizable robot learning. *arXiv preprint arXiv:2504.18904*, 2025.

[9] Ran Gong, Jiangyong Huang, Yizhou Zhao, Haoran Geng, Xiaofeng Gao, Qingyang Wu, Wensi Ai, Ziheng Zhou, Demetri Terzopoulos, Song-Chun Zhu, et al. Arnold: A benchmark for language-grounded task learning with continuous states in realistic 3d scenes. In *ICCV*, 2023.

[10] Huy Ha and Shuran Song. Flingbot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding. In *CoRL*, 2022.

[11] Eric Heiden, Miles Macklin, Yashraj Narang, Dieter Fox, Animesh Garg, and Fabio Ramos. Disect: A differentiable simulation engine for autonomous robotic cutting. *arXiv preprint arXiv:2105.12244*, 2021.

[12] Zhiao Huang, Yuanming Hu, Tao Du, Siyuan Zhou, Hao Su, Joshua B Tenenbaum, and Chuang Gan. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. *arXiv preprint arXiv:2104.03311*, 2021.

[13] Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J Davison. Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026, 2020.

[14] Yunfan Jiang, Ruohan Zhang, Josiah Wong, Chen Wang, Yanjie Ze, Hang Yin, Cem Gokmen, Shuran Song, Jiajun Wu, and Li Fei-Fei. Behavior robot suite: Streamlining real-world whole-body manipulation for everyday household activities. *arXiv preprint arXiv:2503.05652*, 2025.

[15] Chengshu Li, Fei Xia, Roberto Martín-Martín, Michael Lingelbach, Sanjana Srivastava, Bokui Shen, Kent Vainio, Cem Gokmen, Gokul Dharan, Tanish Jain, et al. igibson 2.0: Object-centric simulation for robot learning of everyday household tasks. *arXiv preprint arXiv:2108.03272*, 2021.

[16] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Michael Lingelbach, Jiankai Sun, et al. Behavior-1k: A benchmark for embodied ai with 1,000 everyday activities and realistic simulation. In *Conference on Robot Learning*, pages 80–93. PMLR, 2023.

[17] Sizhe Li, Zhiao Huang, Tao Du, Hao Su, Joshua B Tenenbaum, and Chuang Gan. Contact points discovery for soft-body manipulations with differentiable physics. *arXiv preprint arXiv:2205.02835*, 2022.

[18] Xingyu Lin, Zhiao Huang, Yunzhu Li, Joshua B Tenenbaum, David Held, and Chuang Gan. Diffskill: Skill abstraction from differentiable physics for deformable object manipulations with tools. *arXiv preprint arXiv:2203.17275*, 2022.

[19] Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. In *Conference on Robot Learning*, 2020.

[20] Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. *NeurIPS*, 2023.

[21] Haoran Lu, Ruihai Wu, Yitong Li, Sijie Li, Ziyu Zhu, Chuanruo Ning, Yan Zhao, Longzan Luo, Yuanpei Chen, and Hao Dong. Garmentlab: A unified simulation and benchmark for garment manipulation. *NeurIPS*, 2024.

[22] Yao Mu, Tianxing Chen, Zanxin Chen, Shijia Peng, Zhiqian Lan, Zeyu Gao, Zhixuan Liang, Qiaojun Yu, Yude Zou, Mingkun Xu, Lunkai Lin, Zhiqiang Xie, Mingyu Ding, and Ping Luo. Robotwin: Dual-arm robot benchmark with generative digital twins. In *CVPR*, 2025.

[23] Soroush Nasiriany, Abhiram Maddukuri, Lance Zhang, Adeet Parikh, Aaron Lo, Abhishek Joshi, Ajay Mandlekar, and Yuke Zhu. Robocasa: Large-scale simulation of everyday tasks for generalist robots. *arXiv preprint arXiv:2406.02523*, 2024.

[24] Daniel Seita, Pete Florence, Jonathan Tompson, Erwin Coumans, Vikas Sindhwani, Ken Goldberg, and Andy Zeng. Learning to Rearrange Deformable Cables, Fabrics, and Bags with Goal-Conditioned Transporter Networks. In *ICRA*, 2021.

[25] Haochen Shi, Huazhe Xu, Samuel Clarke, Yunzhu Li, and Jiajun Wu. Robocook: Long-horizon elasto-plastic object manipulation with diverse tools. *arXiv preprint arXiv:2306.14447*, 2023.

[26] Mustafa Shukor, Dana Aubakirova, Francesco Capuano, Pepijn Kooijmans, Steven Palma, Adil Zouitine, Michel Aractingi, Caroline Pascal, Martino Russi, Andres Marafioti, et al. Smolvla: A vision-language-action model for affordable and efficient robotics. *arXiv preprint arXiv:2506.01844*, 2025.

[27] SIGRobotics-UIUC. Lekiwi. https://github.com/SIGRobotics-UIUC/LeKiwi, 2025.

[28] Andrew Szot, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, Oleksandr Maksymets, et al. Habitat 2.0: Training home assistants to rearrange their habitat. *NeurIPS*, 2021.

[29] Stone Tao, Fanbo Xiang, Arth Shukla, Yuzhe Qin, Xander Hinrichsen, Xiaodi Yuan, Chen Bao, Xinsong Lin, Yulin Liu, Tse-kai Chan, et al. Maniskill3: Gpu parallelized robotics simulation and rendering for generalizable embodied ai. *arXiv preprint arXiv:2410.00425*, 2024.

[30] Chenrui Tie, Yue Chen, Ruihai Wu, Boxuan Dong, Zeyi Li, Chongkai Gao, and Hao Dong. Et-seed: Efficient trajectory-level se(3) equivariant diffusion policy, 2025.

[31] Gaotian Wang and Zhuoyi Lu. Xlerobot: A practical low-cost household dual-arm mobile robot design for general manipulation. https://github.com/Vector-Wangel/XLeRobot, 2025.

[32] Yian Wang, Juntian Zheng, Zhehuan Chen, Zhou Xian, Gu Zhang, Chao Liu, and Chuang Gan. Thin-shell object manipulations with differentiable physics simulations. In *ICLR*, 2023.

[33] Yuran Wang, Ruihai Wu, Yue Chen, Jiarui Wang, Jiaqi Liang, Ziyu Zhu, Haoran Geng, Jitendra Malik, Pieter Abbeel, and Hao Dong. Dexgarmentlab: Dexterous garment manipulation environment with generalizable policy. *arXiv preprint arXiv:2505.11032*, 2025.

[34] Zhou Xian, Bo Zhu, Zhenjia Xu, Hsiao-Yu Tung, Antonio Torralba, Katerina Fragkiadaki, and Chuang Gan. Fluidlab: A differentiable environment for benchmarking complex fluid manipulation. *arXiv preprint arXiv:2303.02346*, 2023.

[35] Han Xue, Yutong Li, Wenqiang Xu, Huanyu Li, Dongzhe Zheng, and Cewu Lu. Unifolding: Towards sample-efficient, scalable, and generalizable robotic garment folding. *CoRL*, 2023.

[36] Sriram Yenamandra, Arun Ramachandran, Karmesh Yadav, Austin Wang, Mukul Khanna, Theophile Gervet, Tsung-Yen Yang, Vidhi Jain, Alexander William Clegg, John Turner, et al. Homerobot: Open-vocabulary mobile manipulation. *CoRL*, 2023.

[37] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[38] Bingyang Zhou, Haoyu Zhou, Tianhai Liang, Qiaojun Yu, Siheng Zhao, Yuwei Zeng, Jun Lv, Siyuan Luo, Qiancai Wang, Xinyuan Yu, Haonan Chen, Cewu Lu, and Lin Shao. Clothesnet: An information-rich 3d garment model repository with simulated clothes environment, 2023.