UniGarment: A Unified Simulation and Benchmark for Garment Manipulation

Haoran Lu*[2,1] Yitong Li*[3,1] Ruihai Wu*[1,4] Chuanruo Ning [2,1] Yan Shen [1,4] Hao Dong [1,4]
[1]CFCS, School of CS, PKU [2]School of EECS, PKU [3]Weiyang College, THU
[4]National Key Laboratory for Multimedia Information Processing, School of CS, PKU

***Abstract*—Manipulating garments and fabrics has long been a critical endeavor in the advancement of home-assistant robots, serving as a focal point for researchers in the fields of vision and robotics. However, due to complex dynamics and topological structures, garment manipulations pose significant challenges. Recent successes in reinforcement learning or vision-based methods offer promising avenues for learning garment manipulation. Nevertheless, these approaches are severely constrained by current benchmarks, which exhibit unrealistic simulation behavior and offer only a limited number of tasks. So we present UniGarment, a benchmark designed for deformable object and garment manipulation within realistic 3D indoor scenes. Our benchmark encompasses a diverse range of garment types, robotic systems and manipulators including dexterous hands. The multitude of tasks included in the benchmark enables further exploration of the interactions between garments, deformable objects, rigid bodies, fluids, and avatars. Furthermore, by incorporating multiple simulation methods, including FEM and PBD, along with our proposed sim-to-real techniques, we aim to significantly narrow the sim-to-real gap. We evaluate state-of-the-art vision methods, reinforcement learning (RL), and imitation learning (IL) techniques on these tasks, highlighting the challenges faced by current algorithms, notably their limited generalization capabilities. Our comprehensive analysis and provision of open-source environments lay the foundation for future research in garment manipulation, unlocking the full potential of these methodologies.**

***Index Terms*—Garment Manipulation, Simulation, Benchmark**

I. Introduction

The next-generation indoor assistant robots should possess not only the abilities to directly manipulate a wide variety of objects, including rigid, articulated [1], and deformable objects [2], but also the capability to leverage interactions among those physical media, including flow and fluids, in order to assist humans [3]. For instance, washing clothes entails the interaction between garments and fluids, while dressing up requires collaboration between robots and humans. Among all the tasks proposed in previous work [1], [4], [5], garment manipulation stands out as one of the most challenging yet crucial and extensively discussed tasks in the robotics and computer vision community due to its scientific and practical significance.

Garment Manipulation task presents three following challenges. First, each individual garment, characterized by its unique topological structure and inherent flexibility, possesses an extensive range of self-deformation states and exhibits complex kinematic and dynamic properties. *Therefore, it is crucial for models to comprehend the various forms of garments.***(C1)**. Secondly, apart from different garments necessitating diverse simulation techniques, they also interact with various types of objects, ranging from rigid (e.g., clothes hanger) to articulated (e.g., wardrobe), as well as fluids and people. *Consequently, enabling models to understand these physical properties and interactions presents significant challenges.***(C2)** Finally, considering that strategies for manipulating garments are often highly complex, and visual input of garments is more challenging due to their diverse patterns and materials, *manipulating garments faces a greater sim2real gap.***(C3)** [6], [7].

Fig. 1. **UniGarment Overview** We propose 20 novel tasks in UniGarment Benchmarks to make further exploration in physical interaction between objects and evaluate state-of-the-art deformable and garment manipulation algorithms.

Training such a powerful agent requires a vast amount of data encompassing interactions between robots and objects, making it impractical to directly collect data from the real world. Researchers therefore, have long been pursuing benchmarks for garment manipulation tasks [8]–[11]. However, current deformable simulations suffer from various drawbacks, such as missing garment meshes [8] or lacking support for physics engines [11]. Additionally, they offer a very limited range of tasks, which discourages further research endeavors.

Therefore, we present **UniGarment** (Fig:1) , a unified simulation and benchmark for garment manipulation. **UniGarment** have three novel components to address the demands for diversity and realism: The powerful **UniGarment Engine**, which built on the top of Omniverse Isaac Sim [12], sup-

* Equal Contribution.

ports variety of physical simulation method. The simulator not only supports Particle-Base-Dynamic(PBD) [13], Finite-Element-Method(FEM) [14], to simulate garments, fluid and deformable objects but also makes integration with ROS [15] to provide an efficient teleoperation pipeline for data collection. **UniGarment Assets** is a large-scale indoor dataset comprising 1) garments models covering 11 categories of daily garments from ClothesNet [16] 2) various kinds of robot end-effector including gripper, suction and dexterous hands. 3) high-quality 3D assets including 20 scenes and 9000+ object models from ShapeNet [17].With all these realistic simulation features and rich assets, we propose **UniGarment Task** containing 20 tasks divided into 5 groups to evaluate state-of-art vision-based and reinforcement learning base algorithm.

As a platform designed to address the above challenges for researchers, our simulator has four following characteristics: 1) **Rich**. The richness of our simulator can be categorized into two aspects: the diversity of **simulation content** offered by UniGarment Assets and the depth of **physical interaction** facilitated by UniGarment Engine. It is noteworthy that we specifically emphasize exploration in multi-physics simulation, encompassing rigid-articulated, deformable-garment, fluid dynamics, and flow, along with their interactions. This focus is vital for training agents capable of comprehending real-world physical properties. [18](addressing **C2**) 2) **Real** As the sim-to-real gap emerges as the main obstacle in developing embodied agents, UniGarment Engine surpasses Omniverse capabilities by providing mature sim-to-real algorithms, such as ADR [19], predominantly utilized in the RL field, and the Visual Sim-Real Alignment Algorithm, primarily employed in perception algorithms.(addressing **C3**) 3) **Efficient** Given the highly large and nearly infinite state and action spaces for garment manipulation, a substantial volume of data forms the foundation for models to understand the structure and deformation of garments. As the result, our GPU-based simulator is highly parallelized, which show particularly advantageous during the training process. Larger batch sizes can significantly enhance RL-based algorithms [20], while high data collection speeds can reduce the training time of perception-based algorithms.(addressing **C1**) 4) **Multifunctional** With the rise of algorithms such as imitation learning, the field demands increasing diversity in simulator functionality. Our simulator support ROS [15] for teleoperation and MoveIt for motion planning.This can also narrow the sim2real gap [6](addressing **C3**)

Our benchmark experiments indicate that even a seemingly simple task, such as unfolding in the UniGarment Task, poses significant challenges for current algorithms. Specifically, these difficulties stem from a lack of comprehension of physical interactions and high-dimensional states, particularly evident in complex deformable manipulation scenarios. Additionally, we highlight that current vision-based algorithms exhibit limited generalization capabilities, with their performance significantly impacted by the initial state of objects. Moreover, RL-based algorithms demonstrate poor performance on tasks requiring long-horizon planning. These analyses have the potential to guide the development of improved methods for garment and deformable object manipulation.

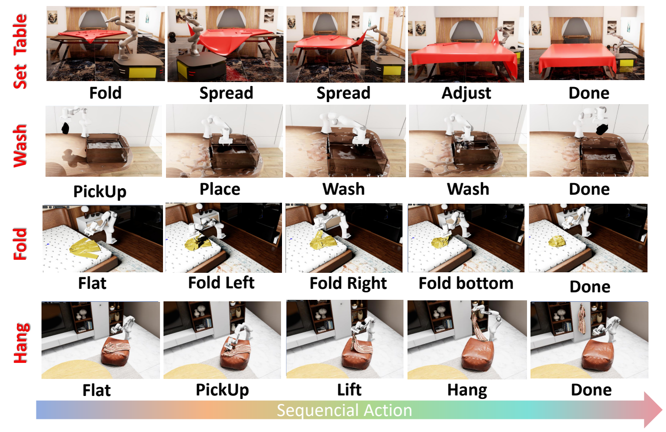

Fig. 2. **UniGarment Tasks Sequence** The picture provides 4 demo of our tasks. Our task types are very diverse, ranging from mobile tasks, dexterous tasks to tasks that require manipulating multiple physical mediums.

In summary, **UniGarment** makes the following contributions:

- A **realistic indoor interactive environment** for garment manipulation featuring **diverse scenes, a plethora of objects, garments, and avatars**, combined with mature **Sim2Real Algorithms**, facilitating the learning and evaluation of garment and deformable manipulation.
- The first benchmark concentrates on the **physical materials and interactions** of garments, deformable objects, fluids, flows, and rigid objects, laying the groundwork for training an agent capable of comprehending real-world physical object behaviors.
- **Extensive experiments and detailed analyses** of state-of-the-art deformable and manipulation algorithms, revealing their strengths and weaknesses in promoting future research on multi-material and multi-physics manipulation task.

II. UniGarment Benchmark

UniGarment aims to integrate state-of-the-art physical simulation methods, high-quality and diverse simulation assets and novel-proposed, rich-interaction tasks into a unified framework. As shown in Fig:3, we will introduce the three main components in detail: UniGarment Engine, UniGarment Assets and UniGarment Tasks.

A. *UniGarment Engine*

The UniGarment Engine supports simulation for various objects, including flow and fluid dynamics. It also offers user-friendly sim2real methods and ROS integration, ideal for end-to-end robotic research like teleoperation and motion planning.

To ensure physically realistic simulation, we employ various methods tailored to different objects. Firstly, Particle-Based Dynamics (PBD) [13] is utilized for simulating large pieces of garments (such as tops, dresses, trousers, and skirts), as well as fluid dynamics. Secondly, for smaller elastic garments

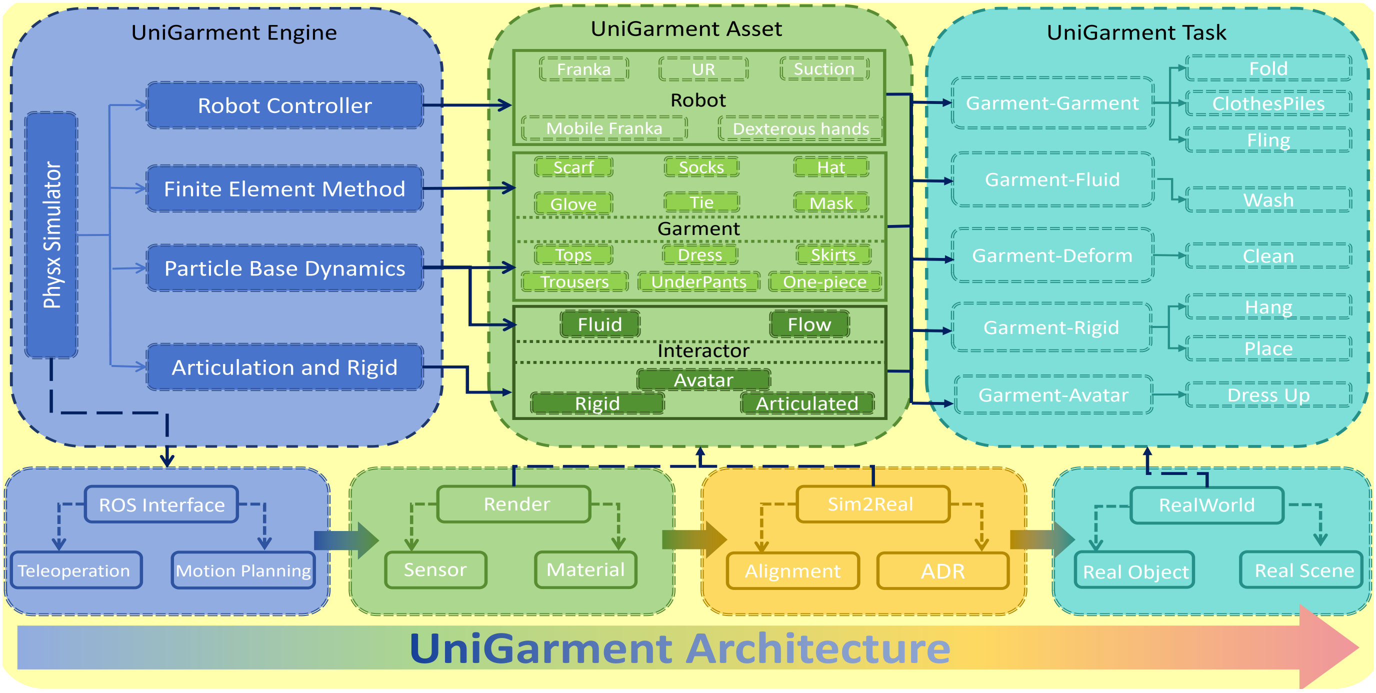

Fig. 3. **The Architecture of UniGarment (Left)** Based on PhysX5, our simulation supports a variety of physical materials. **(Middle)** Our simulator can deliver physically realistic simulations of robots, garments, and interactions between fluids, flows, and avatars. **(Right)** Subsequently, we can utilize these assets to construct tasks across various categories. **Bottom** The diagram illustrates our sim-to-real deployment pipeline. By leveraging ROS, we can efficiently collect data from the real world. Then, using photo-realistic rendering techniques along with our proposed sim-to-real algorithm, we can dramatically reduce the sim-to-real gap.

like gloves and socks, as well as common daily life objects such as toys or sponges, the Finite Element Method (FEM) [14] is applied. Thirdly, Human simulation involves articulated skeletons with bones connected via rotational joints, coupled with a surface skin mesh for high-fidelity rendering. Robot simulation employs the PhysX articulation system, specifically designed for robot simulation, supporting precise force control, P-D control, and inverse dynamics. **It is noteworthy that different physical material parameters are assigned to objects for simulation, including but not limited to surface tension and cohesion for fluids, stiffness and particle contact for garments, and modulus for deformable objects.**

B. UniGarment Assets

The UniGarment Asset comprises simulation content consisting of meshes or URDF files compatible with various simulation methods. Below are the main components of the UniGarment Asset.

- **Garment and Cloth** We select garments from ClothesNet [16], a large-scale dataset of 3D clothes objects with information-rich annotations. Our selection encompasses 11 categories, such as hats, ties, masks, gloves, and socks.
- **Robot** We incorporate a diverse range of robots into our system, including a Franka, a UR5 with suction capabilities, a RidgebackFranka featuring wheels on its base, and a UR10e with ShadowHands mounted on it.
- **Interactor** We also import avatars and articulated objects to create long-horizon tasks for garment manipulation. Since our primary focus lies in interactions with various physical media, we also introduce fluid dynamics and flow simulations into the simulator.

C. UniGarment Task

To fully exploit the model's capability in understanding physical interactions and conduct comprehensive evaluations of current algorithms, we categorize 20 tasks into 5 groups. Examples of task sequences are provided in Fig. 2.

- **Garment-Garment** This category focuses on fundamental garment manipulation, including tasks like folding and unfolding single garments, as well as interactions between multiple garments such as retrieving items from clothes piles.
- **Garment-Fluid** Tasks in this group concentrate on the interaction between garments and fluids, where trajectory dynamics play a crucial role.
- **Garment-Deform** Exploration of tasks involving deformable interactions, such as using a sponge to clean dirt off clothes or packing hats and tops together, is ongoing.
- **Garment-Rigid** Common interactions between clothing and rigid bodies, such as hanging clothes or putting them into a washing machine, require precise grasp point selection and trajectory planning.
- **Garment-Avatar** Dressing tasks pose the greatest challenge, as they require understanding human intention and safe collaboration with humans.

References

[1] Ruihai Wu, Yan Zhao, Kaichun Mo, Zizheng Guo, Yian Wang, Tianhao Wu, Qingnan Fan, Xuelin Chen, Leonidas Guibas, and Hao Dong. VAT-mart: Learning visual action trajectory proposals for manipulating 3d ARTiculated objects. In *International Conference on Learning Representations*, 2022.

[2] Ruihai Wu, Chuanruo Ning, and Hao Dong. Learning foresightful dense visual affordance for deformable object manipulation. In *IEEE International Conference on Computer Vision (ICCV)*, 2023.

[3] Xavi Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta Desai, Alexander William Clegg, Michal Hlavac, So Yeon Min, Vladimir Vondruvs, Theophile Gervet, Vincent-Pierre Berges, John Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Devendra Jitendra Malik, Singh Chaplot, Unnat Jain, Dhruv Batra, † AksharaRai, and † RoozbehMottaghi. Habitat 3.0: A co-habitat for humans, avatars and robots. *ArXiv*, abs/2310.13724, 2023.

[4] Yan Zhao, Ruihai Wu, Zhehuan Chen, Yourong Zhang, Qingnan Fan, Kaichun Mo, and Hao Dong. Dualafford: Learning collaborative visual affordance for dual-gripper object manipulation. *International Conference on Learning Representations (ICLR)*, 2023.

[5] Kaichun Mo, Leonidas J Guibas, Mustafa Mukadam, Abhinav Gupta, and Shubham Tulsiani. Where2act: From pixels to actions for articulated 3d objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6813–6823, 2021.

[6] Han Xue, Yutong Li, Wenqiang Xu, Huanyu Li, Dongzhe Zheng, and Cewu Lu. Unifolding: Towards sample-efficient, scalable, and generalizable robotic garment folding. *ArXiv*, abs/2311.01267, 2023.

[7] Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B Tenenbaum, and Antonio Torralba. Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. In *ICLR*, 2019.

[8] Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. In *Conference on Robot Learning*, pages 432–448. PMLR, 2021.

[9] Hugo Bertiche, Meysam Madadi, and Sergio Escalera. Cloth3d: clothed 3d humans. In *European Conference on Computer Vision*, pages 344–359. Springer, 2020.

[10] Alper Canberk, Cheng Chi, Huy Ha, Benjamin Burchfiel, Eric Cousineau, Siyuan Feng, and Shuran Song. Cloth funnels: Canonicalized-alignment for multi-purpose garment manipulation. In *International Conference of Robotics and Automation (ICRA)*, 2022.

[11] Han Xue, Wenqiang Xu, Jieyi Zhang, Tutian Tang, Yutong Li, Wenxin Du, Ruolin Ye, and Cewu Lu. Garmenttracking: Category-level garment pose tracking. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21233–21242, 2023.

[12] Zhehua Zhou, Jiayang Song, Xuan Xie, Zhan Shu, Lei Ma, Dikai Liu, Jianxiong Yin, and Simon See. Towards building ai-cps with nvidia isaac sim: An industrial benchmark and case study for robotics manipulation. *arXiv preprint arXiv:2308.00055*, 2023.

[13] Jan Bender, Matthias Müller, and Miles Macklin. Position-based simulation methods in computer graphics. In *Eurographics*, 2015.

[14] Philippe G. Ciarlet. The finite element method for elliptic problems. In *Classics in applied mathematics*, 2002.

[15] Morgan Quigley. Ros: an open-source robot operating system. In *IEEE International Conference on Robotics and Automation*, 2009.

[16] 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. *ICCV*, 2023.

[17] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, L. Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *ArXiv*, abs/1512.03012, 2015.

[18] 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.

[19] Zhenjia Xu, Cheng Chi, Benjamin Burchfiel, Eric Cousineau, Siyuan Feng, and Shuran Song. Dextairity: Deformable manipulation can be a breeze. *RSS*, 2022.

[20] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu based physics simulation for robot learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.