DexGarmentLab: Dexterous Garment Manipulation Environment with Generalizable Policy

Yuran Wang[1*] Ruihai Wu[1*] Yue Chen[1*]
Jiarui Wang[1] Jiaqi Liang[1] Ziyu Zhu[1] Haoran Geng[2]
Jitendra Malik[2] Pieter Abbeel[2] Hao Dong[1†]
[1]Peking University [2]University of California, Berkeley

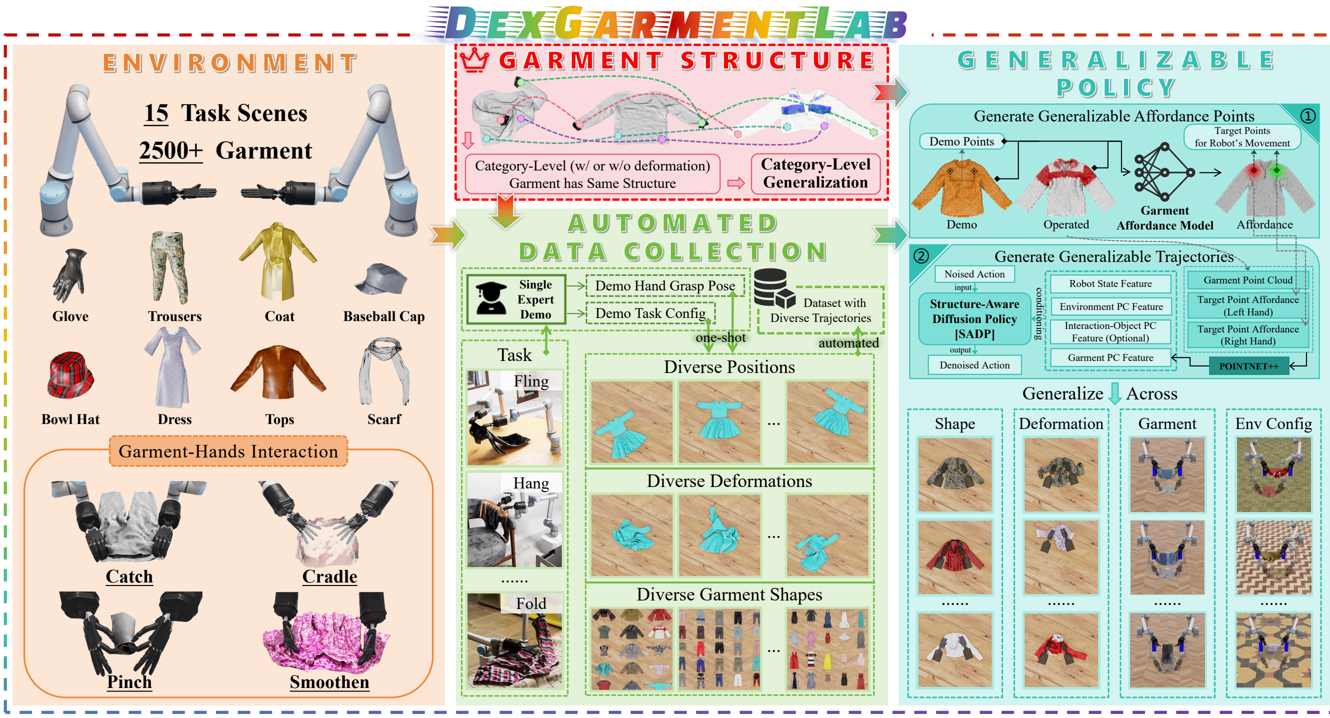

Fig. 1: **Overview.** DexGarmentLab includes three major components: **Environment**, **Automated Data Collection** and **Generalizable Policy**. Firstly, we propose Dexterous Garment Manipulation Environment with 15 different task scenes (especially for bimanual coordination) based on 2500+ garments. Because of the same structure of category-level garment, category-level generalization is accessible, which empowers our proposed automated data collection pipeline to handle different position, deformation and shapes of garment with task config (including grasp position and task sequence) and grasp hand pose provided by single expert demonstration. With diverse collected demonstration data, we introduce **H**ierarchical g**A**rment manipu**L**ation p**O**licy (**HALO**), combining affordance points and trajectories to generalize across different attributes in different tasks.

Abstract— **Garment manipulation is a critical challenge due to the diversity in garment categories, geometries, and deformations. Despite this, humans can effortlessly handle garments, thanks to the dexterity of our hands. However, existing research in the field has struggled to replicate this level of dexterity, primarily hindered by the lack of realistic simulations of dexterous garment manipulation. Therefore, we propose DexGarmentLab, the first environment specifically designed for dexterous (especially bimanual) garment manipulation, which features large-scale high-quality 3D assets for 15 task scenarios, and refines simulation techniques tailored for garment modeling to reduce the sim-to-real gap. Previous data collection typically relies on teleoperation or training expert reinforcement learning (RL) policies, which are labor-intensive and inefficient. In this paper, we leverage garment structural correspondence to automatically generate a dataset with diverse trajectories using only a single expert demonstration, significantly reducing manual intervention. However, even extensive demonstrations cannot cover the infinite states of garments, which necessitates the exploration of new algorithms. To improve generalization across diverse garment shapes and deformations, we propose a Hierarchical gArment-manipuLation pOlicy (HALO). It first identifies transferable affordance points to accurately locate the manipulation area, then generates generalizable trajectories to complete the task. Through extensive experiments and detailed analysis of our method and baseline, we demonstrate that HALO consistently outperforms existing methods, successfully generalizing to previously unseen instances even with significant variations in shape and deformation where others fail. Our project page is available at: https://wayrise.github.io/DexGarmentLab/.**

I. DexGarmentLab Environment

In this section, we present the construction of DexGarmentLab, the first environment specifically designed for dexterous (especially bimanual) garment manipulation and built upon IsaacSim 4.5.0.

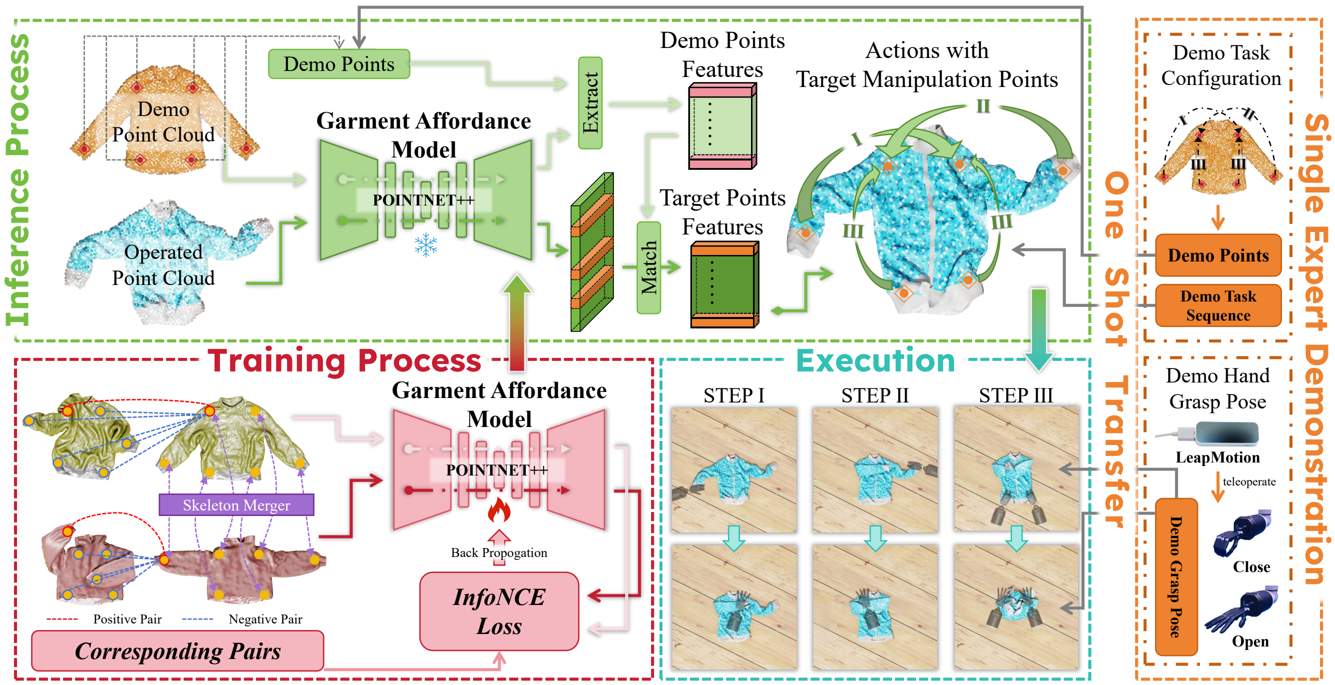

Fig. 2: **Automated Data Collection Pipeline.** Given a single expert demonstration, we can get demo points, demo task sequences and demo grasp poses for the specific task. Category-level garment (w/ or w/o deformation) has almost the same structure, base on which we can train Garment Affordance Model (GAM) with category-level generalization. With GAM (refer Supp.I), we match demo points from the demo garment point cloud O to a new garment point cloud O' and control robot to execute the specific task based on the demo task sequences (through trajectory retargeting) with dexhands' movement guided by demo hand grasp poses (through PD controller based on joint positions). 'Fold Tops' task is shown as example in this figure.

A. *DexGarmentLab Physcial Simulation*

Simulation Method. To achieve realistic simulation, we employ methods tailored to the physical properties of garments. Large garments (e.g., tops, dresses, trousers, etc.) are simulated using Position-Based Dynamics (PBD) [3], while small, elastic items (e.g., gloves, hats) are modeled via the Finite Element Method (FEM) [1]. We provide detailed introduction and selection reason about PBD and FEM in Supp.III. Human avatars are represented by articulated skeletons with rotational joints and a skinned mesh for lifelike rendering.

Key Design for Physical Garment Simulation. PBD is widely used for simulating most garments, but its loosely connected particles often allow grippers to penetrate the garment without achieving effective lifting. GarmentLab introduces attach blocks to address this, enabling garment-gripper attachment (Website Fig.2, left). However, this approach fails to capture realistic interactions, resulting in unnatural sagging when applied to dexterous hands (Website Fig.2, middle). Moreover, even minimal contact—such as a single finger block touching the garment—can establish attachment and lift the garment, which is clearly unreasonable.

Therefore, we introduce adhesion (*between particle and rigid*), friction (*between particle and rigid*) and particle-scale (*between particles*) parameters to enhance realism. Benefiting from friction and adhesion, dexhands can grasp and lift garments based on physical force without attach blocks (Website Fig.2, right), while particle-adhesion (or -friction)-scale stabilizes the particle system, preventing excessive self-collisions between particles which cause garments to become disorganized (Website Fig.3). We provide more details in Supp.III.

B. *Asset Selection and Annotation*

We use garment models from the ClothesNet dataset [6], which contains over 2,500 garments across 8 categories (e.g., tops, coats, trousers, dresses, etc.), and build environment-interaction assets (such as hangers, pothooks, humans, etc.). We provide plain meshes customizable with colors and textures for garments to support both realistic and controlled experimental setups. Controlled randomness in placement for both garments and environment-interaction assets—through limited rotations and translations—maintains task feasibility while enhancing generalization in policy learning.

C. *DexGarmentLab Tasks*

Dexterous (especially bimanual) garment manipulation is vital for domestic applications, yet it has not been thoroughly explored in existing research. To address this, we introduce 15 tasks across 8 garment categories (shown in our website Fig.4). Further details on these tasks are available in Supp.XIV.

II. Automated Data Collection

Collecting data through teleoperation or RL is highly labor-intensive, especially for dexterous garment manipulation tasks, due to the diverse shapes and deformations of garments and the high-dimensional action space of dexterous hands. This makes automated data collection essential, with the key challenges being: 1) identifying appropriate manipulation points across different garment configurations, and 2) generating task-specific hand poses accordingly.

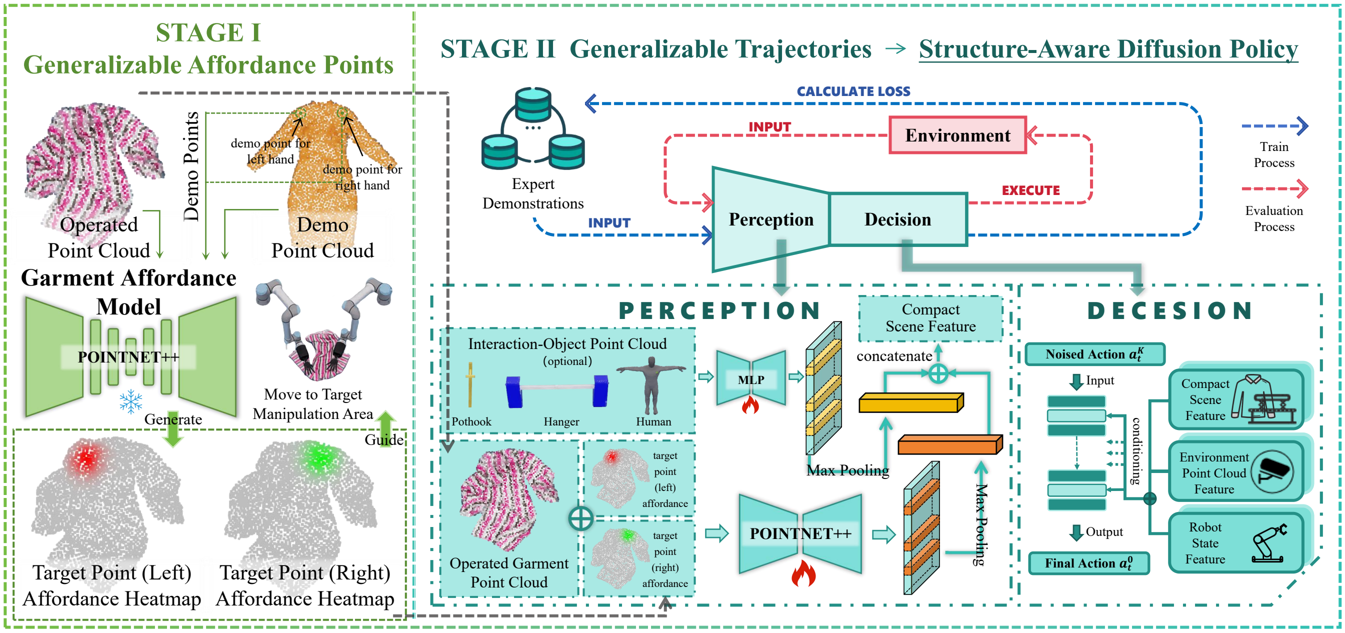

Fig. 3: **Generalizable Policy.** We adopt hierarchical structure to implement Generalizable Policy. Firstly, we use GAM to generate generalizable affordance points, which will be used for robots to locate and move to target area. Secondly, we introduce Structure-Aware Diffusion Policy (SADP), which extracts features from garment point cloud (with left and right point affordances as binding features), interaction-object point cloud, environment point cloud and robot joint states as condition to generate joint actions (including 24 DOF for each hand and 6 DOF for each arm, totally 60 DOF).

In our proposed automated data collection pipeline, for a given task, we begin with a single expert demonstration to extract key information: hand grasp poses, task sequences, and demo grasp points on the garment. Leveraging the Garment Affordance Model, we use affordance to identify target grasp points on novel garments with diverse deformations corresponding to demo grasp points. Then, the pipeline executes the task sequence based on inferred points and hand grasp poses, thereby enabling efficient and scalable data collection. Details about GAM and whole procedure can be found in Supp.I.

III. Generalizable Policy

When dealing with garments, which exhibit highly complex deformation states, current mainstream imitation learning (IL) algorithms (e.g. Diffusion Policy [2], Diffusion Policy 3D [5]) show relatively poor generalization (as evidenced by our experimental results shown in Supp. Tab.I). The main issue is that IL-based trajectories fail to accurately reach the target manipulation points on garments with new shapes and deformations, while also being unable to generate suitable trajectories based on the garment's own shape and structure, ultimately leading to manipulation failures.

To address this, we propose **H**ierarchical g**A**rment manipu**L**ation p**O**licy (**HALO**), a generalizable policy to solve the manipulation of garments with complex deformations and uncertain states. **HALO** is decomposed into two major stages, as shown in Fig. 3.

In the first stage, we use GAM to accurately locate the manipulation area of the garment, addressing the limitation of previous IL in grasping brand-new clothes at the correct position. Refer Supp.I for GAM's details. We next focus on the design of the Structure-Aware Diffusion Policy (SADP).

Due to the poor generalization ability of current mainstream methods such as DP and DP3 for complex and variable garment manipulation scenarios, we propose **Structure-Aware Diffusion Policy (SADP)**, a garment-environment-generalizable diffusion policy that improves the generalization for different garment shapes and scene configurations, thereby enabling the smooth generation of subsequent trajectories after moving to the target manipulation area guided by GAM.

SADP fundamentally follows the framework of Diffusion Policy [2], with the primary distinction lying in its observation representation, denoted as s, which is elaborated below.

With operated garment point cloud and left / right target point affordances generated by GAM, we concatenate them together and use PointNet++ [4] to extract garment feature $F_{garment}$, while using MLP-based Feature Extractors to extract interaction-object feature F_{object}. $F_{garment}$ and F_{object} are concatenated into a compact scene feature F_{scene}. At each timestep, the full environment point cloud $O_{environment}$ and the robot state O_{state} are encoded using MLP and fused with F_{scene} to form the denoising condition s for SADP. As for Garment-Self-Interaction tasks without interaction-object point cloud, we only use $F_{garment}$ to be F_{scene}, which means interaction-object point cloud is optional. Here, $F_{garment}$ captures current garment state (position, shapes, structure, etc.), while F_{object} reflects current interaction-object state (position, etc.).

Through experimental validation, HALO exhibits better generalization capabilities. We will further illustrate this advantage with experimental results in Supp.II. Training details can be found in Supp.XI.

REFERENCES

[1] Klaus-Jürgen Bathe. Finite element method. *Wiley encyclopedia of computer science and engineering*, pages 1–12, 2007.

[2] 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.

[3] Matthias Müller, Bruno Heidelberger, Marcus Hennix, and John Ratcliff. Position based dynamics. *Journal of Visual Communication and Image Representation*, 18(2):109–118, 2007.

[4] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[5] Yanjie Ze, Gu Zhang, Kangning Zhang, Chenyuan Hu, Muhan Wang, and Huazhe Xu. 3d diffusion policy. *arXiv preprint arXiv:2403.03954*, 2024.

[6] 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.