The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Real Garment Benchmark (RGBench): A Comprehensive Benchmark for Robotic Garment Manipulation Featuring a High-Fidelity Scalable Simulator

Wenkang Hu[1*], **Xincheng Tang**[1*], **Yanzhi E**[2], **Yitong Li**[1], **Zhengjie Shu**[1], **Wei Li**[3†], **Huamin Wang**[2], **Ruigang Yang**[1†]

[1] Shanghai Jiao Tong University
[2] Style3D Research
[3] Nanjing University

Abstract

While there has been significant progress to use simulated data to learn robotic manipulation of rigid objects, applying its success to deformable objects has been hindered by the lack of both deformable object models and realistic non-rigid body simulators. In this paper, we present *Real Garment Benchmark* (RGBench), a comprehensive benchmark for robotic manipulation of garments. It features a diverse set of over 6000 garment mesh models, a new high-performance simulator, and a comprehensive protocol to evaluate garment simulation quality with carefully measured real garment dynamics. Our experiments demonstrate that our simulator outperforms currently available cloth simulators by a large margin, reducing simulation error by 20% while maintaining a speed of 3 times faster. We will publicly release RGBench to accelerate future research in robotic garment manipulation.

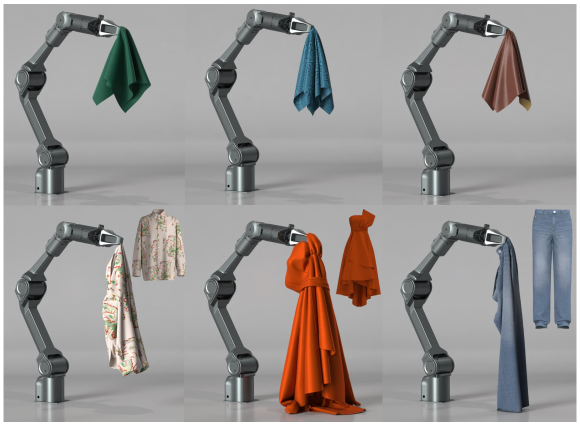

Figure 1: Robotic Manipulation of Garments and Fabrics with Diverse Materials in RGBench

Introduction

Robotic manipulation of deformable objects—most notably the challenge of handling garments—stands as a critical frontier in robotics research, with wide-ranging applications in household assistance and industrial automation (Sanchez et al. 2018). These two major challenges for handling garments are (a) the vast, virtually infinite-dimensional state space of garments makes their configuration difficult to represent; and (b) the highly non-linear, under-actuated thin-shell dynamics, which makes their behavior difficult to predict (Zhang et al. 2024). The high dimensional property and dynamic complexity are further exacerbated by pervasive contact and self-collision, where the thin fabric's constant folding and sliding create an intricate and rapidly changing landscape of physical interactions.

To overcome these challenges while mitigating the costs and risks of real-world trial-and-error, researchers increasingly turn to physics simulators, which offer a safe and scalable environment for developing manipulation policies. However, prevailing robotic simulators suffer from two critical limitations. First, their physical fidelity is often insufficient. To achieve the necessary computational performance, they rely on simplified models like Position-Based Dynamics (PBD), which are approximations of true continuum mechanics. As a result, they exhibit unrealistic physical behaviors, including distortions, stretching, and frequent self-penetration failures, creating a significant sim-to-real gap. Second, even with these simplifications, the performance of these simulators is often too slow. The computational resources required for large-scale parallel training of modern reinforcement learning algorithms are often unattainable.

In order to address the sim-to-real gap, it is essential to have benchmarks and datasets that systematically evaluate the fidelity of simulation. The few existing benchmarks that attempt to quantify the sim-to-real gap are typically restricted to simple cases such as one-dimensional ropes (Lim et al. 2022) or basic simple fabrics (e.g. handkerchiefs) (Blanco-Mulero et al. 2024).

In this paper, we present Real-Garment Benchmark (RGBench) that includes *GarmentDynamics*,a novel high-speed and high-accuracy cloth simulator, and a diverse set of 3D garment models with physically accurate parameters. The contributions of this paper are:

- **A diverse 3D garment model dataset with measured physical properties and motions**. We introduce a novel public dataset for cloth manipulation. Its core fea-

---

[*] These authors contributed equally.

[†] Corresponding author(s).

ture is a rich collection of garments that span a wide variety of materials and topological complexities. Crucially, this dataset provides high-fidelity 3D ground-truth data for garment configurations resulting from both quasi-static and dynamic real-world robotic manipulations.

- **A novel cloth simulator with accuracy, robustness, and efficiency**. We present *GarmentDynamics*, a novel physics-based simulator specifically engineered to overcome key limitations of existing robotics simulators in handling complex garment dynamics. Its design philosophy focuses on achieving physical accuracy, numerical stability, and collision robustness, and high computational performance through a combination of advanced physical models, accurate material property acquisition, and GPU acceleration. It will be released.
- **A dedicated benchmark for evaluating the sim-to-real gap of cloth simulators**. We introduce Real-Garment Benchmark(RGBench), the benchmark designed to enable the rigorous evaluation and comparison of current mainstream physics simulators on the challenging task of garment manipulation. It provides a standardized framework to quantify the sim-to-real gap of any given simulator, featuring the richest collection of real-world garments and robotic actions to date.

Related Work

Deformable Object Manipulation Benchmark

To systematically advance the field of robotic manipulation of deformable objects, benchmarks are essential to assess algorithm performance, define limitations, and establish standardized comparisons (Longhini et al. 2024). The majority of these evaluation platforms are situated within simulated environments, with mainstream robotics simulators like MuJoCo (Todorov, Erez, and Tassa 2012), PyBullet (Coumans and Bai 2016–2021), and Isaac Sim (NVIDIA Corporation 2021–2024) being common choices. Representative works include SoftGym (Lin et al. 2021), and more recent GarmentLab (Lu et al. 2024) and DexGarmentLab (Wang et al. 2025) were built in IssacSim, and Daxbench (Chen et al. 2022), which is based on a differentiable simulator. These benchmarks often utilize models from large open-source simulated datasets, such as ClothesNet (Zhou et al. 2023) and Cloth3D (Bertiche, Madadi, and Escalera 2020), to create diverse evaluation scenarios.

Despite the sophistication of simulation-based benchmarks, the "sim-to-real" gap remains a central challenge. Unlike rigid bodies that can be simulated with high fidelity, the extreme deformability of objects like cloth makes accurate modeling exceptionally difficult, as policies trained in simulation often fail when transferred to the real world. This has motivated a parallel line of research focused on curating real-world datasets, from large-scale image repositories like DeepFashion (Liu et al. 2016), household cloth collections with an RGB-D dataset from (Garcia-Camacho et al. 2022).

To benchmark the robotic manipulation for the deformable object, more targeted efforts have sought to measure the reality gap. (Lim et al. 2022) quantified this gap for 1D cables in simulators like PyBullet and Isaac Sim. (Blanco-Mulero et al. 2024) evaluated the performance difference of various simulators in dynamic and quasi-static fling motions on simple cloth towels, chequered, and linens. While these efforts represent valuable progress, their focus has largely been confined to structurally simple 1D cables and basic simple fabrics. A notable gap remains in the analysis of more complex garments, which we address by providing a benchmark with ground truth to measure the sim-to-real gap for common robotic clothing manipulation tasks: grasping, flinging, and folding.

Physics-Based Cloth Simulation

A physics-based cloth simulator can represent cloth and its deformation in three main ways: as a network of springs (Choi and Ko 2002), a collection of elements (Müller et al. 2005; Volino, Magnenat-Thalmann, and Faure 2009), or a cluster of yarns (Kaldor, James, and Marschner 2008; Cirio et al. 2014). Among these, the element-based representation has become increasingly popular due to its balance between physical accuracy and computational cost.

Given a cloth representation, the key question is how to advance its dynamics over time. Explicit time integration (Bridson, Fedkiw, and Anderson 2002; Bridson, Marino, and Fedkiw 2003) is conceptually simple, but requires very small time steps for stability, becoming computationally expensive. Implicit time integration (Baraff and Witkin 1998) improves stability and allows larger time steps, but remains costly because it requires solving large linear systems. Position-based dynamics (PBD) (Müller 2008; Müller et al. 2014) was introduced as an alternative, emphasizing simplicity and robustness at the expense of physical accuracy and scalability, particularly for high-resolution meshes. Projective dynamics (Bouaziz et al. 2014) later unified PBD and implicit integration under a common constraint-based framework, demonstrating their close relationship. This insight has inspired a line of research on fast cloth simulation (Wang 2015; Peng et al. 2018; Chen et al. 2024), highlighting the potential of GPUs to accelerate implicit methods for real-time cloth simulation.

Efficient and robust collision handling remains one of the most challenging problems in cloth simulation. Over the years, researchers have explored various aspects of this problem — often leveraging GPUs — including collision culling (Tang et al. 2011), collision detection (Brochu, Edwards, and Bridson 2012), penetration untangling (Baraff, Witkin, and Kass 2003), and collision response methods (Tang et al. 2016). More recently, potential-based contact formulations (Li, Kaufman, and Jiang 2021; Wu et al. 2020) have shown strong promise for robust collision handling, and recent work has focused on accelerating these methods on GPUs (Lan et al. 2024; Li et al. 2023).

Beyond the core engine, realism depends critically on accurate physical parameters. These can be acquired through optimization-based methods that match observed motion (Miguel et al. 2012) or learning-based approaches that infer parameters directly (Rasheed et al. 2021). However, the more fundamental distinction lies in how the measurement data is collected: either from videos of uncontrolled or unconstrained cloth motion (Yang, Liang, and Lin 2017), or from carefully controlled experiments (Wang, O'Brien, and Ramamoorthi 2011; Feng et al. 2022). We argue that controlled experiments are more reliable for isolating and precisely measuring individual material properties. For this reason, we also adopt strategies based on dedicated measurement devices.

The RGBench Framework

Figure 2 presents an overview of the RGBench framework, which integrates a diverse garment dataset, dual-arm robotic setups, and the GarmentDynamics simulation system, covering three core tasks to bridge real-world and simulated garment interaction research.

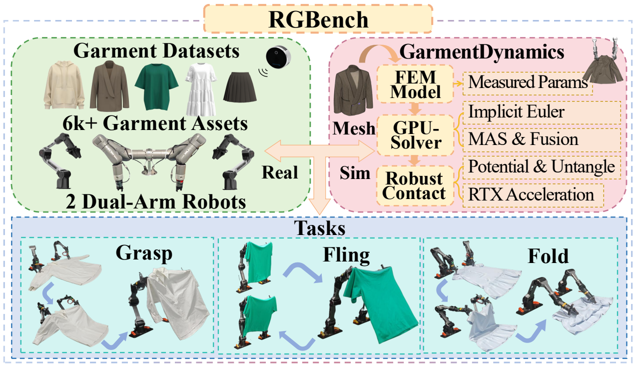

Figure 2: Overview of RGBench framework

RGBench Dataset

Diverse Garment Assets Our dataset encompasses a diverse range of garments with deliberate variations in size, style, and material to ensure comprehensive coverage of real-world manipulation scenarios shown in Figure 3. Specifically, the dataset encompasses a total of 6k+ garment models, consisting of two components: 4k+ self-collected industry-level production-ready assets, which feature high-quality meshes, producible features (including elasticity, folding edges, and seaming lines), and physically accurate parameters (such as stretch stiffness, bending stiffness, and density); and 2k+ garment meshes sourced from ClothesNet (Zhou et al. 2023). Specifically, the collection comprises garments of varying sizes (from small to large), diverse styles (fitted, loose-fitting, structured, and flowy), and a range of materials—cotton, linen, wool, polyester, nylon, silk, etc., each with distinct fabric properties. For manipulation tasks, we select 9 types of cloth to obtain the real point cloud during interaction as their ground truth (GT).

Ground Truth Acquisition Our methodology for acquiring ground truth data is twofold, encompassing the static physical properties of each garment and its dynamic behavior during robotic manipulation.

- **Garment Physical Ground Truth**: To create physically-grounded digital twins, we ensure both geometric and physical accuracy. For geometric accuracy, it is achieved by importing production-grade DXF pattern files to build 1:1 scale models, with seam lines defined to mirror the physical assembly. For physical accuracy, the Tensile Tester (SST1000) and Bending Tester (SBE1000) is used to profile the core mechanical properties of each fabric (e.g., stretching, bending), as shown in part 3 by Figure 3. This measurement protocol adheres to established ASTM standards (D1388 (ASTM International 2018), D3107 (ASTM International 2017)). This process yields a physically-grounded asset for simulation.

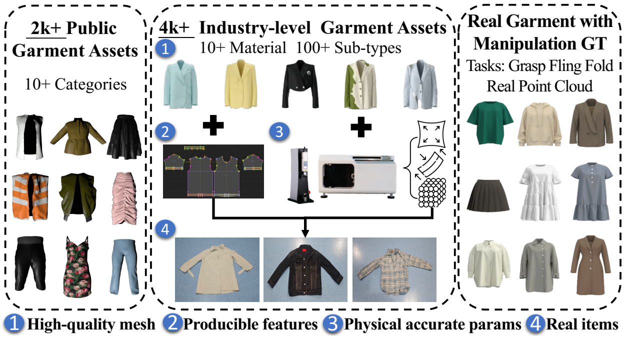

Figure 3: RGBench garment dataset

- **Manipulation Ground Truth** We employ the dual-robotic setup (AGILEX Piper or JAKA K1 with DH PGC-50-35 grippers) to manipulate the garments. An Intel RealSense L515 camera records the entire process, capturing the complex deformations of the garment as a stream of RGB point cloud data, which are segmented by Grounded-SAM (Ren et al. 2024). This captured point cloud serves as the definitive ground truth (P_{real}) against which all simulation experiments are compared.

RGBench Evaluation

Eval Tasks: Robotic Manipulation Our benchmark is built on Grasp, Fling, and Fold —three foundational primitives for robotic garment manipulation. These primitives systematically address the core challenges of the field: A robot must be able to reliably pick up a garment (Grasp), reconfigure it dynamically in-midair (Fling), and precisely arrange it through contact-rich interactions (Fold). Each primitive is specifically designed to stress-test a distinct and critical aspect of the underlying physics simulation.

- **Grasp**: The task begins with a flat garment, with predefined grasp points near the shoulder region. The robotic arms then synchronously grasp these points and lift the garment vertically. This primitive primarily evaluates the simulator's ability to model initial contact dynamics, frictional forces, and gravity-induced deformation.
- **Fling**: Beginning with the garment held aloft by the grippers, the dual arms execute a rapid forward-then-backward trajectory to induce aerial fluttering of the garment. This task is specifically designed to challenge the simulator's handling of high-speed kinematics, inertial effects, aerodynamic drag, and large-scale deformation.
- **Fold**: Starting again with a flat-laid garment, the dual arms grasp the shoulder sections, lift slightly, and trans-

late forward to the bottom edge. This procedure tests the simulator's capacity to handle complex, evolving self-contact, inter-surface friction, and the settling of the fabric into a stable, folded configuration.

Eval Preprocess To ensure high-fidelity correspondence between our physical experiments and simulations, we perform a rigorous three-step alignment process. First, for initial state alignment, we standardize the garment's pose using a physical template of its outline. Second, the relative pose between the robot and camera is determined by hand-eye calibration or the Iterative Closest Point algorithm, which ensures high-precision registration. Finally, to achieve temporal synchronization, we apply a fixed, optimal time delay to the data streams to compensate for system latency.

Eval Metrics We utilize Chamfer Distance(CD) and Hausdorff Distance(HD) as the core metrics for evalution. CD quantifies the average dissimilarity between two point sets. For a simulated mesh with vertices $\mathcal{V}_t$ and a real-world point cloud $\mathcal{P}_t$ at time t, the sim-to-real CD is defined as:

$$CD_{s2r}\left(\mathcal{V}_t, \mathcal{P}_t\right) := \frac{1}{|\mathcal{V}_t|} \sum_{v \in \mathcal{V}_t} \min_{p \in \mathcal{P}_t} \|v - p\|_1 \tag{1}$$

Conversely, the unidirectional real-to-sim CD reverses the point set order, calculated as:

$$CD_{r2s}\left(\mathcal{P}_t, \mathcal{V}_t\right) := \frac{1}{|\mathcal{P}_t|} \sum_{p \in \mathcal{P}_t} \min_{v \in \mathcal{V}_t} \|p - v\|_1 \tag{2}$$

where $|\mathcal{V}_t|$ and $|\mathcal{P}_t|$ denote the number of vertices in the corresponding mesh, and $\| \cdot \|_1$ is the Manhattan distance. This dual formulation accounts for the disparity in point density between real and simulated data, ensuring both perspectives of alignment are captured.

HD focuses on capturing the worst-case alignment error. The unidirectional HD is:

$$HD_{s2r}\left(\mathcal{V}_t, \mathcal{P}_t\right) := \max_{v \in \mathcal{V}_t} \min_{p \in \mathcal{P}_t} \|v - p\|_1 \tag{3}$$

$$HD_{r2s}\left(\mathcal{P}_t, \mathcal{V}_t\right) := \max_{p \in \mathcal{P}_t} \min_{v \in \mathcal{V}_t} \|p - v\|_1 \tag{4}$$

These metrics highlight extreme deviations from both the simulated and real-world perspectives, which is crucial for identifying critical failures in dynamic cloth manipulation, such as large deformations or self-occlusions.

Unlike prior work using only sim-to-real metrics, we use both sim-to-real and real-to-sim metrics. Simulators, especially particle-based ones, have unstable garment simulations (e.g., model expansion/explosion) that inflate sim-to-real metrics, making them useless for comparison. Real-to-sim metrics show how well real garment surfaces match simulated meshes. We analyze both, focusing more on real-to-sim errors as they better reflect simulation fidelity to real observations and are less affected by simulator instabilities.

Simulator

To rigorously evaluate the simulation-to-reality gap, our benchmark is built upon an engine-agnostic framework. Any simulator integrated into this framework must support several key functionalities: the simulation of both rigid and deformable bodies, robust collision handling, variable integration timesteps, and the ability to query garment vertex states. To provide a comprehensive baseline, we incorporate three mainstream robotics simulators that meet these criteria: MuJoCo, PyBullet, and NVIDIA Isaac Sim.

GarmentDynamics Design

Core Components The following sections elucidate its architecture, detailing the design choices behind its GPU-accelerated implementation. We further analyze the mechanisms engineered to achieve superior accuracy, robustness, and computational efficiency for complex garment manipulation tasks.

Continuum-Based FEM Model In our simulator, cloth is treated as a continuous elastic surface discretized into a triangular mesh composed of linear triangular elements. The constitutive behavior is modeled in both in-plane and out-of-plane modes using an anisotropic extension of the model proposed in (Baraff and Witkin 1998), allowing us to account separately for the stretching and bending responses along the warp, weft, and diagonal directions of the fabric. Furthermore, we introduce nonlinearity into the out-of-plane behavior by defining the bending resistance as a quadratic function of the bending curvature. This formulation enables our model to more accurately capture cloth wrinkling, which strongly depends on bending resistance.

Physical Properties and Interactions The realism of a physics-based cloth simulator depends not only on the underlying physical models, but also on how accurately the material properties are measured and how effectively self-collisions and cloth–environment interactions are handled. Leveraging the physically grounded nature of our model, our simulator directly uses fabric properties measured from real-world samples, including mass density, thickness, elastic moduli, and surface roughness for friction. For collision handling, we combine potential-based contact forces with a collision untangling mechanism (Volino and Magnenat-Thalmann 2006). This hybrid strategy achieves high performance while maintaining robustness, preventing penetrations even under frequent and severe contact conditions.

GPU-Accelerated Implicit Time Integration Finally, we advocate using an implicit time-integration scheme in our solver. Our solver leverages recent advances in GPU acceleration — such as multiresolution preconditioning (Wu, Wang, and Wang 2022) and potential-based contact handling — to efficiently simulate cloth dynamics, even for meshes with a large number of degrees of freedom. Compared with the position-based dynamics used in Isaac Sim and explicit integration methods, implicit integration offers higher physical accuracy and improved numerical stability, even with large time steps and significant deformations.

GarmentDynamics GPU Implementation

GPU-Based Solver To achieve high performance, our simulator executes all computation stages entirely on the

GPU, combining algorithmic advances with hardware-level optimizations. Cloth dynamics are solved using implicit Euler integration on the GPU, where each time step is formulated as an energy minimization of the following objective:

$$\mathcal{L}(\mathbf{x}) = \frac{1}{2h^2}(\mathbf{x}-\mathbf{x}^t)\mathbf{M}(\mathbf{x}-\mathbf{x}^t) + E(\mathbf{x}), \tag{5}$$

where h is the time step, $\mathbf{x} \in \mathbb{R}^{3N}$ denotes the unknown positional vector of the N mesh vertices, $\mathbf{x}^t$ is the positional vector from the previous time step, $\mathbf{M} \in \mathbb{R}^{3N\times 3N}$ is the lumped mass matrix, and $E(\mathbf{x})$ represents the potential energy. Unlike (Baraff and Witkin 1998), our simulator solves Eq. 5 using a small number of inexact Newton iterations, each involving the solution of a linear system via a fixed number of preconditioned conjugate gradient (PCG) sub-iterations. Thanks to GPU acceleration, our solver achieves both fast convergence and efficient runtime.

Multilevel Preconditioning To further accelerate the convergence of our PCG solver, we adopt an algebraic multigrid preconditioner within the multilevel additive Schwarz (MAS) framework (Wu, Wang, and Wang 2022). This preconditioner is distinctive in that it approximates the system matrix with a block-diagonal inverse constructed from many small, non-overlapping subdomains across multiple resolution levels. At the start of the linear solve, the preconditioner computes the inverse of each block using Gauss–Jordan elimination and stores these inverses in GPU memory. During runtime, applying the preconditioner to a residual vector becomes a simple conflict-free, per-block sparse matrix–vector multiplication, with each block handled by a GPU thread. As demonstrated in (Wu, Wang, and Wang 2022), this preconditioner effectively reduces the condition number of the system, enabling the PCG solver to converge more rapidly within a fixed number of iterations.

Collision Detection Another major factor contributing to simulation performance is collision detection. To accelerate it, we develop a GPU-based bounding volume hierarchy structure, in which each query is efficiently executed in parallel using NVIDIA RTX ray intersection features. Compared with the general-purpose implementations in Isaac Sim or Bullet, our approach is significantly faster, capable of handling millions of broad-phase collision tests per second. Our implementation further employs optimized sparse-matrix data structures and CUDA kernel fusion to maximize memory coalescing and further boost raw throughput.

Experiments and Results

Garment Manipulation Mode: Our framework provides two garment manipulation modes: robot interaction mode and pseudo interaction mode. (1) The robot mode utilizes URDF files and joint angles to simulate complete robotic behavior and garment dynamics simultaneously. (2) The pseudo mode achieves cloth manipulation simulation by directly controlling the movement of cloth intersection vertices, thereby simplifying the simulation of robotic interaction, which focuses on the garment dynamics itself.

Deformable Parameters: Garment simulation involves a diverse set of parameters. For parameters with clear physical meanings that can be directly measured, we set their values based on experimental measurements using specialized instruments. For environmental parameters such as friction and damping, we first select a representative garment to perform parameter optimization. These optimized parameters are then applied uniformly across all garments, ensuring that environmental influences are consistent.

Validation on a Foundational Cloth Benchmark

To validate the fundamental accuracy of GarmentDynamics, we replicate the dynamic and quasi-static cloth deformation experiments on the rectangle chequered rag dataset from a widely accepted public benchmark (BCM) (Blanco-Mulero et al. 2024). The garment state generated by GarmentDynamics is compared against the results from Mujoco and FLEX, which have the best performance in this dataset.

Mode	Metric	Mujoco	FLEX	Ours
Dynamic	CD_{s2r}	0.067 ± 0.026	0.164 ± 0.134	$\mathbf{0.062 \pm 0.028}$
	HD_{s2r}	0.154 ± 0.035	0.280 ± 0.180	$\mathbf{0.150 \pm 0.051}$
Quasi static	CD_{s2r}	0.076 ± 0.025	0.072 ± 0.019	$\mathbf{0.0389 \pm 0.006}$
	HD_{s2r}	0.186 ± 0.055	0.171 ± 0.024	$\mathbf{0.094 \pm 0.022}$

Table 1: Simulation results on BCM benchmark

As shown in Table 1, GarmentDynamics demonstrates distinct strengths in quasi-static scenarios: it outperforms the state-of-the-art (FLEX) by $\mathbf{46\%}$ in CD and $\mathbf{45\%}$ in HD, validating its accuracy in capturing cloth deformations under slow, contact-dominated motions.

For dynamic tasks, we identified that a key factor affecting performance is the notable noise within the benchmark's input anchor point trajectories. To rigorously evaluate our simulator's resilience, we applied it directly to the raw data without the pre-processing steps. Despite that, GarmentDynamics still achieves the best performance, which highlights its superior robustness to imperfect inputs.

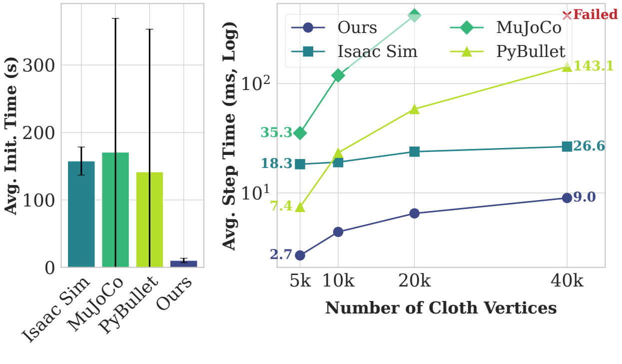

Figure 4: Simulator efficiency comparison (left) Average Initial Time; (right) Average Step Time;

Simulation Efficiency and Scalability

To evaluate the performance and scalability of our proposed simulator, we conduct a comprehensive benchmark against

Cloth	Action	**CD**$_{r2s}$			**HD**$_{r2s}$		
		pybullet	isaacsim	ours	pybullet	isaacsim	ours
Cakeskirt	Fling	0.0588	0.0598	**0.0406**	0.1767	0.1872	**0.1725**
Cakeskirt	Fold	0.0266	0.0308	**0.0179**	0.1082	0.1082	**0.0962**
Cakeskirt	Grasp	0.0487	0.0517	**0.0269**	0.1573	0.1719	**0.1222**
Coat	Fling	0.1071	0.0815	**0.0379**	0.2737	0.2406	**0.1239**
Coat	Fold	0.0309	0.0346	**0.0279**	0.1251	0.1182	**0.1083**
Coat	Grasp	0.0573	0.0605	**0.0320**	0.1817	0.1395	**0.1053**
Dress	Fling	0.0889	0.0928	**0.0687**	0.2007	0.2228	**0.1765**
Dress	Fold	0.0331	0.0459	**0.0187**	0.1229	0.1338	**0.0800**
Dress	Grasp	0.0433	0.0490	**0.0221**	0.1468	0.1563	**0.0953**
Hoodie	Fling	0.0310	0.0352	**0.0256**	0.1307	0.1493	**0.1254**
Hoodie	Fold	0.0240	0.0302	**0.0225**	**0.0952**	0.0966	0.1493
Hoodie	Grasp	0.0275	0.0308	**0.0217**	0.0947	0.1008	**0.0882**
Pleat Skirt	Fling	0.0540	0.0388	**0.0326**	0.1236	0.1089	**0.0736**
Pleat Skirt	Fold	0.0256	0.0255	**0.0126**	0.1383	0.1199	**0.0857**
Pleat Skirt	Grasp	0.0241	0.0201	**0.0159**	0.0878	0.0908	**0.0707**
L-Sleeves	Fling	0.0518	0.0600	**0.0422**	0.1770	0.2058	**0.1547**
L-Sleeves	Fold	0.0280	0.0308	**0.0243**	0.1252	0.1171	**0.0982**
L-Sleeves	Grasp	0.0347	0.0312	**0.0280**	0.1394	0.1123	**0.1106**
T-shirt	Fling	0.0532	0.0567	**0.0419**	0.1522	0.1710	**0.1197**
T-shirt	Fold	0.0227	0.0314	**0.0132**	0.0778	0.0966	**0.0512**
T-shirt	Grasp	0.0348	0.0341	**0.0226**	0.1109	0.1058	**0.0798**

Table 2: Metrics Across Garment & Action in Pseudo mode

three widely used physics simulation environments: Isaac Sim, MuJoCo, and PyBullet. The benchmark task consisted of simulating a deformable cloth model with progressively increasing complexity, specifically with 5k, 10k, 20k, and 40k vertices. Two key performance metrics were measured: (1) Initialization Time, the one-time cost in seconds to load the scene and assets at 5k, 10k, 20k, and 40k vertices. (2) Average Step Time, the average computation time in milliseconds per simulation step, calculated over 100 iterations after the simulation reached a stable state.

The experimental results are shown in Figure 4, where our simulator shows a significant improvement in performance. Among the baseline simulators, IsaacSim is the most competitive performer, maintaining a respectable step time of 26.6 ms even at a high complexity of 40k vertices. However, its performance is less efficient at lower complexities. In contrast, our simulator consistently outperforms IsaacSim by **3.0**× to **7.0**×, achieving a minimum step time of just **2.7** ms. Conversely, both PyBullet and MuJoCo experience a dramatic increase in initialization and simulation time as the number of vertices grows. Our simulator runs approximately **65.0**× faster than MuJoCo at 20k vertices and **16.0**× faster than PyBullet at 40k. Furthermore, MuJoCo fails entirely at the 40k vertex level due to computational complexity, highlighting the superior robustness of our approach. In addition to runtime execution, our simulator also exhibits superior initialization efficiency, reducing setup overhead by about **92%** compared to PyBullet, and by over **93%** on average compared to IsaacSim and MuJoCo.

Experimental results confirm our simulator's superior computational efficiency over SOTA methods: algorithmically, it integrates multilevel preconditioning, inexact Newton iterations, and hybrid collision handling for faster convergence on high-resolution meshes, while implementation-wise, advanced GPU acceleration (e.g., RTX ray intersection, kernel fusion) maximizes performance.

Sim-to-Real Gap Across Manipulation Tasks

To evaluate the consistency and robustness of GarmentDynamics in handling diverse physical interactions from quasi-static to dynamic modes, we design a standardized rubric of three core manipulation actions for each garment.

A qualitative comparison is present in Figure 5, which visualizes the sim-to-real performance for folding a dress. The left column shows the real-world manipulation sequence; The center column visualizes the point-wise distances between the simulated and real point clouds; and the right column shows the process of three evaluated simulators. Notably, Isaac Sim exhibits instability upon gripper contact with the garment, often resulting in exaggerated twisting or inflation. PyBullet fails to generate valid grasping visuals, as its robot arm cannot reliably grasp the garment, only providing images of end-effector points. In contrast, GarmentDynamics remains stable throughout the entire sequence, owing to its robust hybrid collision-handling strategy. Most importantly, the point cloud generated by our simulator shows the closest alignment with real-world data, further validating its ability to replicate authentic garment behavior.

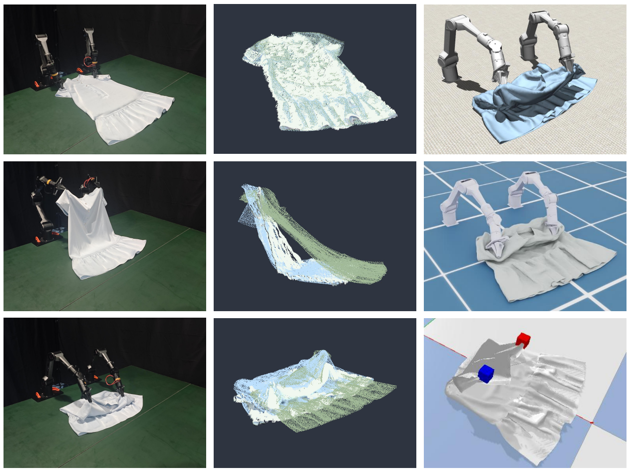

Figure 5: **Real-to-Sim Comparison for Folding** (left column) Real-world robotic folding sequence. (middle column) Point cloud comparison: real-world (white), our simulator (blue), best of other simulators (Grey). (right column) Garment folding state in all simulators.

We extend this analysis quantitatively across the three manipulation tasks in Figure 6, using the T-shirt as a representative example. Full results for all garments are included in the Appendix. Across tasks, GarmentDynamics consistently achieves the smallest sim-to-real discrepancies, with narrow error bars highlighting its repeatability and stability. For grasp and fold tasks, GarmentDynamics exhibits a strong ability to accurately model quasi-static contact mechanics, where precise prediction of deformations is critical, with CD_{r2s} error reduction up to 35% and 58%. Meanwhile, we observe that the sim-to-real gap widens for the

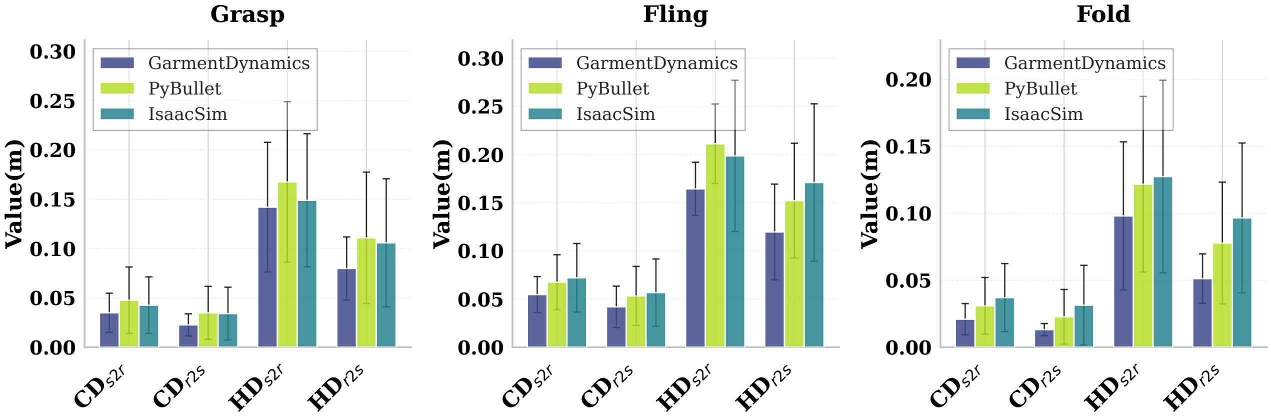

Figure 6: Qualitative results of different actions for a T-shirt

dynamic manipulation task fling as the high-speed interaction. Despite this, GarmentDynamics still maintains a clear advantage in these dynamic scenarios, improving CD_{r2s} and HD_{r2s} by over 20%. This versatility stems from its refined physical-based material modeling and robust GPU-based solver, which better accounts for real-world complexities like contact friction and inertial effects.

Sim-to-Real Gap Across Diverse Garments

To further validate the superior material modeling capabilities and generalizability of GarmentDynamics, we expanded our experiments to include 7 distinct garment types in the pseudo and robot model. These garments span a wide range of variations in materials, structural designs, and topological complexities. The results are summarized in Table 2.

GarmentDynamics demonstrates the lowest error across nearly all garment types and manipulation actions, thereby bridging the sim-to-real gap less than baseline simulators. To conduct a concrete comparison, we selected the grasp task as a representative to contrast how each simulator performs in Figure 7. Additionally, we also compare the two operation modes, providing a more comprehensive assessment of the simulators' behaviors in this experiment. We find the sim-to-real gap increases for thicker and longer garments, due to complex surface friction caused by spatial constraints. The result shows GarentDynamics achieves the lowest mean CD_{r2s} errors, with consistent reductions of more than 37% compared to PyBullet and Isaac Sim in pseudo mode. Its advantage is particularly pronounced for topologically complex items. For instance, in Cakeskirt manipulation, it reduces errors by up to 44% in pseudo mode and 77% in the robot mode — a substantial improvement that underscores its ability to handle intricate garment structures.

Manipulation Mode (Robot *vs*. Pseudo): In robot mode, modeling inaccuracies can lead to slippage, insufficient grip strength, or erratic deformation, resulting in larger sim-to-real gaps compared to pseudo mode. PyBullet fails critically here: it struggles with failed grasps, severe penetration (grippers passing through garments), and overly thin cloth rendering, creating stark visual mismatches. These issues highlight challenges in modeling contact and material properties, while emphasizing our simulator's superior fidelity in capturing physical interactions and visual realism.

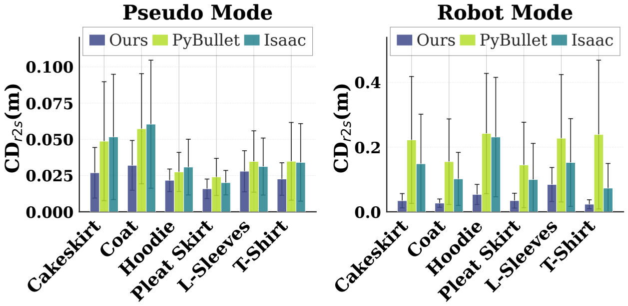

Figure 7: Results of different garment types in grasp task

Conclusion and Discussion

In this work, we addressed the dual challenge of the sim-to-real gap and performance limitations in robotic garment manipulation by Real-Garment Benchmark (RGBench), which includes a diverse real-world dataset and GarmentDynamics, a new high-fidelity, high-performance simulator. We followed the protocol defined in RGBench to evaluate performance across fundamental tasks such as grasping, flinging, and folding. The results clearly show that GarmentDynamics is the new state-of-the-art. It reduces the sim-to-real gap by over 20% on average and by as much as 77% for topologically complex garments. In terms of speed, it is 3.0x faster than its closest competitor, Isaac Sim, while robustly handling high-complexity scenes where others fail and slashing initialization time by over 90% Furthermore, we confirmed its fundamental accuracy on a prior benchmark for simple fabrics, where it surpassed the SOTA by over 45%. Future work will focus on enhancing the benchmark's reliability and extending its capabilities, ultimately aiming to train general-purpose manipulation policies with RGBench and apply GarmentDynamics to a wider range of scenarios and objects.

References

ASTM International. 2017. Standard Test Methods for Stretch Properties of Fabrics Woven from Stretch Yarns. ASTM Standard D3107 / D3107M-17. Available from: www.astm.org.

ASTM International. 2018. Standard Test Method for Stiffness of Fabrics. ASTM Standard D1388-18. Available from: www.astm.org.

Baraff, D.; and Witkin, A. 1998. Large Steps in Cloth Simulation. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, 43–54.

Baraff, D.; Witkin, A.; and Kass, M. 2003. Untangling Cloth. *ACM Trans. Graph. (SIGGRAPH)*, 22(3): 862–870.

Bertiche, H.; Madadi, M.; and Escalera, S. 2020. Cloth3d: clothed 3d humans. In *European Conference on Computer Vision*, 344–359. Springer.

Blanco-Mulero, D.; Barbany, O.; Alcan, G.; Colomé, A.; Torras, C.; and Kyrki, V. 2024. Benchmarking the Sim-to-Real Gap in Cloth Manipulation. *IEEE Robotics and Automation Letters*, 9(3): 2981–2988.

Bouaziz, S.; Martin, S.; Liu, T.; Kavan, L.; and Pauly, M. 2014. Projective Dynamics: Fusing Constraint Projections for Fast Simulation. *ACM Trans. Graph. (SIGGRAPH)*, 33(4): 154:1–154:11.

Bridson, R.; Fedkiw, R.; and Anderson, J. 2002. Robust Treatment of Collisions, Contact and Friction for Cloth Animation. *ACM Trans. Graph. (SIGGRAPH)*, 21(3): 594–603.

Bridson, R.; Marino, S.; and Fedkiw, R. 2003. Simulation of Clothing with Folds and Wrinkles. In *Proceedings of SCA*, 28–36.

Brochu, T.; Edwards, E.; and Bridson, R. 2012. Efficient Geometrically Exact Continuous Collision Detection. *ACM Trans. Graph. (SIGGRAPH)*, 31(4): 96:1–96:7.

Chen, A. H.; Liu, Z.; Yang, Y.; and Yuksel, C. 2024. Vertex Block Descent. *ACM Trans. Graph. (SIGGRAPH)*, 43(4).

Chen, S.; Xu, Y.; Yu, C.; Li, L.; Ma, X.; Xu, Z.; and Hsu, D. 2022. Daxbench: Benchmarking deformable object manipulation with differentiable physics. *arXiv preprint arXiv:2210.13066*.

Choi, K.-J.; and Ko, H.-S. 2002. Stable but Responsive Cloth. *ACM Trans. Graph. (SIGGRAPH)*, 21(3): 604–611.

Cirio, G.; Lopez-Moreno, J.; Miraut, D.; and Otaduy, M. A. 2014. Yarn-Level Simulation of Woven Cloth. *ACM Trans. Graph. (SIGGRAPH Asia)*, 33(6): 207:1–207:11.

Coumans, E.; and Bai, Y. 2016–2021. PyBullet, a Python module for physics simulation for games, robotics and machine learning. http://pybullet.org.

Feng, X.; Huang, W.; Xu, W.; and Wang, H. 2022. Learning-Based Bending Stiffness Parameter Estimation by a Drape Tester. *ACM Trans. Graph. (SIGGRAPH Asia)*, 41(6).

Garcia-Camacho, I.; Borràs, J.; Calli, B.; Norton, A.; and Alenya, G. 2022. Household cloth object set: Fostering benchmarking in deformable object manipulation. *IEEE Robotics and Automation Letters*, 7(3): 5866–5873.

Kaldor, J. M.; James, D. L.; and Marschner, S. 2008. Simulating Knitted Cloth at the Yarn Level. *ACM Trans. Graph. (SIGGRAPH)*, 27(3): 65:1–65:9.

Lan, L.; Lu, Z.; Long, J.; Yuan, C.; Li, X.; He, X.; Wang, H.; Jiang, C.; and Yang, Y. 2024. Efficient GPU Cloth Simulation with Non-distance Barriers and Subspace Reuse. *ACM Trans. Graph. (SIGGRAPH Asia)*, 43(6).

Li, M.; Kaufman, D. M.; and Jiang, C. 2021. Codimensional Incremental Potential Contact. *ACM Trans. Graph. (SIGGRAPH)*, 40(4).

Li, X.; Fang, Y.; Lan, L.; Wang, H.; Yang, Y.; Li, M.; and Jiang, C. 2023. Subspace-Preconditioned GPU Projective Dynamics with Contact for Cloth Simulation. In *SIGGRAPH Asia 2023 Conference Papers*, SA '23.

Lim, V.; Huang, H.; Chen, L. Y.; Wang, J.; Ichnowski, J.; Seita, D.; Laskey, M.; and Goldberg, K. 2022. Real2sim2real: Self-supervised learning of physical single-step dynamic actions for planar robot casting. In *2022 International Conference on Robotics and Automation (ICRA)*, 8282–8289. IEEE.

Lin, X.; Wang, Y.; Olkin, J.; and Held, D. 2021. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. In *Conference on Robot Learning*, 432–448. PMLR.

Liu, Z.; Luo, P.; Qiu, S.; Wang, X.; and Tang, X. 2016. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1096–1104.

Longhini, A.; Wang, Y.; Garcia-Camacho, I.; Blanco-Mulero, D.; Moletta, M.; Welle, M.; Alenyà, G.; Yin, H.; Erickson, Z.; Held, D.; et al. 2024. Unfolding the literature: A review of robotic cloth manipulation. *Annual Review of Control, Robotics, and Autonomous Systems*, 8.

Lu, H.; Wu, R.; Li, Y.; Li, S.; Zhu, Z.; Ning, C.; Zhao, Y.; Luo, L.; Chen, Y.; and Dong, H. 2024. Garmentlab: A unified simulation and benchmark for garment manipulation. *Advances in Neural Information Processing Systems*, 37: 11866–11903.

Miguel, E.; Bradley, D.; Thomaszewski, B.; Bickel, B.; Matusik, W.; Otaduy, M. A.; and Marschner, S. 2012. Data-Driven Estimation of Cloth Simulation Models. *Comput. Graph. Forum (Eurographics)*, 31(2): 519–528.

Müller, M. 2008. Hierarchical Position Based Dynamics. In *Workshop on Virtual Reality Interaction and Physical Simulation (VRIPHYS)*.

Müller, M.; Chentanez, N.; Kim, T. Y.; and Macklin, M. 2014. Strain Based Dynamics. In *Proceedings of SCA*, 149–157.

Müller, M.; Heidelberger, B.; Teschner, M.; and Gross, M. 2005. Meshless Deformations Based on Shape Matching. *ACM Trans. Graph. (SIGGRAPH)*, 24(3): 471–478.

NVIDIA Corporation. 2021–2024. NVIDIA Isaac Sim. https://developer.nvidia.com/isaac/sim.

Peng, Y.; Deng, B.; Zhang, J.; Geng, F.; Qin, W.; and Liu, L. 2018. Anderson Acceleration for Geometry Optimization and Physics Simulation. *ACM Trans. Graph. (SIGGRAPH)*, 37(4): 42:1–42:14.

Rasheed, A. H.; Romero, V.; Bertails-Descoubes, F.; Wuhrer, S.; Franco, J.-S.; and Lazarus, A. 2021. A Visual Approach to Measure Cloth-Body and Cloth-Cloth Friction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1.

Ren, T.; Liu, S.; Zeng, A.; Lin, J.; Li, K.; Cao, H.; Chen, J.; Huang, X.; Chen, Y.; Yan, F.; Zeng, Z.; Zhang, H.; Li, F.; Yang, J.; Li, H.; Jiang, Q.; and Zhang, L. 2024. Grounded SAM: Assembling Open-World Models for Diverse Visual Tasks. arXiv:2401.14159.

Sanchez, J.; Corrales, J.-A.; Bouzgarrou, B.-C.; and Mezouar, Y. 2018. Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey. *The International Journal of Robotics Research*, 37(7): 688–716.

Tang, M.; Manocha, D.; Yoon, S.-E.; Du, P.; Heo, J.-P.; and Tong, R.-F. 2011. VolCCD: Fast Continuous Collision Culling between Deforming Volume Meshes. *ACM Trans. Graph.*, 30(5): 111:1–111:15.

Tang, M.; Wang, H.; Tang, L.; Tong, R.; and Manocha, D. 2016. CAMA: Contact-Aware Matrix Assembly with Unified Collision Handling for GPU-Based Cloth Simulation. *Comput. Graph. Forum (Eurographics)*, 35(2): 511–521.

Todorov, E.; Erez, T.; and Tassa, Y. 2012. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ international conference on intelligent robots and systems*, 5026–5033. IEEE.

Volino, P.; and Magnenat-Thalmann, N. 2006. Resolving Surface Collisions through Intersection Contour Minimization. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, 1154–1159.

Volino, P.; Magnenat-Thalmann, N.; and Faure, F. 2009. A Simple Approach to Nonlinear Tensile Stiffness for Accurate Cloth Simulation. *ACM Trans. Graph.*, 28(4): 105:1–105:16.

Wang, H. 2015. A Chebyshev Semi-Iterative Approach for Accelerating Projective and Position-Based Dynamics. *ACM Trans. Graph. (SIGGRAPH Asia)*, 34(6): 246:1–246:9.

Wang, H.; O'Brien, J. F.; and Ramamoorthi, R. 2011. Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Trans. Graph. (SIGGRAPH)*, 30(4): 71:1–71:12.

Wang, Y.; Wu, R.; Chen, Y.; Wang, J.; Liang, J.; Zhu, Z.; Geng, H.; Malik, J.; Abbeel, P.; and Dong, H. 2025. DexGarmentLab: Dexterous Garment Manipulation Environment with Generalizable Policy. *arXiv preprint arXiv:2505.11032*.

Wu, B.; Wang, Z.; and Wang, H. 2022. A GPU-Based Multilevel Additive Schwarz Preconditioner for Cloth and Deformable Body Simulation. *ACM Trans. Graph. (SIGGRAPH)*, 41(4).

Wu, L.; Wu, B.; Yang, Y.; and Wang, H. 2020. A Safe and Fast Repulsion Method for GPU-Based Cloth Self Collisions. *ACM Trans. Graph.*, 40(1).

Yang, S.; Liang, J.; and Lin, M. C. 2017. Learning-Based Cloth Material Recovery from Video. In *IEEE International Conference on Computer Vision*, 4393–4403.

Zhang, Y.; Liu, F.; Liang, X.; and Yip, M. 2024. Achieving autonomous cloth manipulation with optimal control via differentiable physics-aware regularization and safety constraints. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 9931–9938. IEEE.

Zhou, B.; Zhou, H.; Liang, T.; Yu, Q.; Zhao, S.; Zeng, Y.; Lv, J.; Luo, S.; Wang, Q.; Yu, X.; et al. 2023. Clothesnet: An information-rich 3d garment model repository with simulated clothes environment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 20428–20438.