

Dressing in Motion: A Human Motion-Aware Diffusion Policy for Robot-Assisted Dressing

Haoxiang Sun¹, Fangyuan Wang¹, Songhao Huang¹, Justina Y. W. Liu²,
Jihong Zhu³, Peng Zhou⁴, and David Navarro-Alarcon¹

Abstract—Robotic dressing assistance is a promising solution for supporting older adults with physical impairments in daily living. However, dressing under human motion remains challenging, as complex garment–human contact and occlusions make it difficult to generate actions aligned with arm movements. In this letter, we propose a visuomotor policy that learns dressing skills from static expert demonstrations and generalizes to dynamic user-motion scenarios. A diffusion policy tailored to garment–human interaction geometry learns from partially observed point clouds with varied arm postures. We then introduce an object-centric representation based on PDE diffusion to capture the axial distribution of the arm. By sampling motion-relevant regions and registering them across consecutive observations, the proposed method approximates arm motion and reactively adapts the executed trajectory. We evaluate our method in simulation and a real-world human study involving nine participants, three garment types, and six arm-motion patterns. Results show that our method outperforms baselines in dressing progress, freedom of movement, and user comfort. The project website is <https://anonymous.4open.science/w/dressing-in-motion/>.

I. INTRODUCTION

Dressing is an essential component of activities of daily living (ADLs). Among ADLs, dressing is a physically intensive and repetitive task involving frequent body interaction. For older adults and individuals with physical impairments, dressing often requires assistance from professional caregivers. However, global population aging has intensified the shortage of trained caregivers, making robotic dressing a promising solution for autonomous assistance. In practical scenarios, users may naturally adjust their arm posture due to fatigue or discomfort, requiring the robot to adapt to human motion during assistance. In this work, we study a human motion-aware policy for robotic dressing to support a less restrictive and more user-accommodating assistance process.

In the context of garment manipulation, robotic dressing assistance remains particularly challenging. On the one hand, fabrics are deformable objects that lack compact state representations, and deformation estimation is difficult due to the infinite degrees of freedom and nonlinear dynamics. On the other hand, compared with conventional fabric manipulation tasks such as garment grasping [1] and folding [2], dressing is more contact-rich and interaction-intensive, as the garment usually needs to slide closely along the

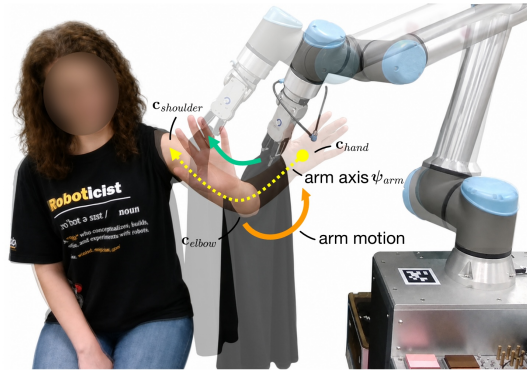


Fig. 1. Robot-assisted dressing under human arm motion. The robot advances the garment along the arm axis while adapting to time-varying arm geometry.

body surface. Unpredictable occlusions substantially reduce the accuracy of sleeve positioning. Meanwhile, excessive garment entanglement, tension, or potential contact between the gripper and human body may cause discomfort and even safety risks.

Many prior studies on robot-assisted dressing have made substantial progress by using vision and/or force feedback to explicitly estimate human–garment contact states or implicitly learn action distributions [3], [4]. In parallel with these advances, reliable dressing behaviors remain difficult to fully specify through low-dimensional state and underparameterized reward functions, making learning from physical demonstrations a more natural paradigm for acquiring dressing skills. Recent advances in data-driven learning have shown great potential for deformable object manipulation, with representative examples including robotic foundation models [5] and generative AI models [6] that learn transferable policy distributions from high-quality expert datasets.

However, most existing robotic dressing approaches are conducted under a static-arm assumption, where the subject is required to keep the arm fixed throughout the dressing process [7], [8], [9], [10]. This assumption provides a strong prior for handling visual occlusions, since the human-skeleton topology, including the hand, elbow, and shoulder [11], or even the full arm point-cloud distribution [12], can be inferred from the initial configuration in some settings. Nevertheless, this simplification inevitably compromises the user experience. In an ideal dressing process, users should be allowed to freely adjust their arm posture when they feel fatigue or need to move, while the robot should compliantly adapt to human motion and complete the dressing task accordingly.

¹Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China. ²School of Nursing, The Hong Kong Polytechnic University, Hong Kong SAR, China. ³School of Physics, Engineering and Technology, University of York, UK. ⁴School of Advanced Engineering, Great Bay University, Dongguan, Guangdong, China. Corresponding author: dna@ieee.org

In this work, we address the problem of robotic dressing under human arm motion. Our key idea is to combine a visuomotor diffusion policy learned from static-arm demonstrations with a real-time adaptive control strategy for dynamic human postures. The proposed dressing diffusion policy is designed to be sensitive to human–garment interaction geometry by utilizing partially observed point-cloud inputs. We introduce an object-centric local-frame representation [13] for the arm surface and use it to automatically sample motion-relevant regions from the arm point cloud. These sampled regions are registered online to estimate user motion and project the learned dressing actions, enabling the policy to adapt to human dynamics during execution. We note that comparing visual and force-based sensing modalities is not the focus of this paper, while incorporating force feedback can serve as a future extension of the proposed method.

The contributions of this paper are as follows:

- A diffusion policy tailored to garment–human interaction geometry for robotic dressing.
- A PDE-diffused scalar field that forms an arm-axis representation to localize motion-relevant regions from partially observed arm point clouds.
- An online trajectory adaptation scheme that estimates arm motion through point registration and projects actions for motion-compatible execution.
- Comprehensive simulation and real-world evaluations across diverse garment types and human motion patterns, demonstrating improved dressing progress, motion tolerance, and user ergonomics.

II. RELATED WORK

A. Robot-Assisted Dressing

Robot-assisted dressing has received considerable research attention. Early works studied user modeling to enable personalized dressing [12], [14], [15]. Subsequent studies explored the incorporation of visual observations and/or force feedback as inputs, either by using model-based methods to learn garment–human dynamics and integrate them with control strategies [10], [16], or by directly inferring action policies within reinforcement learning [3], [9], [17] and imitation learning frameworks [18], [19], [20]. However, these approaches invariably rely on the common assumption that the human arm remains static during dressing. While this assumption simplifies the inference of human-skeleton topology under occlusion, it also limits user comfort and natural interaction. More recent work [21] has begun to address dressing assistance for humans with arm motion in unstructured scenarios, by fine-tuning a simulation-trained reinforcement learning policy with real-world force feedback. However, real-time adaptation to human motion remains limited. Our work differs by contributing an imitation learning framework that learns from static demonstrations while maintaining online awareness of arm-motion variability during execution.

B. Learning Deformable Object Manipulation Skills

Dressing is inherently a contact-rich deformable object manipulation (DOM) task involving close human–robot interaction, where accurate state estimation is challenging due to the nonlinear dynamics and infinite degrees of freedom of deformable materials. For this human-mastered skill, reinforcement learning methods require faithful digital twins and reliable sim-to-real transfer, both remain challenging in practice. In contrast, learning from expert demonstrations provides a promising paradigm for acquiring action policies while bypassing explicit physical modeling. Learning manipulation skills from demonstrations has evolved from model-based regression [22] and behavior cloning [23] to generative modeling approaches [24], [25], which have demonstrated strong capability in learning complex data distributions. Among them, diffusion models have emerged as a powerful tool for constructing visuomotor policies that capture multimodal action distributions in DOM [6], including tissue shape control [26] and fabric folding [27].

In this work, we focus on robot-assisted dressing with human motion, an object-centric DOM task that involves physical interaction with the arm surface and requires adaptation to variations in the spatial position and local curvature of the contact surface [28]. Improper contact can lead to entanglement around the hand and elbow. However, prior work studied shape control of deformable objects in structured environments, where environmental contact is typically static or ignored. In addition, such physical interaction results in cluttered visual observations under occlusion. For this reason, our approach is tailored to garment–human interaction geometry while modeling the target distribution of desired dressing actions under the guidance of arm-motion topology.

III. PROBLEM FORMULATION AND ASSUMPTIONS

We formulate robotic dressing under human arm motion as a visuomotor policy learning problem modeled as a partially observable Markov decision process (POMDP). Let $\mathbf{c}_{\text{sleeve},t}$ and $\mathbf{c}_{\text{hand},t}$ denote the sleeve-opening center and hand position at time t , respectively, and let $\psi_{\text{arm},t}$ denote the curvilinear arm centerline extending from the hand toward the shoulder through the medial axis of the approximately cylindrical arm geometry, as shown in Fig. 1. Differing from approaches that begin with the arm already inserted into the sleeve, the robot starts from an initial pose while grasping the garment in front of the arm, without requiring pre-alignment between $\mathbf{c}_{\text{sleeve},t}$ and $\mathbf{c}_{\text{hand},t}$. It first reduces the hand–sleeve alignment error $\|\mathbf{c}_{\text{sleeve},t} - \mathbf{c}_{\text{hand},t}\|$ and orients the sleeve for insertion onto the forearm, thereby resolving the alignment and orientation constraints during the pre-insertion phase.

The task objective is to progressively advance the garment sleeve over the hand and elbow until it reaches the shoulder while accommodating time-varying arm geometry. Given a partial point cloud $\mathbf{P}_t$, the goal is to learn a policy π_θ that maps the observation embedding $\mathbf{o}_t$, constructed from $\mathbf{P}_t$ and the robot state, to an action distribution $\mathbf{a}_t \sim \pi_\theta(\cdot | \mathbf{o}_t)$. Let $\mathcal{Y}_{h,t} = \{\mathbf{c}_{\text{shoulder},t}, \mathbf{c}_{\text{elbow},t}, \mathbf{c}_{\text{hand},t}\}$ denote the arm landmarks, and let $\chi_{h,t}$ denote the arm joint configuration. The arm

pose is represented as $\mathbf{q}_{h,t} = (\mathcal{V}_{h,t}, \mathcal{X}_{h,t})$, and t_{ins} denotes the sleeve insertion time. During the pre-insertion phase, the arm remains static, such that $\mathbf{q}_{h,t} = \mathbf{q}_{h,0}$ for $t < t_{\text{ins}}$. After insertion, the participant may vary $\mathbf{q}_{h,t}$ for $t \geq t_{\text{ins}}$, resulting in time-varying arm geometry during garment advancement.

IV. METHOD

Method Overview: As shown in Fig. 2, our method follows a hierarchical framework. At the high level, a diffusion model learns a general motion policy from static demonstrations covering variations in arm poses and garment types. To further address dynamic dressing, we construct a diffused scalar field from occluded point clouds to obtain a topologically continuous coordinate along the arm axis, and perform point registration to estimate transformations of the arm pose, which are then used to align the generated motion with the current arm configuration and reactively respond to non-Markovian arm motions. During execution, the diffusion policy updates action chunks at a lower frequency, while trajectory adaptation (TA) runs at a higher rate to generate end-effector commands.

A. Diffusion Policy Learning from Demonstrations

The first core of this work is to learn a visuomotor policy for robot-assisted dressing from expert demonstrations. Visual observations in this task inevitably contain entanglement and occlusions between the garment and the arm. We introduce the dressing diffusion policy, which learns a distribution over multimodal actions by training a diffusion policy that iteratively denoises actions $\mathbf{a}$ conditioned on occluded visual features $\mathbf{s}$. For the observation space $\mathcal{O}$, let $\mathcal{P}$ denote the human–garment point space, and $\mathbf{P}_t \subset \mathcal{P}$ is the arm–garment point set cropped at time t from a rectangular region surrounding the dressing-side arm. The policy observation $\mathbf{o}_t$ is defined as a latent feature that fuses the point-cloud feature extracted from $\mathbf{P}_t$ with the encoded end-effector pose $\mathbf{e}_t$. The points $\mathbf{P}_t$ are extracted from a rectangular region around the dressing-side arm and contains a mixture of human and garment points. At each time step t , the points are downsampled by farthest point sampling (FPS) to preserve spatial coverage with a fixed point-set size, without additional segmentation.

To encode human–garment interaction geometry and extract fine-grained 3D features, the sampled points are processed by a lightweight encoder consisting of an EdgeConv layer [29]. For each point $\mathbf{p}_i$, the EdgeConv feature is computed as:

$$\mathbf{h}_i = \max_{j \in \mathcal{N}_K(i)} \phi_\eta([\mathbf{p}_i, \mathbf{p}_j - \mathbf{p}_i]) \quad (1)$$

where $\mathcal{N}_K(i)$ denotes the K -nearest-neighbor set of $\mathbf{p}_i$ in the sampled point cloud, and ϕ_η is a shared MLP applied to each edge feature. This operation combines the point position $\mathbf{p}_i$ with local offsets $\mathbf{p}_j - \mathbf{p}_i$ over the neighborhood graph to capture local interaction geometry between the human arm and garment. After max aggregation over neighboring edge features, the EdgeConv layer produces a point-feature tensor of shape $N \times d_h$, where N is the number of sampled

points and d_h is the output feature dimension. The resulting point-cloud feature is projected and fused with the encoded end-effector pose feature to obtain $\mathbf{o}_t$, which conditions a diffusion model that iteratively denoises Gaussian noise $\mathbf{a}^k$ into actions $\mathbf{a}^0$:

$$\mathbf{a}^{k-1} = \sqrt{\bar{\alpha}^{k-1}} \hat{\mathbf{a}}_\theta(\mathbf{o}_t, \mathbf{a}^k, k) + \sqrt{1 - \bar{\alpha}^{k-1}} \epsilon_\theta(\mathbf{o}_t, \mathbf{a}^k, k) \quad (2)$$

where ϵ_θ is the noise prediction network and

$$\hat{\mathbf{a}}_\theta = \frac{\mathbf{a}^k - \sqrt{1 - \bar{\alpha}^k} \epsilon_\theta(\mathbf{o}_t, \mathbf{a}^k, k)}{\sqrt{\bar{\alpha}^k}}. \quad (3)$$

Here, DDIM [30] is used as the sampling scheduler for the reverse denoising process. The term $\bar{\alpha}^k = \prod_{s=1}^k \alpha_s$ denotes the cumulative product of the predefined variance schedule. The network is trained with the standard noise-prediction objective:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{a}_0, \mathbf{o}_t, k, \epsilon} [\|\epsilon - \epsilon_\theta(\mathbf{o}_t, \mathbf{a}^k, k)\|_2^2]. \quad (4)$$

B. Arm-Centric Diffused Scalar Field

The diffusion model outputs each action chunk over a prediction horizon $T_p = 8$ at 15 Hz for downstream processing. During dressing, the subject is allowed to adjust the arm pose, which requires the policy to respond at over 50 Hz with an action execution horizon $T_a = 4$. Unlike prior approaches that reconstruct the full arm geometry under occlusion, which is not assumed under changing arm configurations. Instead, we extract an effective region from partially observed point clouds for motion adaptation. This is based on the practical premise that the dressing outcome is mainly governed by the interaction between the robot and the uncovered arm region $\mathcal{U} \subset \mathcal{P}$. In particular, the region of interest $\mathcal{R} \subset \mathcal{U}$ is defined as the outermost uncovered arm points closest to the current garment boundary along the arm axis $\psi_{\text{arm},t}$, which largely determines the sleeve–arm insertion behavior. Guided by this premise, the motion-adaptation problem is reformulated as identifying $\mathcal{R}$ from cluttered point clouds and using it to approximate inter-frame arm motion.

Specifically, accurately sampling $\mathcal{R}$ from occluded visual observations entangled with multiple objects is nontrivial. For this reason, we construct a scalar field to obtain a geometry-aware axial distribution of the observed arm surface by introducing a point sampling method based on PDE diffusion [13]. The scalar field $u(\cdot, \tau) : \mathcal{P} \rightarrow \mathbb{R}$ is defined over the observed points, where τ denotes the diffusion time. The smoothness of u follows the Dirichlet energy principle, which penalizes large spatial gradients and encourages neighboring arm-surface points to have consistent scalar values. In the continuous setting, the energy minimizer under boundary conditions satisfies Laplace’s equation:

$$\Delta u(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega \quad (5)$$

where Ω denotes the underlying arm surface. The gradient flow of this energy yields the continuous diffusion process, which is discretized on the observed point cloud using the point-cloud Laplacian:

$$\partial_\tau u(\mathbf{p}, \tau) = \Delta_P u(\mathbf{p}, \tau), \quad \mathbf{p} \in \mathcal{P} \quad (6)$$

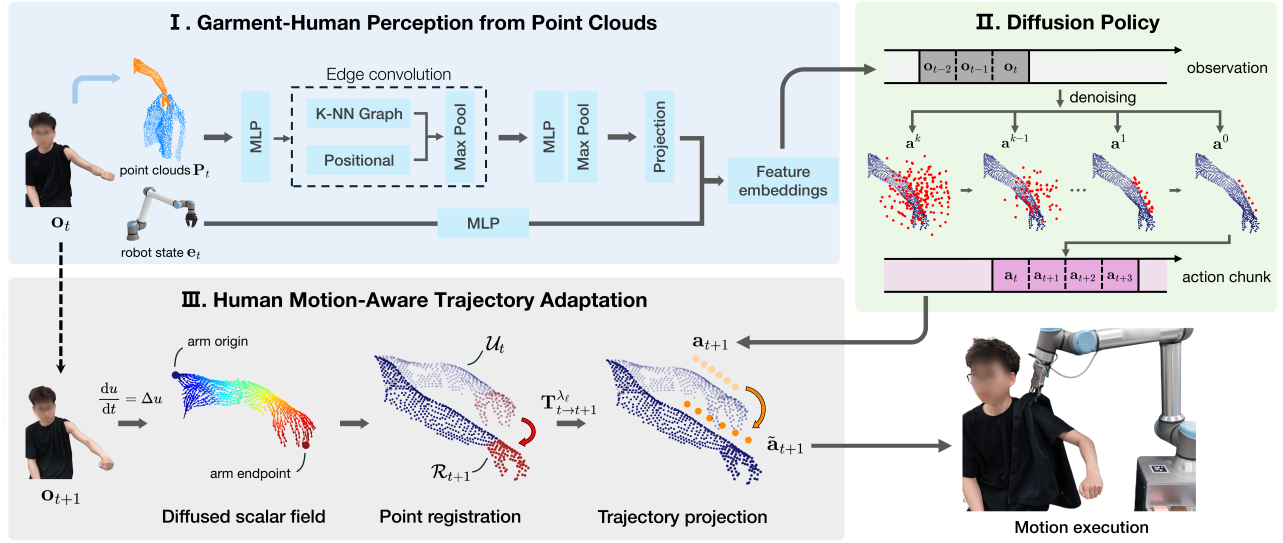


Fig. 2. Overview of the proposed method. *Top left*: Visual encoder extracts point-cloud features and fuses them with the robot state to obtain the observation $\mathbf{o}_t$. *Top right*: A diffusion policy denoise Gaussian noise to predict an action chunk. *Bottom*: Diffused scalar field identifies the region $\mathcal{R}_{t+1}$, which is registered to the reference region $\mathcal{R}_t$ to estimate arm motion. The estimated transformation is applied to generate the adapted execution trajectory.

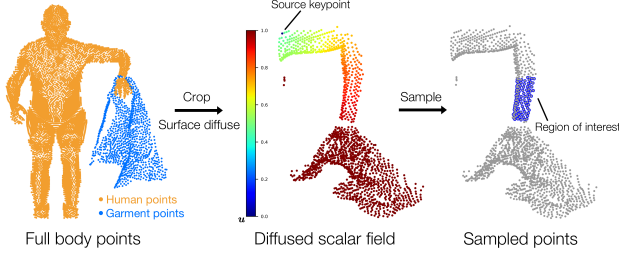


Fig. 3. Diffused scalar-field construction and ROI sampling. Surface diffusion assigns each arm point a scalar value u encoding its geodesic position along the arm axis. The uncovered arm region $\mathcal{U}$ is extracted, from which the region of interest $\mathcal{R}$ is sampled.

where $\Delta_{\mathcal{P}}$ denotes the point-cloud Laplacian. To propagate the local axial distribution of the arm, we specify the shoulder vertex as a source keypoint $\mathbf{s}$ acting as a physical repulsor, with the initial condition $u(\mathbf{s}, 0) = 1$ and $u(\mathbf{p}, 0) = 0$ for all $\mathbf{p} \in \mathcal{P} \setminus \mathbf{s}$. The diffusion equation is solved by implicit time stepping as:

$$\mathbf{u}_{\tau} = (\mathbf{M} - \tau \mathbf{C})^{-1} \mathbf{M} \mathbf{u}_0 \quad (7)$$

where $\mathbf{u}_0, \mathbf{u}_{\tau} \in \mathbb{R}^{|\mathcal{P}|}$ denote the point-wise scalar values before and after diffusion. $\mathbf{C}$ is the weak Laplacian matrix constructed from cotangent weights of local tangent-plane triangulations [31], and $\mathbf{M}$ is the diagonal mass matrix determined by the areas of triangles adjacent to each point.

The sampling procedure based on the diffused scalar field is illustrated in Fig. 3. We adjust τ to perform short-time diffusion, which yields a scalar field sensitive to geodesic distance. By imposing continuity and smoothness constraints on the diffused scalar field along the arm axis from the shoulder source vertex, we reject garment points and wrinkle-induced artifacts, yielding the uncovered arm region $\mathcal{U}$.

C. Arm-Motion-Aware Trajectory Adaptation

The action chunk $\mathbf{A}_t = \{\mathbf{a}_t, \dots, \mathbf{a}_{t+T_a-1}\}$ is predicted at time t and contains desired end-effector trajectories ex-

pressed in the arm-centric frame associated with $\mathbf{o}_t$. At each execution step $\ell = 0, \dots, T_a - 1$, the diffused scalar field is used to sample the current region of interest $\mathcal{R}_{t+\ell}$. We then register the reference region $\mathcal{R}_t$, obtained at prediction time, to the current region $\mathcal{R}_{t+\ell}$ using generalized iterative closest point (GICP). Let $\mathcal{M}$ denote the set of correspondences between $\mathbf{p}_i^t \in \mathcal{R}_t$ and $\mathbf{p}_j^{t+\ell} \in \mathcal{R}_{t+\ell}$. For each $(i, j) \in \mathcal{M}$, the residual is defined as $\mathbf{r}_{ij} = \mathbf{p}_j^{t+\ell} - (\mathbf{R}\mathbf{p}_i^t + \delta)$. GICP estimates the inter-frame transformation by solving

$$\mathbf{T}_{t \rightarrow t+\ell} = \arg \min_{\mathbf{T} \in SE(3)} \sum_{(i,j) \in \mathcal{M}} \mathbf{r}_{ij}^{\top} \left(\Sigma_j^{t+\ell} + \mathbf{R} \Sigma_i^t \mathbf{R}^{\top} \right)^{-1} \mathbf{r}_{ij} \quad (8)$$

where $\mathbf{T} = (\mathbf{R}, \delta)$ comprises the rotation $\mathbf{R} \in SO(3)$ and translation $\delta \in \mathbb{R}^3$, and Σ_i^t and $\Sigma_j^{t+\ell}$ are the local covariance matrices of the matched points. Restricted to the motion-relevant arm region, $\mathbf{T}_{t \rightarrow t+\ell}$ provides a local rigid estimate of the inter-frame arm displacement. We use this transformation matrix to project each pending position target through a bounded correction scheme:

$$\begin{bmatrix} \tilde{\mathbf{d}}_{t+\ell} \\ 1 \end{bmatrix} = \mathbf{T}_{t \rightarrow t+\ell}^{\lambda_{\ell}} \begin{bmatrix} \mathbf{d}_{t+\ell} \\ 1 \end{bmatrix} \quad (9)$$

where

$$\begin{aligned} \mathbf{T}_{t \rightarrow t+\ell}^{\lambda_{\ell}} &= \text{Exp}(\lambda_{\ell} \text{Log}(\mathbf{T}_{t \rightarrow t+\ell})) \\ \lambda_{\ell} &= \min \left(1, \frac{\rho}{\|\text{Log}(\mathbf{T}_{t \rightarrow t+\ell})\| + \varepsilon} \right). \end{aligned} \quad (10)$$

The projected action is

$$\tilde{\mathbf{a}}_{t+\ell} = \begin{bmatrix} \tilde{\mathbf{d}}_{t+\ell} \\ \varphi_{t+\ell} \end{bmatrix}, \quad (11)$$

where $\mathbf{d}_{t+\ell}$ and $\varphi_{t+\ell}$ denote the predicted nominal end-effector position and orientation, respectively. The scalar λ_{ℓ} bounds the correction magnitude by the hyperparameters ρ and ε to avoid abrupt motion updates. This projection adapts the translational component of the end-effector target according to the estimated local arm displacement while preserving the predicted orientation.

V. RESULTS

A. Numerical Simulation

We evaluate our method in the Assistive Gym simulation environment [32]. Four human body models spanning genders and body-shape variations are configured, corresponding to male/female and slim/heavy morphologies. For each model, the arm posture is controlled by two revolute joints located at the shoulder and elbow, each with a motion range of 45° . Three garment types with different sleeve configurations are cropped from the Cloth3D dataset [33] and imported into the simulator. Representative simulation trials are shown in Fig. 4.

In terms of arm motion, we evaluate both static and dynamic conditions in simulation by defining ten arm motion patterns. Eight motions are generated from planar arm displacements in different directions, while the remaining two correspond to clockwise and counterclockwise rotational arm movements. In addition, we evaluate three arm motion speed profiles determined by the joint angular velocity v : static motion with $v = 0$, moderate-speed motion with $v = 1.0$ rad/s, and fast motion with $v = 2.0$ rad/s.

We compare the proposed method and its ablated variant against several baselines:

- **Ours**: the full proposed framework.
- **Ours w/o TA**: an ablated variant of our method without the human motion-aware trajectory adaptation module.
- **DP3** [34]: a point-cloud-conditioned diffusion policy for general robotic manipulation, used to evaluate the effect of our interaction-aware point-cloud encoding.
- **Diff-MPC** [10]: a diffusion garment-dynamics MPC baseline with its original one-hot action encoding replaced by end-effector action feature encoding for fair comparison.
- **DP-image** [6]: an image-conditioned diffusion policy.
- **BC-LSTM** [35]: a recurrent behavior cloning baseline that models temporal dependencies with an LSTM.

Training. In training, all imitation learning methods are trained on the dataset containing 180 trajectories obtained through teleoperation in simulation. For Diff-MPC, we follow its original model-based reinforcement learning setting and train it with 750 trajectories. During evaluation, each trial is limited to a maximum of 500 execution steps. Performance is evaluated using two metrics: sleeve insertion success and dressing ratio, defined as the ratio of the dressed arm length to the total arm length. For each benchmark and arm-motion speed profile, we conduct 70 trials with randomized garment-human initializations. Results are reported across all combinations of arm motions, garment types, and human body models.

Fig. 5 shows the simulation results over all 1470 trials. Under static conditions, several methods achieve reasonable performance, with point-cloud conditioned diffusion policies outperforming image-based policies by using richer 3D geometric cues. As arm motion is introduced and the motion speed increases, most baselines show degraded performance. Since these methods lack explicit motion adaptation, hand

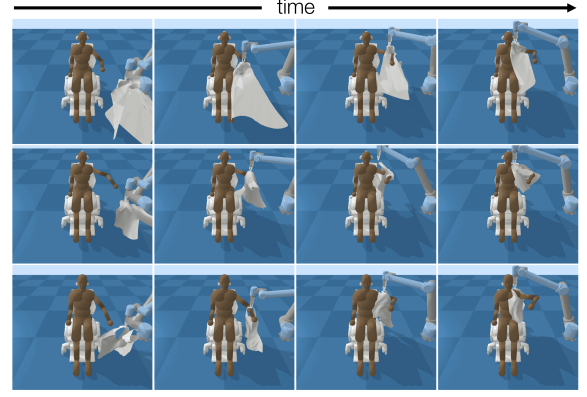


Fig. 4. Examples of dressing process with different garment types and human models in the simulation environment.

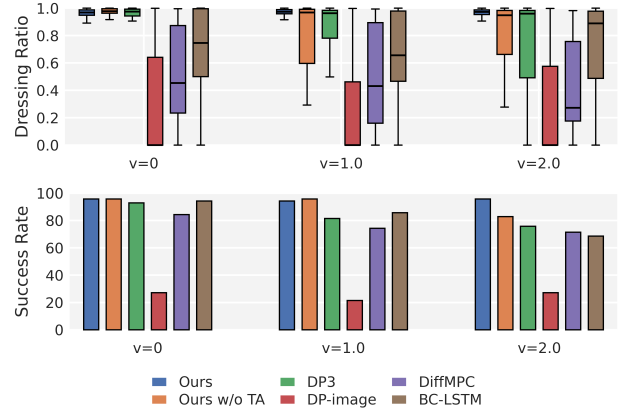


Fig. 5. Simulation performance under different arm-motion speeds. For each benchmark and speed profile, results are computed over 70 randomized trials. *Top*: dressing ratio. *Bottom*: sleeve insertion success rate.

motion can pull the arm away from the garment opening during alignment, causing insertion failure or elbow entanglement. The comparison between Ours and Ours w/o TA further shows that trajectory adaptation provides consistent gains across different movement types, indicating its role in handling arm-motion variability. In addition, the comparison between Ours w/o TA and DP3 demonstrates the benefit of the proposed interaction-aware point-cloud encoding for generalization across garments and arm poses. The full proposed method remains robust to human motion, achieving an average dressing ratio and sleeve insertion success rate above 0.95.

Diffusion-Time Robustness. We evaluate the robustness of the diffused scalar field and point-cloud registration process under different diffusion times τ , as shown in Fig. 6. The evaluation is conducted over 50 randomized trials. Scalar RMSE quantifies the alignment consistency of the rank-normalized diffusion scalar field by comparing matched current-reference arm points after registration. Registration RMSE measures geometric alignment error using the symmetric nearest-neighbor RMS distance between the registered current arm cloud and the reference arm cloud. In principle, increasing τ produces a smoother scalar field that is resilient to noise, whereas short-time diffusion better

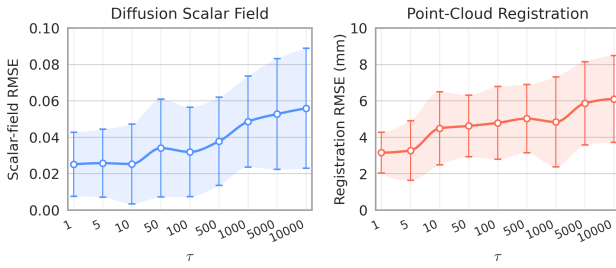


Fig. 6. Error bar plots of the diffused scalar field and point-cloud registration under different diffusion times τ . The plots report the mean RMSE over 50 independent randomized trials, with error bars indicating standard deviations.

preserves gradient information along the arm axis. The mean RMSE results indicate that the proposed point sampling and registration process remains robust across diffusion times.

B. Hardware Experiments

Fig. 7 shows the real-world experimental setup¹. We use a UR10e robot for dressing, with the participant seated to its left. A RealSense D435i camera is mounted above and in front of the participant to capture visual observations. The garments include a vest, a shirt, and a windbreaker. The human study comprises a non-scripted motion test for reactive adaptation and a canonical motion test under representative daily arm-motion patterns. Simulation and real-world experiments use independently collected datasets and separately trained policies, with no sim-to-real transfer.

Canonical motion test. In this test, we evaluate all benchmarks defined in Sec. V-A under six canonical arm-motion patterns: up-and-down, side-to-side, back-and-forth, waving, phone use, and object receiving. During data collection, we use a 3Dconnexion SpaceMouse for teleoperation and collect 210 expert demonstrations. These demonstrations do not involve arm motion; all trajectories are collected from static subjects with varied arm configurations. To reduce contact risk during execution, we set safety thresholds of 20 kg · m/s for joint momentum and 120 N for tool force.

We recruit nine participants, including three females and six males, with heights ranging from 162 cm to 185 cm. Considering potential physical fatigue, each participant performs 28 trials with a random assignment of two garment types and two arm-motion patterns, resulting in 36 trials per method with balanced assignments across garments and motion patterns. For each assigned condition, the same motion sequence is repeated across all methods. Each trial consists of two stages: an initial static insertion phase and a subsequent canonical motion phase. The trial is terminated when invalid garment-body contact occurs or when the robot remains stalled for more than 5 seconds. We then measure the covered arm region to compute the dressing ratio. Example demonstrations of dressing with the proposed method are shown in Fig. 8.

Subjective Ratings. Two subjective ratings are assessed: freedom of movement and comfort. Both criteria are mea-

¹This study was approved by the institutional ethics review board. Approval details are omitted for anonymous review and will be provided in the final version.



Fig. 7. *Top left*: Real-world experimental setup. *Top right*: garments with different sleeve lengths. *Bottom*: motion patterns performed by the participants.

TABLE I

DRESSING RATIO ACROSS GARMENTS

Method	Vest	Shirt	Windbreaker
Ours	0.92	0.89	0.83
Ours w/o TA	0.93	0.92	0.50
DP3	0.79	0.58	0.67
DP-image	0.45	0.33	0.25
Diff-MPC	0.14	0.30	0.27
BC-LSTM	0.58	0.50	0.46

TABLE II

DRESSING RATIO ACROSS MOTION PATTERNS

Method	Up-down	Side2side	Back-forth	Wave	Phone	Object
Ours	1.00	1.00	1.00	1.00	0.77	0.50
Ours w/o TA	0.83	0.83	0.83	0.83	0.76	0.60
DP3	0.50	0.83	0.50	1.00	0.75	0.50
DP-image	0.31	0.17	0.50	0.50	0.33	0.25
Diff-MPC	0.29	0.42	0.20	0.20	0.12	0.20
BC-LSTM	1.00	0.17	0.50	0.92	0.49	0.00

sured on a six-level ordinal scale, with scores uniformly assigned from 1.0 to 0.0 in increments of 0.2, where higher scores indicate better performance. Freedom of movement evaluates the degree to which the subject’s arm motion is constrained during dressing, as restrictions may arise from robot motion or garment pulling. A policy with better awareness of human motion is expected to achieve a higher score. Comfort evaluates the perceived contact quality during dressing, mainly reflecting the contact forces between the human body, the gripper, and the garment. More accurate dressing trajectories generally lead to smoother and more comfortable contact.

Fig. 9 reports the real-world dressing outcomes. Tables I and II summarize the dressing ratios of all methods across garment types and motion patterns, and Fig. 10 reports the subjective ratings of freedom of movement and comfort. From these results, we draw the following conclusions. Consistent with the simulation results, our method outperforms all baselines in dressing progress, tolerance to human motion, and user-perceived comfort. This demonstrates that our policy successfully deployed in dynamic dressing scenarios that mostly have distribution shift, achieving a dressing ratio of 0.88 ± 0.105 and a success rate of $89\% \pm 10.4\%$ over all 252 trials. Moreover, experimental results indicates that the trajectory adaptation module adaptively adjusts the demon-

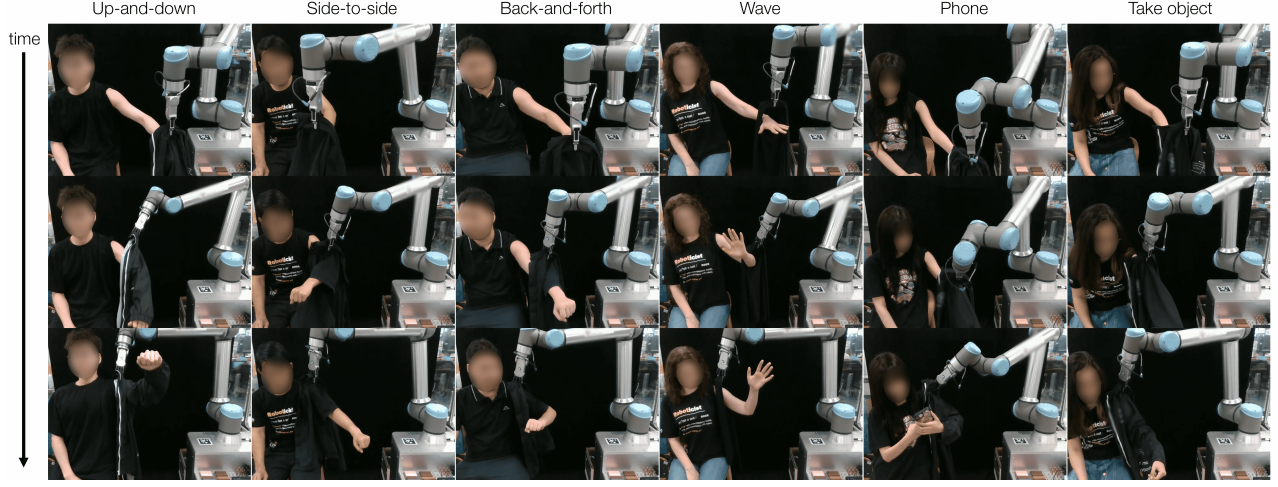


Fig. 8. Snapshots of real-world dressing trials using the proposed method with different participants. Participants select varied initial arm configurations, including arm orientations and shoulder and elbow joint angles. The policy first aligns the garment opening with the hand and then adaptively updates the execution trajectory in response to subsequent arm motion.

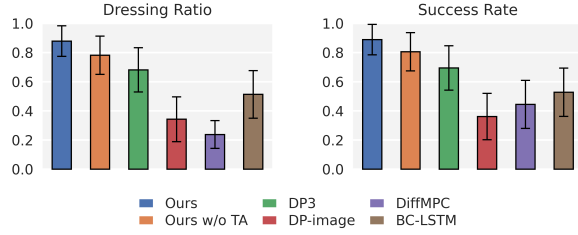


Fig. 9. Dressing outcomes of the real-world human study. *Left*: dressing ratio. *Right*: sleeve insertion success rate.

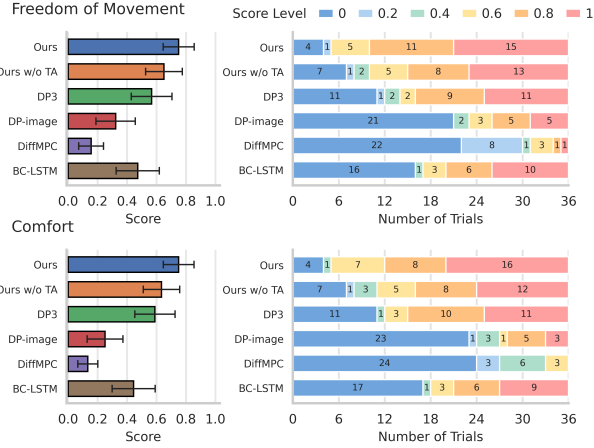


Fig. 10. Subjective rating results from the real-world human study. *Top*: freedom of movement. *Bottom*: comfort. *Left*: box plots of the rating results. *Right*: full distributions of participant ratings.

strated trajectories to arm motion and exhibits compliant-like behavior in most trials. As reflected by the participants, this reduces the perceived physical burden during dressing. In addition, Fig. 8 provides examples in which the learned policy exhibits dressing behaviors conditioned on human morphology and pose rather than a fixed motion pattern, including backward tilting of the gripper when the shoulder is raised in the up-and-down motion and oblique gripper adjustment for upward draping over the shoulder during phone use. In contrast, the baselines reveal different lim-

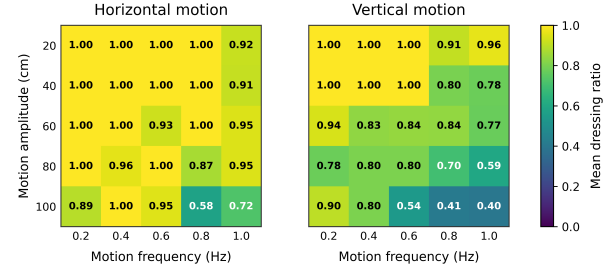


Fig. 11. Results of dressing with different arm motion amplitudes and speeds. *Left*: horizontal motion. *Right*: vertical motion.

itations. Behavior cloning methods tend to learn averaged motions, while Diff-MPC is affected by the sim-to-real gap in garment dynamics and usually produces excessive elbow contact during full-sequence execution. DP3 achieves the closest performance to our method among the baselines, but still lacks effective generalization to dynamic human motion.

Capability test. To evaluate the motion-tolerance envelope of the proposed method, we conduct an additional real-world test by sweeping arm-motion amplitude and frequency in two canonical motion planes. Specifically, we evaluate horizontal and vertical arm swings across five amplitudes of 20–100 cm and five frequencies of 0.2–1.0 Hz, resulting in 150 trials. Fig. 11 reports a clear performance degradation appears in the high-intensity region: when the arm-swing amplitude reaches 80–100 cm and the frequency reaches 0.8–1.0 Hz, the policy’s ability to generate motion-compatible actions starts to degrade, leading to reduced dressing progress.

Runtime. In our experiments, all runtime measurements are obtained on a NVIDIA RTX 5070 Laptop GPU. Following an asynchronous inference stack similar to [36], DDIM completes visual processing and action-chunk inference in less than 50 ms on average. In addition, for each frame, a 96 points region of interest $\mathcal{R}_t$ is sampled from the 1024 arm points $\mathcal{U}_t$ and registered to $\mathcal{U}_{t+1}$; the lightweight `small_gicp` [37] implementation completes this registration in less than 10 ms on average. This enables the overall action-generation pipeline to run online with

reactive updates.

Limitations. One limitation is that arm motion before garment-body interaction may perturb opening alignment with the hand and compromise sleeve insertion. This partly arises from the mismatch between limited control frequency and variable human motion. Second, mapping arm motion to end-effector offsets inherently introduces irregular displacement. The bounded correction in action projection must balance jitter suppression with the representation of the underlying arm-motion trend, which inevitably limits the tightness with which rapid or large-amplitude human motions can be tracked.

VI. CONCLUSION

In this letter, we propose a new visuomotor policy for robot-assisted dressing that accommodates human arm motion. A diffusion policy tailored to human-garment interaction is proposed for learning dressing actions from expert demonstrations conditioned on static arm configurations. We further introduce an arm-centric diffused scalar field based on PDE surface diffusion to characterize the local axial distribution of the arm. Based on this representation, a point-cloud sampling and registration method is developed to approximate arm motion through coordinate transformation and project actions online for motion-aligned trajectory adaptation. We evaluate the proposed method in simulation and real-world experiments involving nine participants, three garment types and six arm-motion patterns. The results show that our method generalizes effectively to novel arm configurations and user motions, outperforming six baselines in dressing progress, sleeve insertion success, freedom of movement, and user comfort.

REFERENCES

- [1] X. Zhu, H. Zhong, Y. Wu, S. Luo, and Y. Gao, "Guiding Robotic Cloth Grasping in Darkness: Infrared Semantic Segmentation and Grasping Position Selection," *IEEE Robot. Autom. Lett.*, vol. 11, no. 1, pp. 546–553, 2026.
- [2] F. Zhang and Y. Demiris, "Visual-Tactile Learning of Garment Unfolding for Robot-Assisted Dressing," *IEEE Robot. Autom. Lett.*, vol. 8, no. 9, pp. 5512–5519, 2023.
- [3] Y. Wang, Z. Sun, Z. Erickson, and D. Held, "One Policy to Dress Them All: Learning to Dress People with Diverse Poses and Garments," in *Proc. Robot.: Sci. Syst.*, 2023.
- [4] S. Kotsovolis and Y. Demiris, "Model Predictive Control with Graph Dynamics for Garment Opening Insertion during Robot-Assisted Dressing," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2024, pp. 883–890.
- [5] B. Ai, A. Amin, R. Aniceto, A. Balakrishna, G. Balke, K. Black *et al.*, " $\pi_{0.7}$: A steerable generalist robotic foundation model with emergent capabilities," 2026, *arXiv:2604.15483*.
- [6] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *Int. J. Robot. Res.*, vol. 44, no. 10–11, pp. 1684–1704, 2024.
- [7] E. Pignat and S. Calinon, "Learning adaptive dressing assistance from human demonstration," *Rob. Auton. Syst.*, vol. 93, pp. 61–75, 2017.
- [8] Z. Erickson, H. M. Clever, G. Turk, C. K. Liu, and C. C. Kemp, "Deep Haptic Model Predictive Control for Robot-Assisted Dressing," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 4437–4444.
- [9] Z. Sun, Y. Wang, D. Held, and Z. Erickson, "Force-constrained visual policy: Safe robot-assisted dressing via multi-modal sensing," *IEEE Robot. Autom. Lett.*, vol. 9, no. 5, pp. 4178–4185, 2024.
- [10] S. Kotsovolis and Y. Demiris, "Garment Diffusion Models for Robot-Assisted Dressing," *IEEE Robot. Autom. Lett.*, vol. 10, no. 2, pp. 1217–1224, 2025.
- [11] J. Zhu, M. Gienger, G. Franzese, and J. Kober, "Do You Need a Hand? – A Bimanual Robotic Dressing Assistance Scheme," *IEEE Trans. Robot.*, vol. 40, pp. 1906–1919, 2024.
- [12] F. Zhang, A. Cully, and Y. Demiris, "Probabilistic Real-Time User Posture Tracking for Personalized Robot-Assisted Dressing," *IEEE Trans. Robot.*, vol. 35, no. 4, pp. 873–888, 2019.
- [13] C. Bilaloglu, T. Löw, and S. Calinon, "Object-centric task representation and transfer using diffused orientation fields," *Science Robotics*, vol. 11, no. 113, p. eaea1762, 2026.
- [14] Y. Gao, H. J. Chang, and Y. Demiris, "User modelling for personalised dressing assistance by humanoid robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 1840–1845.
- [15] G. Chance, A. Jevtić, P. Caleb-Solly, G. Alenyà, C. Torras, and S. Dogramadzi, "“Elbows Out” — Predictive Tracking of Partially Occluded Pose for Robot-Assisted Dressing," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3598–3605, 2018.
- [16] T. Tsukakoshi, T. Miyake, T. Ogata, Y. Wang, T. Akaishi, and S. Sugano, "Close-Fitting Dressing Assistance Based on State Estimation of Feet and Garments With Semantic-Based Visual Attention," *IEEE Robot. Autom. Lett.*, vol. 11, no. 4, pp. 3923–3930, 2026.
- [17] A. Clegg, Z. Erickson, P. Grady, G. Turk, C. C. Kemp, and C. K. Liu, "Learning to Collaborate From Simulation for Robot-Assisted Dressing," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2746–2753, 2020.
- [18] D. Blanco-Mulero, J. Borras, and C. Torras, "Evaluating the Pre-Dressing Step: Unfolding Medical Garments via Imitation Learning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2025, pp. 17791–17796.
- [19] G. Franzese, R. Prakash, C. D. Santina, and J. Kober, "Generalizable Motion Policies Through Keypoint Parameterization and Transportation Maps," *IEEE Trans. Robot.*, vol. 41, pp. 4557–4573, 2025.
- [20] J. Zhao, Y. Lian, A. M. Tyrrell, M. Gienger, and J. Zhu, "Bimanual Robot-Assisted Dressing: A Spherical Coordinate-Based Strategy for Tight-Fitting Garments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2025, pp. 3328–3335.
- [21] A. Y. Hao, Y. Wang, N. S. Ravie, B. Hegde, D. Held, and Z. Erickson, "Force-modulated visual policy for robot-assisted dressing with arm motions," in *Proc. Conf. Robot Learn.*, 2025, pp. 5483–5505.
- [22] J. Zhu, M. Gienger, and J. Kober, "Learning Task-Parameterized Skills From Few Demonstrations," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4063–4070, 2022.
- [23] P. Florence, L. Manuelli, and R. Tedrake, "Self-Supervised Correspondence in Visuomotor Policy Learning," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 492–499, 2020.
- [24] Z. Weng, H. Lu, D. Kragic, and J. Lundell, "DexDiffuser: Generating Dexterous Grasps With Diffusion Models," *IEEE Robot. Autom. Lett.*, vol. 9, no. 12, pp. 11834–11840, 2024.
- [25] F. Zhang and M. Gienger, "Affordance-based robot manipulation with flow matching," 2024, *arXiv:2409.01083*.
- [26] P. M. Scheikl, N. Schreiber, C. Haas, N. Freymuth, G. Neumann, R. Lioutikov, and F. Mathis-Ullrich, "Movement Primitive Diffusion: Learning Gentle Robotic Manipulation of Deformable Objects," *IEEE Robot. Autom. Lett.*, vol. 9, no. 6, pp. 5338–5345, 2024.
- [27] M. Moletta, M. C. Welle, and D. Kragic, "Preference Aligned Visuomotor Diffusion Policies for Deformable Object Manipulation," *IEEE Robot. Autom. Lett.*, vol. 11, no. 4, pp. 4985–4992, 2026.
- [28] A. Longhini, Y. Wang, I. Garcia-Camacho, D. Blanco-Mulero, M. Moletta, M. Welle, G. Alenyà, H. Yin, Z. Erickson, D. Held, J. Borras, and D. Kragic, "Unfolding the Literature: A Review of Robotic Cloth Manipulation," *Annu. Rev. Control Robot. Auton. Syst.*, vol. 8, no. 1, pp. 295–322, 2025.
- [29] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic Graph CNN for Learning on Point Clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, 2019.
- [30] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, *arXiv:2010.02502*.
- [31] N. Sharp and K. Crane, "A Laplacian for Nonmanifold Triangle Meshes," *Comput. Graph. Forum.*, vol. 39, no. 5, pp. 69–80, 2020.
- [32] Z. Erickson, V. Gangaram, A. Kapusta, C. K. Liu, and C. C. Kemp, "Assistive Gym: A Physics Simulation Framework for Assistive Robotics," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 10169–10176.
- [33] H. Bertiche, M. Madadi, and S. Escalera, "CLOTH3D: Clothed 3D Humans," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 344–359.

- [34] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, “3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations,” in *Proc. Robot.: Sci. Syst.*, 2024.
- [35] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, “What Matters in Learning from Offline Human Demonstrations for Robot Manipulation,” in *Proc. Conf. Robot Learn.*, 2022, pp. 1678–1690.
- [36] F. Capuano, C. Pascal, A. Zouitine, T. Wolf, and M. Aractingi, “Robot Learning: A Tutorial,” 2025, *arXiv:2510.12403*.
- [37] K. Koide, “Small_gicp: Efficient and parallel algorithms for pointcloud registration,” *J. Open Source Softw.*, vol. 9, no. 100, p. 6948, 2024.