Generative Action-Chunk Sampling for Adaptive Stiffness Control in Physical Human–Robot Collaboration

Aoi Otake, Ferdinand Hartmann, Ko Igari, and Shingo Murata, *Member, IEEE*

***Abstract*—Physical human–robot collaboration requires a robot to provide assistance when human intention is clear while remaining compliant when several future motions are plausible. We present an adaptive stiffness framework based on generative action-chunk sampling. Conditioned on an RGB image and external joint-torque estimates, the policy samples multiple future action chunks from an observation-conditioned prior. Variation among the sampled action chunks is used to continuously adapt joint stiffness and damping. Greater variation makes the robot more compliant to facilitate human guidance, whereas lower variation provides firmer assistance. In a real-world collaborative transport task with four possible directions, the proposed method achieved an average success rate of 0.95, compared with 0.83 for a fixed-stiffness ablation and 0.69 for a deterministic baseline. Near direction determination, variation among the sampled action chunks increased and the controller accordingly reduced stiffness. These results suggest that variation among actions sampled by a generative policy can serve as an online control signal for balancing assistance and compliance in physical human–robot interaction.**

***Index Terms*—Adaptive stiffness control, generative action chunking, imitation learning, physical human–robot interaction, vision–force integration.**

I. Introduction

PHYSICAL human–robot interaction (pHRI) combines human judgment with robotic precision and strength in collaborative transport, assembly, and assistance [1], [2]. A robot must nevertheless respond to a partner whose motion and intention change during a task and vary across repetitions [3], [4]. At decision points, several future motions can remain plausible. The controller must therefore provide useful assistance when the intended motion is clear while remaining compliant enough to accept human guidance when it is ambiguous. Maintaining high stiffness throughout the task can cause the robot to resist an unanticipated human motion, whereas uniformly low stiffness limits the physical support that the robot can provide. Effective collaboration consequently requires the balance between assistance and compliance to change online rather than being fixed before execution.

Imitation learning enables robots to acquire complex manipulation policies from demonstrations [5], [6]. Action Chunking with Transformers (ACT) is trained as a conditional variational autoencoder (CVAE) to predict action chunks, but its standard inference uses the mean of a fixed standard-normal prior and therefore produces a single deterministic action chunk [7]. Contact-rich pHRI also requires information beyond vision because contact can be visually subtle or occluded [8], [9]. Force Torque-Aware Action Chunking Transformer (FTACT) and Force-Attending Curriculum Training for Contact-Rich Policy Learning (FACTR) condition action-chunk prediction on force or torque observations together with visual observations [10], [11]. These force-aware policies improve contact reasoning but do not generate multiple action alternatives online or connect their variability directly to robot impedance. ACT, conversely, learns a latent representation of demonstrated variation but does not use repeated sampling during standard inference to characterize variation among future actions. Thus, the potential of a force-aware generative policy to provide an online compliance signal remains underexplored.

We address this gap by extending a FACTR-style multimodal policy with ACT-style CVAE training while replacing ACT's fixed prior with an observation-conditioned prior. Given an RGB image and external joint-torque estimates, the policy samples multiple future action chunks. Variation among the sampled action chunks is used to continuously adapt stiffness and damping, serving as a proxy for uncertainty associated with alternative demonstrated motions. Greater variation lowers stiffness to facilitate human guidance, whereas lower variation increases stiffness to provide firmer assistance. The framework thereby links multimodal observation, generative action-chunk prediction, and low-level stiffness control in one closed loop. An overview of the proposed framework is shown in Fig. 1.

We evaluate the framework on a real-world collaborative transport task in which a human and a 7-DoF robot jointly move an object in one of four directions. The proposed method outperformed both the same generative policy with fixed stiffness and deterministic FACTR, while the observed reduction in stiffness near direction determination illustrated the intended adaptive response.

Our main contributions are as follows:

- We develop a generative action-chunking policy conditioned on visual and external joint-torque observations that samples multiple plausible future action chunks.
- We propose an adaptive stiffness controller that continuously adjusts stiffness and damping based on variation among the sampled actions.
- We evaluate the proposed framework on a real robot against fixed-stiffness and deterministic FACTR baselines.

Manuscript received [date]; revised [date]. This work was supported by JST PRESTO (JPMJPR22C9), JST Moonshot R&D (JPMJMS263F), JSPS KAKENHI (JP24K03012, JP26H02515), and the Kayamori Foundation of Informational Science Advancement. (Corresponding author: Shingo Murata.)

This work involved human subjects in its research. The authors confirm that all human subject research procedures and protocols are exempt from review board approval.

The authors are with the School of Engineering and Design, Graduate School of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan (e-mail: murata@elec.keio.ac.jp).

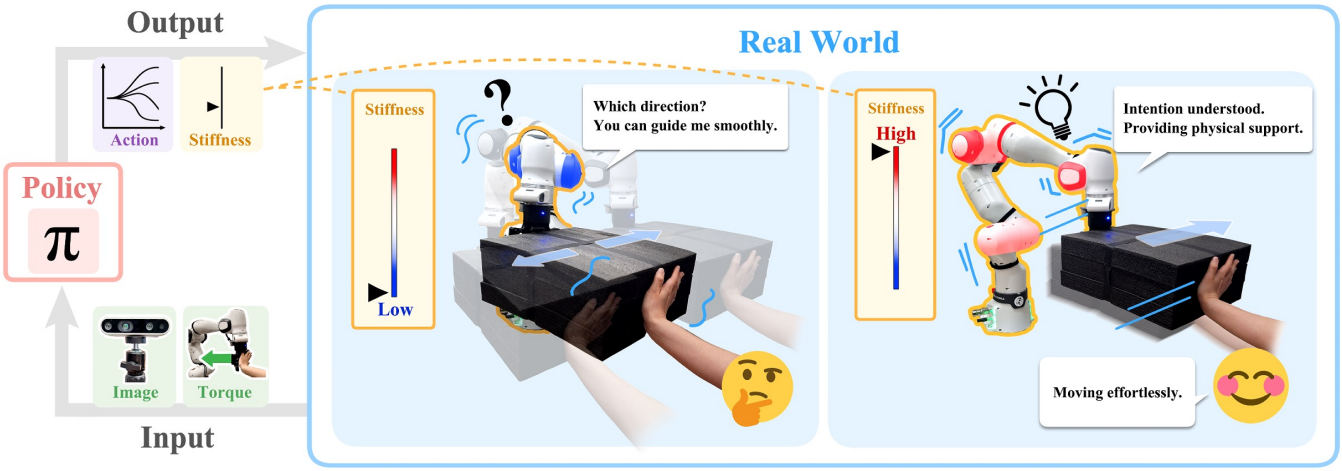


Fig. 1. Overview of the proposed adaptive stiffness control framework using generative action-chunk sampling. RGB images from the camera and external joint-torque estimates are provided to a generative policy, which samples multiple future action chunks. Variation among the sampled actions is used to continuously adjust robot stiffness and damping. When the variation is high, such as at a directional bifurcation, stiffness is lowered to provide compliance and accept human guidance. When the variation is low after the intended direction has been established, stiffness is increased to assist the transport motion while mitigating external joint-torque conflicts.

II. Related Work

A. Learning Multimodal Policies for Contact-Rich Interaction

Behavior cloning offers a direct means of learning manipulation policies from demonstrations, but small deviations during execution can lead the learned policy to visit states outside the demonstrated distribution, where subsequent deviations may compound in closed loop [12]–[14]. Sequence-level prediction mitigates this problem by generating temporally coherent action segments rather than isolated commands. ACT combines action chunking and temporal ensembling with a CVAE-based latent representation that can encode variation across demonstrated behaviors [7]. At test time, however, its encoder is discarded and the mean of its fixed standard-normal prior is used; standard ACT inference therefore produces one deterministic action chunk rather than sampling multiple action alternatives online. Related sequence models further investigate real-time execution of generative action chunks [15]. These developments provide a basis for modeling multiple plausible futures, but their observations are primarily visual and proprioceptive and do not explicitly exploit interaction forces.

Force and tactile sensing provide complementary information in contact-rich manipulation, particularly when contact events are visually subtle or the scene is occluded. Self-supervised vision–touch learning has demonstrated that shared multimodal representations can improve contact reasoning [8]. More recent policies, including FTACT and FACTR, encode force or torque measurements together with visual observations for action-chunk prediction [10], [11]. FACTR additionally uses curriculum training to discourage reliance on a single modality. These approaches establish the value of force-aware representations, but they do not directly use variability among multiple sampled action chunks to regulate the robot's physical compliance. The present work therefore combines visual and external joint-torque observations with conditional generative action-chunk prediction.

B. Modeling Intention Ambiguity in Human–Robot Interaction

Human motion is redundant and variable, and the same partial observation can precede different task outcomes. Representing the future as a distribution is therefore useful when a robot must respond before the human's intention is fully observable. Probabilistic Movement Primitives (ProMPs) represent distributions over trajectories [16], and they have been used to infer human intention during physical interaction [4]. Interaction primitives similarly model coupled human–robot behavior for cooperative tasks [3]. Latent-variable sequence policies extend this probabilistic view to high-dimensional observations and action chunks [7], [17].

Samples from a distribution over future actions can also provide information about ambiguity: small variation among sampled action chunks indicates that the observations support similar futures, whereas large variation indicates that multiple futures remain plausible. In this study, this variation is used as an online proxy for uncertainty associated with demonstrated motion alternatives. It should not be interpreted as a calibrated probability of the human's true intention or as epistemic uncertainty about out-of-distribution inputs. The unresolved question is how such an internal signal can influence physical robot behavior rather than being used only for intention classification or trajectory selection.

C. Variable Impedance and Role Adaptation

Impedance control regulates the dynamic relationship between motion and interaction force and is a standard foundation for physical robot interaction [18]. Variable impedance control extends this principle by changing stiffness and damping according to task requirements, demonstrations, or estimated risk [19]–[21]. In pHRI, lower stiffness can reduce resistance to human corrections, whereas higher stiffness can improve tracking and physical assistance when the desired motion is established [22]. Uncertainty-aware minimal-intervention control and phase-

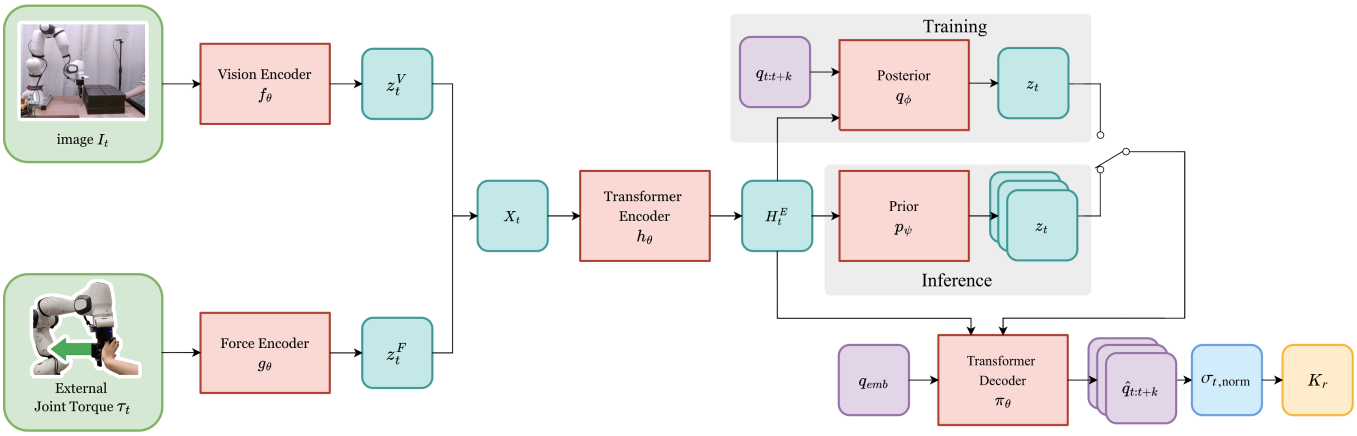


Fig. 2. Architecture of the proposed generative action-chunking policy and its connection to adaptive stiffness control. An RGB image I_t and external joint-torque estimates τ_t are encoded and integrated by the Transformer encoder h_θ to obtain the multimodal representation H_t^E. During training, the approximate posterior q_ϕ is conditioned on H_t^E and the ground-truth action chunk $q_{t:t+k}$, while the observation-conditioned prior p_ψ is learned through KL alignment. During inference, multiple latent variables are sampled from p_ψ. The Transformer decoder π_θ maps H_t^E and each latent sample to a future action chunk $\hat{q}_{t:t+k}$. Variation among the sampled actions is used to obtain the normalized action-deviation signal $\sigma_{t,\text{norm}}$, which is mapped to the adaptive joint-stiffness matrix K_r after slew-rate limiting. Here, r indexes low-level control steps executed at a higher rate than the policy-inference events indexed by t.

dependent interaction strategies likewise adapt how strongly the robot intervenes [23], [24].

Recent methods infer impedance from richer contextual representations. For example, OmniVIC uses vision–language in-context learning to select variable-impedance behavior [25], while role-allocation approaches explicitly estimate whether the human or robot should lead the interaction [26]. Such methods typically introduce a separate risk, semantic, phase, or role model. By contrast, the proposed method derives its control signal from the same generative policy that produces the target action chunk. It occupies the intersection of three research directions: force-aware multimodal policy learning, stochastic modeling of human-motion alternatives, and variable impedance control. Its distinguishing feature is the continuous adaptation of robot stiffness and damping based on variation among actions sampled by the generative policy, allowing the robot to yield when the variation is high and assist when it is low.

III. Methodology

A. Overview

The proposed framework extends FACTR's multimodal action-chunking architecture with ACT-style CVAE training and replaces ACT's fixed prior with an observation-conditioned prior. By sampling multiple action chunks from this prior, the framework obtains an online action-deviation signal that is used to adapt robot stiffness and damping during physical human–robot interaction. The architecture of the proposed framework is illustrated in Fig. 2.

B. Formulation of Proposed Architecture

The policy uses a Transformer architecture to predict a chunk of future joint-position actions conditioned on multimodal observations. At time t, it receives an RGB image I_t and external joint-torque estimates τ_t and outputs the action chunk $\hat{q}_{t:t+k}$, which contains $k+1$ joint-position actions from time t through $t+k$. First, modality-specific feature extraction is performed. The RGB image I_t is encoded by a pre-trained Vision Transformer (ViT) [27], [28]. We extract the learnable CLS token, which aggregates global image features, as the visual latent representation $z_t^V \in \mathbb{R}^{1\times d}$. Meanwhile, the external joint-torque estimates τ_t are processed by a multilayer perceptron (MLP) encoder into a force latent representation $z_t^F \in \mathbb{R}^{1\times d}$. These representations are concatenated to form the input token sequence X_t for the Transformer:

$$X_t = [z_t^V; z_t^F] \in \mathbb{R}^{2\times d}. \tag{1}$$

The Transformer encoder h_θ applies a multi-head self-attention mechanism to compute a contextualized feature representation $H_t^E = h_\theta(X_t)$ that captures the interdependencies between visual and force tokens.

To enable stochastic action-chunk generation, the proposed architecture uses ACT-style CVAE training together with a learned observation-conditioned prior. The contextualized encoder output $H_t^E \in \mathbb{R}^{2\times d}$ serves as the condition for generating the latent variable z. Throughout this section, q denotes joint-position actions that have already been z-score normalized per action dimension using the mean and standard deviation of the training set, and $\hat{q}$ denotes predictions in the same normalized action space. During training, a Transformer-based posterior network takes the ground-truth action chunk $q_{t:t+k}$ and H_t^E to estimate μ_{q_ϕ} and $\log\sigma^2_{q_\phi}$ for the approximate posterior distribution:

$$q_\phi(z|H_t^E, q_{t:t+k}) = \mathcal{N}\left(\mu_{q_\phi}, \mathrm{diag}(\sigma^2_{q_\phi})\right). \tag{2}$$

During inference, and for KL-divergence computation during training, an MLP-based prior network takes only H_t^E and estimates μ_{p_ψ} and $\log\sigma^2_{p_\psi}$ for the observation-conditioned prior distribution:

$$p_\psi(z|H_t^E) = \mathcal{N}\left(\mu_{p_\psi}, \mathrm{diag}(\sigma^2_{p_\psi})\right). \tag{3}$$

Here, all means and log variances are vectors in $\mathbb{R}^{d_z}$, and the variance vectors are obtained elementwise by exponentiating the corresponding log variances. During training, the reparameterization $z = \mu_{q_\phi} + \exp\left(\frac{1}{2}\log\sigma^2_{q_\phi}\right) \odot \epsilon, \epsilon \sim \mathcal{N}(0, I)$, is used. The latent variable is sampled from q_ϕ during training and from p_ψ during inference. Finally, the Transformer decoder π_θ takes the original encoder output H_t^E, learnable position embeddings q_{emb} as queries, and the sampled latent variable z to compute the decoded features H_t^D. These features are projected into the action space via an MLP to obtain the predicted action chunk:

$$\hat{q}_{t:t+k} = \text{MLP}(H_t^D) \in \mathbb{R}^{(k+1)\times d_a}, \tag{4}$$

where d_a denotes the dimension of the robot's action space. Before execution, the selected normalized predictions are transformed back to joint-position coordinates using the training-set statistics.

C. Training Procedure

1) Loss Function: The network parameters are optimized by minimizing a linear combination of a reconstruction loss $\mathcal{L}_{\text{rec}}$ and a Kullback-Leibler (KL) divergence loss $\mathcal{L}_{\text{KL}}$:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{KL}}, \tag{5}$$

where β is a weighting coefficient. The reconstruction loss is the mean absolute error between the predicted and ground-truth action chunks in the normalized action space:

$$\mathcal{L}_{\text{rec}} = \frac{1}{(k+1)d_a}\sum_{i=t}^{t+k}\|\hat{q}_i - q_i\|_1. \tag{6}$$

The ℓ_1 norm is averaged over the action dimensions and mini-batch in the implementation. The KL divergence term is defined as the statistical distance between the approximate posterior q_ϕ and the prior p_ψ:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(z|H_t^E, q_{t:t+k}) \parallel p_\psi(z|H_t^E)). \tag{7}$$

Minimizing this term forces the prior distribution, which relies solely on current observations, to align with the approximate posterior formed using the ground-truth action chunk. This enables the model to sample appropriate latent variables from the prior distribution during inference when ground-truth data is unavailable.

2) Curriculum Learning: To prevent overfitting to visual inputs and effectively integrate force data, we adopt the curriculum learning strategy proposed in FACTR [11].

D. Adaptive Stiffness Control Based on Action Deviation

1) Concept of Action-Deviation-Based Stiffness Control: The proposed method changes robot stiffness using an action-deviation signal derived from action chunks sampled from the observation-conditioned prior. When the observations are compatible with multiple future motions, the sampled actions tend to vary, producing a large signal that serves as a proxy for uncertainty at behavioral bifurcations. The robot then lowers its stiffness to accept human guidance. When the samples support similar actions, the signal is small and the robot increases its stiffness to track the selected motion more firmly. This model-derived control signal is not a calibrated probability of the human's intention or an estimate of epistemic uncertainty.

2) Quantification of Empirical Action Variance and Deviation: Because the proposed model generates action chunks stochastically, variation among alternative actions can be quantified by drawing multiple samples during inference and calculating their empirical variance. Given the input H_t^E, we sample M latent variables $\{z^1, \ldots, z^M\}$ from the prior distribution $p_\psi(z|H_t^E)$. These are passed through the decoder π_θ to generate M corresponding action chunks $\{\hat{q}^i_{t:t+k}\}_{i=1}^M$.

All calculations are performed directly on $\hat{q}$ in the normalized action space defined above. At chunk offset $\kappa \in \{0, \ldots, k\}$, we first calculate the aggregate empirical action variance $\sigma^2_{t+\kappa}$ and then take its square root to obtain the aggregate action-deviation magnitude $\sigma_{t+\kappa}$:

$$\begin{aligned}\sigma^2_{t+\kappa} &= \frac{1}{M-1}\sum_{i=1}^{M}\|\hat{q}^i_{t+\kappa} - \bar{\hat{q}}_{t+\kappa}\|^2,\\ \sigma_{t+\kappa} &= \sqrt{\sigma^2_{t+\kappa} + \epsilon}.\end{aligned} \tag{8}$$

Here, $\bar{\hat{q}}_{t+\kappa} = \frac{1}{M}\sum_{i=1}^{M}\hat{q}^i_{t+\kappa}$ is the mean normalized action at time $t+\kappa$, and $\|\cdot\|^2$ denotes the squared Euclidean norm. The denominator $M-1$ applies Bessel's correction, yielding the sum of the component-wise unbiased sample variances across the M generated actions. A small constant $\epsilon = 10^{-8}$ is added for numerical stability near zero. The square-root transformation is applied before the temporal weighting described below so that small changes in empirical variance produce more pronounced changes in the resulting action-deviation magnitude, particularly in the low-variance regime.

To emphasize variation in the later part of the predicted chunk, we define the temporally weighted action-deviation signal $\sigma_{t,\text{W}}$ as the weighted average across all chunk offsets:

$$\sigma_{t,\text{W}} = \frac{\sum_{\kappa=0}^{k} w_\kappa \sigma_{t+\kappa}}{\sum_{\kappa=0}^{k} w_\kappa}, \tag{9}$$

where the weight w_κ is linearly interpolated from w_{start} at the starting point ($\kappa = 0$) to w_{end} at the endpoint ($\kappa = k$):

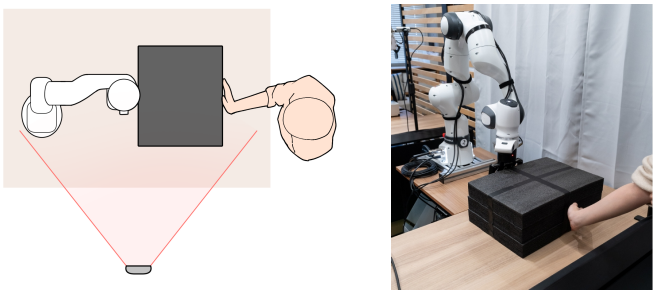
$$w_\kappa = w_{\text{start}} + (w_{\text{end}} - w_{\text{start}})\frac{\kappa}{k}. \tag{10}$$

A uniform average produced only small changes around the task bifurcation in preliminary design analysis. We therefore assigned larger weights to later predictions so that emerging alternatives near the end of the horizon contribute more strongly to the control signal. This weighting is a design choice and may also reflect horizon-dependent prediction error.

To map the action-deviation signal to stable control parameters, we mitigate the effect of observation noise by applying an exponential moving average (EMA) to obtain the smoothed signal $\sigma_{t,\text{EMA}}$:

$$\sigma_{t,\text{EMA}} = (1-\gamma)\sigma_{t-1,\text{EMA}} + \gamma\sigma_{t,\text{W}}, \tag{11}$$

where $\gamma \in (0, 1]$ is the smoothing factor. In our implementation, we set $\gamma = 0.1$, meaning the system retains 90% of the historical state, effectively dampening sudden spikes. The EMA is initialized using the weighted action-deviation signal from the first inference event, $\sigma_{0,\text{EMA}} = \sigma_{0,\text{W}}$. This smoothed value is then normalized to a range of $[0, 1]$ using Min–Max scaling:



(a) Schematic overhead view (b) Experimental workspace

Fig. 3. Experimental environment for the collaborative transport task. (a) A schematic overhead view of the setup. A human collaborator and the robot arm face each other across a desk, sandwiching a black box from opposite sides. An RGB camera is positioned at the front (bottom of the view) to capture the workspace. (b) An actual photograph of the experimental workspace, taken from the bottom-right perspective of the schematic view.

$$\sigma_{t,\mathrm{norm}} = \mathrm{clip}\left(\frac{\sigma_{t,\mathrm{EMA}} - \sigma_{\min}}{\sigma_{\max} - \sigma_{\min}}, 0, 1\right), \tag{12}$$

where $\sigma_{\min}$ and $\sigma_{\max}$ are the empirical lower and upper bounds of the smoothed action-deviation signal determined from the offline validation data, as detailed in Section IV-D2.

3) Calculation of Adaptive Stiffness Parameters: Based on the normalized action-deviation signal $\sigma_{t,\mathrm{norm}}$, a stiffness blending coefficient $\hat{\alpha}_t$ is derived to determine the mixing ratio between high and low stiffness:

$$\hat{\alpha}_t = 1.0 - \sigma_{t,\mathrm{norm}}. \tag{13}$$

Here, $\hat{\alpha}_t = 1$ corresponds to the lower action-deviation bound and high stiffness, whereas $\hat{\alpha}_t = 0$ corresponds to the upper action-deviation bound and low stiffness. To mitigate abrupt physical responses caused by sudden changes in the action-deviation signal, a slew-rate limiter is applied at each low-level control step to obtain the actual control coefficient α_r:

$$\alpha_r = \alpha_{r-1} + \mathrm{sgn}(\hat{\alpha}_t - \alpha_{r-1}) \cdot \min(|\hat{\alpha}_t - \alpha_{r-1}|, \Delta\alpha_{\max}), \tag{14}$$

where t denotes the most recent policy-inference event, whereas r indexes low-level control steps executed at a higher rate; $\hat{\alpha}_t$ is held constant between inference events. In our implementation, $\Delta\alpha_{\max} = 0.005$ per low-level control step and $\alpha_0 = 1.0$. Finally, the stiffness matrix K_r and damping matrix D_r are determined via linear interpolation between predefined high-stiffness $(K_{\mathrm{H}}, D_{\mathrm{H}})$ and low-stiffness $(K_{\mathrm{L}}, D_{\mathrm{L}})$ parameters:

$$K_r = \alpha_r K_{\mathrm{H}} + (1 - \alpha_r) K_{\mathrm{L}}, \tag{15}$$

$$D_r = \alpha_r D_{\mathrm{H}} + (1 - \alpha_r) D_{\mathrm{L}}. \tag{16}$$

IV. Experiments

We evaluate the system through a collaborative transport task in which a human collaborator and the robot jointly manipulate an object in one of four target directions: forward, backward, left, and right.

A. Experimental Environment and Task Details

The setup of the experimental environment is illustrated in Fig. 3. The left side shows a schematic top-down view, while the right side displays the actual workspace from a bottom-right perspective. The experiments were conducted on a workbench with sufficient workspace for the robotic arm and a human collaborator to face each other and perform collaborative tasks.

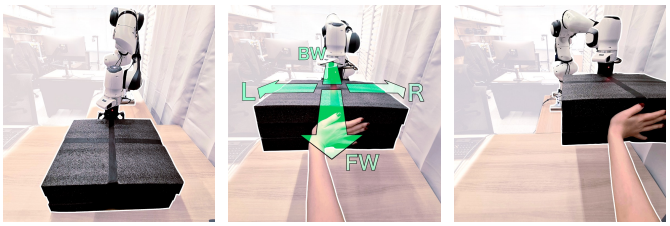
The manipulated object was a box-shaped item constructed from multiple layers of black sponge material. A key characteristic of this object is its extreme lightness, allowing it to be held purely via friction. Consequently, the human and the robot can lift and transport the object by merely pressing against its opposite sides, without the need for finger grasping.

The sequence of the task, as shown in Fig. 4, proceeds as follows:

1) **Approach and Lifting (Fig. 4a):** The robot arm moves from its initial position to the side of the object. The human and the robot sandwich the object from opposite sides and collaboratively lift it vertically to a specified height. Since the direction of movement is fixed during this phase, the sampled action chunks are expected to show little variation, and the robot is expected to maintain high stiffness, which would actively assist the lifting motion.
2) **Direction Determination and Guidance (Fig. 4b):** After lifting, the human conveys the intended transport direction (one of the four arrows shown in Fig. 4b) by applying force to the object. During this intermediate phase—from when the human decides the direction until the actual transport begins—multiple transport directions can remain plausible, so variation among the sampled action chunks is expected to increase. The proposed method is expected to lower the robot's stiffness accordingly. This induced compliance is expected to allow the human to guide the robot in the desired direction while mitigating external joint-torque conflicts.
3) **Transport and Assistance (Fig. 4c):** Once the movement direction (e.g., rightward in Fig. 4c) is established and transport begins, variation among the sampled action chunks is expected to decrease, and stiffness is intended to increase again. By maintaining high stiffness, the robot tracks the selected action chunk and assists the intended transport motion.

B. Hardware Setup

The proposed framework was implemented using a 7-DoF articulated robot arm, the Franka Research 3. The force input comprised external joint-torque estimates provided by the manufacturer's API, which uses the robot's internal dynamics model to account for gravity and friction; it did not comprise direct Cartesian contact-force measurements. For visual observation, an Intel RealSense D435 front camera was mounted to capture the overall environment, providing RGB images at a resolution of 640×480 pixels. The control frequency of the robot was set to 500 Hz, allowing for real-time, responsive stiffness adjustment. The robot's built-in torque safety limits stopped a trial when excessive joint torque was detected.



(a) Approach and Lifting (b) Direction Determination (c) Transport and Assistance

Fig. 4. Sequence of the collaborative transport task. (a) The robot arm approaches the side of the box and collaboratively lifts it upward with the human. (b) The human applies force to the object to indicate the intended transport direction. The four possible directions are denoted as Left (L), Right (R), Forward (FW), and Backward (BW). (c) The robot and human cooperatively transport the object in the determined direction.

C. Data Collection

The robot's observations consisted of RGB images from the camera and external joint-torque estimates. Data were recorded at a frequency of 30 Hz, with each trial lasting 420 time steps. The captured RGB images were resized to 224×224 pixels before being fed into the model. The sole human collaborator in both demonstration collection and real-world evaluation was the first author.

Data were collected under varying conditions by combining the control parameter settings ("Low" and "High" stiffness shown in Table I) with the four directional movement variations shown in Fig. 4b. For each of these conditions, 15 trials were recorded, yielding a total of 120 demonstration trials. The data were manually stratified into 100 training trials and 20 offline validation trials; no random split or random seed was used. The Low-stiffness training subset comprised 13 Right, 13 Forward, 12 Left, and 12 Backward trials, whereas the High-stiffness subset comprised 12 Right, 12 Forward, 13 Left, and 13 Backward trials. Thus, the training set contained 50 trials per stiffness setting and 25 trials per direction. The remaining 20 trials were held out from training for offline analysis and controller-threshold calibration; they were not the 180 online robot-evaluation trials reported in Section V-C.

The stiffness (K) and damping (D) values in Table I were empirically tuned based on preliminary experiments. Using the "Medium" setting as a baseline, we subjectively evaluated physical followability during contact. The parameters for "Low" and "High" stiffness were consequently selected to provide sufficient responsiveness while mitigating excessive force conflicts under the present experimental conditions.

TABLE I
DIAGONAL JOINT-SPACE STIFFNESS K (N m/rad) AND DAMPING D (N m s/rad) PARAMETERS USED IN THE EXPERIMENTS.

Parameter		J1	J2	J3	J4	J5	J6	J7
Low	K	70	70	230	90	30	20	10
	D	7	7	12	6	2	1	1
Medium	K	200	200	370	220	100	40	30
	D	20	20	30	18	8	3	3
High	K	300	300	450	310	200	110	100
	D	40	40	50	30	20	10	10

D. Implementation Details

1) Training Details: We primarily adopt the base architecture and training hyperparameters from FACTR [11]. Following FACTR, we apply its latent-space curriculum with an initial start scale of 7. Furthermore, for the newly introduced CVAE framework, the prior network is parameterized by a 3-layer MLP with GELU activations, while the posterior network is implemented as a 3-layer Transformer encoder with 8 attention heads. The dimension of the latent variable z is set to 16, and the KL divergence coefficient β is set to 1.

2) Inference and Control Settings: The base observation and action clock was 30 Hz. Action-chunk inference was performed every 4 observation steps (7.5 Hz, approximately 133 ms). Each inference generated $M = 10$ action chunks spanning $\hat{q}_{t:t+k}$, where $k = 99$; each chunk therefore contained $k + 1 = 100$ action steps (approximately 3.33 s). All M samples were used to compute the empirical action variance and the resulting action-deviation signal, whereas one sampled action chunk was used as the control candidate. Averaging samples can produce an invalid intermediate action in multimodal situations; using one sample preserves a coherent hypothesis, while the stiffness controller reduces physical resistance when the action-deviation signal is high. The selected normalized action chunk was transformed back to joint-position coordinates, and its next 30 steps (1.0 s) were stored for temporal ensembling. Commands were dispatched in 12-step segments every 400 ms (2.5 Hz) and were subsequently interpolated by the 500-Hz joint controller.

Specifically, the target joint command $q_{\text{target}}(t)$ at time t is calculated from N_{buf} temporally overlapping predictions $\hat{q}_i^{\text{buf}}(t)$. The weight w_i for each prediction is defined using a decay factor m:

$$w_i = \exp(-m \cdot i) \tag{17}$$

where i denotes the age of the prediction ($i = 0$ is the oldest and $i = N_{\text{buf}} - 1$ is the newest, consistent with the buffer implementation). In this experiment, we set $m = 0.25$. The target command is then normalized by the sum of all weights $W = \sum_{i=0}^{N_{\text{buf}}-1} w_i$:

$$q_{\text{target}}(t) = \frac{1}{W} \sum_{i=0}^{N_{\text{buf}}-1} w_i \hat{q}_i^{\text{buf}}(t). \tag{18}$$

Under this setting, when the buffer reaches its maximum capacity ($N_{\text{buf}} = 8$), the oldest prediction ($i = 0$) retains a contribution weight of approximately 25.6%, whereas the

newest prediction ($i = 7$) contributes only about 4.4%. Following ACT [7], the exponential weighting assigns greater weight to older predictions, promoting continuity among temporally overlapping action chunks.

At each chunk offset, the aggregate empirical action variance was converted to an action-deviation magnitude by the square-root transformation in (8). These magnitudes were then temporally weighted and smoothed during inference. The weight coefficient linearly scales from $w_{\text{start}} = 0.1$ to $w_{\text{end}} = 0.9$. The weighted action-deviation signal and its EMA were updated synchronously with each inference result at 7.5 Hz. To suppress sudden temporal spikes, an EMA with a smoothing factor $\gamma = 0.1$ was applied to obtain $\sigma_{t,\text{EMA}}$.

The scaling thresholds $\sigma_{\min}$ and $\sigma_{\max}$ were determined by analyzing transitions of the smoothed action-deviation signal across 10 trials from the offline validation set, which was strictly separated from the training data. The averages of the maximum and minimum values in each trial were calculated. Based on these averages, margins were applied for controller calibration in the real-world environment. Specifically, 10% was subtracted from the minimum side and 40% from the maximum side. This adjustment was designed to suppress unintended stiffness drops caused by minor prediction noise while retaining sensitivity near task bifurcations. The final values used in this experiment were $\sigma_{\min} = 0.061$ and $\sigma_{\max} = 0.216$.

During execution, the stiffness and damping parameters were continuously adjusted by changing the blending ratio between the Low and High profiles based on the normalized action-deviation signal.

E. Baselines and Evaluation Metrics

Comparative experiments were conducted under three conditions. The *Proposed* condition used the stochastic policy and adaptive stiffness control. The *Ablation (fixed stiffness)* condition used the same trained stochastic policy but fixed the stiffness and damping at the Medium setting in Table I. The *FACTR baseline* used a separately trained deterministic FACTR model with the same Medium setting. Neither model was retrained or fine-tuned during the evaluation. Thus, the Proposed–Ablation comparison isolates the present adaptive controller relative to a statically tuned Medium setting, whereas the comparison with FACTR includes the difference in policy architecture.

For each condition, 15 trials were conducted in each of the four directions, yielding 180 online evaluation trials in total. Directions were presented in a repeating Right–Forward–Left–Backward cycle. A trial was defined as successful when the object reached the prespecified displacement in the instructed direction: at least 20 cm in the leftward, rightward, and backward directions and at least 15 cm in the forward direction, where the available workspace was smaller. A torque-limit shutdown or movement in an unintended direction was classified as failure. No time limit was imposed, and no object drop occurred.

V. Results

A. Analysis of Latent-Space Variance

We evaluated the temporal dynamics of the diagonal variance parameters in the 16-dimensional latent space using all 20 offline validation sequences. At each time step, the prior and approximate posterior networks output log-variance vectors; these values were exponentiated elementwise to obtain $\sigma^2_{p_\psi}$ and $\sigma^2_{q_\phi}$. Fig. 5 plots these network-predicted variance parameters for each sequence and their across-sequence mean.

The approximate posterior variance remained close to zero throughout the task, indicating that the distribution became highly concentrated when the ground-truth future action chunk was provided. In contrast, the prior variance was larger and varied over time because the prior was conditioned only on the current observations.

The prior variance tended to increase around time step 250, particularly in latent dimensions 8, 10, and 11. This interval was associated with direction determination in the recorded task sequence. The observed association suggests that these dimensions respond to changes occurring near the behavioral bifurcation, although task phase and elapsed time were not independently controlled.

B. Analysis of Sampled Action Chunks

Next, we evaluated how variation under the learned prior affects the generated action chunks. Fig. 6 illustrates action chunks sampled from the prior distribution alongside the ground-truth (GT) joint trajectories. Variation among the sampled action chunks indicates that the decoder responds to the sampled latent variables.

Focusing on the action chunks generated using the prior distribution, we observed pronounced variation across the joint trajectories, particularly between steps 250 and 300. This variation was especially visible in Joints 1, 2, 4, and 7, suggesting that predictions for these joints differed around the behavioral bifurcation.

The temporal co-occurrence of increased prior variance and greater variation among the generated joint trajectories indicates that changes in the latent distribution were reflected in the action predictions in these sequences.

C. Real-World Evaluation

Finally, we evaluated the proposed framework in a real-world collaborative transport task. To examine the performance of generative action-chunk prediction with adaptive stiffness control, we compared three previously defined conditions: the proposed generative policy with adaptive stiffness, the same policy with fixed stiffness (ablation), and deterministic FACTR with fixed stiffness (baseline). Quantitative evaluation of the success rates revealed several key trends, as summarized in Table II. Specifically, the proposed method achieved an overall average success rate of 0.95, compared with 0.83 for the ablation condition and 0.69 for the baseline.

The FACTR baseline recorded fewer successes than the proposed method in all four directions. This difference may partly reflect its deterministic architecture, but the comparison also includes other architectural differences and therefore does not isolate the effect of conditional-prior sampling.

The Backward and Left success rates were similar for the proposed and ablation conditions. The largest numerical difference among the three methods occurred in the Forward direction. In

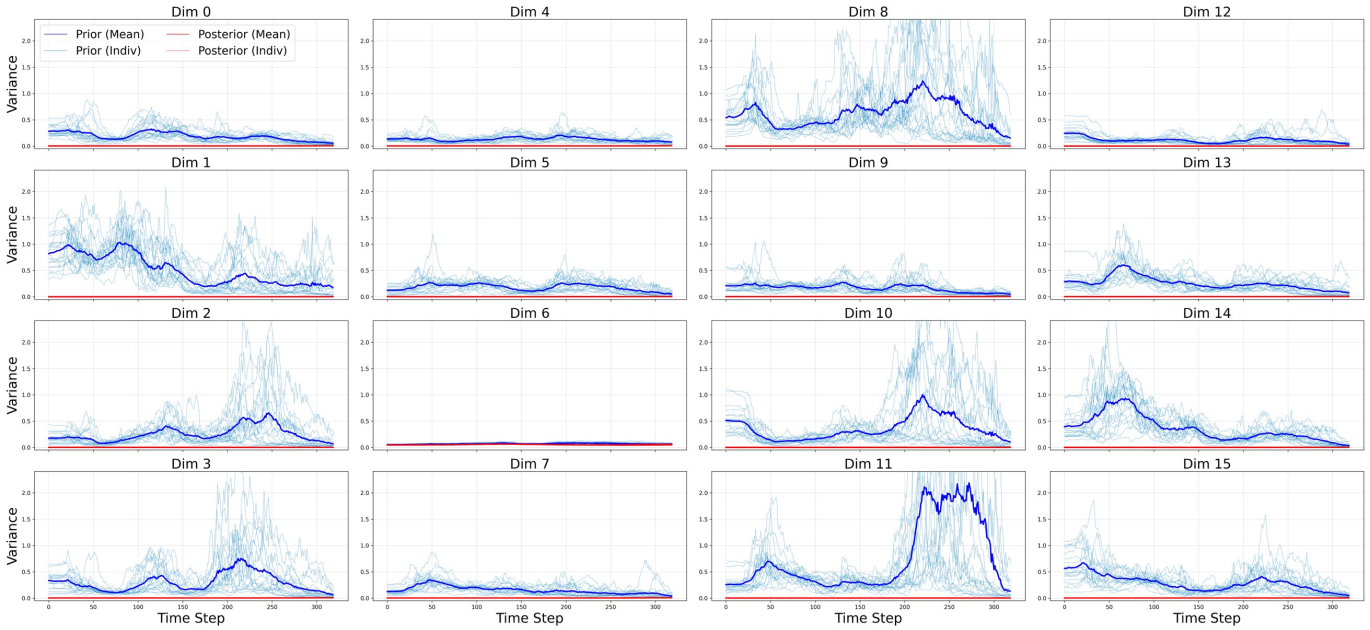


Fig. 5. Temporal transitions of the diagonal variance parameters for each dimension of the 16-dimensional latent representation across the 20 offline validation sequences. The plotted values are obtained by exponentiating the log variances output by the prior and approximate posterior networks. Thin lines show individual sequences, and thick lines show their mean at each time step.

TABLE II
COMPARISON OF TASK SUCCESS RATES ACROSS DIFFERENT TRANSPORT DIRECTIONS.

Method	Forward	Backward	Left	Right	Avg. Rate
Proposed	14/15	15/15	14/15	14/15	0.95
Ablation	10/15	14/15	14/15	12/15	0.83
FACTR	4/15	13/15	12/15	12/15	0.69

the Right direction, all methods achieved relatively high success rates, although the ablation and FACTR baseline each recorded 12/15 successes compared with 14/15 for the proposed method. The following sections examine representative control behaviors in the Forward and Right directions.

1) Analysis of Forward Transport: The largest numerical difference was observed in the forward direction. Fig. 7 illustrates representative experimental behaviors and recorded values for each method during this task. The FACTR baseline success rate was 0.27, compared with 0.93 for the proposed method.

In the representative proposed-method trial, the normalized action-deviation signal increased and stiffness decreased around the bifurcation interval (time steps 250–300). In the displayed FACTR trial, external joint-torque estimates increased until the torque safety limit stopped the system. The displayed ablation trial completed the task but exhibited sustained external joint-torque estimates around the same interval. These traces illustrate the control behavior observed in representative trials.

2) Analysis of Rightward Transport: Rightward transport yielded relatively high success rates for all three conditions (Table II): 14/15 for the proposed method and 12/15 for both the ablation and FACTR. Inspection of the recorded failures indicated that some ablation and FACTR trials transitioned into backward motion after rightward transport. A likely cause was the visual similarity between observations near direction determination and those near task completion, which led the terminal state to be interpreted as another direction-determination phase. In the proposed method, even when the model similarly treated the end of the task as a bifurcation, the reduced stiffness allowed the participant to correct the resulting return motion manually. This physical adjustability limited excessive backward motion and helped maintain the target position. The observed behavior illustrates how adaptive stiffness control can mitigate the physical consequences of visual ambiguity.

VI. DISCUSSION

The results indicate that variation among sampled actions can be used to regulate the trade-off between action tracking and physical compliance in pHRI. Near the direction-determination phase, greater variation among the generated action chunks was accompanied by reduced stiffness. In the representative forward trials, the fixed-stiffness conditions exhibited larger external joint-torque estimates. Inspection of failed rightward trials further suggested that visually similar observations may trigger unintended return motion. The proposed method achieved a 0.95 average success rate, compared with 0.83 for the fixed-stiffness ablation and 0.69 for deterministic FACTR. Because the generative policy was shared by the proposed method and the ablation, their numerical difference suggests an empirical advantage of adaptive stiffness over the statically tuned Medium setting in this task. However, access to the wider Low–High parameter range, rather than adaptation alone, may partially contribute to this difference.

The observed behavior may also be interpreted as a primitive form of implicit role adaptation. Humans can increase limb

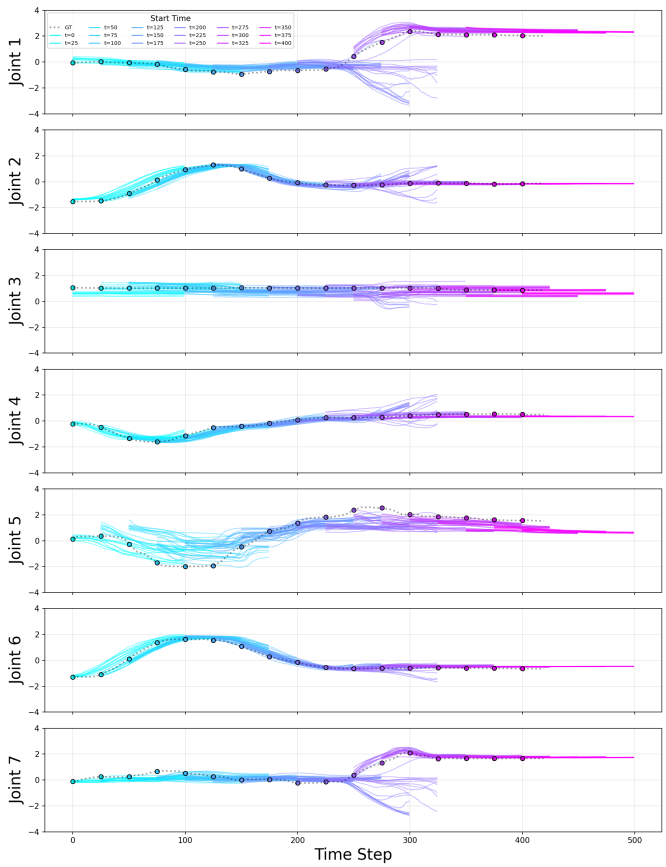


Fig. 6. Comparison of ground-truth (GT) joint trajectories and action chunks sampled from the prior distribution. Each graph illustrates the normalized target angles for the seven joints of the robot. Ten action chunks are sampled from the prior at intervals of 25 time steps to visualize variation among the predictions.

impedance when interacting with uncertain dynamics [29]; complementary robot compliance can therefore help prevent both partners from resisting each other. Although leader–follower roles were not measured or labeled in this experiment, this interpretation is consistent with reports that delayed recognition of human leadership under fixed-gain control can produce excessive force and oscillatory behavior [30]. Unlike approaches that use a separate semantic model or explicit leader–follower optimization [25], [26], our controller uses the distribution already produced for action generation: the robot yields when the action-deviation signal is high and assists when it is low. This direct connection avoids an additional phase or role classifier, although it does not replace such methods when semantic task reasoning is required.

Several limitations qualify these findings. First, the scaling thresholds $\sigma_{\min}$ and $\sigma_{\max}$ were selected empirically and may require retuning for different tasks or robot hardware. Second, the action-deviation signal is derived from variation among actions sampled from the learned conditional prior; it is not a calibrated probability of the human's intention and does not identify epistemic uncertainty under out-of-distribution observations. Finally, evaluation in the present collaborative transport setting does not establish generalization beyond the tested task and directional outcomes. Future work should automate calibration of the action-deviation-to-stiffness mapping, evaluate the framework in broader collaborative settings, and combine the present measure with explicit epistemic uncertainty and fail-safe compliance.

VII. Conclusion

This paper presented an adaptive stiffness control framework for physical human–robot interaction based on a generative action-chunking policy conditioned on RGB images and external joint-torque estimates. The policy uses an observation-conditioned prior to sample multiple future action chunks, and variation among the sampled actions is used to continuously adjust stiffness and damping to balance action tracking and compliance. In the evaluated collaborative transport task, the proposed method achieved an average success rate of 0.95, compared with 0.83 for the fixed-stiffness ablation and 0.69 for the deterministic FACTR baseline. The observed relationship between sampled-action variation and stiffness, together with representative external joint-torque traces, indicates that the proposed framework can provide a useful control mechanism in this experimental setting.

Two directions are particularly important for future work. First, to improve robustness under out-of-distribution conditions, we plan to incorporate explicit epistemic uncertainty measures and fail-safe compliance strategies. Second, we aim to extend the current reactive framework toward active bidirectional collaboration, such as dynamic leader-follower role switching, for more intuitive human–robot cooperation.

References

[1] P. Maurice, M. E. Huber, N. Hogan, and D. Sternad, "Velocity-curvature patterns limit human–robot physical interaction," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 249–256, 2017.

[2] M.-L. Lee, X. Liang, B. Hu, G. Onel, S. Behdad, and M. Zheng, "A review of prospects and opportunities in disassembly with human–robot collaboration," *Journal of Manufacturing Science and Engineering*, vol. 146, no. 2, p. 020902, 2024.

[3] H. B. Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2831–2837.

[4] O. Dermy, A. Paraschos, M. Ewerton, J. Peters, F. Charpillet, and S. Ivaldi, "Prediction of intention during interaction with icub with probabilistic movement primitives," *Frontiers in Robotics and AI*, vol. 4, p. 45, 2017.

[5] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 39, pp. 1–40, 2016.

[6] O. Kroemer, S. Niekum, and G. Konidaris, "A review of robot learning for manipulation: Challenges, representations, and algorithms," *Journal of Machine Learning Research*, vol. 22, no. 30, pp. 1–82, 2021.

[7] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," *Arxiv Preprint Arxiv:2304.13705*, 2023.

[8] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8943–8950.

[9] H. Yang, W. Sun, J. Liu, J. Zheng, J. Xiao, and A. Mian, "Occlusion-aware 3d hand-object pose estimation with masked autoencoders," *Arxiv Preprint Arxiv:2506.10816*, 2025.

[10] R. Watanabe, M. Alvarez, P. Ferreiro, P. Savkin, and G. Sano, "Ftact: Force torque aware action chunking transformer for pick-and-reorient bottle task," *Arxiv Preprint Arxiv:2509.23112*, 2025.

[11] J. J. Liu, Y. Li, K. Shaw, T. Tao, R. Salakhutdinov, and D. Pathak, "Factr: Force-attending curriculum training for contact-rich policy learning," *Arxiv Preprint Arxiv:2502.17432*, 2025.

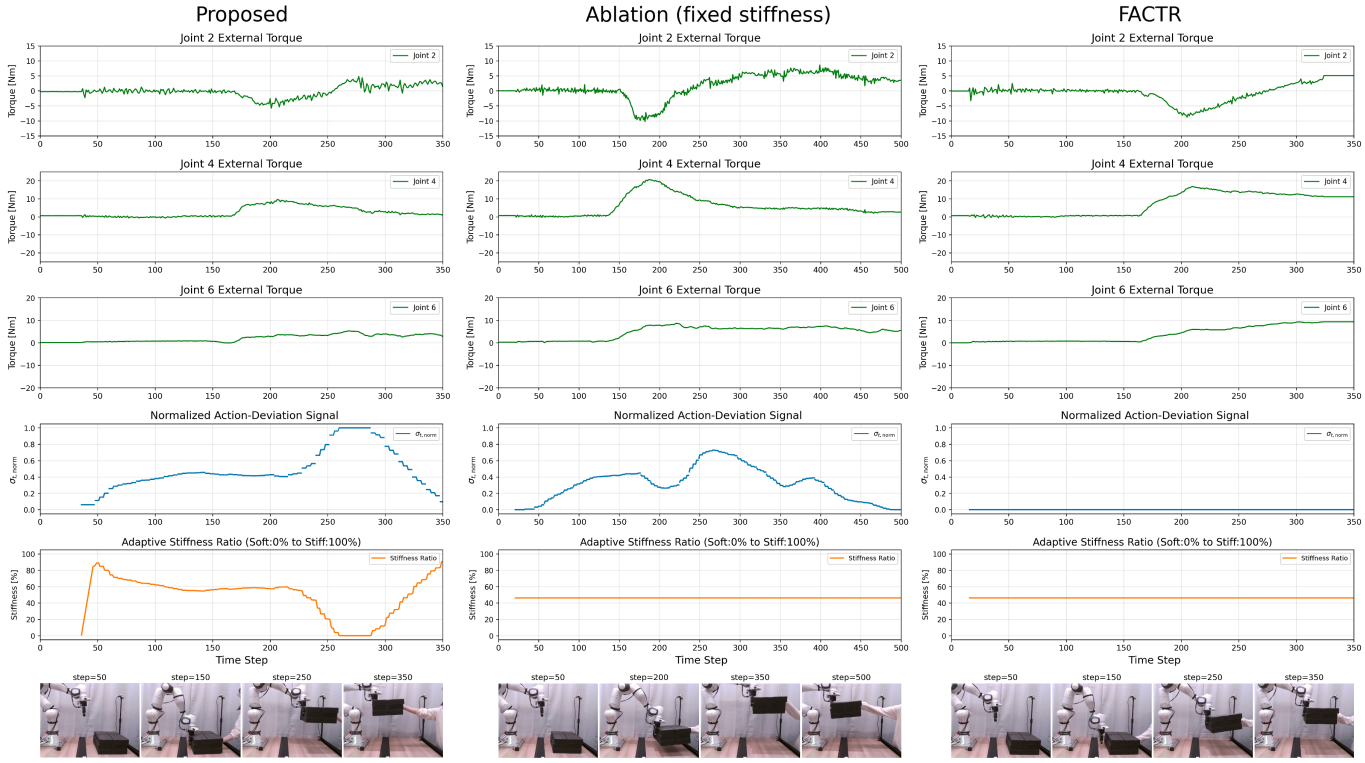


Fig. 7. Control behavior and execution results during the forward transport task. The columns, from left to right, illustrate the trials for the proposed method (Success), the ablation model with fixed stiffness (Success), and the FACTR baseline (Failure). The rows, from top to bottom, display the temporal transitions of the external joint-torque estimates at Joints 2, 4, and 6; the normalized action-deviation signal $\sigma_{t,\mathrm{norm}}$; the stiffness blending ratio; and representative RGB camera images captured at key time steps during the task.

[12] S. Ross and D. Bagnell, "Efficient reductions for imitation learning," in *Proceedings of The Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 661–668.

[13] P. D. Haan, D. Jayaraman, and S. Levine, "Causal confusion in imitation learning," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[14] S. A. Mehta, Y. U. Ciftci, B. Ramachandran, S. Bansal, and D. P. Losey, "Stable-bc: Controlling covariate shift with stable behavior cloning," *IEEE Robotics and Automation Letters*, 2025.

[15] K. Black, M. Y. Galliker, and S. Levine, "Real-time execution of action chunking flow policies," *ArXiv Preprint ArXiv:2506.07339*, 2025.

[16] A. Paraschos, C. Daniel, J. R. Peters, and G. Neumann, "Probabilistic movement primitives," *Advances in neural information processing systems*, vol. 26, 2013.

[17] D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling, "Semi-supervised learning with deep generative models," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[18] N. Hogan, "Impedance control: An approach to manipulation," in *1984 American Control Conference*. IEEE, 1984, pp. 304–313.

[19] J. Buchli, F. Stulp, E. Theodorou, and S. Schaal, "Learning variable impedance control," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 820–833, 2011.

[20] J. R. Medina, D. Lee, and S. Hirche, "Risk-sensitive optimal feedback control for haptic assistance," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 1025–1031.

[21] F. J. Abu-Dakka and M. Saveriano, "Variable impedance control and learning—a review," *Frontiers in Robotics and AI*, vol. 7, p. 590681, 2020.

[22] V. Duchaine and C. M. Gosselin, "General model of human-robot cooperation using a novel velocity based variable impedance control," in *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*. IEEE, 2007, pp. 446–451.

[23] J. Silvério, Y. Huang, L. Rozo *et al.*, "An uncertainty-aware minimal intervention control strategy learned from demonstrations," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 6065–6071.

[24] M. Bdiwi, I. A. Naser, J. Halim, S. Bauer, P. Eichler, and S. Ihlenfeldt, "Towards safety4. 0: A novel approach for flexible human-robot-interaction based on safety-related dynamic finite-state machine with multilayer operation modes," *Frontiers in Robotics and AI*, vol. 9, p. 1002226, 2022.

[25] H. Zhang, W.-H. Huang, G. Solak, and A. Ajoudani, "Omnivic: A self-improving variable impedance controller with vision-language in-context learning for safe robotic manipulation," *Arxiv Preprint Arxiv:2510.17150*, 2025.

[26] H. Liu, Y. Tong, and Z. Zhang, "Dtrt: Enhancing human intent estimation and role allocation for physical human-robot collaboration," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 16 312–16 318.

[27] S. Dasari, M. K. Srirama, U. Jain, and A. Gupta, "An unbiased look at datasets for visuo-motor pre-training," in *Conference on Robot Learning*. PMLR, 2023, pp. 1183–1198.

[28] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[29] E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, no. 6862, pp. 446–449, 2001.

[30] K. H. Allen, C. Rogers, and E. S. Short, "Haptic communication in human-human and human-robot co-manipulation," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2025, pp. 2535–2542.