PANTACT: TEACHING WHOLE-BODY DEXTERITY WITH A WEARABLE CONTACT INTERFACE

Anonymous authors
Paper under double-blind review

ABSTRACT

Humanoid robots have many degrees of freedom and contact surfaces, yet most loco-manipulation research restricts manipulation to the hands and avoids other body contacts with the environment. To extend manipulation beyond the hands, we present PanTact[1], a system that learns humanoid *whole-body dexterity*, the capability to manipulate with any suitable body surface, directly from robot-free human demonstrations. Our key idea is to use contact as manipulation-centric information that transfers across embodiments. Contact locations and timing specify the interactions needed to accomplish a task, while the robot chooses how to coordinate its body to realize them within its own physical constraints. Specifically, PanTact captures, learns, and executes whole-body contact via: 1) a wearable, robot-free data collection interface that augments egocentric vision and bimanual UMI grippers with whole-body modular contact patches to record body contact locations and states; 2) a visuomotor policy that mitigates the egocentric observation gap and predicts whole-body contact goals; and 3) contact-aware whole-body control formulations that realize active contact goals using suitable robot surfaces. We deploy PanTact on a Unitree G1 humanoid without additional body contact sensors, demonstrating whole-body dexterity involving body support, pushing, sweeping, and load-sharing. Website: https://pantacticlr.github.io.

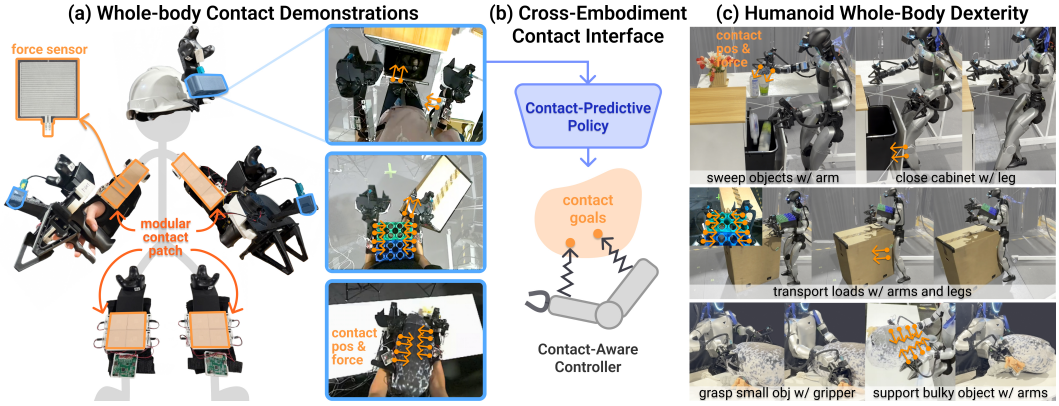

Figure 1: **PanTact gives a humanoid whole-body dexterity, learned from robot-free human demonstrations.** (a) The data collection interface captures whole-body contacts alongside vision and motion. (b) Contact provides an interface for human-to-robot transfer. The visuomotor policy predicts contact goals, and the whole-body controller realizes them using suitable robot surfaces. (c) PanTact enables diverse whole-body dexterous skills on a humanoid.

1 INTRODUCTION

When humans interact with the physical world, we use more than just our hands: our arms and torso embrace bulky objects, our forearms sweep multiple items or brace against surfaces, and our knees push objects when our hands are full. We call this capability **whole-body dexterity** (Xu et al., 2025a): the ability to accomplish manipulation tasks using any suitable surface of the body.

[1]The name PanTact combines *pan-* (all) with *tact-* (contact, tactile), reflecting a humanoid that manipulates with any body surface.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Humanoid robots share the human body's overall morphology, with dozens of actuated degrees of freedom and potential contact surfaces distributed across the body. Yet much of loco-manipulation research limits manipulation to the hands, reserving the rest of the body for locomotion and balance and typically avoiding other body contact with the environment. This leaves much of the humanoid body's potential for manipulation unused.

Realizing this potential requires coordinating whole-body motion with contact sequences across body surfaces. Human demonstrations provide natural examples of this coordination. We therefore ask: *how can a humanoid learn whole-body dexterity directly from human demonstrations?*

Advances in imitation learning and demonstration interfaces (Wu et al., 2024; Cheng et al., 2025; Chi et al., 2024; Xu et al., 2026) have made learning manipulation from humans increasingly scalable. However, these pipelines capture and imitate either end-effector (EEF) or joint trajectories, and do not explicitly carry body-contact information from demonstration through policy prediction to robot execution. Extending these pipelines to whole-body dexterity introduces three connected challenges:

- **Capturing body contact in human demonstrations.** Existing interfaces record only EEF motions or joint angles and do not directly measure where and when body contact occurs.
- **Preserving contact in human-to-robot policy transfer.** EEF targets leave body contacts unspecified, while retargeting whole-body joints can alter contact locations because humans and robots differ in shape and proportions.
- **Executing contact with whole-body control.** Controllers designed to track motion references or EEF targets lack an explicit objective for achieving contact goals on the body. They must incorporate contact goals while maintaining motion tracking and balance.

To address the above challenges, our key observation is that contact provides *manipulation-centric* information that can be separated from embodiment-specific motion. Contact location and timing describe the task-level physical interaction, whereas the body surface and joint configuration used to produce it depend on the robot embodiment. Contact also identifies which parts of the demonstrated body motion need to be preserved during transfer, allowing the robot to learn task-critical contacts without necessarily matching free-space body motions. We therefore use **contact as a cross-embodiment interface**: *measuring* contacts on the human, *predicting* their desired states and locations with the policy, and *tracking* active contact goals with the controller. Tracking only active contact goals without assigning them to particular robot links allows the controller to choose a body configuration suited to the robot to make contacts and maintain balance.

We present PanTact, a system that learns whole-body dexterity from robot-free human demonstrations and deploys it on an off-the-shelf humanoid through three components:

- **Contact-complete, robot-free data collection system.** We augment egocentric sensing and bimanual UMI grippers with *modular contact patches*, reconfigurable force-sensing tiles covering the demonstrator's forearms and knees that provide contact timing, intensity and location.
- **Contact-predictive whole-body manipulation policy.** The policy observes embodiment agnostic observations with body portions geometrically masked out from the egocentric view, and jointly predicts EEF trajectories and per-sensor contact goals, each specifying a contact state and position.
- **Contact-aware whole-body controller.** We propose three formulations to achieve active contact goals using suitable robot surfaces under physical constraints: upper-body contact-aware IK, whole-body IK executed by a dense tracker, and direct sparse contact tracking.

We deploy PanTact on a Unitree G1 humanoid without adding body contact sensors. Our tasks combine grasping and body support to move differently sized objects simultaneously, sweeping objects with arm surfaces, using the legs for non-prehensile manipulation and locomotion at the same time, and sequentially switching contact across different body parts. Result videos are best viewed on https://pantacticlr.github.io. We will open-source the hardware designs, data, and code.

2 Related Work

Whole-body sensing for robots. Distributed range, force, tactile, and vision sensing has been used for collision avoidance, contact detection, and compliant whole-body interactions. Examples include range sensing for safe navigation and interaction (Tanaka et al., 2015; Qi et al., 2022), force and pressure sensing for compliant contact (Kollmitz et al., 2018; Goncalves et al., 2022; Xu et al., 2025b), tactile sensing for contact feedback (Liu et al., 2023; Dean-Leon et al., 2019; Jiang & Wong,

2024; Murooka et al., 2025; Jang et al., 2026), and whole-body vision on serial and soft growing robots (Xu et al., 2025a; Qin et al., 2026). These systems primarily integrate whole-body sensing into the robot and acquire demonstrations through teleoperation. PanTact incorporates whole-body tactile sensing into a wearable interface, and uses human contact measurements to supervise policy contact goal actions, which the robot realizes without additional body sensors.

Data collection interfaces for robot learning. Teleoperation provides robot-native demonstrations, but it is slow, costly, and hard to deploy on humanoids in diverse environments (Wu et al., 2024; Xiong et al., 2025). UMI (Chi et al., 2024) enables robot-free data collection with a portable hand-held gripper, and its extensions add a head-mounted camera (Xu et al., 2026), gripper-centric tactile and force sensing (Choi et al., 2026), while humanoid-oriented interfaces additionally capture body or keypoint trajectories (Nai et al., 2026). These motion-centric interfaces capture EEF and keypoint trajectories or joint angles, but none of them measure body contact signals. PanTact adds contact sensing to preserve contact interactions during human to robot transfer.

Learning loco-manipulation from demonstrations. A common architecture combines a high-level policy that predicts task-space commands with a whole-body controller that executes them (Nai et al., 2026; Zhao et al., 2026). Controllers use optimization (Ferrolho et al., 2023; Xu et al., 2026) or learned motion tracking (He et al., 2024; Dugar et al., 2024; Weng et al., 2025), with demonstrations obtained from wearable devices (Nai et al., 2026; Zhao et al., 2026), human video (Bahety et al., 2025), or simulation (Li et al., 2025), but the policy-to-controller interface typically specifies only motion targets. Recent work introduces contact-conditioned motion tracking (Chen et al., 2026; Li et al., 2026), but assigns contacts to predefined body parts and does not integrate contact prediction by a high-level visuomotor policy. PanTact integrates human contact sensing, a visuomotor policy that predicts contact goals, and contact-aware whole-body control into a complete system.

3 PanTact Data Collection Interface

PanTact's data collection interface combines egocentric-augmented bimanual UMIs (Chi et al., 2024; Xu et al., 2026) with modular contact patches (Fig. 2), capturing vision and motion together with body-contact signals that identify where and when contact occurs. We record images from a head-mounted stereo camera (`ZED X Mini`) and two fisheye wrist cameras (`ZED X One S`) at 60 fps, poses from HTC Vive trackers on the head and both wrists at 250 Hz, and gripper widths at 1 kHz. All streams share a monotonic clock and are aligned to 60 Hz during post-processing.

Existing whole-body tactile skins are typically tailored to a particular robot embodiment. To accommodate different demonstrators and body regions, we design the sensing surface from small, identical *modules* that can be assembled into wearable *patches* of varying shapes and sizes.

Contact sensing and electronics. Each module is a $44 \times 44\,\mathrm{mm}$ unit carrying a single force-sensitive resistor (FSR). A custom $50 \times 50\,\mathrm{mm}$ PCB simultaneously streams up to 16 FSR channels at 1 kHz. Each FSR is connected in series with a fixed $1\,\mathrm{k}\Omega$ resistor to form a voltage divider, sampled by a 12-bit ADC with $8\times$ oversampling. Modules communicate over a daisy-chained, multi-drop RS-485 bus using a request-and-response protocol with CRC checks, and also support direct USB readout.

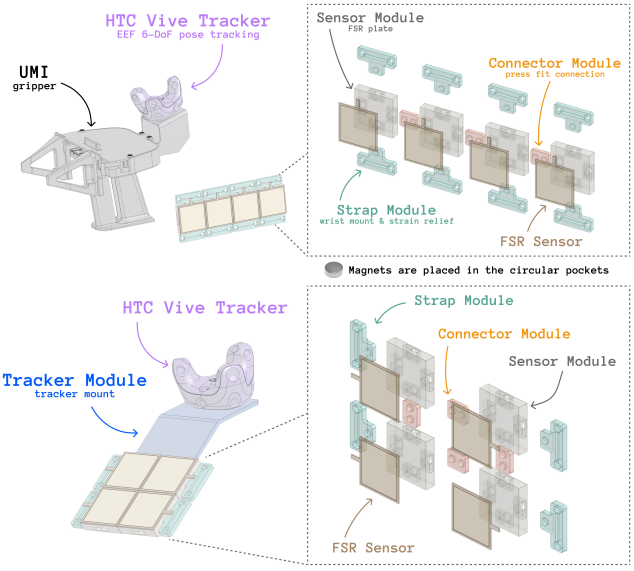

Figure 2: **PanTact modular contact patch.** Sensor, connector, and strap modules snap together into patches of arbitrary shape and extent. We showcase arm and knee patches.

Modular magnetic tiling. Each sensing module has a connector slot on each side, which accepts press-fit magnetic connectors, allowing modules to snap into patches of varying shapes and sizes.

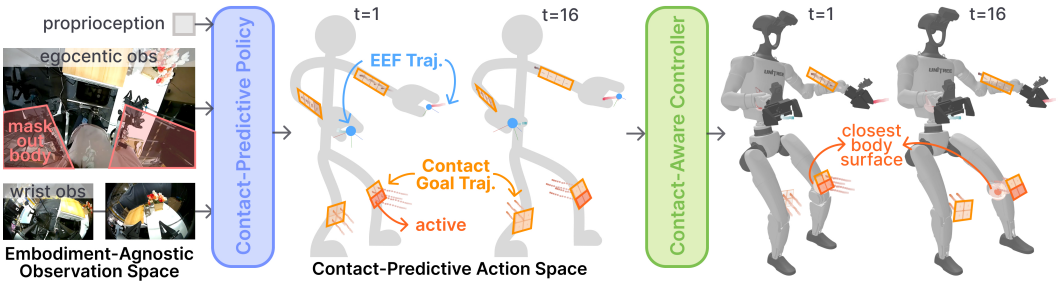

Figure 3: **Cross-Embodiment Contact Interface.** The *Contact-Predictive Whole-Body Manipulation Policy* uses body-masked 3D egocentric observations, wrist images, and proprioception to predict EEF trajectories and contact goals (position and binary state). The *Contact-Aware Whole-Body Controller* coordinates whole-body motion to track EEF commands and realize active contact goals using suitable robot surfaces.

We design three connector types: a full connector that bridges adjacent sensing modules; a strap connector that secures the patch to the body and provides cable strain relief; and a tracker connector that mounts a HTC Vive tracker to track the patch's 6-DoF pose.

Contact detection and localization. FSR resistance drifts with temperature and mounting pressure, so we binarize contact relative to a per-channel adaptive baseline rather than an absolute threshold. Each patch has an associated tracker, and a rigid transform maps each sensor center into the tracking world frame. At each synchronized timestep, we store an $N \times 4$ contact array, where N is the number of FSR sensors, and each row contains a sensor's binary contact state and 3D position. These sequences supervise the policy's contact-goal predictions (§ 4), with the binary state indicating when the controller should track contact location (§ 5).

4 Contact-Predictive Whole-Body Manipulation Policy

To learn whole-body dexterity from human demonstrations, we employ a policy that predicts whole-body contact goals alongside EEF trajectories (Fig. 3). To mitigate the human-to-robot embodiment gap, the policy masks body regions from the egocentric view (§ 4.1) and predicts world-frame EEF and contact goals as actions (§ 4.2).

4.1 Embodiment-Agnostic Observation Space

Whole-body contacts often occur behind the grippers, outside the views of wrist cameras. An egocentric view provides broader scene context and observes whole-body interactions, but introduces a larger human-to-robot embodiment gap. During demonstrations, the demonstrator's body often appears in the egocentric view, creating a mismatch in appearance between training and deployment. Differences in human and robot height also produce a viewpoint mismatch.

Following HoMMI (Xu et al., 2026), we use a 3D egocentric representation to reduce sensitivity to viewpoint differences and geometrically mask body regions. Each gripper frame and contact patch defines a body-facing half-space, since arms originate behind the grippers, and body surfaces start behind the contact patch surfaces. We approximate the body region by the intersection of these half-spaces and mask the corresponding points in the attention layer of the egocentric vision encoder (Fig. 3). This construction requires no appearance model of either embodiment and scales to any number of contact patches. The policy's overall observation space is

$$\mathbf{o}_t = \left(P^{\text{ego}}_{t-h+1:t},\ I^L_{t-h+1:t},\ I^R_{t-h+1:t},\ \mathbf{x}^L_{t-h+1:t},\ \mathbf{x}^R_{t-h+1:t}\right), \tag{1}$$

where h=2 is the number of history steps, P^{ego} is the masked egocentric pointmap, I^L, I^R are the wrist images, and $\mathbf{x}^L, \mathbf{x}^R$ each contain an EEF pose and gripper width.

4.2 Contact-Predictive Action Space

We use a diffusion policy architecture (Chi et al., 2025) that runs at 10 Hz and predicts action chunks with a horizon of H=16 steps. Each step is

$$\mathbf{a}_t = \left(\boldsymbol{\lambda}_t,\ \mathbf{x}^L_t,\ \mathbf{x}^R_t,\ \{(\hat{c}_{t,i},\ \hat{\mathbf{p}}_{t,i})\}_{i=1}^N\right), \tag{2}$$

comprising a 3D look-at point $\boldsymbol{\lambda}_t$ for the head (Xu et al., 2026), target EEF poses and gripper widths $\mathbf{x}^L_t,\ \mathbf{x}^R_t$, and *contact goals* $\{(\hat{c}_{t,i},\ \hat{\mathbf{p}}_{t,i})\}_{i=1}^N$. $\hat{c}_{t,i}$ is a binary contact state that determines whether

the contact position $\hat{\mathbf{p}}_{t,i}$ is a tracking target: only active goals ($\hat{c}_{t,i} = 1$) request the robot to make contact with the world at $\hat{\mathbf{p}}_{t,i}$. Inactive goals ($\hat{c}_{t,i} = 0$) do not require the robot to reproduce the corresponding sensor's free-space trajectory. The whole-body controller (§ 5) selects a suitable configuration to achieve the intended active contacts while not being over-constrained by inactive contacts. We express the pointmaps, proprioception poses, EEF poses and contact positions across each observation history and action chunk in the same frame, relative to the latest observed left-gripper pose (Xu et al., 2026), reducing sensitivity to differences in human and robot kinematics.

5 Contact-Aware Whole-Body Control

5.1 General formulation

The whole-body controller converts the policy's EEF trajectories and contact goals $\{(\hat{c}_i, \hat{\mathbf{p}}_i)\}$ into robot whole-body motions. It tracks contact locations only when $\hat{c}_i = 1$, prioritizing task-critical interactions without requiring the robot to reproduce demonstrator's free-space body motion. Each active goal can be realized by any suitable robot surface, leaving the controller to coordinate the remaining degrees of freedom under robot-specific constraints.

Any-surface contact objective. Let q denote the robot configuration, $\mathcal{B}$ the set of candidate robot bodies, and $\mathrm{d}_b(\mathbf{p}, q)$ the distance from point $\mathbf{p}$ to body b's collision surface. The proximate distance from a commanded contact $\mathbf{p}$ to the robot is:

$$\mathrm{d}(\mathbf{p}, q) = -T \log \sum_{b \in \mathcal{B}} \exp\big(-\mathrm{d}_b(\mathbf{p}, q)/T\big), \tag{3}$$

where $T = 0.02\,\mathrm{m}$ controls the smoothing. The whole-body controller drives $\mathrm{d}(\hat{\mathbf{p}}_i, q) \to 0$ for active contacts to approach the target.

Virtual contact targets. Reaching a contact location alone does not prescribe a pressing force. Following the virtual-target principle (Hou et al., 2025), we use a *virtual contact target*: a controller setpoint displaced beyond the intended contact location into the object. When the object constrains the robot surface at contact, the residual tracking error generates a pressing effort through the low-level controller. At deployment, we offset active contact targets by:

$$\tilde{\mathbf{p}}_i = \hat{\mathbf{p}}_i + \hat{c}_i\, \delta\, \mathbf{n}_i, \qquad \delta = \min\left(\frac{F_{\mathrm{des}}}{k_{\mathrm{virt}}}, \delta_{\max}\right), \tag{4}$$

where $\mathbf{n}_i$ is a unit pressing direction into the object, F_{des} is a nominal per-site pressing force, k_{virt} is a design stiffness, and $\delta_{\max}$ bounds the offset. The controller tracks $\tilde{\mathbf{p}}_i$ to encourage pressing force.

5.2 Whole-Body Control Strategies

We implement contact-aware tracking objective through three formulations: an upper-body IK, whole-body IK executed by an existing dense motion tracker, and a sparse contact tracker that we train to follow EEF and contact goals directly.

I. Upper-Body IK. A straightforward solution is to add contact tasks to differential IK. At each step, the solver jointly optimizes EEF and contact tracking through the quadratic program (QP):

$$\min_{\dot{q}} \quad \sum_{k \in \mathcal{T}} w_k \big\| J_k(q)\dot{q} - K_k e_k(q) \big\|^2 + \varepsilon \|\dot{q}\|^2 \quad \text{s.t.} \quad \dot{q} \in \mathcal{C}, \tag{5}$$

where q is the joint configuration and $\dot{q}$ is the commanded joint velocity. For task k, e_k is the target-minus-current error, J_k maps joint velocities to task-space velocities, K_k is a feedback gain, and w_k sets the task weight. The task set $\mathcal{T}$ includes EEF position and orientation tracking, nominal-posture regularization, and active contact tracking. The coefficient $\varepsilon > 0$ penalizes joint velocity, and $\mathcal{C}$ enforces joint-position and velocity limits. For each active contact $\hat{\mathbf{p}}_i$, the solver finds the closest surface point $\mathbf{s}_b$ on each candidate body and retains the three nearest bodies. Their contact-task weights are proportional to $\exp(-\mathrm{d}_b/T)$ and normalized over the selected bodies. Each receives a position task that brings $\mathbf{s}_b$ toward the target. The resulting joint velocities are integrated into joint-position targets for the arms, while a standing controller controls the legs. This formulation requires neither training data nor simulation, but is limited to upper-body tasks with fixed legs.

II. WBIK + SONIC. To extend contact tracking to the lower body, we expand the QP to optimize whole-body degrees of freedom and use the existing SONIC motion tracker (Luo et al., 2026) for execution. When the EEF and contact goals are reachable within the current stance, the feet remain

planted while the solver adjusts posture through squatting and leaning. Otherwise, it commands a stepping motion. The resulting full-body configuration trajectory provides a dense reference that SONIC executes by tracking all joint targets. This formulation enables lower-body contacts and locomotion while reusing an existing whole-body tracking controller. However, the dense reference specifies postural details that are not essential to the task, and tracking them can compete with EEF and contact accuracy. Moreover, the IK solver does not model SONIC's execution limits and may generate dynamically infeasible references.

III. Sparse Contact Tracker. To preserve freedom in whole-body coordination during execution, we train a controller that directly tracks sparse EEF and contact goals without an intermediate dense reference. Following masked motion tracking (Tessler et al., 2024), the controller learns to satisfy the provided targets while choosing the remaining body motion.

Controller interface. The tracker runs at $50\,\mathrm{Hz}$ and outputs 29 joint-position targets for PD control. It receives proprioception including joint positions and velocities, root orientation and angular velocity, and the previous action, together with a goal block containing masked future joint targets, EEF targets, and S contact slots. Each slot contains a 3D point in the robot root frame and a contact state flag, with a separate validity mask identifying the slots to use. At deployment, the controller receives EEF targets and active contact goals, leaving other targets masked.

Hindsight contact labeling. Training such a controller requires (goal, motion) pairs with achievable contact goals, which standard motion-capture data do not directly provide. We construct these pairs by rolling out reference motions with an expert tracker in simulation and sampling target points uniformly over contactable body collision surfaces along the achieved trajectories. We apply this labeling to both a general motion-capture dataset used to train SONIC (Luo et al., 2026) for broad whole-body motion coverage, and to our human demonstrations, executed with WBIK+SONIC (§ 5.2) under trajectory perturbations for task-specific coverage. During training, each episode draws S slots from the clip's pool with a random number of active slots, and each selected slot is randomly marked as in-contact (to be tracked tightly) or pre-contact (to be approached softly). Contact targets derived from achieved rollouts thus become the commands the policy learns to satisfy, providing supervision without human annotation and exposing the controller to diverse contact configurations.

Training objective. The reward combines motion tracking, contact tracking, and regularization: $r = r_{\text{track}} + r_{\text{contact}} + r_{\text{reg}}$. For the set $\mathcal{V}$ of valid training slots, the contact reward is

$$r_{\text{contact}} = w \sum_{i \in \mathcal{V}} \left[0.3\, c_i e^{-\mathrm{d}_i^2/\sigma_a^2} + 0.1\,(1 - c_i) e^{-\mathrm{d}_i^2/\sigma_s^2} \right], \tag{6}$$

where $\mathrm{d}_i = \mathrm{d}(\hat{\mathbf{p}}_i, q)$ and c_i distinguishes in-contact from pre-contact training goals. We use $\sigma_a = 0.05\,\mathrm{m}$ for tight tracking and $\sigma_s = 0.25\,\mathrm{m}$ for soft approach. The weight w increases during training to introduce contact tracking after basic motion-tracking competence. The term r_{track} rewards body position, orientation, and velocity tracking alongside EEF tracking, while r_{reg} penalizes abrupt actions, joint-limit violations, and uncommanded contact.

We train with PPO (Schulman et al., 2017) in 4096 parallel simulated environments using hindsight-labeled rollouts generated by SONIC (Luo et al., 2026), with domain randomization over masses, friction, and contact-goal perturbations. This formulation learns whole-body contact realization directly, at the cost of simulation training and dependence on the training distribution.

6 Evaluation

We evaluate whether whole-body dexterity can be learned from robot-free human demonstrations and transferred to a humanoid. Our evaluation probes four capabilities:

- **Cross-embodiment contact transfer:** contacts measured on a human body and predicted by the visuomotor policy must be realized by a humanoid with a different configuration. This capability is required for all tasks.
- **Non-prehensile upper-body manipulation:** supporting bulky objects with the arms (§ 6.1, 6.3) and sweeping objects with the arm surfaces (§ 6.2).
- **Lower-body loco-manipulation:** coordinating the lower body for non-prehensile manipulation, locomotion, and balance at the same time (§ 6.2, 6.3).
- **Whole-body contact sequencing:** establishing, maintaining, and releasing contacts across different body regions as each task progresses, required for all tasks.

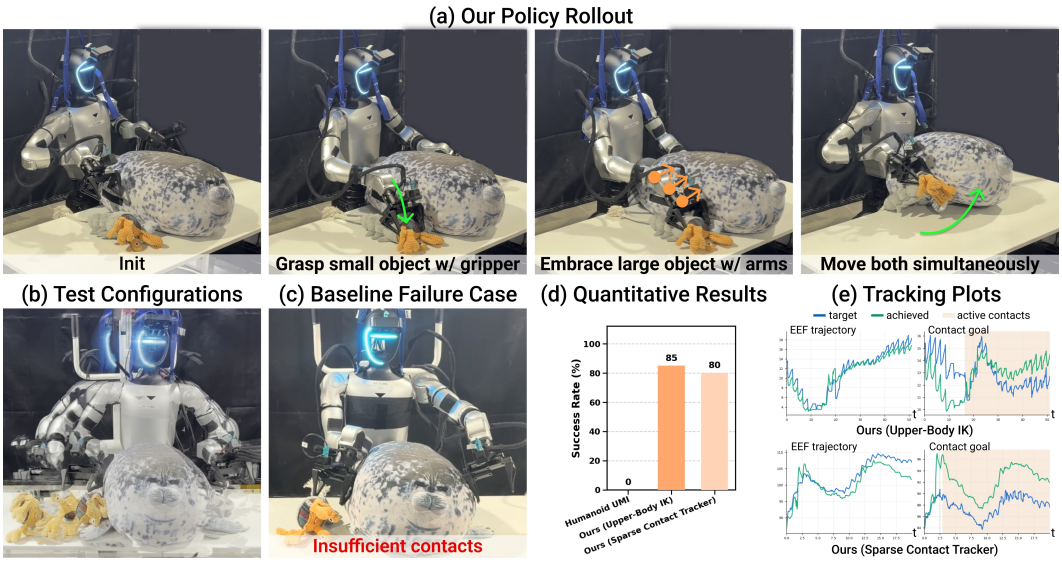

Figure 4: **Grasp-and-Embrace.** (a) Autonomous rollout of our policy deployed with `Upper-Body IK`, demonstrating capability to support a bulky object with arms while grasping small object with a gripper. (b) Overlayed test configurations with varied object placements and initial robot configurations. (c) Typical failure case of `Humanoid UMI` is insufficient arm contact to support the bulky object. (d) Task success rates. (e) Real-world tracking plots.

We compare against two pose-based interfaces that lack explicit body-contact objectives. These baselines test whether EEF trajectories alone, or with additional lower-body pose targets, suffice to transfer the demonstrated interactions without contact information:

- **Humanoid UMI** (Xu et al., 2026; Zhao et al., 2026): vanilla humanoid UMI setup, where a visuomotor policy predicts EEF actions from egocentric pointmap and wrist images and proprioception, and a whole-body controller tracks only EEF trajectories.
- **Humanoid UMI++**: an extension that additionally predicts and tracks knee poses. Each pose target is tied to a fixed robot body part and is tracked continuously.

All hardware experiments use a 29-DoF Unitree G1 equipped with grippers and cameras mirroring the data collection interface. For each task, all methods are trained on the same demonstrations and evaluated from matched test configurations.

6.1 Grasp-and-Embrace

Task: The robot first grasps a small object from a table, then embraces a bulky object with its arms (Fig. 4a). Success requires moving both objects simultaneously. **Capability:** Grasping the small object requires accurate EEF tracking, while supporting and moving the bulky object requires sufficient arm contact force on both sides. **Data Collection:** We collected 100 demonstrations with varied table and goal placements and initial body postures. **Test Scenarios:** We evaluate each method over 20 rollouts with varied initial object placements and robot postures (Fig. 4b). **Performance:** `Ours (Upper-Body IK)` achieves an 85% success rate, and `Ours (Sparse Contact Tracker)` achieves 80% (Fig. 4d). Occasional failures occur when the gripper misses the small object. Fig. 4e shows accurate EEF and active-contact tracking with `Upper-Body IK`, while `Sparse Contact Tracker` exhibits slightly larger errors and tracking dead zones. `Humanoid UMI` baseline consistently fails to support the bulky object because its whole-body controller tracks only EEF commands, leaving arm posture unspecified (Fig. 4c). We also observe the grippers continuing reaching forward until approaching arm joint limits, potentially reflecting accumulated tracking error that takes the observation outside policy's training distribution.

6.2 Sweep-and-Stow

Task: The robot uses arm to sweep one or more tall objects from a counter into an open bin, then uses a lower-body surface to push the cabinet closed (Fig. 5a). The *Sweep Object* score is based on the number of objects deposited in the bin. The *Close Cabinet* score awards 50% credit for partial

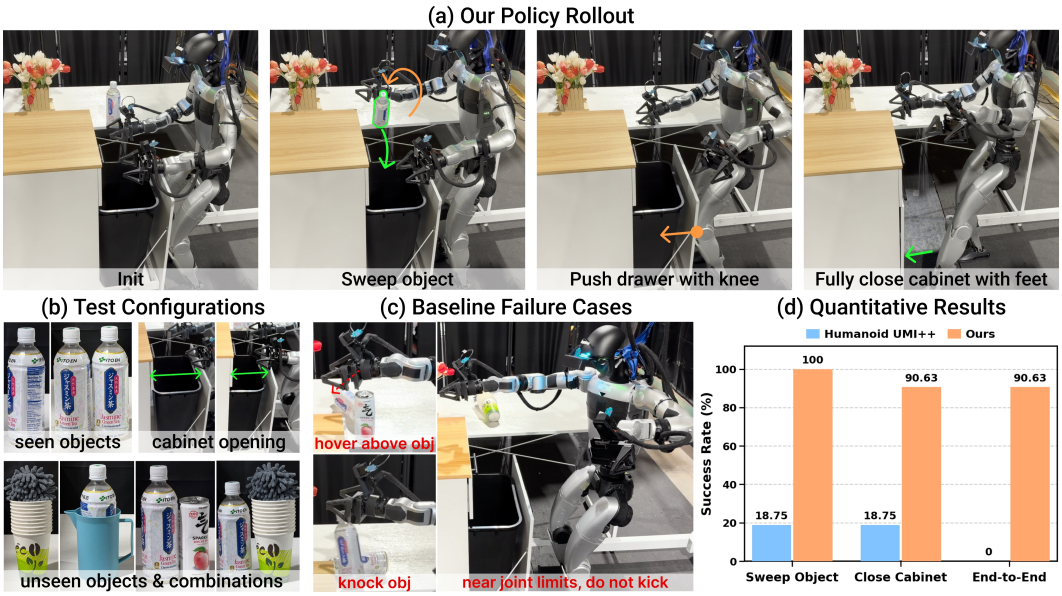

Figure 5: **Sweep-and-Stow.** (a) Our policy rollout with `WBIK+SONIC` tracker, demonstrating arm sweeping and lower-body pushing and locomotion at the same time. (b) Test configurations with seen and unseen object combinations, varied cabinet openings, and different robot-to-cabinet distances. (c) Typical failure case of the `Humanoid UMI++` baseline include the gripper passes above an object or knocks it outside the bin, and the robot fails to advance a leg for cabinet closure as its arms approach joint limits. (d) Stage-wise and end-to-end success rates.

closure below halfway and 100% for closure of more than halfway. The end-to-end score is the product of the two stage scores. **Capability:** Sweeping requires coordinated arm contacts to guide objects into the bin, while cabinet closure requires lower-body pushing coordinated with stepping and balance. Across stages, the policy must predict contact goals appropriate to the current task stage, and the controller must select suitable surfaces to realize them. **Data Collection:** We collected 150 demonstrations with varied object counts, types, and layouts, initial body postures, and cabinet openings. **Test Scenarios:** We evaluate each method over 16 rollouts, divided equally among four conditions: a single seen object, two seen objects, a single unseen object, and two unseen objects (Fig. 5b). The seen-object conditions vary object and cabinet layouts. The single-unseen-object condition includes two objects with two rollouts each, the two-unseen-object condition includes two object pairs with two rollouts each. **Performance:** `Ours (WBIK+Dense Tracker)` achieves an end-to-end success rate of 90.63%. The robot sweeps with its wrist, forearm, or gripper side and closes the cabinet with either knee or a foot, adapting beyond the demonstrated forearm and left-knee contacts. Occasional partial closures occur when the controller prioritizes balance over reaching the contact target. `Humanoid UMI++` often misses objects or knocks them outside the bin due to the lack of arm regularization. The arms also approach joint limits more often, accompanied by failures to advance a leg and close the cabinet (Fig. 5c). This may be due to differences in human and robot proportions, as jointly tracking demonstrated hand and knee poses can require relative placements that are difficult for the robot to realize.

6.3 Carry-and-Push

Task: The robot carries a box in both arms, walks towards a larger box on the floor, and pushes it with its lower-body (Fig. 6a). Success requires retaining the carried box while moving the floor box forward by more than 20 cm. **Capability:** The arms must maintain sufficient contact support throughout locomotion, while the legs coordinate walking, pushing, and balance. **Data Collection:** We collected 110 demonstrations with varied starting positions and box locations. **Test Scenarios:** We evaluate each method over 10 rollouts, with the floor box initially to the robot's right in five rollouts and to its left in the other five, and varied initial robot configurations (Fig. 6b). **Performance:** `Ours (WBIK+Dense Tracker)` achieves a 90% success rate. Failures occur when the robot drops the carried box, which happens more frequently with `Ours (Sparse Tracker)` and `Humanoid UMI++`. As shown in Fig. 6e, `Humanoid UMI++` continuously tracks knee poses, whereas `Ours` imposes leg contact objectives only during active contact phases. The base-

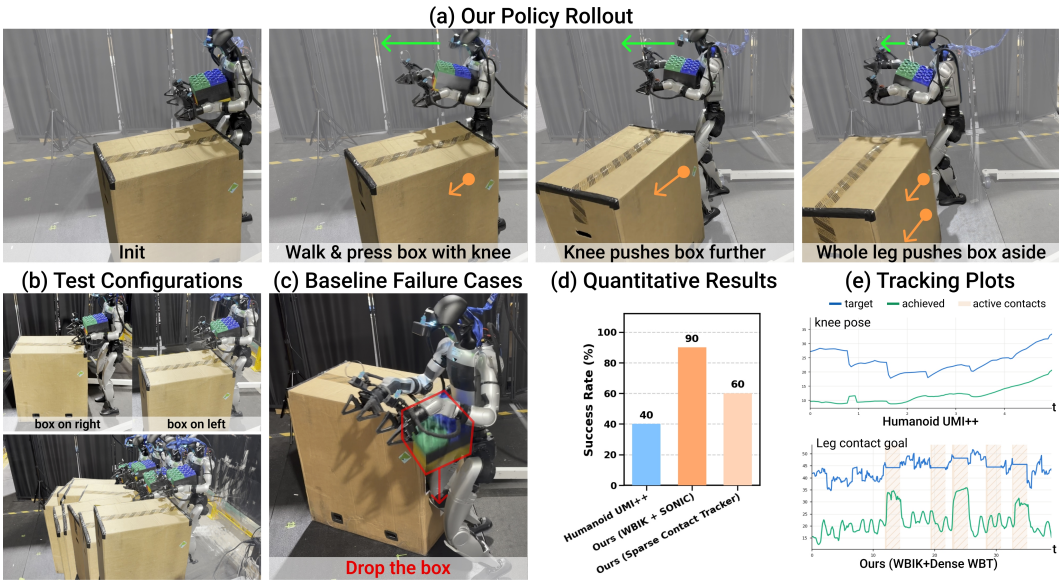

Figure 6: **Carry-and-Push.** (a) Our policy with `WBIK+SONIC` tracker, demonstrating arms load-sharing while walking and pushing with lower body. (b) Test scenarios with varied box placements and robot postures. (c) Failure case of `Humanoid UMI++`, dropping the carried box because of insufficient contact support. (d) Task success rates. (e) Tracking plots of real-world rollouts.

line exhibits substantial knee-pose error, while `Ours` controller follows the commanded leg contact locations more accurately during active contact phases.

6.4 Controller Comparisons

We replay the demonstration trajectories through the controllers in simulation, with and without contact goals. Fig. 7 reports contact surface distance, measured from each commanded contact point to the nearest robot surface, and EEF tracking error. We report medians of per-trajectory median errors over 10 Grasp-and-Embrace trajectories for upper-body evaluation, and 20 trajectories for whole-body evaluation, split equally between Sweep-and-Stow and Carry-and-Push. Adding contact goals reduces contact surface distance for every controller, at little cost to EEF tracking accuracy. With contact goals enabled, `Upper-Body IK` achieves the lowest upper-body EEF error, while `Sparse Contact Tracker` achieves the smallest contact distance and outperforms `WBIK+SONIC` on both metrics. For whole-body tasks, the two trackers exhibit a tradeoff. `Sparse Contact Tracker` achieves a lower contact distance than `WBIK+SONIC`, but a higher EEF error.

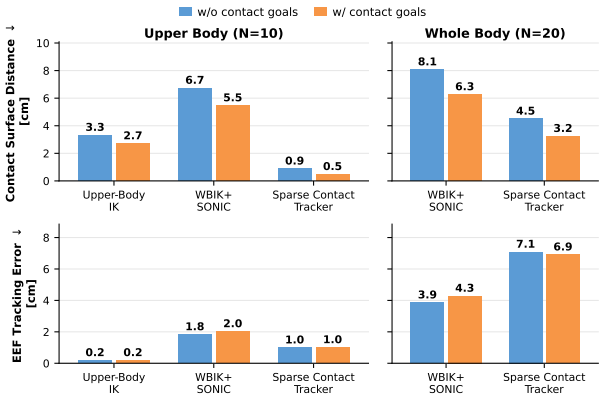

Figure 7: **Controller Comparisons.** We replay the demonstrations in simulation with our three whole-body control formulations, with and without contact goals.

7 Conclusion

We present PanTact, a system that learns humanoid whole-body dexterity from robot-free human demonstrations. The system treats contact as manipulation-centric information being transferred from human to robot with 1) modular contact patches for wearable, reconfigurable, low-cost body-contact sensing; 2) an embodiment-agnostic policy that predicts contact goals alongside end-effector trajectories; and 3) three whole-body controller formulations that realize active contact goals using suitable robot surfaces. We hope the proposed hardware, learning framework, and controller formulations will encourage and facilitate future research on robot whole-body dexterity.

AI USE STATEMENT

We used OpenAI's ChatGPT to assist with manuscript polishing, literature search and summarization, and LaTeX formatting. The authors bear full responsibility for the manuscript's accuracy, originality, and conclusions.

REPRODUCIBILITY STATEMENT

The paper describes the wearable contact interface and demonstration collection procedure, the policy's observation and action representations (§ 4), and the three contact-aware controller formulations (§ 5). Our evaluation specifies the manipulation tasks, demonstration counts, evaluation trials, and performance metrics (§ 6). We plan to fully release the hardware designs, demonstration data, and code to support reproduction and further research.

REFERENCES

Arpit Bahety, Arnav Balaji, Ben Abbatematteo, and Roberto Martín-Martín. Safemimic: Towards safe and autonomous human-to-robot imitation for mobile manipulation. In *RSS 2025 Workshop: Mobile Manipulation: Emerging Opportunities & Contemporary Challenges*, 2025.

Sirui Chen, Shibo Zhao, Zhen Wu, Jiaman Li, Guanya Shi, and C Karen Liu. Scenebot: Contact-prompted general humanoid whole body tracking with scene-interaction. *arXiv preprint arXiv:2606.27581*, 2026.

Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. In *Conference on Robot Learning*, pp. 2729–2749. PMLR, 2025.

Cheng Chi, Zhenjia Xu, Chuer Pan, Eric Cousineau, Benjamin Burchfiel, Siyuan Feng, Russ Tedrake, and Shuran Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

Hojung Choi, Yifan Hou, Chuer Pan, Seongheon Hong, Austin Patel, Xiaomeng Xu, Mark R. Cutkosky, and Shuran Song. In-the-wild compliant manipulation with umi-ft, 2026. URL https://arxiv.org/abs/2601.09988.

Emmanuel Dean-Leon, J Rogelio Guadarrama-Olvera, Florian Bergner, and Gordon Cheng. Whole-body active compliance control for humanoid robots with robot skin. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5404–5410. IEEE, 2019.

Pranay Dugar, Aayam Shrestha, Fangzhou Yu, Bart van Marum, and Alan Fern. Learning multi-modal whole-body control for real-world humanoid robots. *arXiv preprint arXiv:2408.07295*, 2024.

Henrique Ferrolho, Vladimir Ivan, Wolfgang Merkt, Ioannis Havoutis, and Sethu Vijayakumar. Roloma: Robust loco-manipulation for quadruped robots with arms. *Autonomous Robots*, 47 (8):1463–1481, 2023.

Aimee Goncalves, Naveen Kuppuswamy, Andrew Beaulieu, Avinash Uttamchandani, Katherine M Tsui, and Alex Alspach. Punyo-1: Soft tactile-sensing upper-body robot for large object manipulation and physical human interaction. In *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, pp. 844–851. IEEE, 2022.

Zhengmao He, Kun Lei, Yanjie Ze, Koushil Sreenath, Zhongyu Li, and Huazhe Xu. Learning visual quadrupedal loco-manipulation from demonstrations. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 9102–9109. IEEE, 2024.

Yifan Hou, Zeyi Liu, Cheng Chi, Eric Cousineau, Naveen Kuppuswamy, Siyuan Feng, Benjamin Burchfiel, and Shuran Song. Adaptive compliance policy: Learning approximate compliance for diffusion guided control. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4829–4836. IEEE, 2025.

Jaehwi Jang, Zhaoyuan Gu, Alfred Cueva, Zimeng Chai, Junjie Sheng, Thong Nguyen, Himank Galundia, Yifan Wu, Huishu Xue, Isaac Legene, Ojas Mediratta, Davin Doan, Andrew Collins, Sarah Sadegh, KyoungMok Kim, Rishita Dhalbisoi, Zun Chen, and Ye Zhao. Wt-umi: Tactile-based whole-body manipulation via force-supervised contact-aware planning. *arXiv preprint arXiv:2606.13232*, 2026.

Shuo Jiang and Lawson LS Wong. A Hierarchical Framework for Robot Safety using Whole-body Tactile Sensors. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8021–8028. IEEE, 2024.

Marina Kollmitz, Daniel Büscher, Tobias Schubert, and Wolfram Burgard. Whole-body sensory concept for compliant mobile robots. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5429–5435. IEEE, 2018.

Chengshu Li, Mengdi Xu, Arpit Bahety, Hang Yin, Yunfan Jiang, Huang Huang, Josiah Wong, Sujay Garlanka, Cem Gokmen, Ruohan Zhang, et al. Momagen: Generating demonstrations under soft and hard constraints for multi-step bimanual mobile manipulation. In *RSS 2025 Workshop: Mobile Manipulation: Emerging Opportunities and Contemporary Challenges*, 2025.

Xinyao Li, Xialin He, Runpei Dong, and Saurabh Gupta. Contactmimic: Humanoid object interaction via contact control. *arXiv preprint arXiv:2607.08742*, 2026.

Yun Liu, Xiaomeng Xu, Weihang Chen, Haocheng Yuan, He Wang, Jing Xu, Rui Chen, and Li Yi. Enhancing generalizable 6d pose tracking of an in-hand object with tactile sensing. *IEEE Robotics and Automation Letters*, 2023.

Zhengyi Luo, Ye Yuan, Tingwu Wang, Chenran Li, Fernando Castañeda, Sirui Chen, Zi-Ang Cao, Jiefeng Li, David Minor, Qingwei Ben, et al. Sonic: Supersizing motion tracking for natural humanoid whole-body control. *Science Robotics*, 11(117):eaed4592, 2026.

Masaki Murooka, Takahiro Hoshi, Kensuke Fukumitsu, Shimpei Masuda, Marwan Hamze, Tomoya Sasaki, Mitsuharu Morisawa, and Eiichi Yoshida. Tact: Humanoid whole-body contact manipulation through deep imitation learning with tactile modality. *IEEE Robotics and Automation Letters*, 10(8):7819–7826, 2025.

Ruiqian Nai, Boyuan Zheng, Junming Zhao, Haodong Zhu, Sicong Dai, Zunhao Chen, Yihang Hu, Yingdong Hu, Tong Zhang, Chuan Wen, et al. Humanoid manipulation interface: Humanoid whole-body manipulation from robot-free demonstrations. *arXiv preprint arXiv:2602.06643*, 2026.

Keke Qi, Zhibin Song, and Jian S Dai. Safe physical human-robot interaction: A quasi whole-body sensing method based on novel laser-ranging sensor ring pairs. *Robotics and Computer-Integrated Manufacturing*, 75:102280, 2022.

Yimeng Qin, Xiaomeng Xu, William Heap, Aditi Oak, Shuran Song, and Allison Okamura. Panovine: Whole-body visuomotor control for soft growing vine robot. *arXiv preprint arXiv:2606.22923*, 2026.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Motoyasu Tanaka, Kazuyuki Kon, and Kazuo Tanaka. Range-sensor-based semiautonomous whole-body collision avoidance of a snake robot. *IEEE Transactions on Control Systems Technology*, 23 (5):1927–1934, 2015.

Chen Tessler, Yunrong Guo, Ofir Nabati, Gal Chechik, and Xue Bin Peng. Maskedmimic: Unified physics-based character control through masked motion inpainting. *ACM Transactions on Graphics (TOG)*, 43(6):1–21, 2024.

Haoyang Weng, Yitang Li, Nikhil Sobanbabu, Zihan Wang, Zhengyi Luo, Tairan He, Deva Ramanan, and Guanya Shi. Hdmi: Learning interactive humanoid whole-body control from human videos. *arXiv preprint arXiv:2509.16757*, 2025.

Philipp Wu, Yide Shentu, Zhongke Yi, Xingyu Lin, and Pieter Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12156–12163. IEEE, 2024.

Haoyu Xiong, Xiaomeng Xu, Jimmy Wu, Yifan Hou, Jeannette Bohg, and Shuran Song. Vision in action: Learning active perception from human demonstrations. In Joseph Lim, Shuran Song, and Hae-Won Park (eds.), *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 5450–5463. PMLR, 27–30 Sep 2025. URL https://proceedings.mlr.press/v305/xiong25a.html.

Xiaomeng Xu, Dominik Bauer, and Shuran Song. RoboPanoptes: The All-Seeing Robot with Whole-body Dexterity. In *Proceedings of Robotics: Science and Systems*, 2025a.

Xiaomeng Xu, Yifan Hou, Zeyi Liu, and Shuran Song. Compliant residual DAgger: Improving real-world contact-rich manipulation with human corrections. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025b.

Xiaomeng Xu, Jisang Park, Han Zhang, Eric Cousineau, Aditya Bhat, Jose Barreiros, Dian Wang, Jeannette Bohg, and Shuran Song. HoMMI: Learning whole-body mobile manipulation from human demonstrations. In *Proceedings of Robotics: Science and Systems (RSS)*, 2026.

Shibo Zhao, Hengrui Zhang, Peng Wang, Lucas Nogueira, and Sebastian Scherer. Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8729–8736. IEEE, 2021.

Zehui Zhao, Yuxuan Zhao, Gaojing Zhang, Chenxi Liu, Maolin Zheng, and Wenzhao Lian. Halomi: Learning humanoid loco-manipulation with active perception from human demonstrations. *arXiv preprint arXiv:2606.18772*, 2026.

APPENDIX

A DEMONSTRATION INTERFACE DETAILS

A.1 RECORDED DATA

Each demonstration is a synchronized sequence

$$\mathcal{D} = \left\{ \left(I_t^{\text{ego}}, P_t^{\text{ego}}, I_t^L, I_t^R, T_t^H, \mathbf{x}_t^L, \mathbf{x}_t^R, \{(c_{t,i}, f_{t,i}, \mathbf{p}_{t,i})\}_{i=1}^N \right) \right\}_{t=1}^{T_{\text{demo}}}, \tag{7}$$

where $I \in \mathbb{R}^{H_{\text{img}} \times W_{\text{img}} \times 3}$ denotes RGB images and $P^{\text{ego}} \in \mathbb{R}^{H_{\text{img}} \times W_{\text{img}} \times 3}$ contains a 3D point per egocentric pixel. The transform $T^H \in SE(3)$ denotes the head pose, and $\mathbf{x}^L, \mathbf{x}^R$ each contains an EEF pose $T^L, T^R \in SE(3)$ and gripper width $g^L, g^R \in \mathbb{R}$. For sensor i, $c_{t,i} \in \{0, 1\}$, $f_{t,i} \in \mathbb{R}_{\geq 0}$, and $\mathbf{p}_{t,i} \in \mathbb{R}^3$ record binary contact, contact intensity, and location. Here, N is the number of sensors and T_{demo} is the sequence length.

A.2 POST-PROCESSING

Contact binarization. For sensor i, we compute the demonstrated contact state from its raw ADC count $r_{t,i}$ and running baseline β_i:

$$c_{t,i} = \mathbf{1}[\, \beta_i - r_{t,i} > \max(0.35\, \beta_i, r_{\min}) \,], \tag{8}$$

where $r_{\min}$ is a fixed noise floor.

Visual processing. All cameras use a resolution of 960×600. We compute stereo depth offline using the neural depth model provided with the `ZED` camera, back-project it into a pointmap, and apply the body mask described in § 4.1.

Synchronization. All sensor streams are logged on an `NVIDIA Jetson AGX Orin` and share a monotonic clock. During post-processing, we align them to the 60 Hz camera timeline, then sample synchronized sequences at 20 Hz during training. Each training example uses the $h = 2$ observation history and $H = 16$ action horizon defined in § 4.

B POLICY DETAILS

We use a diffusion policy (Chi et al., 2025) with a DiT denoising head (Fig. 3), and finetune a pre-trained DINO-v3 vision transformer (Siméoni et al., 2025) on egocentric and wrist images. For the masked egocentric pointmap, we augment image patch features with sinusoidal encodings of their 3D coordinates (Xu et al., 2026). These observations condition the denoising transformer, trained with ϵ-prediction and sampled with 16 DDIM (Song et al., 2020) steps at inference. Tab. 1 summarizes the training settings. We augment images with color jitter and use an exponential moving average (EMA) of the model weights for evaluation.

Table 1: Policy architecture and training hyperparameters.

Parameter	Value
Transformer depth / width / heads	10 / 768 / 12
Training diffusion steps / inference DDIM steps	100 / 16
Training epochs	500
Effective batch size	48
Base learning rate	7.5×10^{-5}
Encoder learning rate	7.5×10^{-6}
Learning-rate schedule	Cosine decay with warmup
Validation split	5%

C CONTROLLER DETAILS

We provide implementation details for the control strategies in § 5.

C.1 IK AND REFERENCE GENERATION

Task weights and limits. Both upper-body and whole-body IK variants use Eq. (5), with EEF position and orientation weights of 500 and 200, a contact-task weight of 250, and a posture weight of 1 per joint group. Whole-body IK adds pelvis position weights $(80, 80, 40)$, orientation weights $(40, 40, 5)$, and a horizontal center-of-mass weight of 250, encouraging upright posture for stability

Table 2: Sparse contact tracker reward terms.

Group	Term	σ	Weight
Pelvis	position / orientation	0.3 / 0.4	1.0 / 0.5
Body	position / orientation	0.3 / 0.4	1.0 / 1.0
Body	linear / angular velocity	1.0 / 3.14	1.0 / 1.0
EEF	position / orientation	0.3 / 0.4	1.2 / 1.2
EEF	linear / angular velocity	1.0 / 3.14	1.2 / 1.2
EEF fine	position / orientation	0.06 / 0.1	0.3 / 0.3
Contact	Eq. (6)	–	1.0
Regularization	action rate	–	-0.1
Regularization	soft joint limit (0.9 of range)	–	-10
Regularization	uncommanded contact	–	-0.1

while permitting greater variation in pelvis height and yaw. A velocity-damped configuration limit uses gain 0.95, with joint-velocity bounds of $1\,\mathrm{rad/s}$ for the arms and waist and $2\,\mathrm{rad/s}$ for the legs. At each iteration, we recompute the closest surface points, task weights, and Jacobians, then hold them fixed within the QP.

Stepping. In WBIK + SONIC, we impose hard SE(3) constraints on stance foot posture. The solver commands a step when the pelvis remains more than $0.12\,\mathrm{m}$ horizontally from the support midpoint for $0.2\,\mathrm{s}$. The solver releases the trailing foot's constraint and tracks a $0.4\,\mathrm{s}$ cycloidal swing with a peak lift of $0.06\,\mathrm{m}$. The stride follows the pelvis drift direction, with length clamped to 0.10–$0.25\,\mathrm{m}$; swing position and orientation weights are 1000 and 100. During the swing, the stance foot remains fixed and the center-of-mass target shifts to it. At touchdown, the solver fixes the landed foot and enforces a $0.4\,\mathrm{s}$ cooldown before another step.

C.2 Sparse Contact Tracker

Contact slots. Upper-body trackers use $S = 8$ contact slots, four per forearm; whole-body trackers use $S = 16$, adding four per knee. The validity mask selects the slots included in $\mathcal{V}$ in Eq. (6). For each valid slot, $c_i = 1$ requests tight in-contact tracking, while $c_i = 0$ requests a softer pre-contact approach. At episode reset during training, we sample the number of valid slots $|\mathcal{V}|$ uniformly from $\{0, \ldots, S\}$ and draw their targets independently from the rollout's target pool. We permute the slot order and hold the sampled goals fixed throughout the episode. At deployment, contact points are expressed in a gravity-aligned frame centered at the measured robot base and oriented by its yaw.

Reward. Tab. 2 lists the scales and weights for the reward defined in § 5. Motion-tracking terms use exponential kernels e^{-e^2/σ^2}, where e denotes the corresponding global-frame error. EEF terms measure errors at the gripper tool-center points, with additional narrow kernels rewarding precise position and orientation tracking.

Domain randomization. We randomize static friction over $[0.3, 1.6]$, dynamic friction over $[0.3, 1.2]$, restitution over $[0, 0.5]$, and base center-of-mass offsets up to $(2.5, 5, 5)\,\mathrm{cm}$. Joint zero-position biases vary by $\pm 0.015\,\mathrm{rad}$, PD gains by $\pm 10\%$, and link masses by factors of $[0.8, 1.5]$, with an additional base-mass offset of $[-1, 3]\,\mathrm{kg}$. Pelvis velocity perturbations of up to $0.4\,\mathrm{m/s}$ occur every 1.5–$3\,\mathrm{s}$.

PPO settings. We train with PPO (Schulman et al., 2017) for 30,000 iterations in 4096 parallel environments. The actor learning rate starts at 2×10^{-5} and adapts within $[10^{-5}, 2 \times 10^{-4}]$ using a KL target of 0.01; the critic learning rate is 10^{-4}. The entropy coefficient is 0.013. The gradient norm clipping threshold is 0.1, and the initial action noise is 0.05.

C.3 Real-world Deployment

Policy inference and low-level control run asynchronously. Each predicted action chunk is timestamped at its source observation. A scheduler discards expired steps, replaces overlapping targets with the latest prediction, and resamples commands to the tracker's $50\,\mathrm{Hz}$ control rate.

The policy expresses actions in the left-gripper frame at the source observation. For execution, we transform them into the world frame using pose estimates from Super Odometry (Zhao et al., 2021), which leverages the G1's built-in head-mounted LiDAR-inertial unit.