FORGE: Force-Aware Multi-Granularity Control for Efficient Humanoid Loco-Manipulation

Liangjun Xing[1,2,*], Ke Fan[1,4,*], Kailin Li[1,*], Bin Wang[1,3], Xiaojie Niu[1], Qingwei Ben[1,5], Jiafei Cao[1], Wei Yao[2], Yebin Liu[2,†], Jingbo Wang[1,†], Weinan Zhang[1,†]

[1]Shanghai AI Laboratory [2]Tsinghua University [3]Zhejiang University
[4]Shanghai Jiao Tong University [5]The Chinese University of Hong Kong

[*]Equal contribution. [†]Corresponding authors.

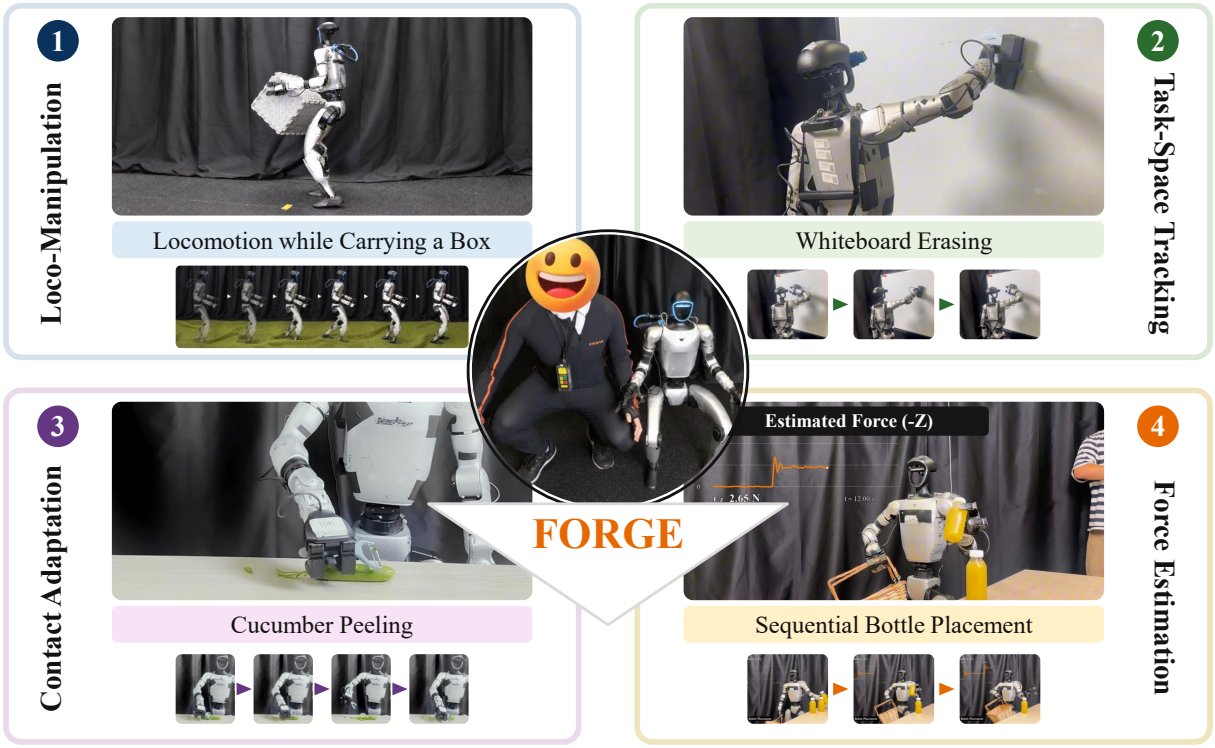


Fig. 1. **FORGE** enables efficient humanoid loco-manipulation via multi-granularity control: velocity navigation, foot placement, single-/dual-arm end-effector tracking, and force-conditioned contact adaptation.

***Abstract*— Humanoid loco-manipulation requires coordinated locomotion, balance regulation, and manipulation under frequent contacts, making teleoperated demonstration collection difficult: a practical controller must simultaneously provide precision, stability, and contact awareness. We propose FORGE, a force-aware multi-granularity control system with a coarse-to-fine command space spanning velocity navigation, planar foot placement, wrist-pose tracking, and optional force-conditioning inputs. Policy training requires no full-body human motion trajectories or joint-angle references; deployment consumes only sparse bilateral wrist poses relative to the pelvis, coarse motion commands, and optional desired wrist-load inputs. For sustained contact, FORGE combines a momentum observer with a learned correction trained using simulator contact-force supervision. At deployment, it estimates low-frequency, wrist-equivalent interaction forces from proprioception and reconstructed low-level arm PD torques, without wrist force/torque sensors. A quasi-static virtual-target loop uses this estimate and an optional desired applied-force input to adapt the wrist target. A frozen locomotion prior and asymmetric residual actions stabilize manipulation-induced whole-body corrections. Experiments evaluate unified-control stability, task-space tracking, quasi-static wrist-equivalent force estimation, squeezing forces during sustained contact with a rigid box, and downstream diffusion-policy execution of a squat-and-lift task across payload conditions. Website: `bb0928.github.io/forge`.**

I. Introduction

Humanoid loco-manipulation requires coordinating manipulation with stepping, leaning, or posture adjustment under complex dynamics and frequent contacts [1]–[3]. Unlike fixed-base dual-arm robots, which benefit from stable support and mature teleoperation systems [4], humanoids must map operator intent to whole-body behavior under underactuation and changing contacts. Balance errors can compromise stability, joint errors can accumulate into end-effector drift, and rigid position control can disrupt sustained interaction. A practical controller must therefore jointly provide *Stability*,

Precision, and *Contact Awareness*, including interaction-force estimation for regulating sustained contact.

Existing approaches satisfy these requirements only in part. Full-body tracking and large-scale motor-control models [5]–[7] provide broad motion repertoires, but dense motion-reference objectives do not directly expose task-space commands for precise foot placement, wrist targeting, and sustained contact. Decoupled locomotion/upper-body stacks [8], [9] preserve locomotion stability but can limit task-dependent whole-body coordination. Force-aware approaches [10]–[12] demonstrate learned force estimation and force-conditioned position control, yet sensorless estimation is difficult: external contact, friction, actuator behavior, and whole-body coupling are entangled in proprioception.

A unified controller must therefore support coarse-to-fine teleoperation: approaching a task region, adjusting stance, tracking wrist trajectories, and adapting to contact. Commands must remain composable as task demands change.

These limitations motivate **FORGE**, which *forges* a **FOR**ce-aware multi-**G**ranularity policy for **E**fficient humanoid loco-manipulation. Its unified coarse-to-fine interface supports base velocity commands, planar foot-placement targets, Cartesian wrist-pose targets, and optional force-conditioning inputs, individually or jointly. Granularity-conditioned masking [13] activates the observations and rewards relevant to each command group. Training needs no full-body human trajectories or human joint-angle references; teleoperation supplies only bilateral wrist poses relative to the pelvis and coarse commands from wearable trackers, handheld VR devices, or vision-based pose estimation.

For sustained contact, a *Physics-Informed Hybrid Force Estimator* learns a simulator-supervised correction to a model-based momentum observer. At deployment, it uses arm joint states and reconstructed low-level PD torques without wrist force/torque sensors. The estimator captures the low-frequency component of the equivalent interaction force at each wrist. We combine this estimate with an optional command for the robot-applied force to adjust wrist position targets during sustained, quasi-static contact.

To stabilize mixed-objective training, we freeze a pre-trained lower-body locomotion expert and learn an asymmetric residual policy that controls the upper body and applies corrective lower-body actions. This reduces exploration difficulty while preserving the locomotion prior and allowing task-dependent whole-body coordination.

Simulation and hardware experiments show that masking and asymmetric residual training improve unified-control stability and task-space tracking over their ablations, while the hybrid estimator reduces force-estimation error under the evaluated quasi-static contacts compared with analytical-only and learned-only estimators. To assess **FORGE** as a whole-body executor for visuomotor learning, we train a diffusion policy [14] on collected demonstrations; it commands **FORGE** to execute squat-and-lift across payloads.

Our contributions are a unified multi-granularity controller trained without full-body human motion references, physics-informed sensorless wrist-force estimation with contact-load adaptation, and asymmetric residual learning for stable whole-body loco-manipulation.

II. Related Work

A. Whole-Body Control and Teleoperation Systems

Humanoid loco-manipulation controllers broadly follow two paradigms. Motion-reference-driven policies learn broad whole-body behaviors from retargeted or captured motions [5]–[7], [15], [16], whereas modular systems separate learned locomotion from upper-body control [8], [9]. The former relies on dense whole-body references; the latter preserves locomotion stability but can restrict task-dependent whole-body coordination. SoFTA explicitly separates slow locomotion and fast upper-body control to stabilize the end effectors during walking [17]. Related work also explores masked motion generation and reusable behavior priors [13], [18]. In contrast, **FORGE** trains from task-space commands without full-body human motion references and uses a frozen locomotion prior with asymmetric residual learning [19], [20]. Its teleoperation input comprises only pelvis-relative wrist poses and coarse commands, whether obtained from wearable, handheld, or vision-based trackers.

Building on 3D Diffusion Policy [21], Improved 3D Diffusion Policy combines upper-body teleoperation with 3D visuomotor learning on a cart-supported humanoid [22]. Our downstream evaluation instead uses a diffusion policy to command a balancing whole-body executor.

B. Force Perception and Compliant Interaction

Humanoid controllers increasingly incorporate interaction-aware or force–position objectives [23], [24], while sensorless estimators infer wrist interaction from proprioception [10]–[12]. UniFP estimates external force from proprioceptive history and maps force–position commands to compensated position targets [10]. **FORGE** retains this separation but adds a momentum-observer prior before learned correction. It estimates low-frequency translational wrist interaction for quasi-static virtual-target adaptation, not full-wrench sensing or dynamic impedance control.

III. Method

FORGE combines a compact multi-granularity command interface with a force-aware contact pathway. Navigation, stance, and wrist-pose commands share one whole-body policy; a **Physics-Informed Hybrid Force Estimator** and virtual-target loop provide sustained-contact adaptation without wrist force/torque sensors. **Asymmetric Residual Actions** anchor this policy on a frozen locomotion prior (Fig. 2).

A. Multi-Granularity Control System

One policy accepts four individually or jointly active command groups: base velocity $\mathbf{v}_{\mathrm{cmd}}$ (**Navigation**), planar left/right footholds $\mathbf{p}_{\mathrm{feet}} \in \mathbb{R}^{2\times 2}$ (**Stance**), bilateral wrist poses $\mathbf{T}_{\mathrm{wrist}} \in \mathrm{SE}(3)^2$ (**Pose**), and optional desired robot-applied wrist forces $\mathbf{f}_{\mathrm{cmd}} \in \mathbb{R}^{2\times 3}$ (**Force**). An operator or upstream model may supply the force command for virtual-target updates (Sec. III-B), not direct torque control.

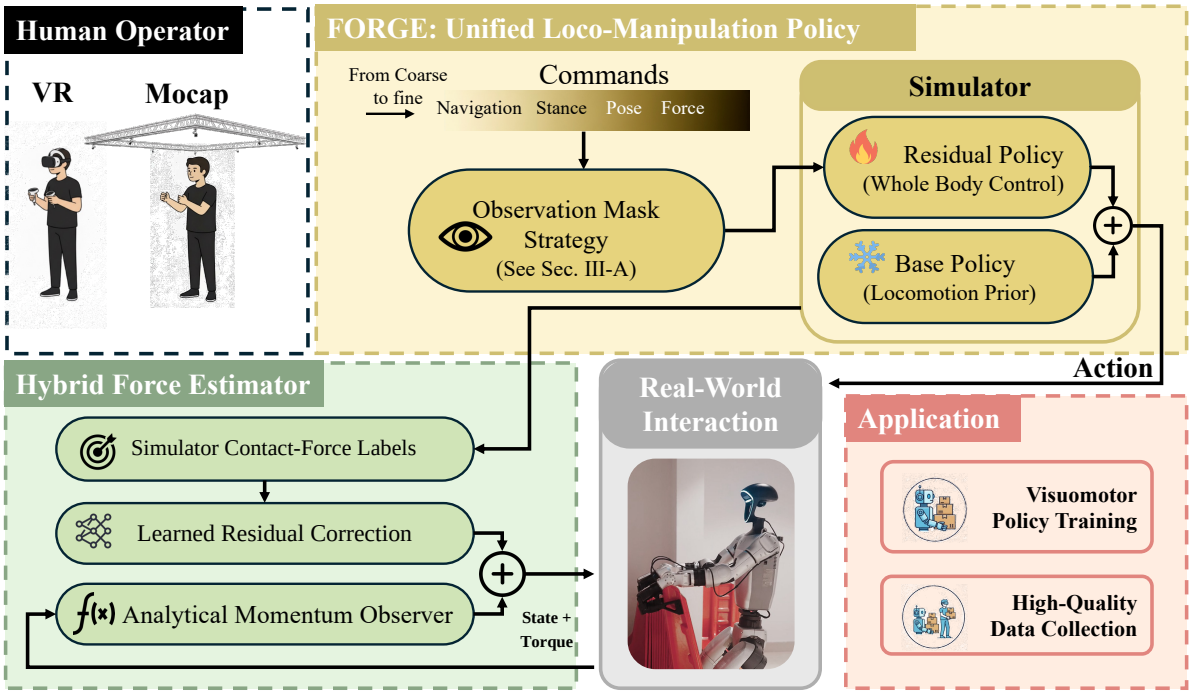


Fig. 2. **Overall pipeline of FORGE.** The framework has three components: multi-granularity control over navigation, stance, pose, and optional force-conditioning inputs; a physics-informed hybrid force estimator and virtual-target contact-adaptation loop; and asymmetric residual actions anchored on a frozen locomotion prior. Policy training uses sparse task-space commands rather than full-body human motion references.

During training, we sample active combinations of these four command groups and their target values. For each inactive group, we replace the associated observation entries with null values and exclude its reward terms from the training objective. Masking both observations and objectives prevents stale targets from conflicting with the active task. Four task-level activation bits expand into a per-group mask supplied with a six-step observation history; compound behaviors activate the required groups without selecting another controller or changing the 29-D action semantics. The deployed one-step signal contains 138 scalars: 3 navigation, 5 stance, 17 pose, and 6 desired-force command entries, together with 6 IMU quantities, 29 joint positions, 29 joint velocities, the previous 29-D action, and the 14-D momentum-observer residual. The explicit semantic-group mask is appended separately and excluded from this count.

As shown in Fig. 3, every step is partitioned into semantic groups for base motion, footholds, wrist poses, desired wrist forces, the 14-D observer residual, and proprioception. After the mask is applied, one shared MLP maps each per-step vector to a token. A two-layer Transformer (128-D latent, four heads, 256-D feed-forward layers) uses a learnable `CLS` token to summarize the sequence as h_{CLS}. In parallel, a 512–256–128 adaptation encoder maps the same six masked observable frames to a 64-D latent z_{adapt}; the 512–256–128 action head consumes $[h_{\mathrm{CLS}}, z_{\mathrm{adapt}}]$. A 128–64 decoder maps z_{adapt} to 15 supervised outputs: a 6-D observer-force correction, bilateral wrist position (6-D), and base linear velocity (3-D). The first six outputs remain active at deployment and refine the observer estimate; wrist position and base velocity are auxiliary training targets. Simulator

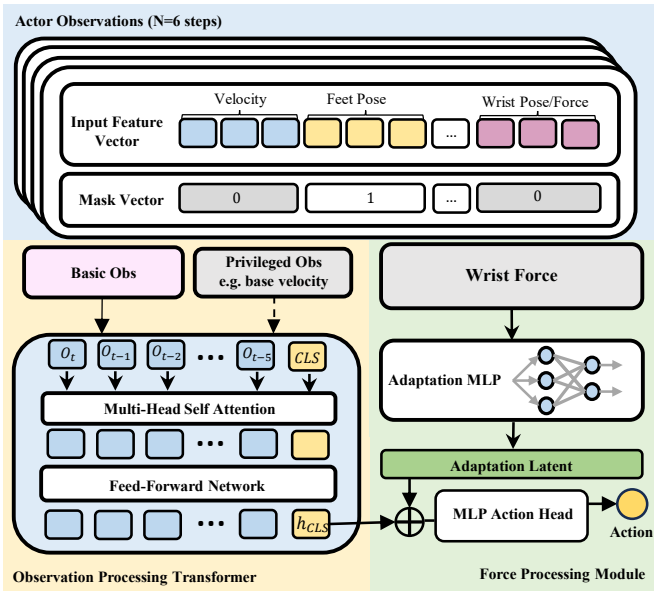
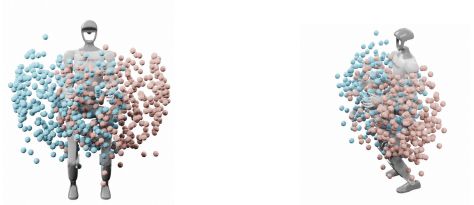


Fig. 3. Actor architecture. Solid deployment paths use only the masked observable history. "Wrist Force" denotes the observer-derived residual history, not a force/torque-sensor reading or simulator force. The adaptation decoder's 6-D correction refines the observer at deployment; the dashed privileged quantities supply training targets only.

contact forces supervise the corrected force estimate, while simulator wrist positions and base linear velocities supervise the auxiliary predictions during training; none of these reference values is supplied to the actor.

Stance targets are sampled within a reachable support region and expressed in the base-yaw frame. For Pose targets, we interpolate valid upper-body joint anchors, map them through wrist forward kinematics, and rotate the resulting



(a) Front view. (b) Side view.

Fig. 4. Kinematics-aware bilateral wrist-target distributions.

Cartesian targets into the base-yaw frame. This procedure produces the reachable distributions in Fig. 4; the actor observes wrist targets rather than the sampled joints and resolves whole-body IK implicitly.

B. Physics-Informed Hybrid Force Estimator

While the Multi-Granularity Control System specifies targets, the *Force* granularity requires contact feedback. For each wrist $i \in \{L, R\}$, we distinguish the optional desired robot-applied force $\mathbf{f}^i_{\text{cmd}}$, the wrist-equivalent environmental reaction $\mathbf{f}^i_{\text{env}}$, and its estimate $\hat{\mathbf{f}}^i_{\text{env}}$. The desired and environmental forces use the same world-frame axes but represent opposite sides of the contact pair; consequently, ideal quasi-static force balance satisfies $\mathbf{f}^i_{\text{env}} = -\mathbf{f}^i_{\text{cmd}}$. Here "wrist-equivalent" denotes the 3D translational force whose Jacobian transpose best explains the observed arm-joint disturbance under the assumed wrist-contact condition; unmodeled link contacts and whole-body coupling are not separately identified. Rather than using dedicated wrist force/torque sensors, we estimate its low-frequency, quasi-static component from arm joint states and reconstructed low-level PD torques. The estimator combines a *model-based momentum observer* [25] with a *learned correction*: the observer supplies a physically structured disturbance estimate, while the learned term compensates repeatable residual errors from the reduced arm model and actuation.

1) Model-Based Momentum Observer: We first estimate a joint-space disturbance residual using an arm momentum observer [25], avoiding direct acceleration differentiation. Let $\mathbf{q}^u \in \mathbb{R}^{14}$ contain the two 7-DoF arm configurations and let $\mathbf{M}^u(\mathbf{q}^u)$ assemble their arm-only mass matrices. With internal momentum state $\hat{\mathbf{p}}$ and observer residual $\mathbf{c} \in \mathbb{R}^{14}$, each control step applies

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{p}}_k + \Delta t\big(\boldsymbol{\tau}_m - \hat{\mathbf{g}} - \hat{\boldsymbol{\tau}}_f + \mathbf{c}_k\big), \tag{1}$$

$$\mathbf{c}_{k+1} = \beta \mathbf{c}_k + \Delta t\, \mathbf{K}\big(\mathbf{M}^u(\mathbf{q}^u)\dot{\mathbf{q}}^u - \hat{\mathbf{p}}_{k+1}\big). \tag{2}$$

where $\boldsymbol{\tau}_m$ is a PD-torque proxy reconstructed from commanded joint targets and measured joint states, rather than motor-current torque; $\hat{\mathbf{g}}$ and $\hat{\boldsymbol{\tau}}_f$ model gravity and smooth Coulomb friction. We set $\beta = 0.9$ and initialize $\hat{\mathbf{p}}_0 = \mathbf{c}_0 = \mathbf{0}$. The assembled arm-only mass matrices, computed with the composite-rigid-body algorithm (CRBA) [26], preserve within-arm configuration coupling but omit the floating base, waist, legs, and cross-arm coordinates. Consequently, velocity-dependent terms, excluded whole-body coupling, and actuation mismatch are not canceled analytically; they remain in the residual learned by the correction pathway. We split $\mathbf{c} = [\mathbf{c}^L; \mathbf{c}^R]$ and project each 7-D arm residual through its translational wrist Jacobian $\mathbf{J}^i_v$:

$$\hat{\mathbf{f}}^i_{\text{obs}} = \big(\mathbf{J}^i_v \mathbf{J}^{i\top}_v + \lambda \mathbf{I}_3\big)^{-1} \mathbf{J}^i_v \mathbf{c}^i, \qquad i \in \{L, R\}, \tag{3}$$

where λ improves numerical stability near singularities. We use diagonal gain $\mathbf{K} = 20\mathbf{I}$, smooth Coulomb coefficient $1.5\,\text{N m}$ with $\epsilon = 0.05\,\text{rad/s}$, joint-residual clipping at 20, and $\lambda = 10^{-3}$. The wrist estimate is norm-clipped at $50\,\text{N}$ and exponentially smoothed with coefficient 0.8; observer update and projection run at the $50\,\text{Hz}$ policy rate.

2) Learned Compensation with Privileged Supervision: The force-correction output of the adaptation decoder is denoted $\Delta\mathbf{f}_\theta(\mathbf{o}_{t-5:t})$. Rather than relearning the complete force mapping, it predicts only the discrepancy between the structured observer output and the simulator contact-force reference from the same six masked frames. The final estimate and training objective are

$$\hat{\mathbf{f}}_{\text{env}} = \hat{\mathbf{f}}_{\text{obs}} + \Delta\mathbf{f}_\theta(\mathbf{o}_{t-5:t}), \quad \mathcal{L}_{\text{est}} = \mathbb{E}\left[\left\|\hat{\mathbf{f}}_{\text{env}} - \mathbf{f}_{\text{sim}}\right\|_2^2\right]. \tag{4}$$

Simulator contact forces $\mathbf{f}_{\text{sim}}$ supervise the corrected estimate and never enter the actor. At deployment, the decoder predicts corrections from observable history; $\mathbf{f}_{\text{cmd}}$ is the target-adaptation command, not the force-estimation target.

3) Force-Conditioned Virtual-Target Adaptation: Following the force–position separation used in UniFP [10], we use the desired applied-force input to bias a position target. For each wrist $i \in \{L, R\}$, the desired environmental reaction is $\mathbf{f}^{i,*}_{\text{env}} = -\mathbf{f}^i_{\text{cmd}}$. Around the nominal wrist target $\mathbf{p}^{i,*}_w$, we define

$$\mathbf{e}^i_f = \hat{\mathbf{f}}^i_{\text{env}} - \mathbf{f}^{i,*}_{\text{env}} = \hat{\mathbf{f}}^i_{\text{env}} + \mathbf{f}^i_{\text{cmd}}, \qquad \mathbf{p}^{i,\text{virt}}_w = \mathbf{p}^{i,*}_w + \mathbf{K}_f^{-1}\mathbf{e}^i_f, \tag{5}$$

where the signed force-balance error vanishes when the desired robot-applied force is balanced by the estimated environmental reaction. The diagonal matrix $\mathbf{K}_f$ has units N/m, and $\mathbf{K}_f^{-1}$ maps the mismatch to a position offset in m/N. The two wrist updates are computed independently and then supplied jointly as bilateral pose commands to the same whole-body policy. Force estimation and target updates operate in world coordinates; policy-facing commands are rotated into the base-yaw frame. The $50\,\text{N}$ norm bound and exponential smoothing are applied to the estimated force before its conversion into the virtual-target offset; they are not a separate hard clip on $\mathbf{p}^{i,\text{virt}}_w$. Thus, the deployment loop is closed entirely through observable arm state and commanded targets: the torque proxy drives the observer, the learned correction refines its environmental-force estimate, and the desired applied-force input together with that estimate updates the wrist target tracked by the policy. This is an algebraic, admittance-like target adaptation.

At each $50\,\text{Hz}$ control step, the hybrid force estimate is updated and used to compute the virtual wrist target. During training, each Cartesian component of $\mathbf{f}_{\text{cmd}}$ is sampled in $[-20, 20]\,\text{N}$ and smoothly ramped, held, and released, while independently generated simulator contact forces supervise correction. We use $K_f = 200\,\text{N/m}$ in force-command mode.

Training uses 2,048 parallel environments, 50 simulation steps/PPO update, and up to 20,000 updates.

C. Asymmetric Residual Whole-Body Control

Mixed locomotion, pose, and contact objectives can destabilize gait learning because unrestricted whole-body exploration can sacrifice support behavior to reduce short-horizon manipulation errors. We therefore train a lower-body expert with upper-body actions fixed to zero, freeze it, and learn a whole-body residual $\Delta\mathbf{a} = [\Delta\mathbf{a}^{\ell}; \Delta\mathbf{a}^{u}] \in \mathbb{R}^{29}$. Freezing the expert prevents gradients from manipulation and contact objectives from rewriting the nominal gait, while the trainable lower-body residual still permits command-dependent foothold and balance corrections. The residual policy receives the active multi-granularity command and learns these bounded leg corrections alongside direct upper-body actions. Summing the frozen action and residual before executing one whole-body action ensures that stance adjustment and manipulation-induced posture changes remain coupled rather than handled by independent controllers. With 15 lower-body and 14 upper-body dimensions, the executed action is

$$\mathbf{a} = \left[\mathbf{a}^{\ell}_{\text{base}} + \Delta\mathbf{a}^{\ell};\, \Delta\mathbf{a}^{u}\right], \tag{6}$$

where $\mathbf{a}^{\ell}_{\text{base}}$ is the frozen locomotion action. PPO exploration uses action-standard-deviation limits of 0.2 for lower-body and 1.0 for upper-body residuals, protecting the gait prior while retaining upper-body exploration.

IV. Experimental Evaluation

We evaluate four aspects of **FORGE**: (i) how its hardware latency and task-space tracking compare with existing humanoid controllers; (ii) whether one policy supports multiple command granularities without destabilizing locomotion; (iii) whether the hybrid estimator improves wrist-equivalent interaction-force estimation and supports contact-load adaptation; and (iv) whether **FORGE** serves as a downstream whole-body executor for visuomotor manipulation. Fig. 5 shows sequential bottle placement and the estimated vertical wrist-force responses around bottle release.

A. Implementation Details

Unless otherwise specified, ablation variants share the training settings and environment randomization, except for the components being ablated. Early training fixes yaw-rate commands at zero, smooths torque commands with EMA ($\alpha = 0.3$), and applies left–right symmetric augmentation. Task-space termination is disabled for the first 300 steps of every episode and until 2,000 global updates; afterward, we use $\delta_{\text{term}} = 0.2\,\text{m}$. The lower-body residual action standard deviation is clipped at 0.2.

The locomotion expert and residual stage use Adam PPO with learning rate 10^{-3} under an adaptive schedule targeting KL 0.01, discount $\gamma = 0.99$, GAE parameter $\lambda = 0.95$, clip ratio 0.2, entropy coefficient 0.01, five learning epochs, four minibatches, and gradient-norm limit 1.0. Both stages are capped at 20,000 updates. Isaac Gym advances at 200 Hz with the TGS solver (four position iterations), $0.01\,\text{m}$ contact offset, and zero rest offset; four-step decimation yields the 50 Hz policy rate. Actions specify joint targets as $\mathbf{q}_{\text{des}} = \mathbf{q}_{\text{default}} + 0.25\mathbf{a}$. The low-level stiffness/damping pairs in N m/rad and N m s/rad are $100/2$ for hips, $150/4$ for knees, $40/2$ for ankles, $300/5$ for the waist, and $40/1$ for shoulder, elbow, and wrist joints.

The actor uses six observable frames, whereas a 512–256–256 critic receives one current privileged frame containing the actor signals plus bilateral simulator wrist force (6-D), wrist position (6-D), virtual target (6-D), and base linear velocity (3-D). Auxiliary losses for wrist force, wrist position, and base velocity use relative weights $1.0/1.0/0.25$, with overall adaptation-loss coefficient 1.0. Domain randomization samples added base mass in $[-5, 10]\,\text{kg}$ and hand payload in $[-0.1, 0.3]\,\text{kg}$, link-mass and PD-gain multipliers in $[0.9, 1.1]$, COM and body-frame offsets in $[-0.1, 0.1]\,\text{m}$, friction in $[0.1, 2.5]$, and initial-joint scaling/offset in $[0.8, 1.2]$/$[-0.1, 0.1]\,\text{rad}$. External base pushes are applied every 4 s with velocity perturbations up to $0.5\,\text{m/s}$ horizontally and $0.2\,\text{m/s}$ vertically.

TABLE I
Comparison with Humanoid Controllers

Method	Latency (ms) ↓	Foot MPE (cm) ↓	Wrist MPE (cm) ↓	Contact force (N) ↓
HOMIE	27	N/A	N/A	15.48
TWIST2	197	7.87	7.76	17.16
SONIC	252	7.35	6.42	15.32
FORGE	**22**	**5.42**	**4.15**	**8.52**

Note: Latency/tracking: 20 hardware trials per method. Contact force: simulated bilateral mean absolute normal reaction under identical wrist-position targets; lower values indicate less squeezing in this test. Latency: time delay from command transmission to robot end-effector reaching the target position.

B. Comparison with Existing Humanoid Controllers

Table I compares four controllers on the same robot and target sequences over 20 trials each: **FORGE**, HOMIE [8], TWIST2 [16], and SONIC [6]. For TWIST2 and SONIC, we use the official implementations and checkpoints and reproduce their training pipelines locally. Control latency is the time from teleoperation-command emission to robot end-effector arrival at the target position. Foot and wrist MPE are temporal mean Euclidean target errors in planar 2D and 3D, respectively, averaged over the two limbs and then across trials. HOMIE directly maps exoskeleton commands to upper-body joints. With no wrist-target or foothold interface, tracking is N/A.

We additionally compare the squeezing force generated during sustained bilateral contact in simulation. All methods maintain bilateral contact with the same rigid $0.4 \times 0.4 \times 0.4\,\text{m}$ box, centered at $(0.25, 0, 0.25)\,\text{m}$ in the pelvis frame, with the same left/right wrist-position targets at $(0.25, \pm 0.15, 0.25)\,\text{m}$. These targets lie $5\,\text{cm}$ inside the corresponding box surfaces. From $2\,\text{s}$ after command onset, we time-average the absolute normal reaction force and average both wrists. This measures each complete controller's squeezing force under identical wrist-position targets.

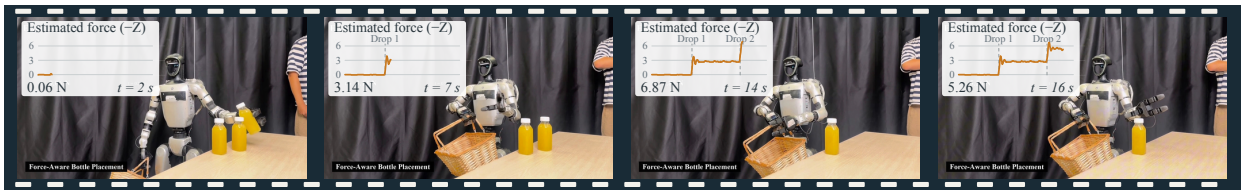


Fig. 5. Sequential bottle placement. Four ordered frames show bottles placed into a basket, with vertical wrist-equivalent force traces at marked releases.

TABLE II
PRINCIPAL REWARD TERMS AND FORMULATIONS

Reward Term	Weight	Formulation / Error Term
Command Tracking		
Vel X / Y Track	2.0 / 1.5	$\exp(-\|\Delta v_{x,y}\|^2/\sigma)$
Ang Vel / Int Track	2.0 / 2.0	$\exp(-\|\Delta\omega_z\|^2/\sigma)$, $\exp(-I_{err}/\sigma_{int})$
Base Height Track	2.0	$\exp(-\|\Delta h\| \cdot 10)$
End-Effector Tasks		
Wrist Pos Track	5.0	$\exp(-\|\Delta\mathbf{p}_w\|/\sigma_w)$
Wrist Quat Track	2.0	$\exp(-\|2\arccos(\|w_{err}\|)\|/\sigma_q)$
Planar Feet Pos Track	5.0	$\exp(-\|\Delta\mathbf{p}_{f,xy}\|/\sigma)$
Force-Conditioned Contact Adaptation		
Virtual-Target Wrist Track	5.0	$\exp(-\|\mathbf{p}_w - \mathbf{p}_{virt}\|/\sigma_w)$
Virtual-Target Offset Penalty	0.01	$-\max(0, \|\Delta\mathbf{p}_{virt}\| - d_{th})$
Virtual-Target Smoothness	0.001	$\exp(-\|\Delta\mathbf{e}_{virt}\|/0.05)$
Stability & Posture		
Orientation	-3.0	$\|\mathbf{g}_{proj,xy}\|^2$
Penalties		
Action Rate / Smooth	-0.01	$\|a_t - a_{t-1}\|^2$ / $\|a_t - 2a_{t-1} + a_{t-2}\|^2$
Feet Slip	-0.25	$\|v_{f,xy}\|$

Note: Weights follow the listed terms. Penalties use negative weights, except the offset term with a negative formula.

TABLE III
MASKING ABLATION RESULTS

Metric	w/o mask	Ours
Valid rollout (s) ↑	15.69 ± 1.32	19.46 ± 0.52
Lin. velocity error (m/s) ↓	0.194 ± 0.003	0.162 ± 0.003
Planar feet MPE (m) ↓	0.182 ± 0.002	0.167 ± 0.002
Wrist MPE (m) ↓	0.115 ± 0.003	0.079 ± 0.004

Note: Mean ± SD over 20 rollouts per variant, each capped at 20 s.

C. Simulation Validation of Unified Control

The training reward is $R_t = \sum_i w_i r_i$, with all weights scaled by the policy period $\Delta t = 0.02$ s. Tracking terms use $\exp(-\|e\|^2/\sigma)$ unless noted otherwise, with $\sigma = 0.25$ and wrist-position scale $\sigma_w = 0.20$. Table II lists the principal reward terms, weights, and formulations.

Unified control, masking, and stability: Fig. 6 and Tab. III evaluate whether a single policy can coordinate velocity navigation, foot-placement control, wrist tracking, and force-conditioned contact adaptation while remaining trainable. An end-to-end 29-DoF policy trained without a locomotion prior fails to learn stable behavior, whereas the base–residual design converges. Removing residual action-std clipping causes late-stage reward collapse. The hard-switching baseline in Fig. 6 corresponds to the "w/o mask" ablation in Tab. III. For inactive command groups, our method masks the corresponding observations and excludes their task-specific reward terms from training. This ablation increases early-termination frequency and velocity, foot, and wrist errors, consistent with interference between inactive command groups and active objectives. Together, the residual prior, lower-body exploration clipping, and masking support stable multi-granularity training under the evaluated command schedule.

Fig. 7 compares tracking-error distributions under different termination thresholds. The 0.2 m setting concentrates errors at low values, whereas 0.3 and 0.4 m yield broader distributions with longer high-error tails; we therefore use $\delta_{\text{term}} = 0.2$ m in our training configuration.

D. Wrist-Equivalent Force Estimation and Contact-Load Adaptation

We compare observer-only, learned-only, and hybrid force estimates under staged wrist loads (Fig. 8). Fig. 9 presents the contact-adaptation test and planar-force evaluation. A fragile-box comparison switches the desired-force-conditioning pathway on or off while keeping the robot, object, nominal wrist trajectory, and task fixed. The contact-force column in Table I separately compares complete controllers under identical wrist-position targets.

Contact-sensitive teleoperation: We lift the same foam box with or without force feedback (Fig. 9a), using identical policy weights, initial pose, nominal wrist trajectory, and lift target. Disabling feedback tracks nominal wrist targets directly; enabling it adapts them using estimated interaction forces. Both complete the lift, but feedback (left) produces less visible deformation than the baseline (right).

Force estimation accuracy: We evaluate force estimation on G1 hardware using manually applied wrist loads measured with a handheld force gauge (Fig. 9b). The gauge is aligned with each evaluated axis, and its readings provide a digitally logged, synchronized reference. The deployed estimator uses arm joint states and reconstructed PD torques without wrist force/torque sensors. Throughout estimator evaluation, $\mathbf{f}_{\text{cmd}} = \mathbf{0}$ and every gauge load is externally applied. We repeat each load condition 10 times for staged 0–20 N loads on all axes, signed planar forces in $[-30, +30]$ N, and held-out vertical loads of 35/40/50 N. Fig. 8 shows representative trials; shading in Fig. 9c shows one standard deviation over ten trials. The learned-only baseline follows UniFP's proprioceptive-history estimator design [10].

The hybrid estimate follows the gauge more closely than observer-only and learned-only variants for staged loads

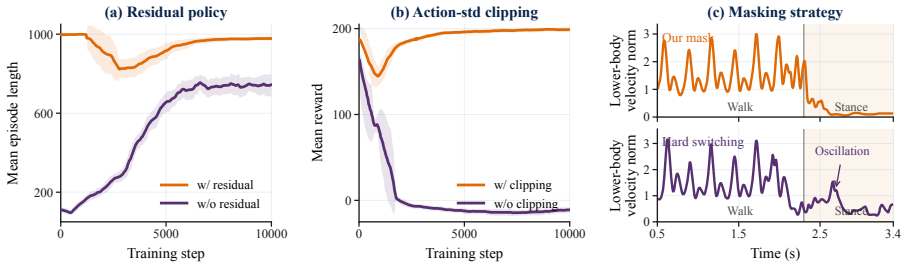


Fig. 6. Unified-control ablations. (a) Asymmetric base+residual converges; end-to-end 29-DoF learning does not. (b) Action-std clipping prevents late collapse. (c) Masking smooths walk→stance transitions.

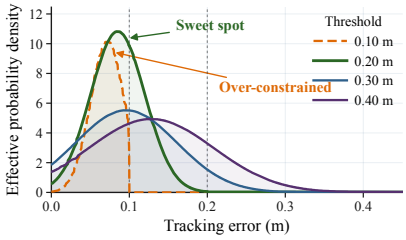


Fig. 7. Precision–stability sweep over $\delta_{\text{term}} \in \{0.1, 0.2, 0.3, 0.4\}$ m.

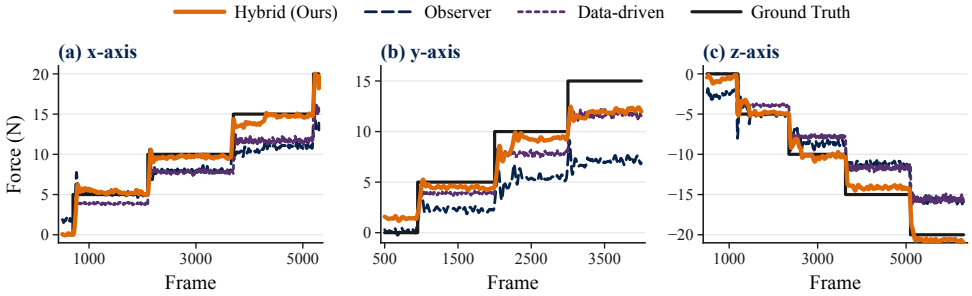
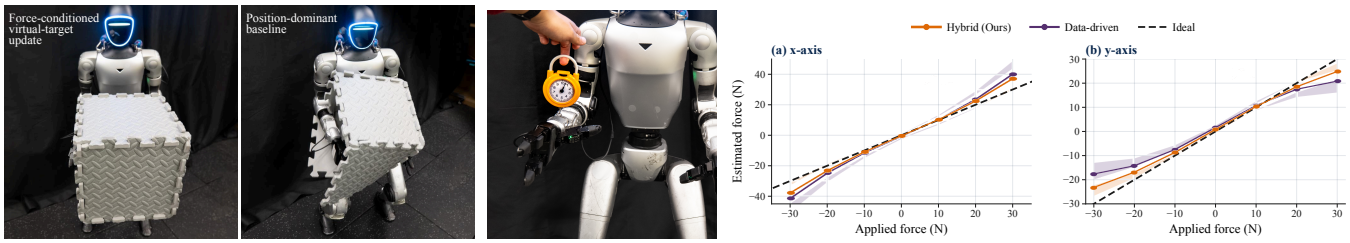


Fig. 8. Representative single-trial real-world $x/y/z$ wrist-equivalent interaction-force estimates at staged $0, 5, 10, 15, 20$ N loads. Each load is repeated 10 times with digitally synchronized force-gauge references.



(a) Foam-box lifting. (b) Force-test setup. (c) Planar x/y loads.

Fig. 9. Contact-load adaptation and wrist-equivalent force estimation. (a) Foam-box lifting: feedback enabled (left) reduces visible deformation relative to feedback disabled (right). (b) Force-test setup. (c) Planar-load evaluation under external forces in $[-30, +30]$ N: the hybrid estimator records lower error than the learned-only estimator. Shading denotes one standard deviation over ten repeats, using synchronized force-gauge references.

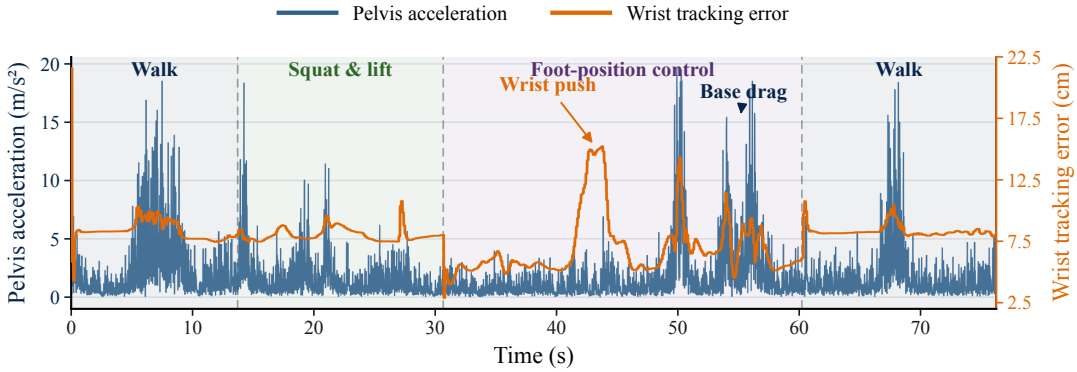


Fig. 10. Walking, squat/lift, and foot-position control: wrist errors and pelvis acceleration briefly deviate and recover under wrist pushes and base drags.

(Fig. 8) and records lower planar error than learned-only estimation (Fig. 9c). Table IV quantifies all three estimators on representative staged-load traces. At held-out vertical loads of 35/40/50 N, hybrid RMSEs are 8.28/9.66/12.35 N.

E. Hardware Tracking and Perturbation Recovery

Real-world tracking and perturbation recovery: Hardware tests cover walking, squat/lift, and commanded planar foot placement; Fig. 10 reports wrist-tracking error and pelvis acceleration in representative runs. A separate dynamic-reaching evaluation yields 3D wrist-position and planar 2D foot-position RMSEs below 8 cm and 10 cm, respectively. Under the evaluated 30 N wrist push and 50 cm body drag, representative traces show transient errors followed by recovery with bounded pelvis acceleration.

TABLE IV
FORCE-ESTIMATION ERRORS UNDER STAGED LOADS

Estimator	x **axis**	y **axis**	z **axis**
Hybrid	$\mathbf{0.81/0.50}$	$\mathbf{1.97/1.57}$	$\mathbf{0.87/0.67}$
Observer-only	$2.83/2.29$	$5.15/4.43$	$3.02/2.58$
Learned-only	$2.43/2.19$	$2.23/1.92$	$2.88/2.47$

Note: Entries are RMSE/MAE in N for the traces in Fig. 8.

TABLE V
CROSS-PAYLOAD BOX-TRANSPORT SUCCESS RATES

Payload	**Empty**	**0.5 kg**	**1 kg**	**2 kg**
Success rate	100%	100%	90%	90%

Note: 20 hardware trials per payload. One cross-payload Diffusion Policy with **FORGE** is trained on 80 demonstrations per payload condition.

F. Demonstration Collection and Downstream Learning

Diffusion Policy as a validation: We use Diffusion Policy as an upstream visuomotor planner and **FORGE** as the only whole-body joint-action executor. One cross-payload model maps head-mounted RGB and task state to multi-granularity command chunks. Training pools 80 demonstrations from each of four payload conditions (320 total).

We evaluate this pipeline over 20 hardware trials for each payload on the squat-and-lift-box transport task with empty, 0.5 kg, 1 kg, and 2 kg boxes. A trial succeeds only when the robot completes the prescribed transport sequence without task termination or human intervention. As shown in Table V, the complete pipeline succeeds in 20/20, 20/20, 18/20, and 18/20 trials, corresponding to success rates of 100%, 100%, 90%, and 90%, respectively. Assuming independent trials, the 95% Wilson intervals are 83.9–100% for 20/20 successes and 69.9–97.2% for 18/20 successes. The same Diffusion Policy–**FORGE** pipeline completes the transport task across the tested 0–2 kg payload range.

V. CONCLUSION

FORGE combines multi-granularity commands, sensorless force estimation, and asymmetric residual whole-body control. Hardware experiments demonstrate task-space tracking, contact-aware manipulation, and execution of Diffusion Policy commands across tested payloads.

REFERENCES

[1] T. He, J. Gao, W. Xiao, Y. Zhang, Z. Wang, J. Wang, Z. Luo, G. He, N. Sobanbabu, C. Pan, Z. Yi, G. Qu, K. Kitani, J. Hodgins, L. J. Fan, Y. Zhu, C. Liu, and G. Shi, "Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills," *arXiv preprint arXiv:2502.01143*, 2025.

[2] Q. Ben, B. Xu, K. Li, F. Jia, W. Zhang, J. Wang, J. Wang, D. Lin, and J. Pang, "Gallant: Voxel grid-based humanoid locomotion and local-navigation across 3d constrained terrains," *arXiv preprint arXiv:2511.14625*, 2025.

[3] B. Xu, H. Weng, Q. Lu, Y. Gao, and H. Xu, "Facet: Force-adaptive control via impedance reference tracking for legged robots," 2025, arXiv:2505.06883.

[4] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," in *Conference on Robot Learning (CoRL)*, 2024.

[5] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi, "Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning," 2024.

[6] Z. Luo, Y. Yuan, T. Wang, C. Li, S. Chen, F. Castañeda, Z.-A. Cao, J. Li, D. Minor, Q. Ben, X. Da, R. Ding, C. Hogg, L. Song, E. Lim, E. Jeong, T. He, H. Xue, W. Xiao, Z. Wang, S. Yuen, J. Kautz, Y. Chang, U. Iqbal, L. Fan, and Y. Zhu, "Sonic: Supersizing motion tracking for natural humanoid whole-body control," *arXiv preprint arXiv:2511.07820*, 2025.

[7] Q. Liao, T. E. Truong, X. Huang, Y. Gao, G. Tevet, K. Sreenath, and C. K. Liu, "Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion," 2025, arXiv:2508.08241.

[8] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang, "HOMIE: humanoid loco-manipulation with isomorphic exoskeleton cockpit," *CoRR*, vol. abs/2502.13013, 2025.

[9] Y. Li, Y. Lin, J. Cui, T. Liu, W. Liang, Y. Zhu, and S. Huang, "CLONE: closed-loop whole-body humanoid teleoperation for long-horizon tasks," *CoRR*, vol. abs/2506.08931, 2025.

[10] P. Zhi, P. Li, J. Yin, B. Jia, and S. Huang, "Learning unified force and position control for legged loco-manipulation," *arXiv preprint arXiv:2505.20829*, 2025.

[11] Q. Lu, Y. Feng, B. Shi, M. Piseno, Z. Bao, and C. K. Liu, "Gentlehumanoid: Learning upper-body compliance for contact-rich human and object interaction," 2025, arXiv:2511.04679.

[12] H. Wang, W. Zhang, R. Yu, T. Huang, J. Ren, F. Jia, Z. Wang, X. Niu, X. Chen, J. Chen, Q. Chen, J. Wang, and J. Pang, "Physhsi: Towards a real-world generalizable and natural humanoid-scene interaction system," 2025, arXiv:2510.11072.

[13] C. Tessler, Y. Guo, O. Nabati, G. Chechik, and X. B. Peng, "Maskedmimic: Unified physics-based character control through masked motion inpainting," *ACM Trans. Graph.*, vol. 43, no. 6, pp. 209:1–209:21, 2024.

[14] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[15] T. He, Z. Luo, W. Xiao, C. Zhang, K. Kitani, C. Liu, and G. Shi, "Learning human-to-humanoid real-time whole-body teleoperation," *arXiv preprint arXiv:2403.04436*, 2024.

[16] Y. Ze, S. Zhao, W. Wang, A. Kanazawa, R. Duan, P. Abbeel, G. Shi, J. Wu, and C. K. Liu, "Twist2: Scalable, portable, and holistic humanoid data collection system," *arXiv preprint arXiv:2511.02832*, 2025.

[17] Y. Li, Y. Zhang, W. Xiao, C. Pan, H. Weng, G. He, T. He, and G. Shi, "Hold my beer: Learning gentle humanoid locomotion and end-effector stabilization control," 2025. [Online]. Available: https://arxiv.org/abs/2505.24198

[18] C. Tessler, Y. Jiang, E. Coumans, Z. Luo, G. Chechik, and X. B. Peng, "Maskedmanipulator: Versatile whole-body manipulation," 2025, arXiv:2505.19086.

[19] N. Rudin, D. Hoeller, P. Reist, and M. Hutter, "Learning to walk in minutes using massively parallel deep reinforcement learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 91–100.

[20] K. Li, P. Li, T. Liu, Y. Li, and S. Huang, "Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, pp. 6991–7003.

[21] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, "3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[22] Y. Ze, Z. Chen, W. Wang, T. Chen, X. He, Y. Yuan, X. B. Peng, and J. Wu, "Generalizable humanoid manipulation with 3d diffusion policies," 2025.

[23] Y. Zhang, Y. Yuan, P. Gurunath, I. Gupta, S. Omidshafiei, A.-a. Aghamohammadi, M. Vazquez-Chanlatte, L. Pedersen, T. He, and G. Shi, "Falcon: Learning force-adaptive humanoid loco-manipulation," *arXiv preprint arXiv:2505.06776*, 2025.

[24] S. Chen, Z. ang Cao, Z. Luo, F. Castañeda, C. Li, T. Wang, Y. Yuan, L. J. Fan, C. K. Liu, and Y. Zhu, "Chip: Adaptive compliance for humanoid control through hindsight perturbation," 2025, arXiv:2512.14689.

[25] A. De Luca and R. Mattone, "Sensorless robot collision detection and hybrid force/motion control," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2005, pp. 999–1004.

[26] R. Featherstone, *Rigid Body Dynamics Algorithms*. Boston, MA: Springer, 2008.