ExecHOI: Planning-grade 3D Simulation of Long-Horizon Human Object Interaction Tasks for Proactive Musculoskeletal Disorder Risk Forecast

Kangrui Ren[1] **and Gaang Lee**[1]

[1]University of Alberta, Canada

kangrui1@ualberta.ca, Gaang@ualberta.ca

Abstract -

Work-related Musculoskeletal Disorders (WMSDs) are a leading cause of worker attrition in construction, yet countermeasures remain largely reactive, relying on post-hoc observations. Traditional WMSD simulation tools based on pre-defined motion templates are suitable for static environments but remain limited in handling selected sources of condition variability in representative construction-style tasks. To proactively assess WMSD risks and optimize work conditions, we need technology that predicts environment-adaptive motions by simulating how a worker's body likely responds to specific site constraints. While advancements in physics-informed AI have enabled 3D simulations of human-environment interactions, achieving prolonged Human-Object Interaction (HOI), which is the nature of most construction tasks, at a planning grade remains a critical hurdle. To address this gap, the paper presents Executable HOI (ExecHOI), a hierarchical draft-to-execution framework that couples task-conditioned kinematic drafting with an unrolled physics-execution loop. Trajectories are refined via smooth surrogate penalties for contact and geometry, improving executability while maintaining task intent. An endogenous four-network control stack (PCRNN, BSNN, LagBioPN, LiMN) schedules phase-dependent objectives, injects an inverse-dynamics biomechanical prior, and modulates impedance through continuous-time memory to improve stability across contact transitions (e.g., grasp/release and foot strike/off). Evaluated in a rigorous physics-driven virtual environment, ExecHOI achieves 86% violation-free task completion while reducing penetration, contact slip, and mechanical work relative to kinematic replay and PD tracking. Robustness is validated under systematic mass and friction perturbations. Overall, ExecHOI supports stable long-horizon HOI execution with fewer contact artifacts and dynamic inconsistencies, enabling more reliable simulation-based what-if comparisons of representative construction-style task variants for proactive WMSD risk forecasting.

Keywords -

Work-related Musculoskeletal Disorders; Human-Object Interaction; Proactive Risk Forecasting; Virtual Planning

1 Introduction

Work-related musculoskeletal disorders (WMSDs) remain a persistent challenge in the construction industry, with substantial consequences for worker safety and productivity. Construction workers are frequently exposed to highly repetitive tasks and physically demanding work under changing site constraints, often requiring non-neutral postures. Surveillance summaries from the Center for Construction Research and Training (CPWR) underscore this burden: more than 33,000 construction workers missed work due to WMSDs in 2021–2022 [1], and these cases resulted in longer absences than other occupational injuries (median 15 vs. 11 days away from work) [2].

Despite the prevalence and cost of WMSDs, prevention in construction remains largely reactive compared with more structured and repeatable work settings. In those settings, task demands and postural requirements are more predictable, and inter-individual variability is often limited, supporting proactive ergonomic design based on anticipated work patterns. In construction, work conditions are difficult to anticipate. Even the same nominal task (e.g., bricklaying or rebar installation) can impose markedly different physical demands across projects or across stages within a single project as surrounding constraints change. As a result, WMSD risk is rarely incorporated into pre-task planning. Instead, ergonomic risks are typically assessed after work has begun or when discomfort is reported, often using observational tools (e.g., RULA [3] and REBA [4]) applied to ongoing tasks [5].

Recent advances in digital-twin-based, physics-enabled simulation suggest a pathway to shift construction WMSD prevention toward earlier, planning-stage intervention [6]. By enabling three-dimensional what-if analyses, these approaches allow planners to explore alternative layouts, methods, and task sequences before field execution. Modern physics engines can execute three-dimensional human motion under gravity, friction, and collisions, producing forces, torques, and contact reactions that support biomechanics-informed risk assessment. In parallel, recent research has made significant progress in simulating long-horizon, multi-stage human–object interaction (HOI)

[7]. These approaches capture task sequences in which workers manipulate tools and materials across successive execution stages, reflecting the inherently multi-stage nature of construction tasks. Collectively, these advances provide important building blocks for construction-grade virtual planning, spanning ergonomic exposure modeling, generative interaction priors, richer HOI datasets, and scene- and language-conditioned motion synthesis. Diffusion-based generative models have improved both the diversity and controllability of synthesized human motion. For example, human-motion diffusion models such as MDM demonstrated classifier-free diffusion with flexible conditioning and contact-aware geometric objectives [8]. HOI diffusion models further capture human–object coupling: CG-HOI jointly generates human–object sequences with contact guidance [9], and related diffusion approaches improve long-horizon interaction consistency by incorporating additional physical and contact-aware constraints. Parallel progress in HOI datasets and data curation has expanded the scope and fidelity of interaction supervision. GeneOH Diffusion enhances sequence quality via denoising and curation [10], while datasets such as BEHAVE and HOI4D provide rich HOI annotations [11, 12]. In addition, scene-aware and language-conditioned motion generation has emerged as a critical direction, particularly for construction environments where geometry strongly constrains feasible work methods. TeSMo demonstrated text-controlled, scene-aware motion generation with goal-reaching constraints [13], and related works bridge text, scenes, and motion via object grounding and affordance representations [14, 15]. Advances in physics-based character control and motion imitation further enable kinematic motions to be executed under contact and dynamic constraints. However, these methods are not typically evaluated for whether long-horizon HOI drafts can be executed with sufficient reliability for planning-oriented ergonomic what-if analysis in representative construction-style tasks. Existing digital human and musculoskeletal approaches support biomechanical assessment but often rely on prescribed motions or structured task templates, whereas motion-synthesis and HOI-generation methods provide richer mechanisms for generating interaction motions but are not always assessed for planning-grade execution reliability under downstream ergonomic use.

Despite the promise of long-horizon HOI simulation, critical gaps remain before these techniques can reliably support planning-stage WMSDs risk assessment in construction. For proactive ergonomic planning, simulations must be physics-executable and maintain planning-grade fidelity beyond visual plausibility. Yet many HOI pipelines remain brittle under physics-based execution, limiting their use for reliable risk forecasting. The technical gaps:

- **(i) Lack of Temporal Continuity.** Frequent phase transitions (e.g., lift–carry–place) can induce boundary discontinuities in pipelines that lack stage/phase consistency. These artifacts destabilize long-horizon trajectories and can mislead time-accumulated evaluations (e.g., ergonomics indicators) by introducing spikes or resets that violate such metrics' continuity assumptions.
- **(ii) Unreliable Interaction Feasibility.** In cluttered, tight-clearance environments, small tracking errors can accumulate into contact failures (slip, penetration, or intermittent contact), causing collisions or constraint violations. Because most pipelines lack explicit scene-consistent contact/clearance enforcement and violation-driven corrective feedback, trajectories may look plausible yet fail under physics-based execution.
- **(iii) Fragile Physics–Energy Plausibility.** Without closed-loop physics regularization (e.g., enforcing energy/impulse consistency), small kinematic/contact deviations can amplify into long-horizon drift as contacts switch or dynamics shift. The result is execution-infeasible trajectories that depend on non-physical impulses, undermining the credibility of downstream risk indicators needed for proactive planning.

To address these gaps, this study develops a draft-to-execution HOI framework that incorporates environment constraints during drafting and performs closed-loop physics–energy refinement during execution. The framework ensures: (i) **Temporal Continuity** across task phase transitions, (ii) robust **Interaction Feasibility** under tight-clearance conditions, and (iii) **Physics and Energy Plausibility** of long-horizon trajectories. These properties support planning-grade simulation and provide a more reliable basis for simulation-based WMSD risk assessment within the evaluated task setting.

2 The Proposed Pipeline: ExecHOI

ExecHOI is a draft-to-execution pipeline that generates long-horizon HOI trajectories conditioned on task intent $\mathcal{I}$ (e.g., action and target) and scene context $\mathcal{S}$ (e.g., geometry, constraints, and contacts). To address the gaps in Sec. 1, ExecHOI adopts three core strategies:

- **Temporal Continuity via Stage-wise Drafting.** Tasks are decomposed into sequential stages, with interaction drafts synthesized using transition regularization. Cross-stage continuity reduces boundary artifacts and stabilizes long-horizon exposure metrics.
- **Interaction Feasibility via Constraints Alignment.** HOI is represented by relative geometry between each active contact frame and the object. By enforcing event-aligned contact timing, relative-pose consistency, and clearance constraints, the framework suppresses slip and

intermittent contact and avoids collisions, preserving load-transfer fidelity under geometric variation.

- **Physics–Energy Plausibility via Loop Refinement.** Execution is refined within an unrolled simulation loop using multi-objective optimization. Structure-informed constraints [16] and neural correctors jointly promote physical compliance and energy consistency, mitigating drift and impulses to enable planning-grade simulation.

ExecHOI realizes these strategies through two coupled modules: **Module A** (Task-Conditioned Kinematic Drafting) and **Module B** (Physics-Conditioned Execution).

- **Module A:** Synthesizes a kinematic draft $\hat{\mathcal{D}}$, including human/object trajectories and a contact schedule, to preserve task intent and cross-stage continuity, serving as a stage-wise initialization for execution refinement.
- **Module B:** Executes $\hat{\mathcal{D}}$ in a physics engine by optimizing feed-forward torques in an unrolled simulation loop with soft feasibility penalties (dynamics, geometry, energy). A proportional–derivative (PD) tracker and bounded impedance modulation stabilize contact switches and improves robustness to perturbations.

2.1 Module A: Task-Conditioned Kinematic Drafting

Module A maps task intent $\mathcal{I}$ and scene context $\mathcal{S}$ to an interaction draft in the scene frame, where $\mathcal{I}$ encodes the current stage objective and target specification, and $\mathcal{S}$ encodes local scene and contact-relevant geometry.

$$\hat{\mathcal{D}} = \left\{ \hat{q}(t),\ \hat{\xi}(t),\ \mathbf{c}(t) \right\}_{t\in[0,T]}. \tag{1}$$

where $\hat{q}(t)$ and $\hat{\xi}(t)$ denote the humanoid configuration and object pose, and $\mathbf{c}(t) \in \{0,1\}^M$ is a contact schedule over M contact keypoints (indexed by $j = 1, \dots, M$).

Algorithm 1: Stage-wise Kinematic Drafting

Input: Task Intent $\mathcal{I}$, Context $\mathcal{S}$, Horizon $[0,T]$
Output: Draft $\hat{\mathcal{D}}$, Interaction Interface $\{\hat{T}_{\mathrm{rel},j}(t)\}$

```
Decompose 𝓘 → {𝓘_i}_{i=1}^N on 0 = t_0 < ··· < t_N = T
Initialize D̂_0^end with initial draft state at t_0
Notation: D̂_i is the draft segment on [t_{i-1}, t_i]
for i ← 1 to N do
    Generate D̂_i ← 𝒢(𝓘_i, 𝒮, D̂_{i-1}^end) on [t_{i-1}, t_i]
    if i > 1 then Stitch D̂_{i-1}, D̂_i via blending α(·)
    Update terminal state D̂_i^end ← D̂_i(t_i)
end
Concatenate all segments to obtain D̂ on [0, T]
Compute T̂_rel,j(t) ← T_j(q̂(t))^{-1} ξ̂(t) for all j
return D̂, {T̂_rel,j(t)}
```

Long-horizon drafts are generated stage-wise by a conditional generator $\mathcal{G}$ and stitched via segment blending to preserve continuity, including the root motion (Alg. 1).

2.2 Module B: Physics-Conditioned Execution

Module B executes the kinematic draft via the four module control stack: **Physics-Conditioned Residual Neural Network (PCRNN)**, **Budgeted Scheduler Neural Network (BSNN)**, **Lagrangian Biomechanical Prior Network (LagBioPN)**, and **Liquid Memory Network (LiMN)**. These are task-specific functional modules instantiated within ExecHOI, rather than off-the-shelf pretrained models or standalone generic architectures. Within an unrolled simulation loop, Module B optimizes a baseline torque sequence to produce a feasible trajectory that is robust to contact switches and scene perturbations, with task/scene information remaining reflected in the refinement objectives and feasibility penalties. The optimized actuation is applied at the articulated joints, while the carried object remains in the rigid-body simulation so that its weight affects the resulting dynamics and required torques. Gradients are obtained by BPTT through the differentiable parts of the unrolled graph (e.g., control synthesis, residual networks, and surrogate contact/geometry penalties), while discrete contact events are treated as non-differentiable and stop-gradient is applied when needed.

2.2.1 Physics-Conditioned Residual Neural Network

PCRNN targets interaction feasibility in cluttered scenes where contact instabilities can accumulate into collision/penetration over long horizons. It performs correction inside the physics loop via bounded residual torques and impedance modulation, enforcing phase-gated contact integrity and clearance-aware feasibility penalties.

Let $x(t)$ denote the state evolved by a physics engine, including $(q, \dot{q})$, object pose, and contact variables. Here $\tau(t) \in \mathbb{R}^{n_q}$ is the joint torque:

$$x(t + \Delta t) = \mathrm{SimStep}\big(x(t), \tau(t); \mathcal{E}\big). \tag{2}$$

The draft contact schedule $\mathbf{c}(t)$ is smoothed into phase gates $\chi_j(t) \in [0,1]$ around each on/off interval to mitigate boundary discontinuities. The gates weight contact terms to suppress jitter, slip, and contact loss, and are stacked as $\chi(t) = [\chi_1(t), \dots, \chi_M(t)]$. Engine-reported contact variables within $x(t)$ evaluate contact integrity, together with surrogate penalties for collision, friction, and budget constraints. Let $q(t)$ and $\xi(t)$ denote the humanoid configuration and object pose. For each contact keypoint j, the relative transform is $T_{\mathrm{rel},j}(t) = T_j\big(q(t)\big)^{-1}\xi(t)$; deviations from the draft $\hat{T}_{\mathrm{rel},j}(t)$ are penalized.

The horizon $[0,T]$ is discretized into H steps ($t_k = k\Delta t$), and Module B parameterizes torque as a baseline plus a bounded neural residual. Let $\Omega = \{\mathrm{t}, \mathrm{c}, \mathrm{g}, \mathrm{p}, \mathrm{e}, \mathrm{s}\}$ (Table 1); $\chi(t_{k+1})$ gates contact-related losses for phase consistency, while geometry terms promote collision-free clearance.

$$J_{\text{base}} = \sum_{k=0}^{H-1} \sum_{\star \in \Omega} w_\star(t_k)\, \mathcal{L}_\star\big(x_{k+1}, \hat{\mathcal{D}}, \mathcal{S}; \chi(t_{k+1})\big)\, \Delta t. \quad (3)$$

Here $w_\star(t_k) = w_{\text{tot}} \pi_\star(t_k)$ is predicted by BSNN from the shared feature $\tilde{z}_k = [z_k, h_k]$ (Sec. 2.2.2–2.2.4).

$$\begin{aligned} \tau_{\text{ff}}(t) &= \bar{\tau}_{\text{ff}}(t) + \Delta\tau_\theta(t). \\ \Delta\tau_\theta(t) &= \epsilon_\tau \tanh\big(\mathrm{PCRNN}_\theta^{\Delta\tau}(\tilde{z}(t))\big). \quad (4) \\ \alpha_d(t) &= \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\, \sigma\big(\mathrm{PCRNN}_\theta^{\alpha}(\tilde{z}(t))\big). \end{aligned}$$

Here $z(t)$ summarizes the simulator state, phase gates, scene context, and the draft, augmented with LiMN memory to form the shared feature for PCRNN. The residual is bounded by construction ($\|\Delta\tau_\theta(t)\|_\infty \le \epsilon_\tau$), which maintains trajectories under mismatch and contact uncertainty.

Table 1. Objective terms used in Module B.

Term	Description
$\mathcal{L}_{\mathrm{t}}$	Track draft keypoints and object pose.
$\mathcal{L}_{\mathrm{c}}$	Phase-gated consistency; suppress tangential slip.
$\mathcal{L}_{\mathrm{g}}$	Collision/clearance via signed distances.
$\mathcal{L}_{\mathrm{p}}$	Friction feasibility and contact-force regularization.
$\mathcal{L}_{\mathrm{e}}$	Power/work budget penalty.
$\mathcal{L}_{\mathrm{s}}$	Smoothness regularization on torque-rate and jerk.

2.2.2 Budgeted Scheduler Neural Network

To address temporal continuity, BSNN generates a smooth, phase-aware schedule of loss weights across contact transitions. Given the shared trajectory feature with phase gates, BSNN outputs temperature-scaled Softmax weights π_k over Ω, scaled by a fixed budget w_{tot} so that weights are nonnegative and sum to w_{tot} at each step.

To discourage degenerate schedules and promote temporal continuity, a regularizer is added:

$$\mathcal{L}_{\text{sch}} = -\lambda_{\text{ent}} \sum_{k=0}^{H-1} \mathcal{H}(\pi_k) + \lambda_{\text{sm}} \sum_{k=1}^{H-1} \big\|\pi_k - \pi_{k-1}\big\|_2^2. \quad (5)$$

where $\mathcal{H}(\pi) = -\sum_{\star\in\Omega} \pi_\star \log \pi_\star$ and $\lambda_{\text{ent}}, \lambda_{\text{sm}} \ge 0$ weight entropy and smoothness, respectively.

2.2.3 Lagrangian Biomechanical Prior Network

To enforce physics–energy plausibility without modifying the simulator, LagBioPN provides a soft actuation prior. Given an interaction-centric embedding $\tilde{z}(t)$ (including $q(t)$), LagBioPN outputs a lower-triangular factor $L(t)$, a potential $V_\psi(q(t))$, and a gate logit $\tilde{\lambda}(t)$. A positive-definite inertia surrogate is defined as:

$$M(q(t)) = L(t)L(t)^\top + \epsilon_M I, \qquad \epsilon_M > 0. \quad (6)$$

Gravity-like terms follow via $g(q(t)) = \nabla_q V_\psi(q(t))$, and the confidence weight is bounded as $\lambda_{\text{Bio}}(t) = \lambda_{\max}\, \sigma(\tilde{\lambda}(t))$. Let $\ddot{q}(t)$ denote the trajectory acceleration; LagBioPN then forms a biomechanical torque prior.

$$\tau_{\text{bio}}(t) = M\big(q(t)\big)\, \ddot{q}(t) + C\big(q(t), \dot{q}(t)\big)\dot{q}(t) + g\big(q(t)\big). \quad (7)$$

LagBioPN is trained offline via teacher distillation on feasible trajectories for the same draft and scene. Given teacher states $\big(q^{\mathrm{T}}(t), \dot{q}^{\mathrm{T}}(t), \ddot{q}^{\mathrm{T}}(t)\big)$, the prior is instantiated on these states to obtain $\tau_{\text{bio}}^{\mathrm{T}}(t)$, and training matches the teacher torque while aligning a positive-power proxy:

$$\begin{aligned} \hat{p}(\tau, \dot{q}) &\triangleq (\tau^\top \dot{q})_{+,\beta}. \\ \delta_p^{\mathrm{T}}(t) &\triangleq \hat{p}\big(\tau_{\text{bio}}^{\mathrm{T}}(t), \dot{q}^{\mathrm{T}}(t)\big) - \hat{p}\big(\tau^{\mathrm{T}}(t), \dot{q}^{\mathrm{T}}(t)\big). \quad (8) \\ \mathcal{L}_{\text{dist}}(t) &= \big\|\tau_{\text{bio}}^{\mathrm{T}}(t) - \tau^{\mathrm{T}}(t)\big\|_2^2 + \lambda_p \big(\delta_p^{\mathrm{T}}(t)\big)^2. \end{aligned}$$

where $\tau^{\mathrm{T}}(t)$ is the teacher torque and $\tau_{\text{bio}}^{\mathrm{T}}(t)$ is the prior evaluated on teacher states. Here $(x)_{+,\beta}$ is a smooth approximation of $\max(x, 0)$ via a softplus: $(x)_{+,\beta} \triangleq \beta^{-1} \log\big(1 + \exp(\beta x)\big)$, where $\beta > 0$ controls sharpness.

During execution, the objective adds a confidence-gated prior loss and a gate-usage regularizer. These terms encourage alignment when confident and prevent gate collapse, with $0 < \lambda_{\min} < \lambda_{\max}$.

$$\begin{aligned} \delta_p(t) &\triangleq \hat{p}\big(\tau(t), \dot{q}(t)\big) - \hat{p}\big(\tau_{\text{bio}}(t), \dot{q}(t)\big). \\ \ell_{\text{bio}}(t) &\triangleq \big\|\tau(t) - \tau_{\text{bio}}(t)\big\|_2^2 + \rho_p\, \delta_p(t)^2. \\ \mathcal{L}_{\text{bio}} &= \sum_{k=0}^{H-1} \lambda_{\text{Bio}}(t_k)\, \ell_{\text{bio}}(t_k)\, \Delta t. \quad (9) \\ \mathcal{L}_\lambda &= \eta_\lambda \sum_{k=0}^{H-1} \big(\lambda_{\min} - \lambda_{\text{Bio}}(t_k)\big)_{+,\beta}^2 \Delta t. \end{aligned}$$

2.2.4 Liquid Memory Network

To capture short-term temporal context (e.g., post-switch transients and accumulated effort), LiMN maintains a latent memory $h(t) \in \mathbb{R}^{d_h}$ that is updated along the trajectory. The shared feature is formed by concatenation $\tilde{z}(t) = [z(t), h(t)]$ and fed to the fabric modules. The memory is initialized at trajectory start and updated with a lightweight latent dynamics model using a forward-Euler step: $h_{k+1} = h_k + \Delta t\, f_\rho(h_k, z_k)$.

Module B executes $\hat{\mathcal{D}}$ in an unrolled simulation loop and optimizes baseline torques and network parameters by backpropagating through differentiable paths (e.g., control synthesis, residual networks, and surrogate contact penalties). Alg. 2 summarizes the procedure.

Algorithm 2: Physics-Conditioned Execution

Input: Draft $\hat{\mathcal{D}}$ (incl. $\mathbf{c}(t)$), context $\mathcal{S}$, horizon H, engine $\mathcal{E}$, frozen LagBioPN prior ψ
Output: Torques $\bar{\tau}_{\mathrm{ff}}$ and trajectory $\{x_k\}_{k=0}^{H}$

Initialize once: $\bar{\tau}_{\mathrm{ff}}$, parameters (θ, ϕ, ρ)
for $r \leftarrow 1$ **to** R **do**
    Reset simulator state x_0 and memory h_0
    $J_{\mathrm{base}}, \mathcal{L}_{\mathrm{ent}}, \mathcal{L}_{\mathrm{sm}}, \mathcal{L}_{\mathrm{bio}}, \mathcal{L}_{\lambda} \leftarrow 0$
    for $k \leftarrow 0$ **to** $H-1$ **do**
        $\chi_{k+1} \leftarrow \chi(t_{k+1}; \mathbf{c}(t))$
        $\tilde{z}_k \leftarrow [z_k, h_k]$
        Predict π_k and $w_\star(t_k)$ via BSNN
        $\tau_k \leftarrow \bar{\tau}_{\mathrm{ff}}$ & PCRNN & PD tracking (Eq. 4)
        Step $x_{k+1} = \mathrm{SimStep}(x_k, \tau_k; \mathcal{E})$ (Eq. 2)
        Accumulate J_{base} (Eq. 3)
        Accumulate $\mathcal{L}_{\mathrm{bio}}, \mathcal{L}_{\lambda}$ (frozen ψ) (Eq. 9)
        $\mathcal{L}_{\mathrm{ent}} \mathrel{+}= \mathcal{H}(\pi_k)$
        if $k > 0$ **then** $\mathcal{L}_{\mathrm{sm}} \mathrel{+}= \|\pi_k - \pi_{k-1}\|_2^2$
        Update memory h_{k+1}
    end
    $\mathcal{L}_{\mathrm{sch}} \leftarrow -\lambda_{\mathrm{ent}}\mathcal{L}_{\mathrm{ent}} + \lambda_{\mathrm{sm}}\mathcal{L}_{\mathrm{sm}}$ (Eq. 5)
    Set $J \leftarrow J_{\mathrm{base}} + \mathcal{L}_{\mathrm{sch}} + \mathcal{L}_{\mathrm{bio}} + \mathcal{L}_{\lambda}$
    Update by BPTT through differentiable paths:
    $(\bar{\tau}_{\mathrm{ff}}, \theta, \phi, \rho) \leftarrow (\bar{\tau}_{\mathrm{ff}}, \theta, \phi, \rho) - \eta \nabla J$
end
return $\bar{\tau}_{\mathrm{ff}}$ and final trajectory x_k

3 Experiments

ExecHOI was evaluated on a representative long-horizon lift–carry–place object-carrying task in construction-like scenes. The evaluation examined (i) the contact-aware feasibility of Module B executions, (ii) the contributions of individual components (PCRNN/BSNN/LagBioPN/LiMN), and (iii) robustness to controlled geometric and physical perturbations.

3.1 Experimental Setup

Experiments were conducted on a desktop equipped with an NVIDIA RTX 5090 (32 GB VRAM) and 64 GB RAM. Isaac Lab (Omniverse Isaac Sim / PhysX) was run at 30 Hz ($\Delta t = \frac{1}{30}$ s, horizon $H = 300$ steps, duration $T = 10$ s) with 2 physics substeps per control step. PhysX contact parameters were fixed at nominal Coulomb friction $\mu = 0.9$, and μ was perturbed only in robustness tests. The reference trajectory was represented in SMPL-X and executed via an articulated rigid-body proxy with a fixed retargeting map and consistent root-frame alignment.

The humanoid grasped and lifted a rigid object, transported it, and placed it on an elevated target surface (Fig. 1). Each trial instantiated an object from a parameterized rigid-body family, start and goal regions, and a nominal transport speed; object mass was varied via density scaling while maintaining consistent inertia. Contact keypoints $j = 1, \ldots, M$ included the feet and the hand keypoints specified by the draft contact schedule.

Figure 1. **Object carrying task overview.** ExecHOI trajectory for placing boxes onto an elevated surface: (a) grasp/lift, (b) transport, (c) elevated placement.

A trial was deemed successful if the object was placed in the goal region and remained within tolerance for $K = 20$ frames ($\|\mathbf{p} - \mathbf{p}^\star\| \leq 0.05$, $\angle(R^\top R^\star) \leq 10°$), while satisfying the feasibility constraints in Sec. 3.2. Failures included object drop, humanoid fall, constraint violation, and time-out; trials were terminated upon failure. Metrics were aggregated over $N_{\mathrm{trial}} = 240$ trials and 5 seeds.

3.2 Baselines and Metrics

The following variants were evaluated under the same scenes, trial distribution, and feasibility criteria:

- **Draft-only (A):** Kinematic replay of $\hat{\mathcal{D}}$ was performed by setting the humanoid and object states at each step.
- **PD-only:** PD execution was performed with $\tau_{\mathrm{ff}} \equiv 0$; the object trajectory arose from the realized grasp contacts.
- **w/o PCRNN:** $\bar{\tau}_{\mathrm{ff}}$ was optimized and executed with PD feedback, with PCRNN disabled ($\Delta\tau_\theta \equiv 0$, $\alpha_d(t) \equiv 1$).
- **w/o BSNN:** The time-varying schedule $w_\star(t)$ was replaced with tuned constants $w_\star$.
- **w/o LagBioPN:** $\mathcal{L}_{\mathrm{bio}}$ and $\mathcal{L}_{\lambda}$ were removed.
- **w/o LiMN:** The memory was disabled by setting $h_k \equiv 0$.

For each trial, the feed-forward torque sequence was updated for $R = 40$ gradient steps by back-propagating through a full-horizon trajectory of $H = 300$ frames. When enabled, the trainable execution modules in Module B (PCRNN, BSNN, LagBioPN, LiMN) were updated jointly with the torque sequence; in ablations where a component was removed, its corresponding outputs and parameters were disabled and excluded from optimization. Optimization used AdamW with learning rate 3×10^{-4} and ℓ_2 gradient clipping ($|\nabla|_2 \leq 1.0$). To ensure fair comparison across methods, this execution-stage optimization uses a fixed update budget rather than an adaptive stopping criterion. Stability is monitored through rollout feasibility outcomes and the success/failure conditions defined in the experimental setup. Mixed precision was enabled to improve throughput. Evaluation metrics are defined in Table 2.

Table 2. Evaluation metrics.

Metric	Definition
Success Rate	Fraction of violation-free goal-reaching trials
Max Penetration	Maximum penetration depth over the trajectory
Contact Slip	Peak tangential slip speed on scheduled contacts
Mechanical Work	Time-integral of positive joint power

4 Results and Discussion

4.1 Overall performance and ablations

Table 3 summarized overall performance on the object-carrying trial set. Relative to kinematic replay and PD tracking, the full system achieved a higher success rate while reducing penetration, contact slip, and mechanical work. Removing individual components degraded one or more metrics, supporting the contribution of each module.

Table 3. Object-carrying results.

Method	**SR** ↑	**Pen** ↓ (cm)	**Slip** ↓ (cm/s)	**Work** ↓ (kJ)
Baselines				
Draft-only (A)	0.46 ± 0.09	1.22 ± 0.38	4.18 ± 1.05	22 ± 5.2
PD-only	0.61 ± 0.08	0.86 ± 0.27	2.98 ± 0.82	18.5 ± 4.1
Ablations				
ExecHOI (full)	$\mathbf{0.86 \pm 0.05}$	$\mathbf{0.36 \pm 0.13}$	$\mathbf{1.63 \pm 0.46}$	$\mathbf{12 \pm 2.4}$
w/o PCRNN	0.69 ± 0.07	0.66 ± 0.21	2.32 ± 0.69	15.8 ± 3.3
w/o BSNN	0.73 ± 0.07	0.74 ± 0.24	3.18 ± 0.78	13.5 ± 2.6
w/o LagBioPN	0.80 ± 0.06	0.46 ± 0.16	2.10 ± 0.57	20 ± 3.9
w/o LiMN	0.77 ± 0.07	0.61 ± 0.20	2.82 ± 0.71	14.5 ± 2.7

Note: ↑ higher is better; ↓ lower is better.

To localize brittle phases, event-aligned analysis was performed around key transitions. Signals were time-aligned to the detected event, with the vertical dashed line marking $\Delta t = 0$, and plotted as mean ± std over rollouts.

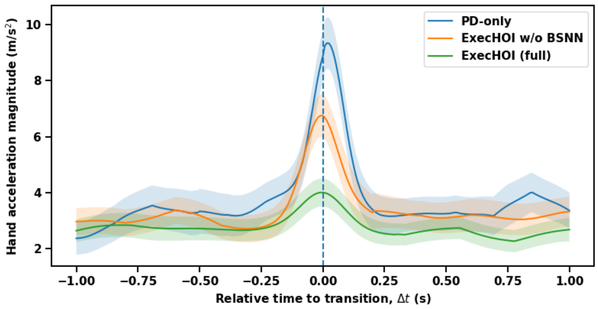

Figure 2. **Hand acceleration at phase transition**.

Fig. 2 plots the magnitude of hand-acceleration over time relative to the transition. Removing BSNN increases both the peak and the variance near $\Delta t = 0$, suggesting less smooth phase transitions and more abrupt corrections.

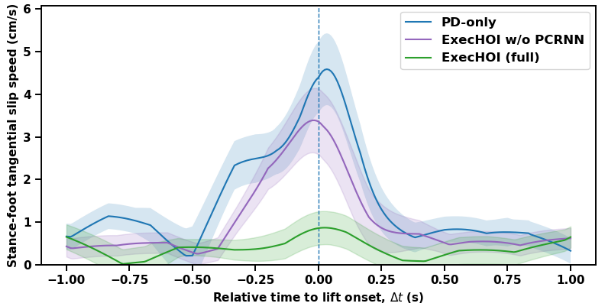

Figure 3. **Stance-foot tangential slip at lift onset**.

Fig. 3 shows the stance-foot tangential slip speed aligned to lift onset. Disabling PCRNN produces larger slip bursts around $\Delta t = 0$, indicating a weaker contact-consistent correction during contact switching.

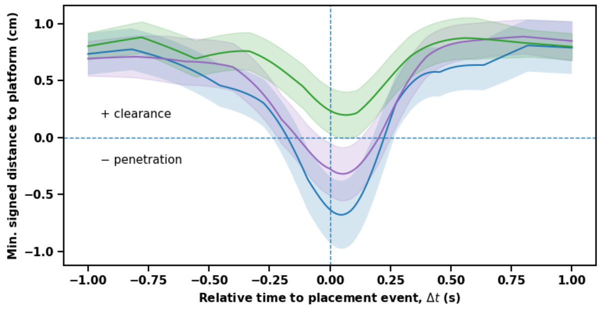

Figure 4. **Clearance/penetration at placement**.

Fig. 4 reports the minimum signed distance to the platform around placement (positive: clearance; negative: penetration). The full system maintains a larger positive margin and recovers faster after the event, whereas PD-only and ablation variants approach or cross zero, consistent with the higher penetration reported in Table 3.

4.2 Robustness under controlled perturbations

To assess robustness, the kinematic draft $\hat{\mathcal{D}}$ (including the contact schedule) was held fixed for each trial, and Module B was re-executed under controlled perturbations to scene and physics parameters (e.g., object mass/density, friction μ, and clearance offsets). Robustness was summarized by the feasibility rate, defined as the fraction of perturbed executions that satisfied the same success and feasibility criteria in Sec. 3.2. A 7×7 grid over mass and friction was evaluated with 10 randomized replays per grid cell, and the resulting feasibility map is shown in Fig. 5.

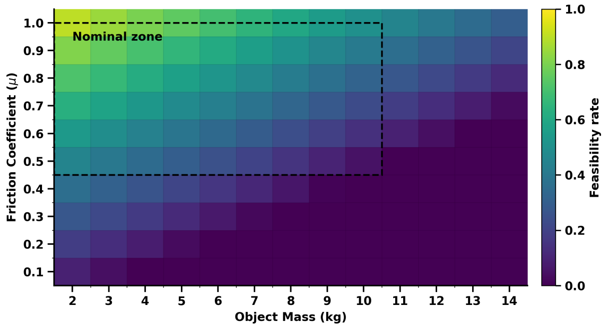

Figure 5. **Robustness heatmap.**

4.3 Discussion and Limitation

The ablation results indicate that ExecHOI helps address the three reliability gaps required for planning-grade HOI simulation—temporal continuity, interaction feasibility, and physics–energy plausibility—through complementary roles of the four-module execution stack.

First, removing **BSNN** disrupts phase-aware objective scheduling, manifesting as stronger event-local acceleration transients at $\Delta t = 0$ (Fig. 2) and a clear drop in overall success (SR 0.86 → 0.73; Table 3), consistent with less smooth stage transitions. Next, disabling **PCRNN** weakens contact-consistent correction during contact switching: slip bursts become more pronounced around lift onset (Fig. 3), and aggregate contact instability increases (Slip 1.63 → 2.32). Moreover, removing **LagBioPN** substantially inflates the positive-work proxy (Work 12 → 20 kJ), suggesting that improved execution performance is not achieved by simply injecting stronger actuation, but rather by enforcing energy-plausible dynamics through the biomechanical prior. Finally, disabling **LiMN** degrades long-horizon stability (SR 0.86 → 0.77), consistent with memory suppressing post-switch transients and preventing error accumulation across stages.

Beyond ablations, the distance trace around placement summarizes how ExecHOI recovers clearance margin and avoids penetration (Fig. 4), while the mass–friction feasibility map characterizes an operating envelope under perturbations (Fig. 5), enabling condition-aware planning rather than reliance on a single nominal setting.

It is important to note that this study emphasizes execution reliability after physics-based execution for ergonomic planning, rather than general superiority over the broader HOI or motion-generation literature in terms of motion realism. While the tests support planning-grade simulation and robustness under perturbations, several limitations remain in how broadly these conclusions generalize:

- **Task scope and interaction diversity.** The current evaluation centers on a single long-horizon template (lift–carry–place) with rigid objects. Although representative, practical planning deployments involve broader task families (e.g., pushing/pulling, tool use, handoffs, and constrained manipulation) and richer contact modes. Follow-up studies will extend the benchmark to multiple construction-relevant tasks with varying stage structures and interaction patterns.
- **Coverage of off-nominal conditions.** Robustness was primarily characterized over mass–friction grids and clearance offsets, which capture two key drivers of feasibility but do not span the full space of site variability. Future tests will broaden the perturbation set to include geometry and layout changes (e.g., corridor topology, obstacle shape and placement), object properties beyond mass (e.g., inertia and size), and additional surface and contact conditions (e.g., compliance or heterogeneous friction), better approximating field-level uncertainty.
- **Execution and modeling assumptions.** Results are obtained using an articulated rigid-body proxy, a fixed retargeting map, and fixed PhysX contact settings. These assumptions simplify deployment but may affect transfer when subject morphology, contact parameters, or retargeting quality varies. In addition, the current evaluation does not yet include validation against experimentally measured human motion or subject-specific biomechanical ground truth, but instead focuses on execution-oriented simulation criteria such as feasibility, contact consistency, stability, and mechanical work proxies. Future work will examine sensitivity to these factors and incorporate validation against measured motion and biomechanical data.

Finally, the differentiable simulation loop is currently aimed at offline planning support rather than real-time deployment. Its computational cost is driven mainly by the unrolled simulation horizon, repeated execution-stage updates, and contact-rich interaction evaluation. Improving throughput will therefore be important for interactive, human-in-the-loop planning and larger-scale deployment.

5 Conclusion

By introducing ExecHOI, this work argues that planning-grade HOI simulation must be assessed by execution reliability under realistic condition shifts rather than visual plausibility alone. Across long-horizon lift-carry-place rollouts, ExecHOI maintains temporal continuity and biomechanical stability even when perturbing key site factors such as mass-friction settings and spatial clearance. These results suggest that planning-grade HOI simulation can move beyond visual plausibility toward more reliable what-if analysis within representative construction-style planning scenarios. This may help future planning tools screen candidate task variants using feasibility and robustness envelopes while diagnosing operational fragility through event-aligned slip and effort proxies to guide targeted plan revisions. Ultimately, ExecHOI can contribute to transforming WMSDs prevention from a reactive post-hoc observation into a proactive, data-driven task plan optimization process. By enabling reliable what-if comparisons of construction plans before field deployment, this framework provides a robust computational foundation for proactive safety management and task design in dynamic work environments.

Acknowledgement

This work was supported by the New Frontiers in Research Fund (NFRF), grant number NFRFE-2023-00174.

References

[1] CPWR – The Center for Construction Research and Training. Musculoskeletal disorders (msds). The Construction Chart Book, 7th Edition (Interactive Dashboard), 2025. Accessed: 2025-08-04.

[2] CPWR – The Center for Construction Research and Training. Data bulletin – june 2025, June 2025. Accessed: 2025-08-04.

[3] Lynn McAtamney and E Nigel Corlett. Rula: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics*, 24(2):91–99, 1993.

[4] Sue Hignett and Lynn McAtamney. Rapid entire body assessment (reba). *Applied ergonomics*, 31(2): 201–205, 2000.

[5] Kangrui Ren, Eren Shahrokhi, Ali Golabchi, and Gaang Lee. Feasibility of an integrated multi-model approach for dynamic musculoskeletal disorder risk assessment. In *Proceedings of the 14th Creative Construction Conference (CCC 2025)*, pages 495–506, 2025.

[6] Vigneshkumar Chellappa and Jitesh Singh Chauhan. Digital twin approach for the ergonomic evaluation of vertical formwork operations in construction. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 40, pages 302–308. IAARC Publications, 2023.

[7] Sören Pirk, Karol Hausman, Alexander Toshev, and Mohi Khansari. Modeling long-horizon tasks as sequential interaction landscapes. *arXiv preprint arXiv:2006.04843*, 2020.

[8] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H Bermano. Human motion diffusion model. *arXiv preprint arXiv:2209.14916*, 2022.

[9] Christian Diller and Angela Dai. Cg-hoi: Contact-guided 3d human-object interaction generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19888–19901, 2024.

[10] Xueyi Liu and Li Yi. Geneoh diffusion: Towards generalizable hand-object interaction denoising via denoising diffusion. *arXiv preprint arXiv:2402.14810*, 2024.

[11] Bharat Lal Bhatnagar, Xianghui Xie, Ilya A Petrov, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. Behave: Dataset and method for tracking human object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15935–15946, 2022.

[12] Yunze Liu, Yun Liu, Che Jiang, Kangbo Lyu, Weikang Wan, Hao Shen, Boqiang Liang, Zhoujie Fu, He Wang, and Li Yi. Hoi4d: A 4d egocentric dataset for category-level human-object interaction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21013–21022, 2022.

[13] Hongwei Yi, Justus Thies, Michael J Black, Xue Bin Peng, and Davis Rempe. Generating human interaction motions in scenes with text control. In *European Conference on Computer Vision*, pages 246–263. Springer, 2024.

[14] Zan Wang, Yixin Chen, Baoxiong Jia, Puhao Li, Jinlu Zhang, Jingze Zhang, Tengyu Liu, Yixin Zhu, Wei Liang, and Siyuan Huang. Move as you say interact as you can: Language-guided human motion generation with scene affordance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 433–444, 2024.

[15] Zhi Cen, Huaijin Pi, Sida Peng, Zehong Shen, Minghui Yang, Shuai Zhu, Hujun Bao, and Xiaowei Zhou. Generating human motion in 3d scenes from text descriptions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1855–1866, 2024.

[16] Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*, 2020.