# HOI-Dyn: Learning Interaction Dynamics for Human-Object Motion Diffusion



Lin Wu[1], Zhixiang Chen[2] and Jianglin Lan[1*]

[1] University of Glasgow
[2] University of Sheffield

AIR Lab

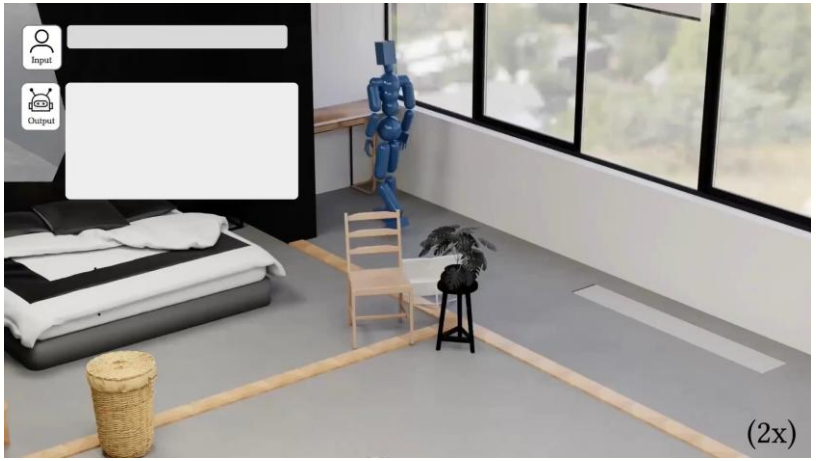
Project Website

*corresponding author

# Background

- Synthesizing complex and realistic 3D human-object interactions (HOIs) is essential for progress in computer animation, and robotics, yet remains a significant challenge.

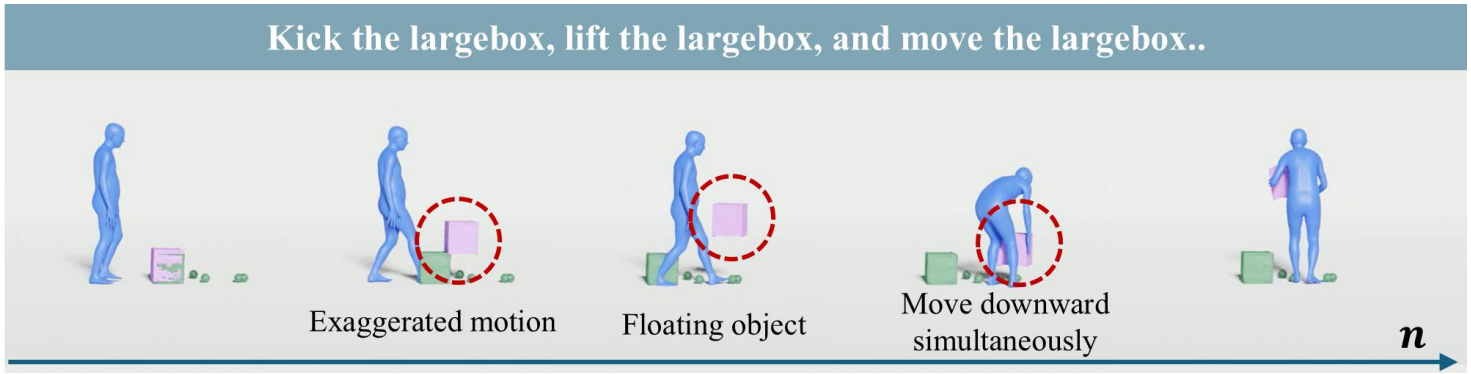

*Jiaman Li et al. CHOIS*
*ECCV 2024*




*Zhen Wu et al. HOI-HLI*
*ICCV 2025*

# Challenges

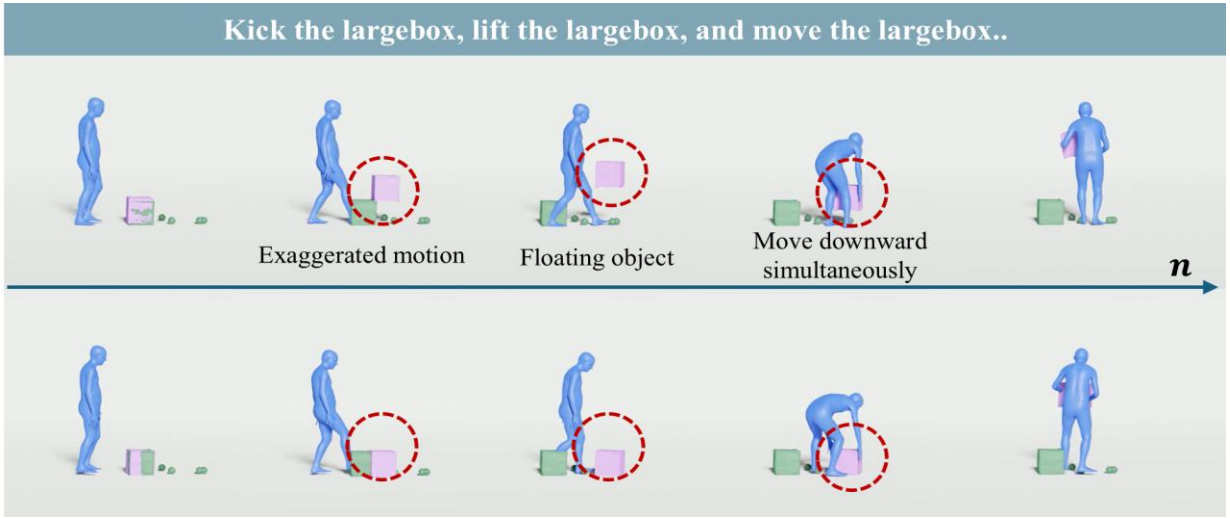
- Directly applying generative models to HOI often leads to **decoupled motions**, resulting in physically unrealistic and causally inconsistent behaviours.
- In contrast, realistic HOI synthesis requires **capturing interaction dynamics** — stable contact, forces, and action–response relationships.



University of Glasgow

# Objective

Our goal is to synthesize **synchronized HOIs** while maintaining **internal causal consistency**, by leveraging controllable signals such as text.



# New perspective: Driver-Responder System

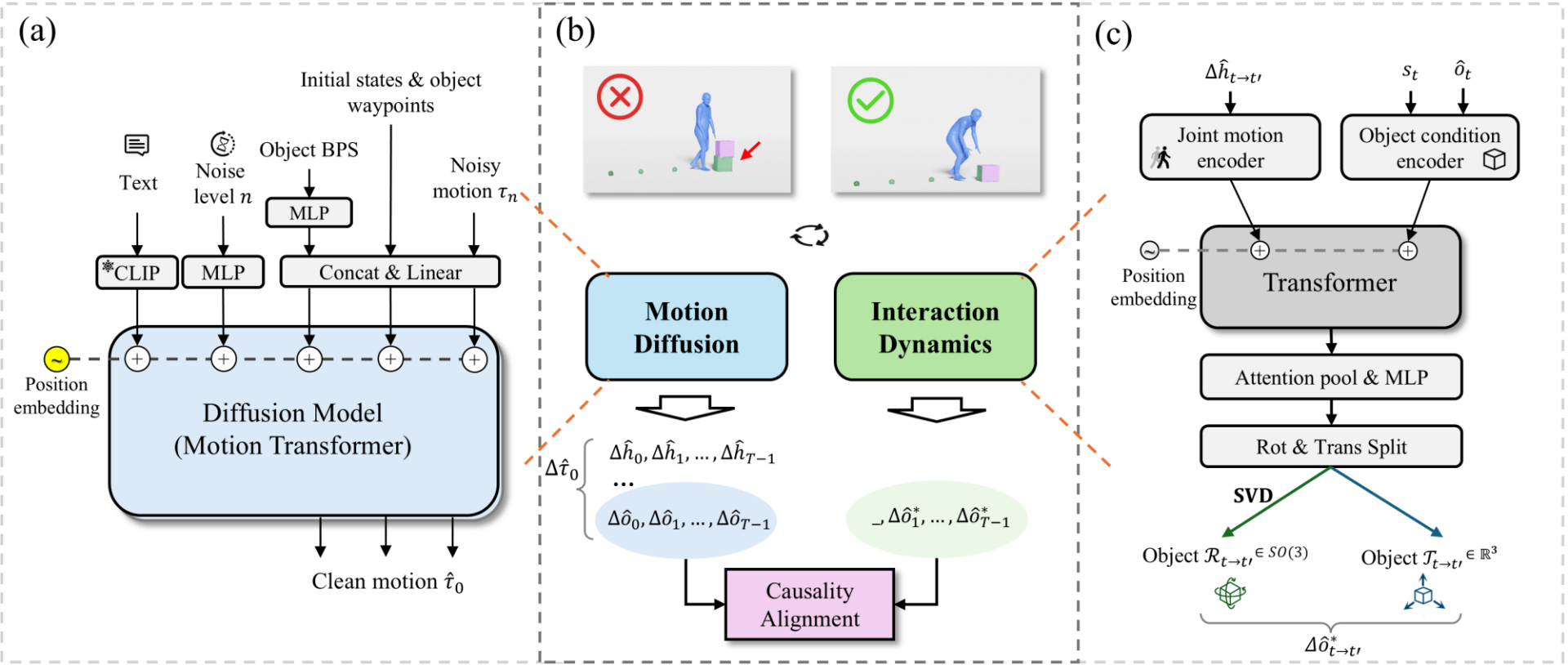
Motivated by classical synchronization control formulation, we frame HOI generation as a **driver-responder system**, where human actions serve as the driver and object respond accordingly.

$$\textbf{Driver (Human)}: \begin{cases} h^{(t+1)} = h^{(t)} + \Delta t \cdot F_h(h^{(t)}) \\ y_h^{(t)} = g_h(h^{(t)}) \end{cases},$$

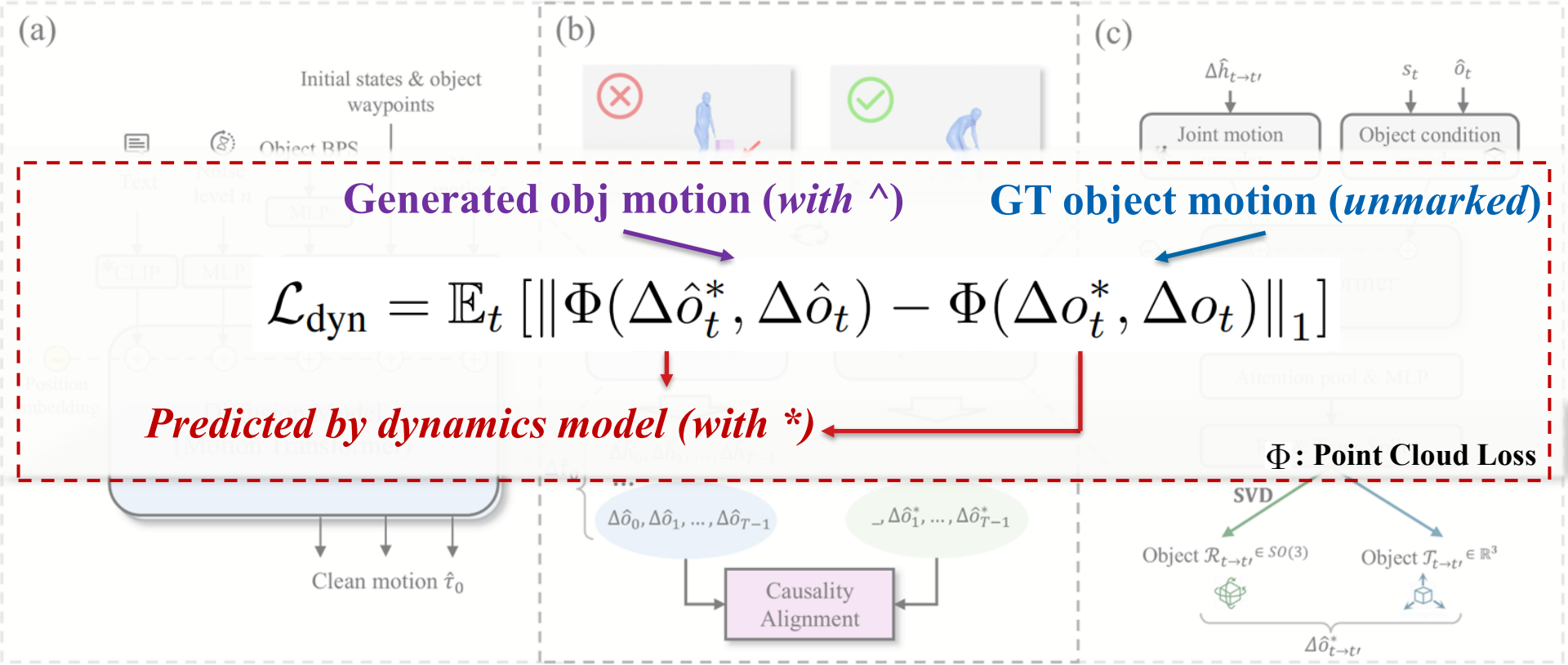
$$\textbf{Responder (Object)}: \begin{cases} o^{(t+1)} = o^{(t)} + \Delta t \cdot F_o(o^{(t)}, s^{(t)}, u^{(t)}) \\ y_o^{(t)} = g_o(o^{(t)}) \end{cases},$$

💡 ***better model the causal dependencies between human actions and object responses in a dynamic and physically consistent manner.***

# Proposed Framework: HOI-Dyn



# Interaction Dynamics Guided Generation




Generated obj motion (*with ^*)

GT object motion (*unmarked*)

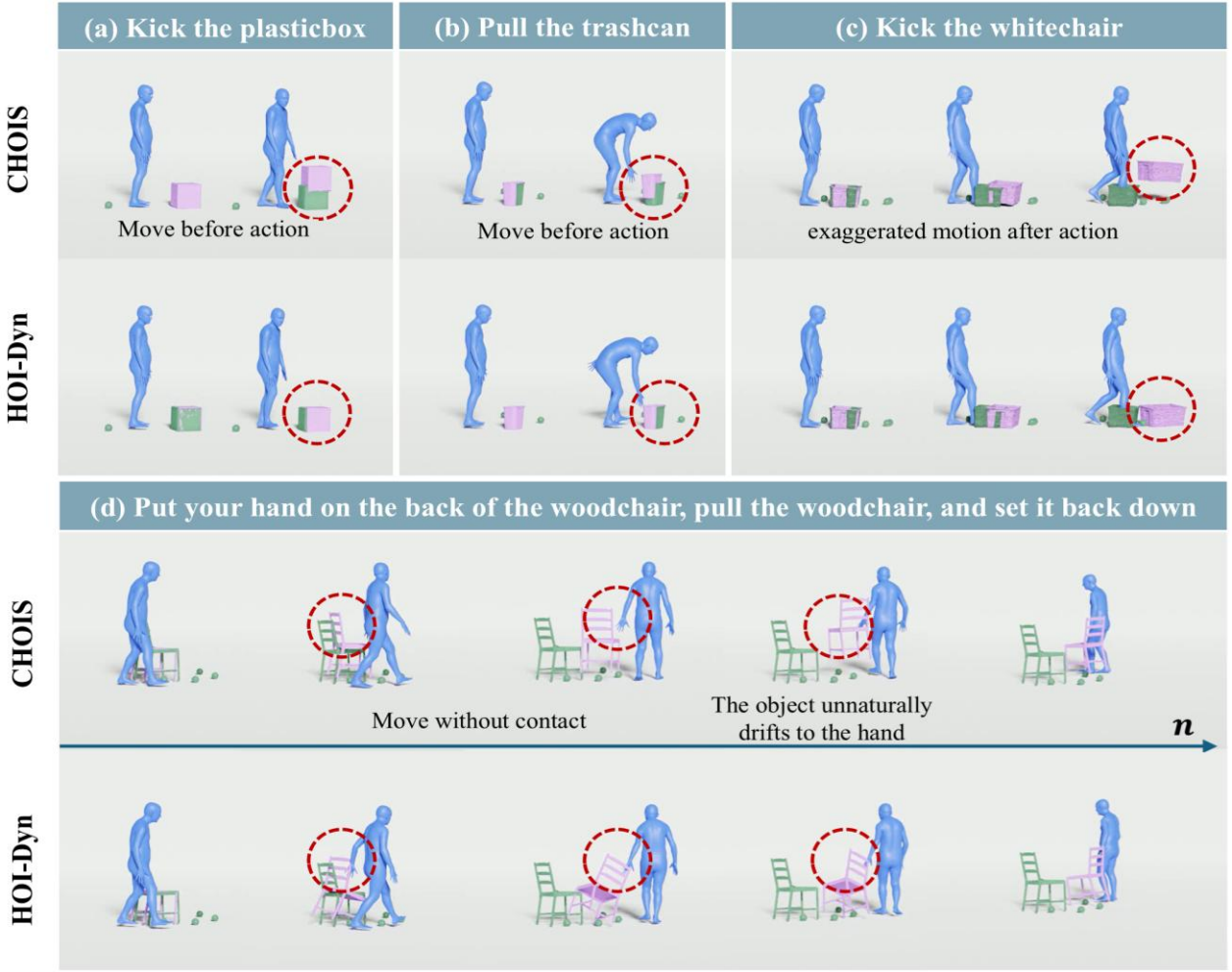
$$\mathcal{L}_{\text{dyn}} = \mathbb{E}_t \left[ \left\| \Phi(\Delta\hat{o}^*_t, \Delta\hat{o}_t) - \Phi(\Delta o^*_t, \Delta o_t) \right\|_1 \right]$$

*Predicted by dynamics model (with *)*

$\Phi$: Point Cloud Loss

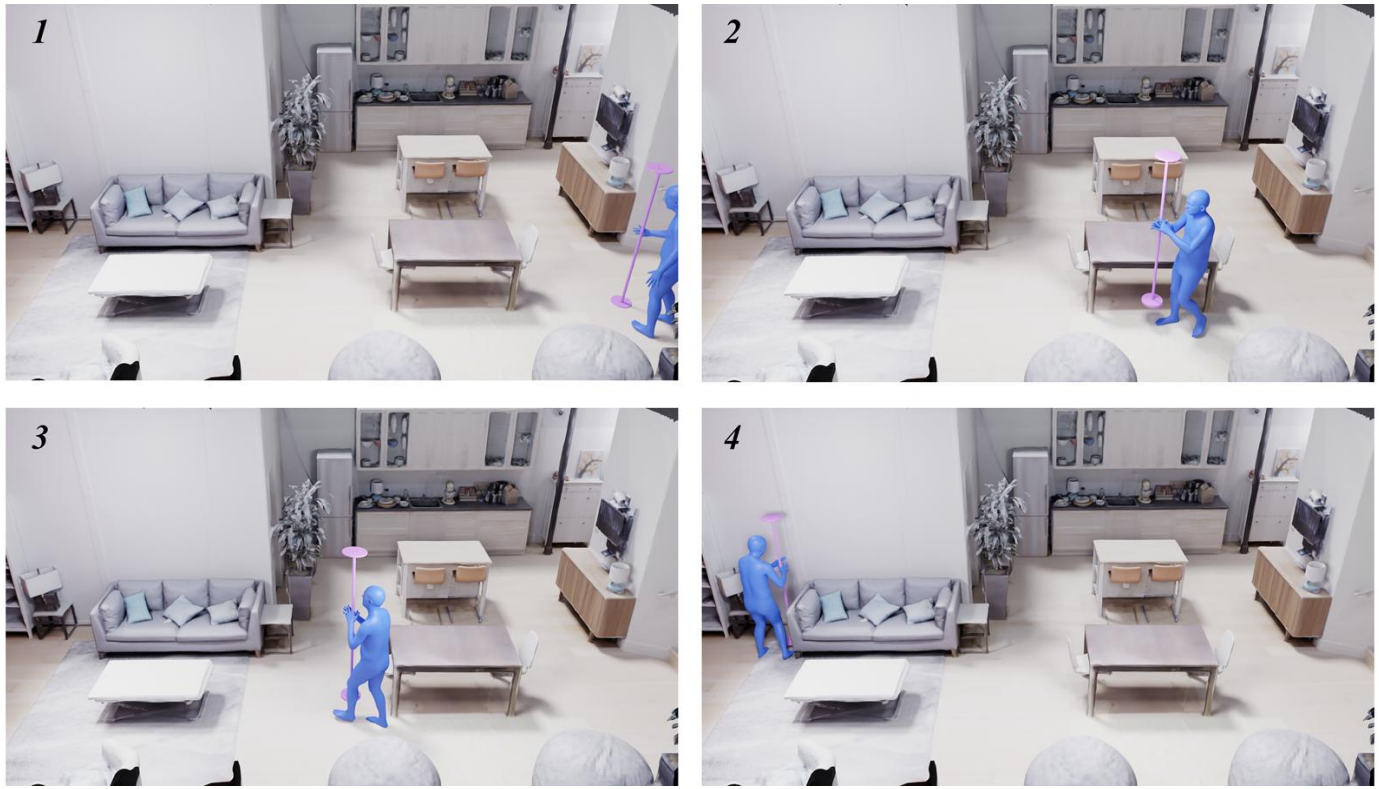
University of Glasgow

# Qualitative Results



University of Glasgow

# HOI generation in 3D Scene



University of Glasgow

# Quantitative Results

*Conducted on FullBodyManipulation Dataset*

Table 1: Comparison of methods across different metrics. Arrows indicate whether lower (↓) or higher (↑) is better, and the same notation applies hereafter.

| Method | Condition Matching | | | Human Motion | | | Interaction | | | GT Difference | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_s \downarrow$ | $T_e \downarrow$ | $T_{xy} \downarrow$ | $H_{\text{feet}} \downarrow$ | FS ↓ | FID ↓ | $C_{F1} \uparrow$ | C% ↑ | $P_{\text{hand}} \downarrow$ | MPJPE ↓ | $T_{\text{root}} \downarrow$ | $T_{\text{obj}} \downarrow$ | $R_{\text{obj}} \downarrow$ |
| Interdiff | 0.00 | 158.84 | 72.72 | 0.90 | 0.42 | 208.0 | 0.33 | 0.27 | 0.55 | 25.91 | 63.44 | 88.35 | 1.65 |
| MDM | 5.18 | 33.07 | 19.42 | 6.72 | 0.48 | 6.16 | 0.53 | 0.43 | 0.66 | 17.86 | 34.16 | 24.46 | 1.85 |
| Lin-OMOMO | 0.00 | 0.00 | 0.00 | 7.21 | 0.41 | 15.33 | 0.57 | 0.54 | 0.51 | 21.73 | 36.62 | 17.12 | 1.21 |
| Pred-OMOMO | 2.39 | 8.03 | 4.15 | 7.08 | 0.40 | 4.19 | 0.66 | 0.62 | 0.58 | 18.66 | 28.39 | 16.36 | 1.05 |
| GT-OMOMO | 0.00 | 0.00 | 0.00 | 7.10 | 0.41 | 5.69 | 0.67 | 0.59 | 0.55 | 15.82 | 24.75 | 0.00 | 0.00 |
| CHOIS | 2.10 | 6.16 | **3.03** | 3.39 | 0.41 | 0.87 | 0.66 | 0.54 | **0.61** | 16.01 | 24.33 | 14.29 | 0.99 |
| HOI-Dyn (Ours) | **1.75** | **5.58** | 3.26 | **3.07** | **0.37** | **0.48** | **0.71** | **0.60** | 0.64 | **15.60** | **23.90** | **12.47** | **0.90** |

✨ **better human motion fidelity and improved interaction quality.**

University of Glasgow

# Quantitative Results

*Conducted on 3D-FUTURE Dataset*

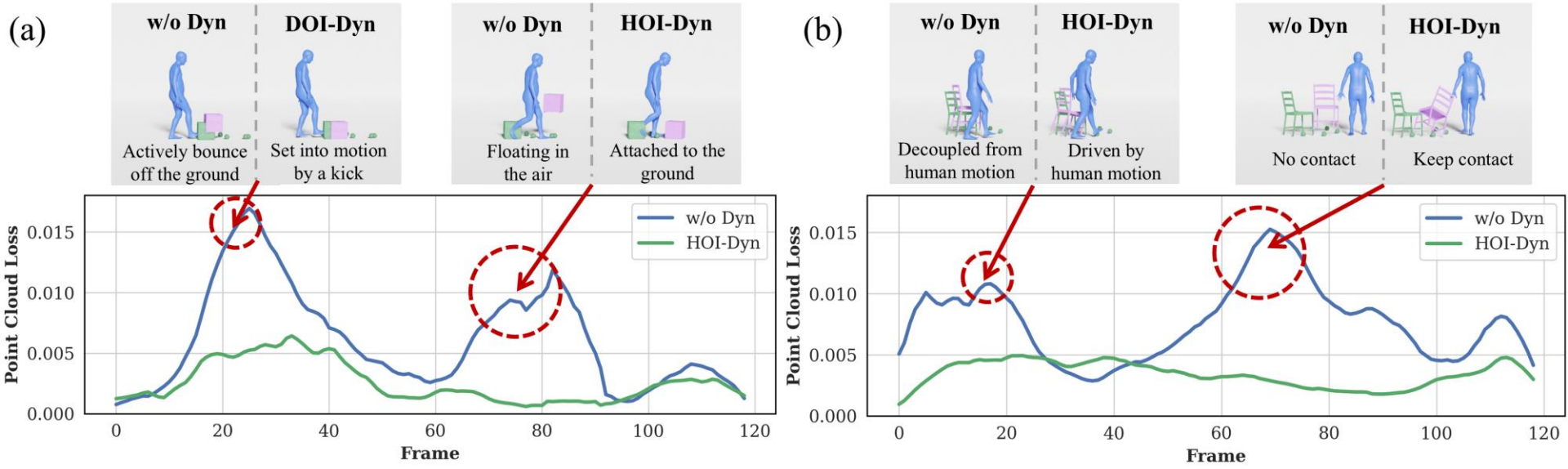
Table 2: Interaction synthesis results on the 3D-FUTURE dataset

| Method | Condition Matching | | | Human Motion | | | Interaction | |
|---|---|---|---|---|---|---|---|---|
| | $T_s \downarrow$ | $T_e \downarrow$ | $T_{xy} \downarrow$ | $H_{\text{feet}} \downarrow$ | FS $\downarrow$ | FID $\downarrow$ | C% $\uparrow$ | $P_{\text{hand}} \downarrow$ |
| InterDiff | 0.00 | 161.26 | 72.77 | -0.26 | 0.42 | 207.3 | 0.24 | 0.11 |
| MDM | 12.58 | 40.55 | 28.72 | 7.02 | 0.49 | 8.50 | 0.34 | 0.26 |
| Lin-OMOMO | 0.00 | 0.00 | 0.00 | 6.32 | 0.42 | 23.17 | 0.44 | 0.11 |
| Pred-OMOMO | 4.15 | 9.03 | 3.89 | 6.08 | 0.40 | 3.74 | 0.50 | 0.18 |
| CHOIS | **3.23** | 6.21 | 2.99 | 2.95 | 0.42 | 1.67 | 0.47 | **0.19** |
| HOI-Dyn (Ours) | 4.60 | **6.17** | **2.95** | **2.56** | **0.37** | **1.62** | **0.54** | 0.26 |

✨ **generalizes well to unseen objects on the 3DFUTURE dataset.**

University of Glasgow

# Dynamics as Causality Indicator




✨ **The interaction dynamics serves as a surrogate evaluator of causal consistency.**

University of Glasgow

# Conclusion & Future Work

- We propose a novel **driver-responder framework** that explicitly models object **Interaction Dynamics** and integrates with existing HOI motion diffusion techniques to achieve more realistic HOIs.

- **Future Work:** We also plan to explore the extension of our framework to handle multi-human and multi-object scenarios to assess scalability.

University of Glasgow

University of Sheffield

NEURAL INFORMATION PROCESSING SYSTEMS

# Thanks for your attention!

AIR Lab

Project Website

Paper