



Paper ID
7308

Scaling Up Dynamic Human-Scene Interaction Modeling

Nan Jiang^{1,2*}, Zhiyuan Zhang^{1,2*}, Hongjie Li¹, Xiaoxuan Ma³, Zan Wang⁴,
Yixin Chen², Tengyu Liu², Yixin Zhu¹✉, Siyuan Huang²✉

¹Institute for AI, Peking University ²National Key Lab of General AI, BIGAI ³School of Computer Science, CFCS, Peking University

⁴Beijing Institute of Technology *Equal contributors ✉ yixin.zhu@pku.edu.cn, syhuang@bigai.ai

<https://jnnan.github.io/trumans/>



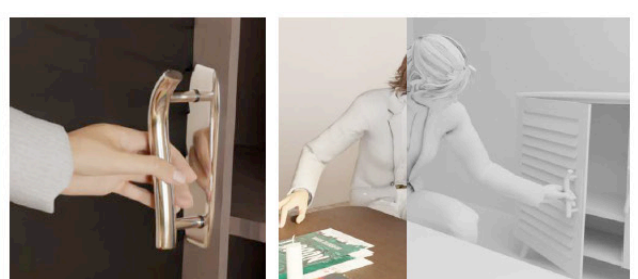
Project Page



Contributions

- ✓ A **dataset** with motion-captured indoor human activities
- ✓ A **method** for generating human-scene interaction motions

15-Hour MoCap
Dataset



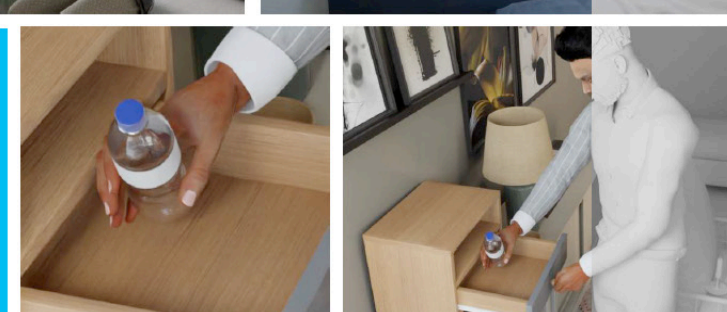
Scene-Aware HOI
Motion Synthesis

100 Scenes

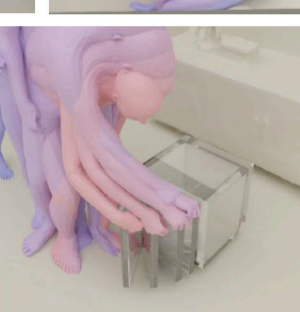


Natural
Collision
Avoidance

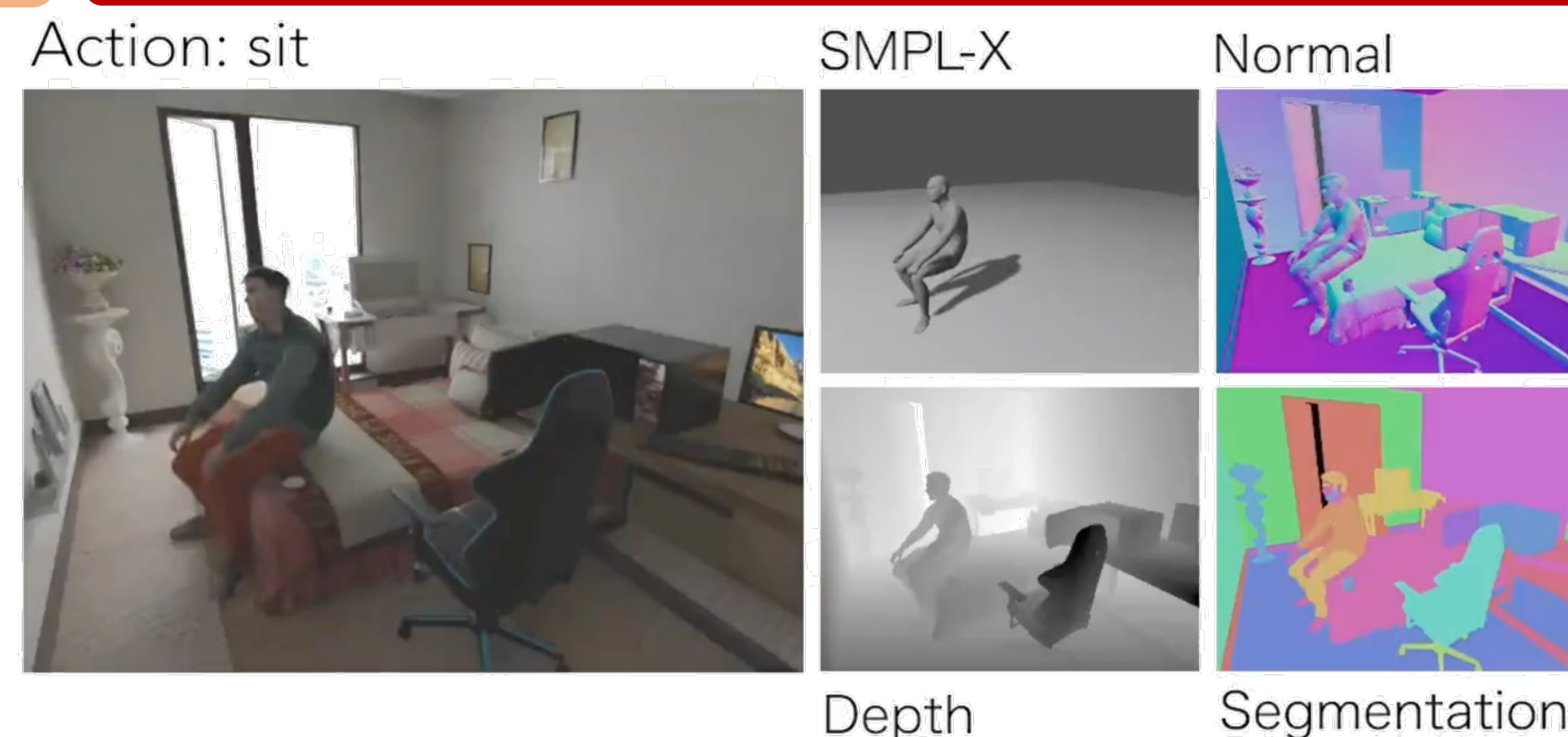
Dynamic
Object



Generate
Motions of
Infinite length

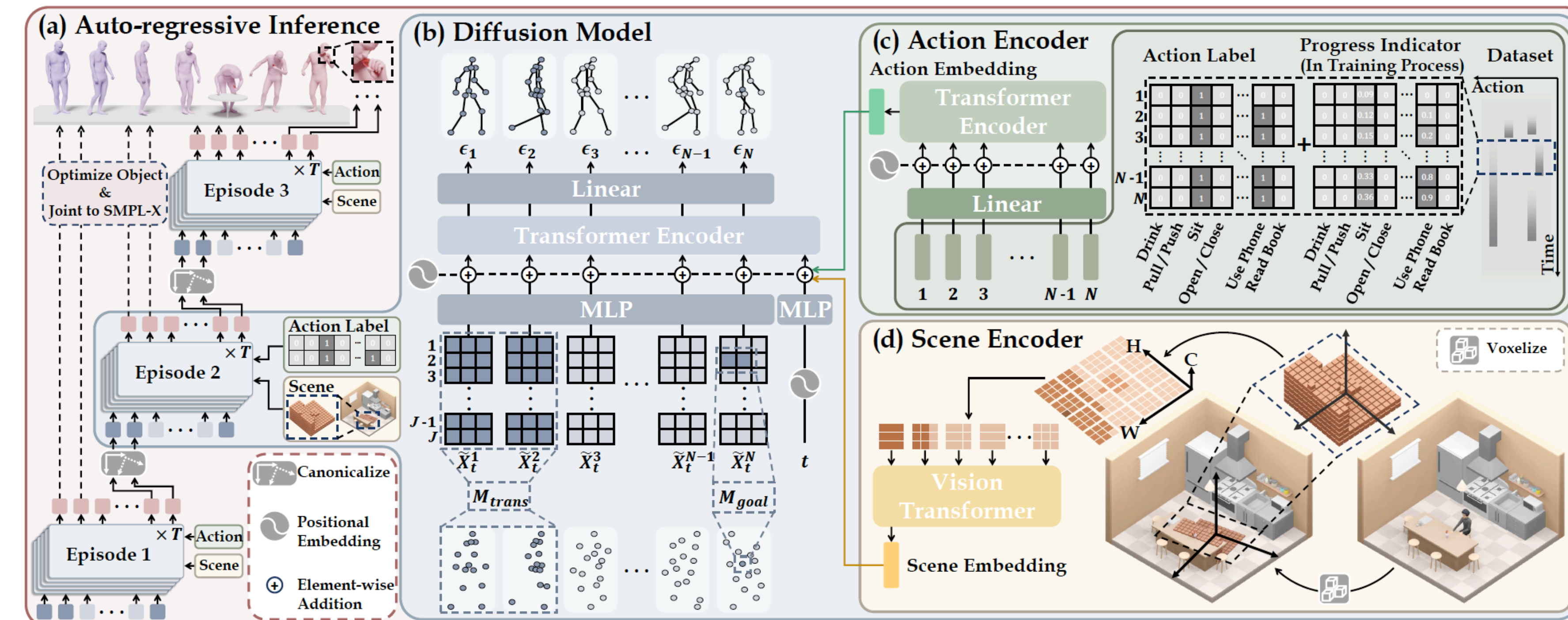


Proposed Dataset

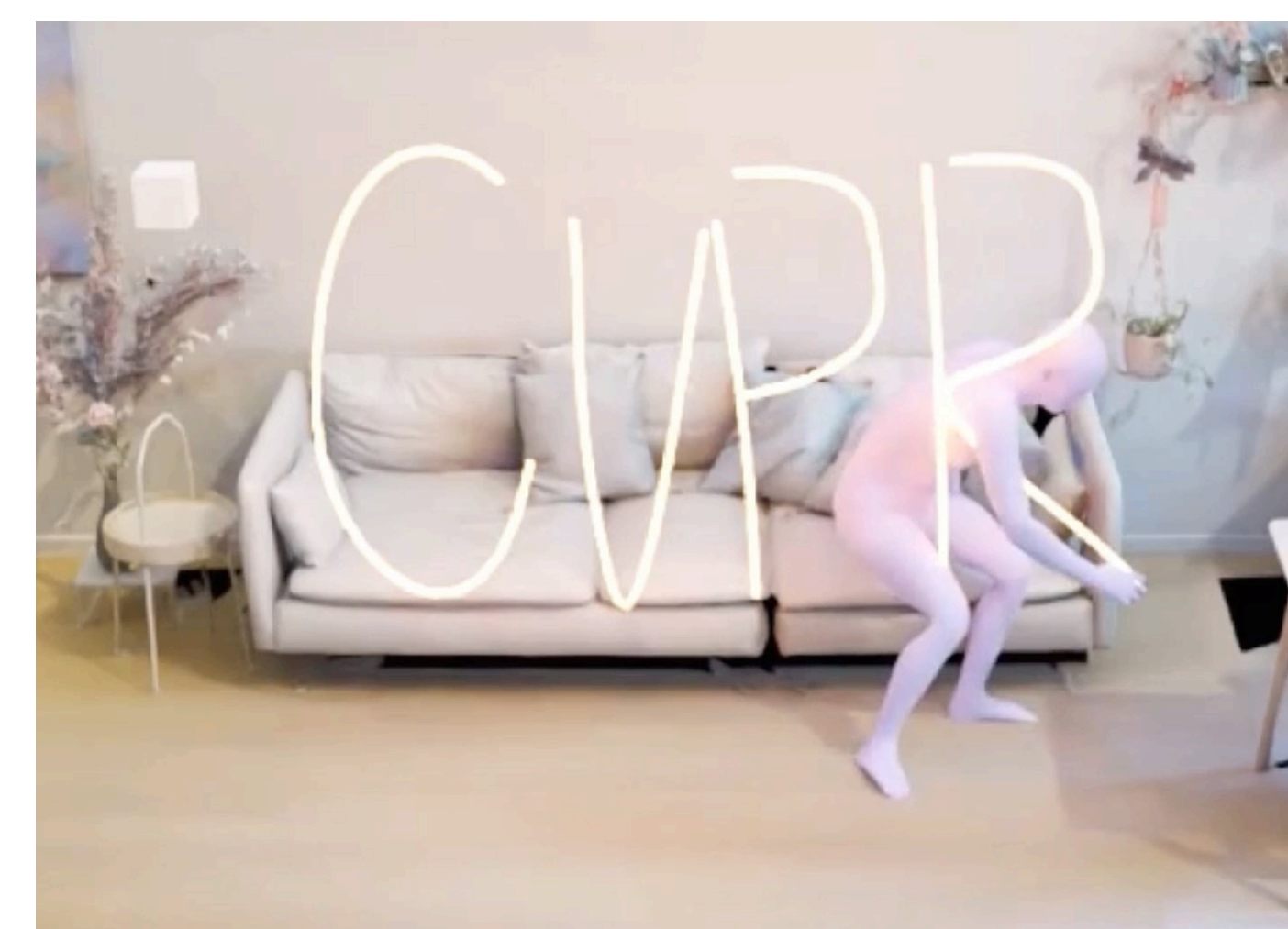
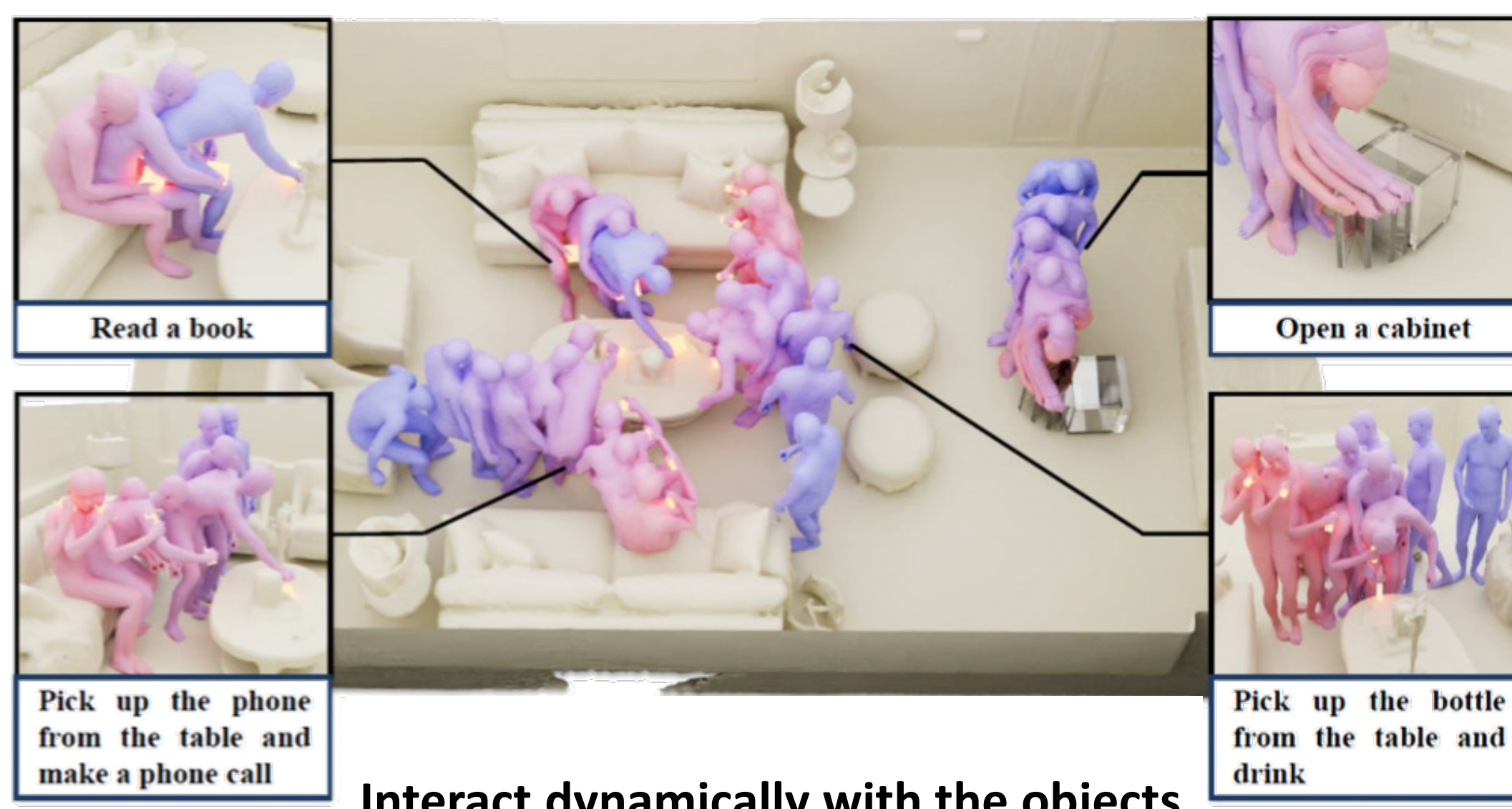


- ✓ **TRUMANS**, an extensive **motion-captured** human-scene interaction dataset across **100 indoor scenes**
- ✓ Comprises **15 hours** of diverse **interaction sequences**
- ✓ With **20 types** of rigid and **articulated** objects **dynamically** involved in human-object interaction

Motion Synthesis Method



Motion Synthesis Results



Quantitative Results

Motion generation

Table 2. Evaluation of locomotion and scene-level interaction. We compare performances on **TRUMANS** and **PROX** [16].

Method	Cont.↑	Pene _{mean} ↓	Pene _{max} ↓	Dis. suc.↓
Wang et al. [53]	0.969	1.935	14.33	0.581
SceneDiff [21]	0.912	1.691	17.48	0.645
GMD [23]	0.931	2.867	21.30	0.871
Ours	0.992	1.820	11.74	0.258
Ours w/o aug.	0.991	2.010	15.52	-

Human mesh estimation

Table 4. Performance of Ma et al. [29] trained on 3DPW [51] combined with **TRUMANS** in different ratios.

Training Data	MPVE↓	MPJPE↓	PA-MPJPE↓
3DPW [51]	101.3	88.2	54.4
3DPW+T (2:1)	88.8	77.2	46.4
3DPW+T (1:1)	78.5	78.5	46.4

Table 3. Evaluation of object-level interaction. We compare performances on **TRUMANS** and **GRAB** [47]. The definition of “Real” follows the one defined in Tevet et al. [49].

Method	FID↓	Diversity→	Pene _{scene} ↓	Dis. suc.↓
Real-TRUMANS	-	2.734	-	-
GOAL [48]	0.512	2.493	34.10	0.801
IMoS [11]	0.711	2.667	37.48	0.774
Ours	0.313	2.693	11.74	0.226
Ours - \mathcal{A}_{ind}	2.104	1.318	10.62	1.000

Human-scene contact estimation

Table 5. Performance of **BSTRO** [20] trained on **RICH** [20] combined with **TRUMANS**

Training Data	Prec↑	Rec↑	F1↑	geo err↓
RICH [20]	0.6823	0.7427	0.6823	10.27
R+T (2:1)	0.7087	0.7370	0.6927	9.593
R+T (1:1)	0.7137	0.7286	0.6923	9.459