From Egocentric Experience to Interactive Intelligence: Building the World in First Person



Ziwei Liu

Nanyang Technological University

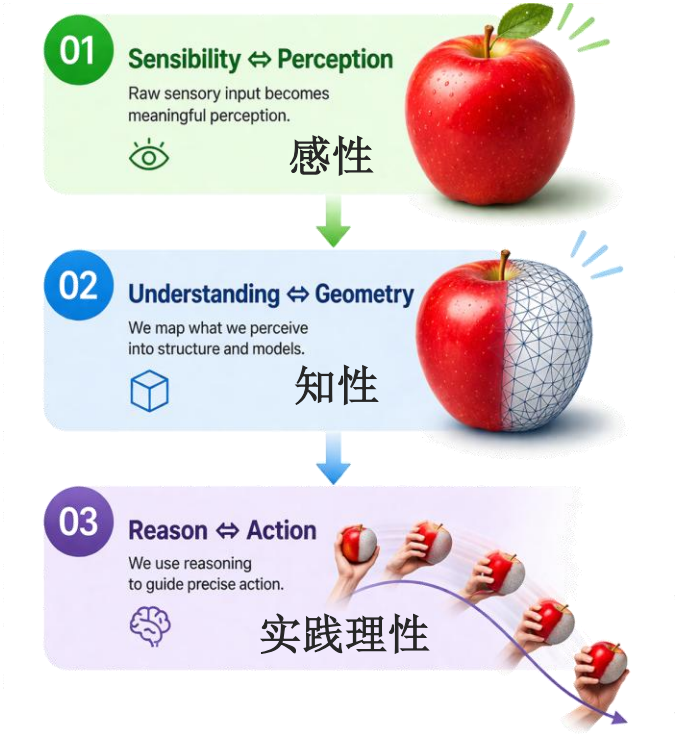
Ropedia

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

S-LAB FOR ADVANCED INTELLIGENCE

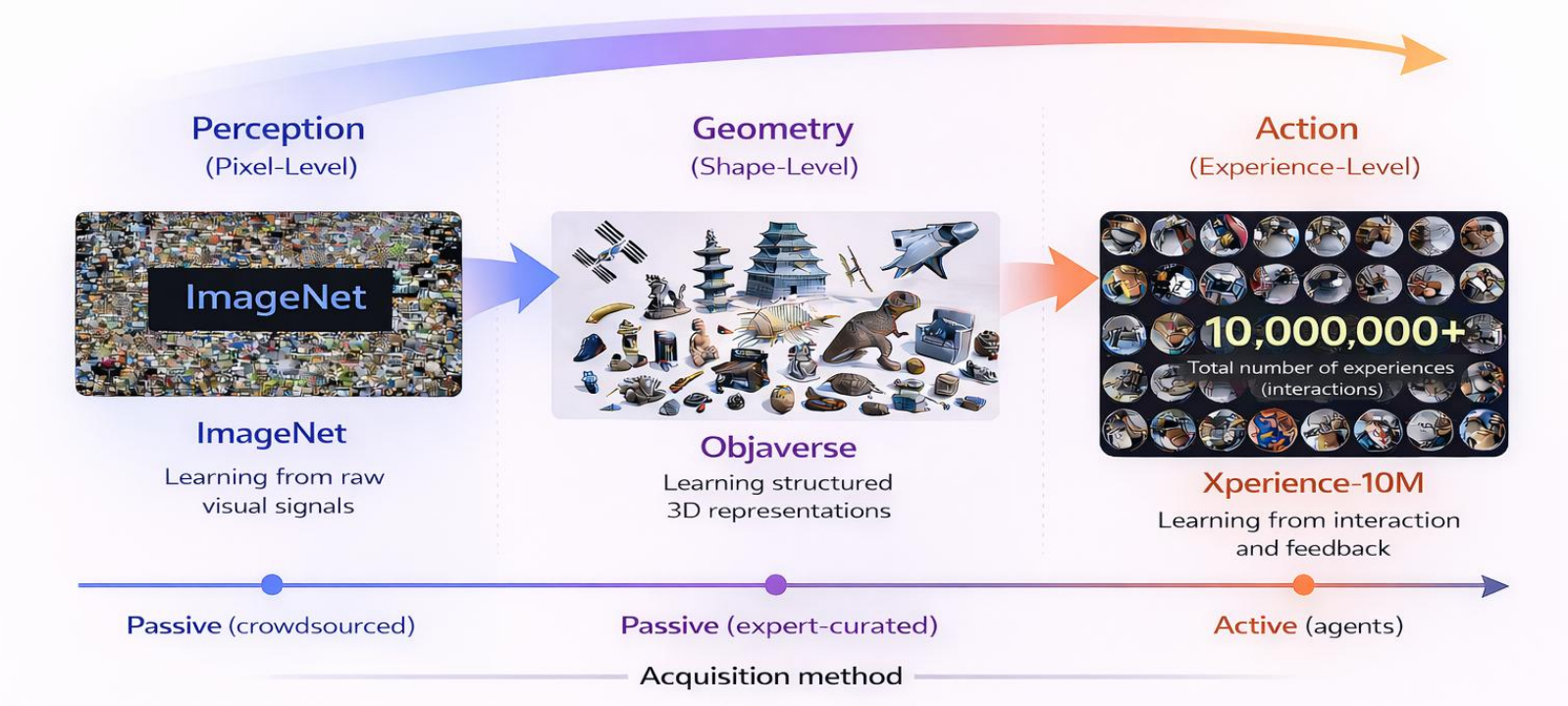
Three Dimensions of Kant's Laws of Nature





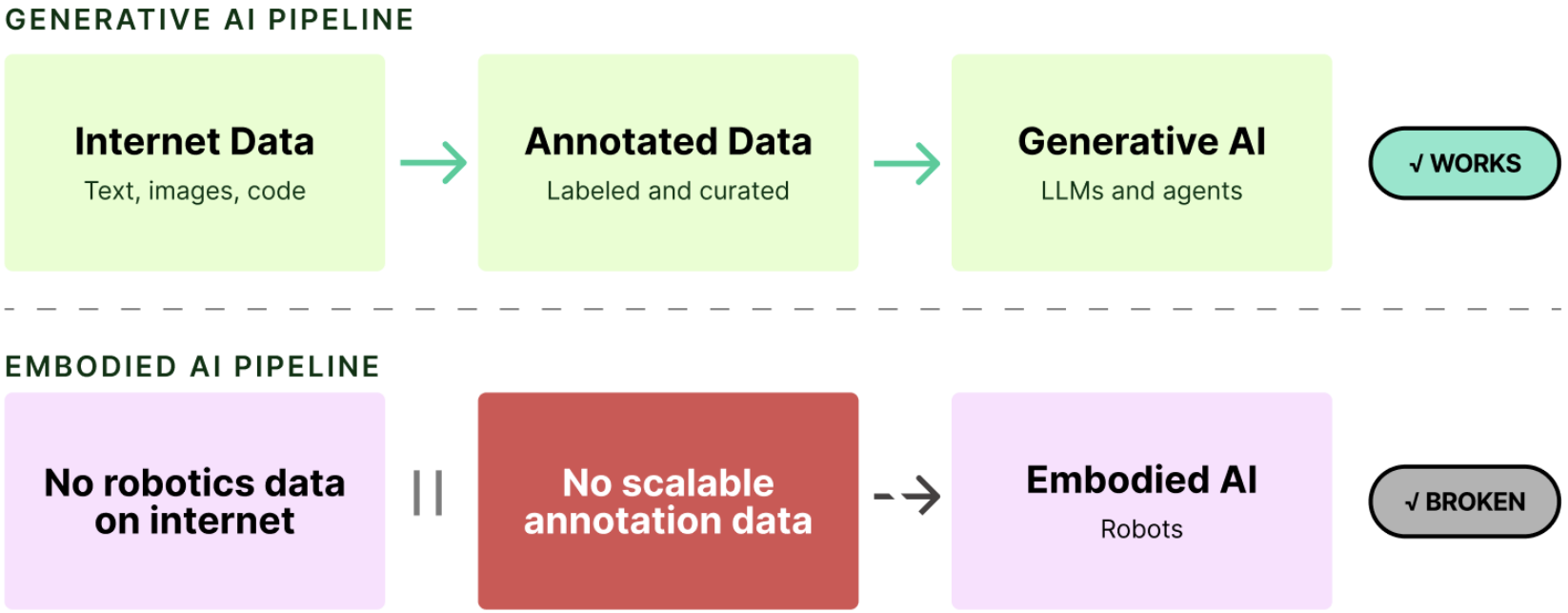
"纯粹理性批判" - 康德

From Perception to Action: The Evolution of Visual Intelligence



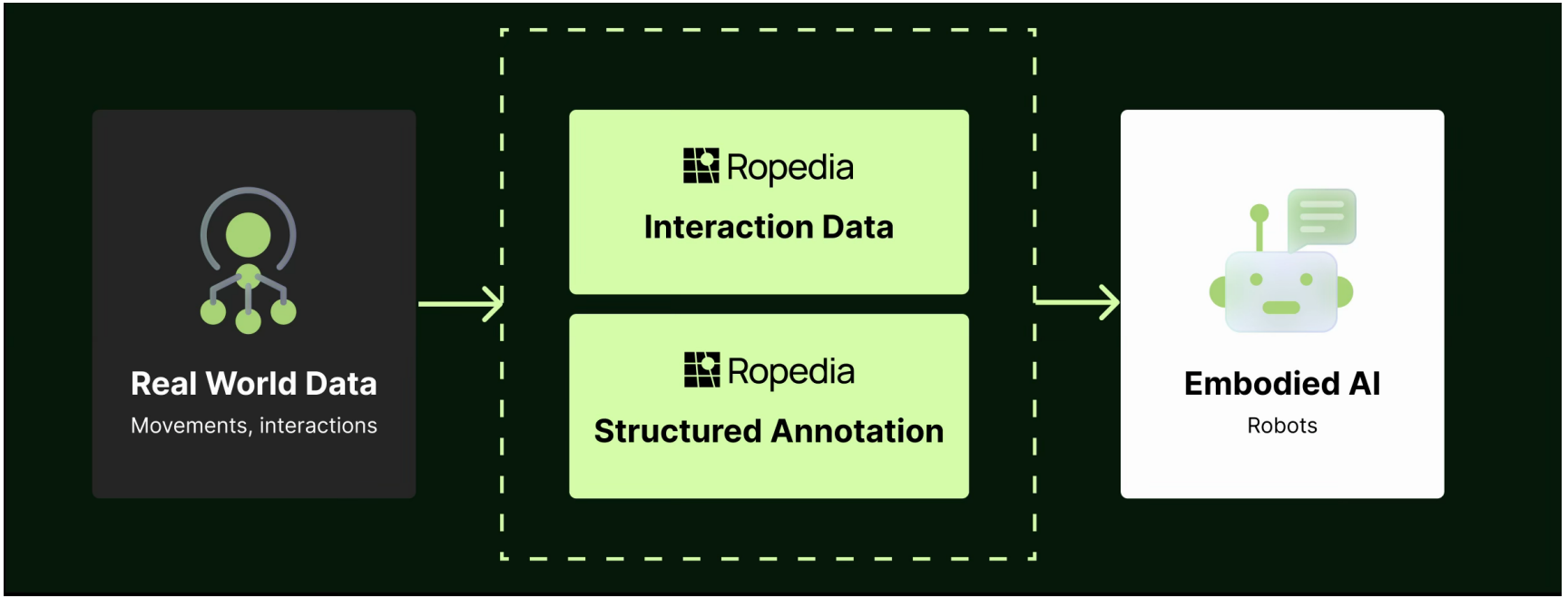
The Real-World Data Bottleneck

The same scaling law that powered LLMs is now entering the physical world.



Our Solution

We capture structured real-world interaction at **scale** and **low cost**.



Full-stack Low-cost Portable Data Capture

HOMIE

320g
Super lightweight

50μs
Synchronization

360°
Field of View

13h
Storage Duration

10x
Faster Deployment

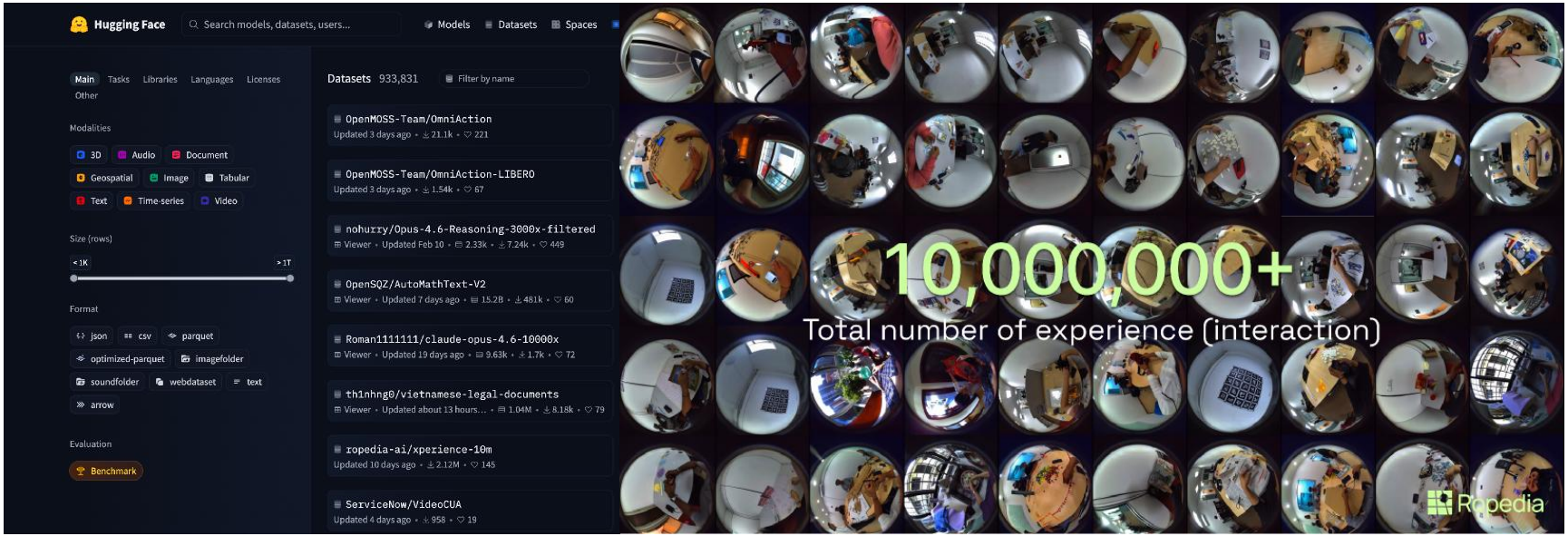
12.5x
Cost Reduction



Our Dataset: In-the-Wild Human Xperience-10M

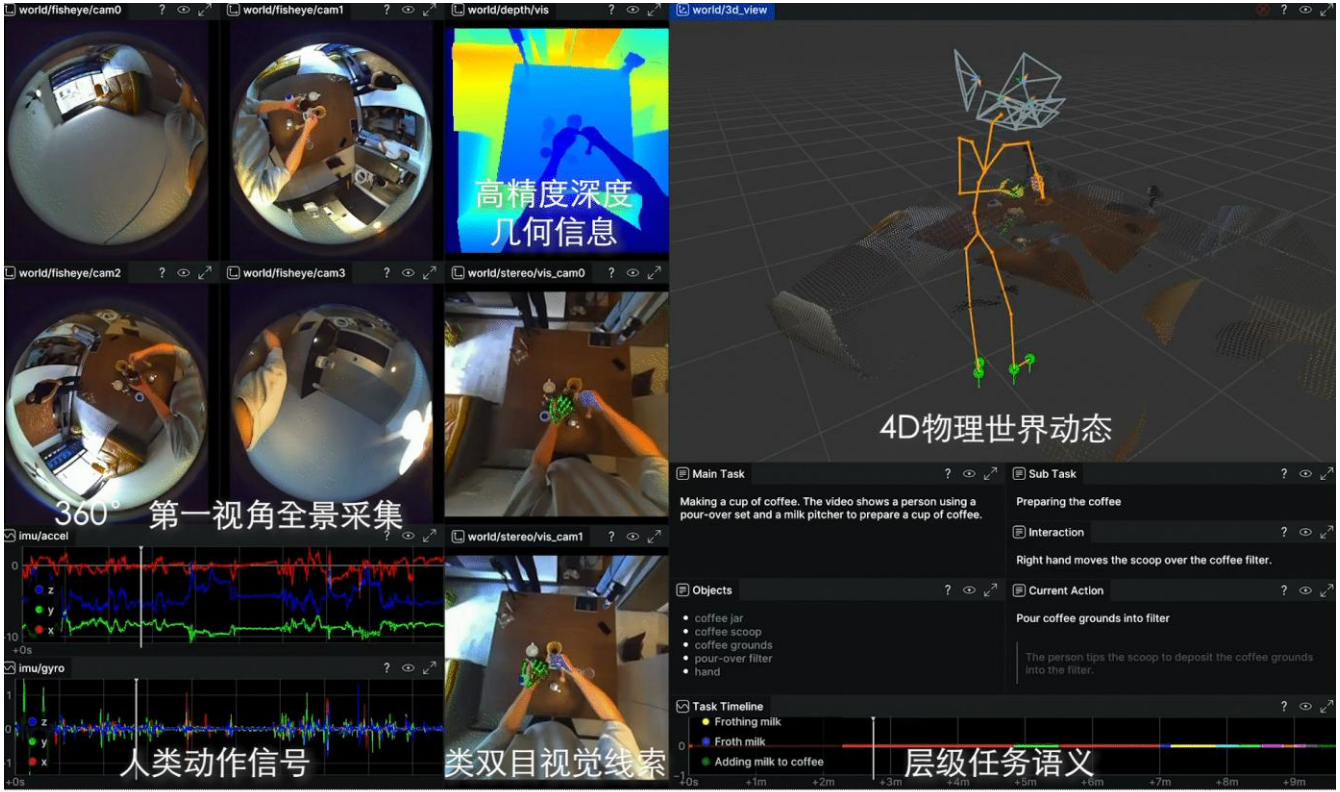
2.1M Total Downloads in 1 Week

Top 3 Over ALL Datasets on Hugging Face



Ropedia

High-Quality Data from Anywhere by Anybody



INTERACTION EPISODES

10M

VIDEO W/0 AUDIO

10,000h

RGB FRAME NUMBER

2.88B

MOCAP FRAME

576M

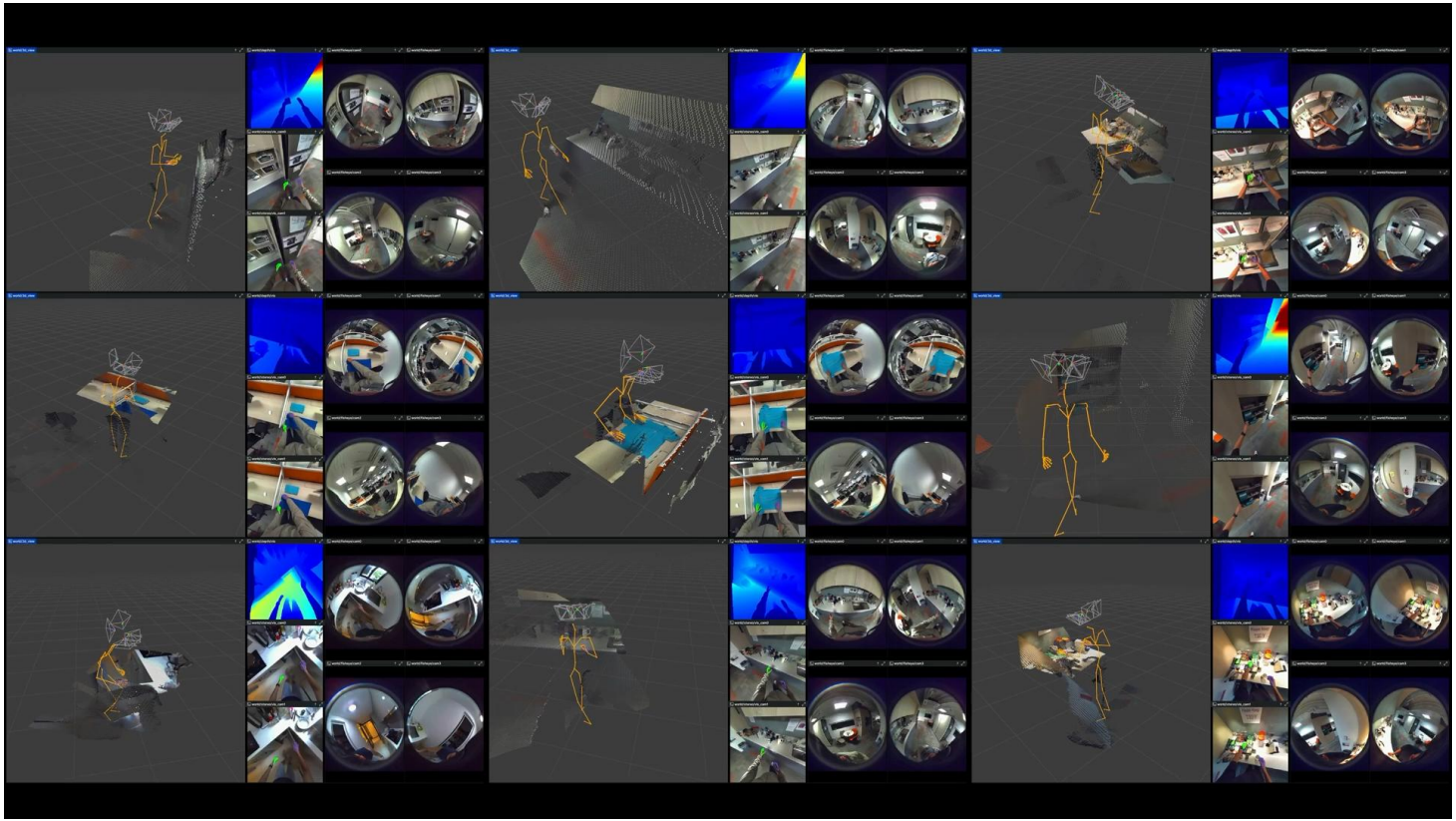
NO. OF OBJECTS

350K

TOTAL STORAGE

~1PB

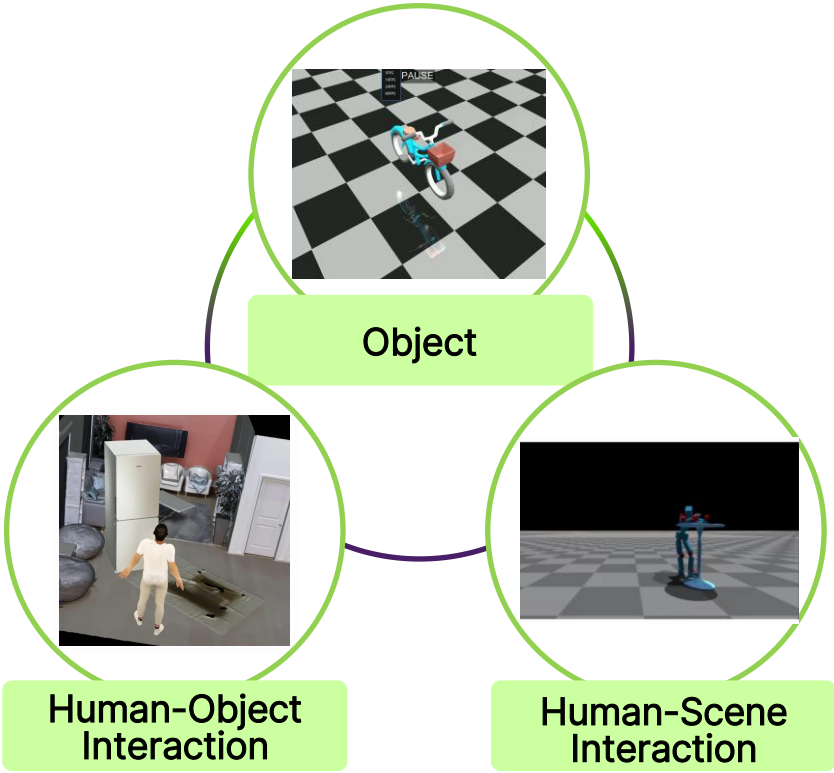
Demo: In-the-wild Data Capture



From **Interactive Experience** to **Simulation-Ready** Modeling



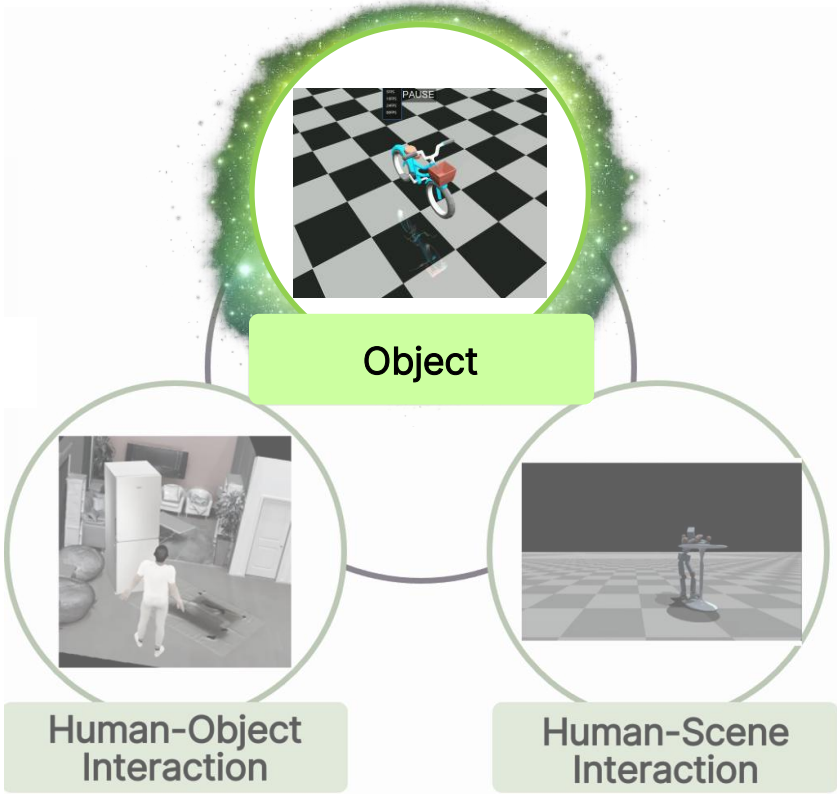
Xeperience-10M



From **Interactive Experience** to **Simulation-Ready** Modeling



Xeperience-10M



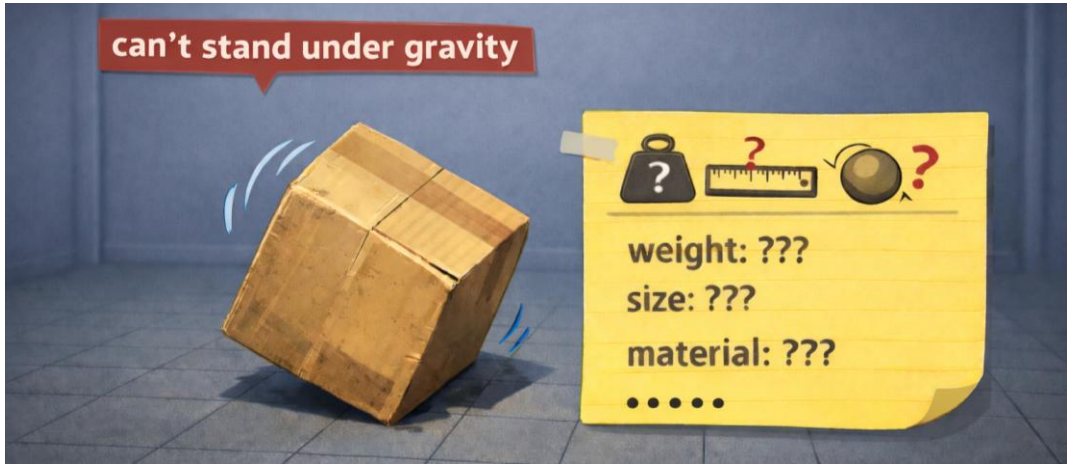
PhysX-Anything: **Simulation-Ready** Physical 3D Assets from Single Image

Ziang Cao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, Ziwei Liu.

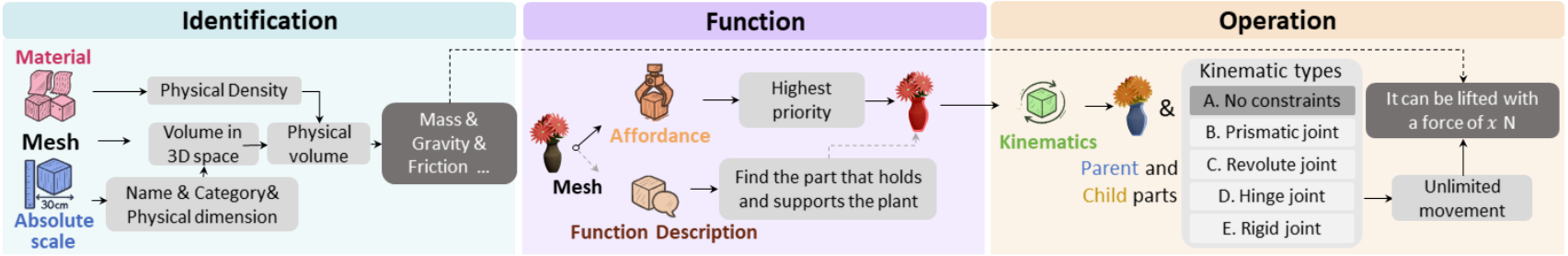


Challenge

- Existing 3D generation primarily emphasizes geometries and textures while neglecting physical-grounded modeling, hampering their real-world application in physical domains like simulation and embodied AI.

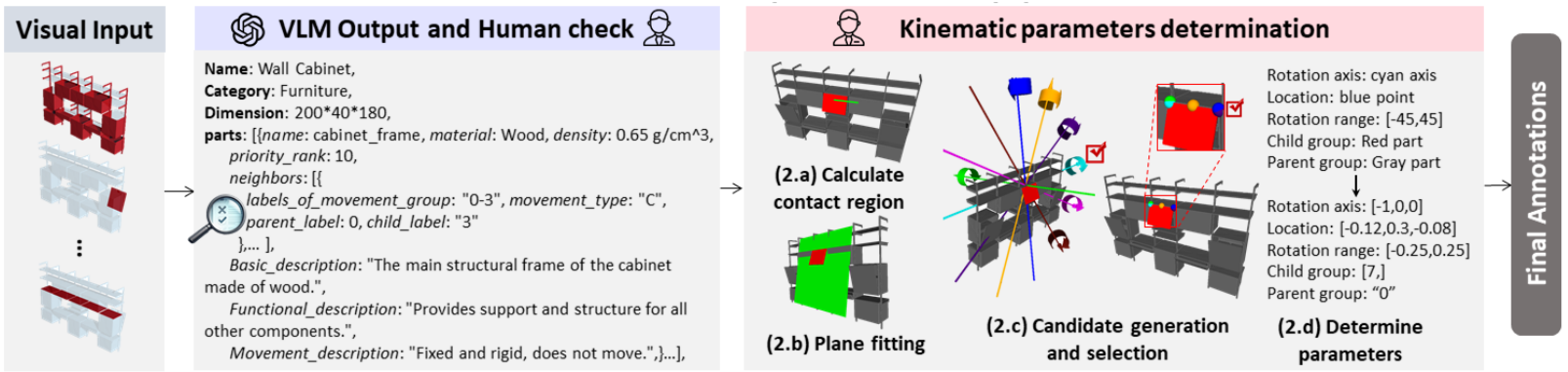


Definition of Physical Properties



- **Identification** – Determining the basic nature of the object
  - Absolute scaling and material (Young's modulus, Poisson's ratio, and density)
- **Function** – Understanding its potential applications
  - Functional affordance analysis and function descriptions
- **Operation** – Detailed usage methodologies
  - Kinematic Parameter

Human-in-the-loop Annotation Pipeline



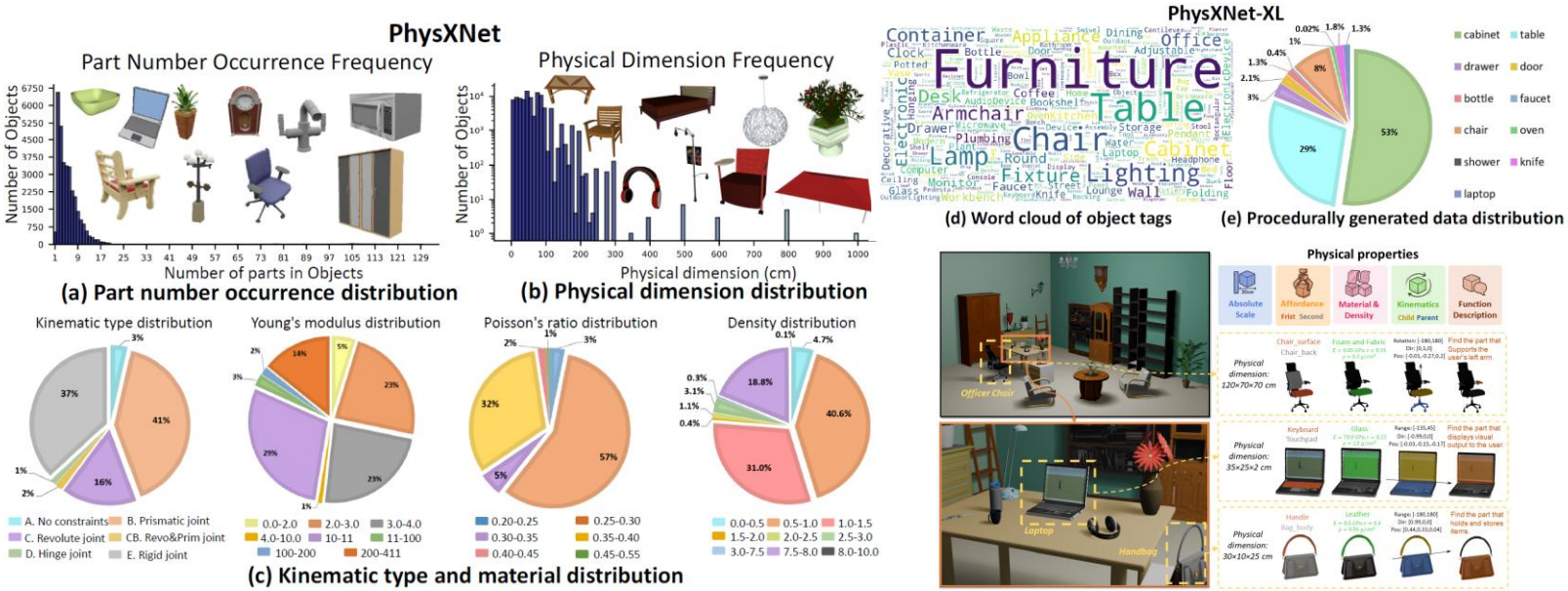
Preliminary Data Acquisition

GPT-4o → Human Check

Kinematic Parameter Determination

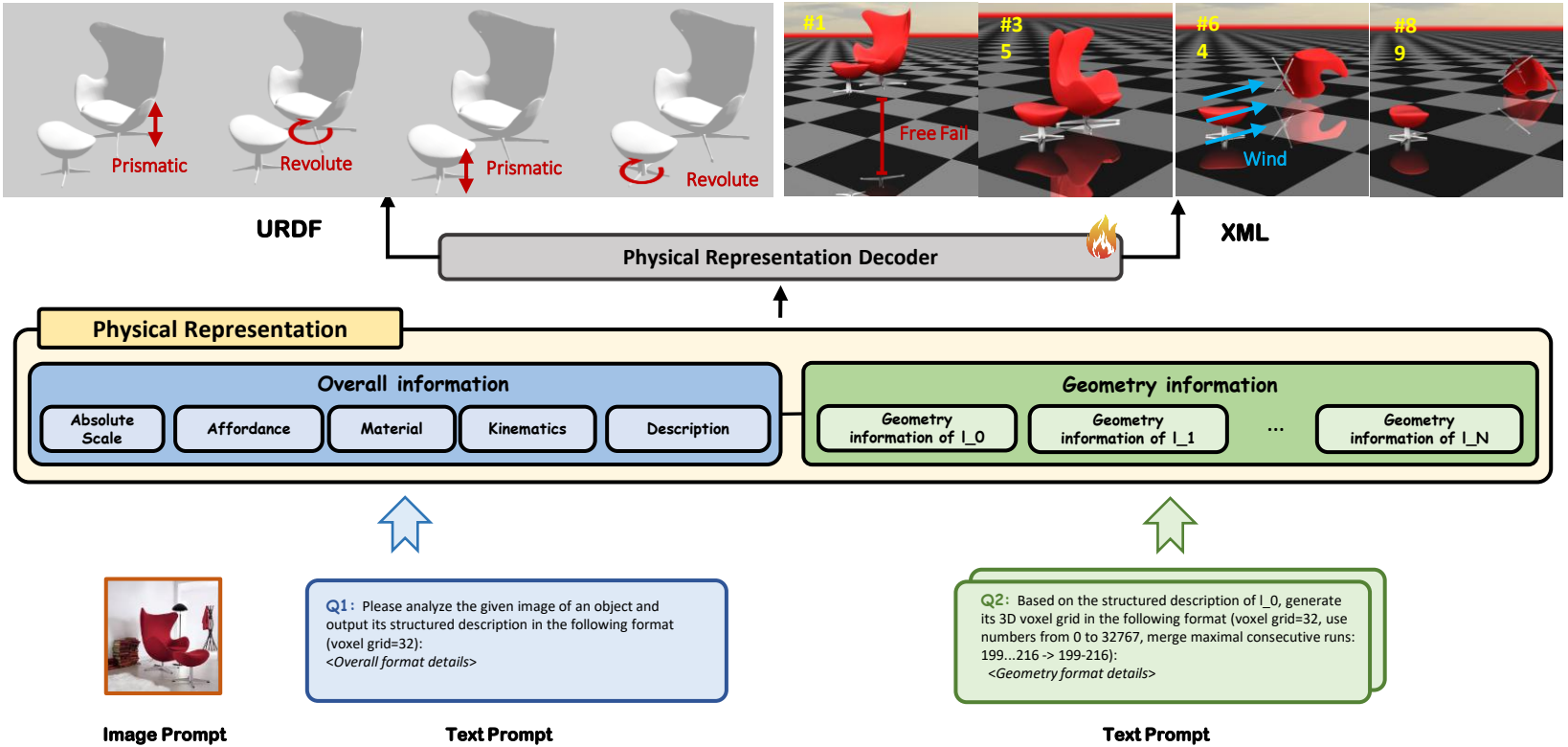
Contact Region → Plane Fitting → Candidate Generation and Selection → Kinematic Parameter Determination

PhysXNet & PhysXNet-XL

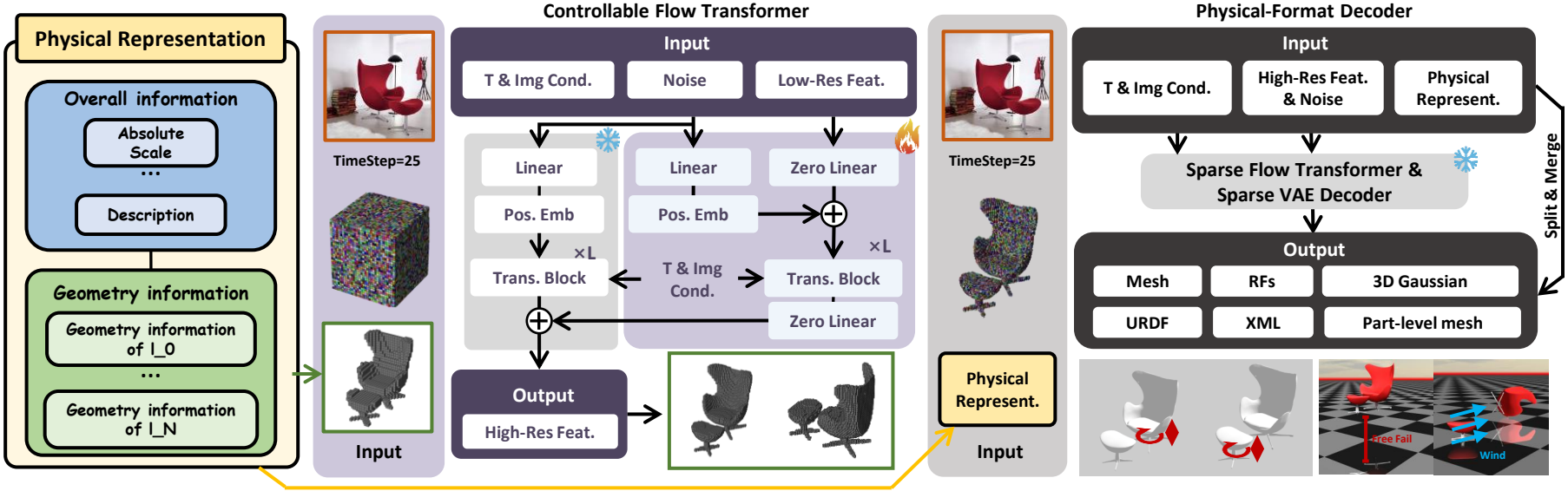


We propose PhysXNet and PhysXNet-XL – the first comprehensive physical 3D dataset with over 26K and 6M richly annotated 3D objects

Key Idea: Interactive 3D Object as **Physical Code**



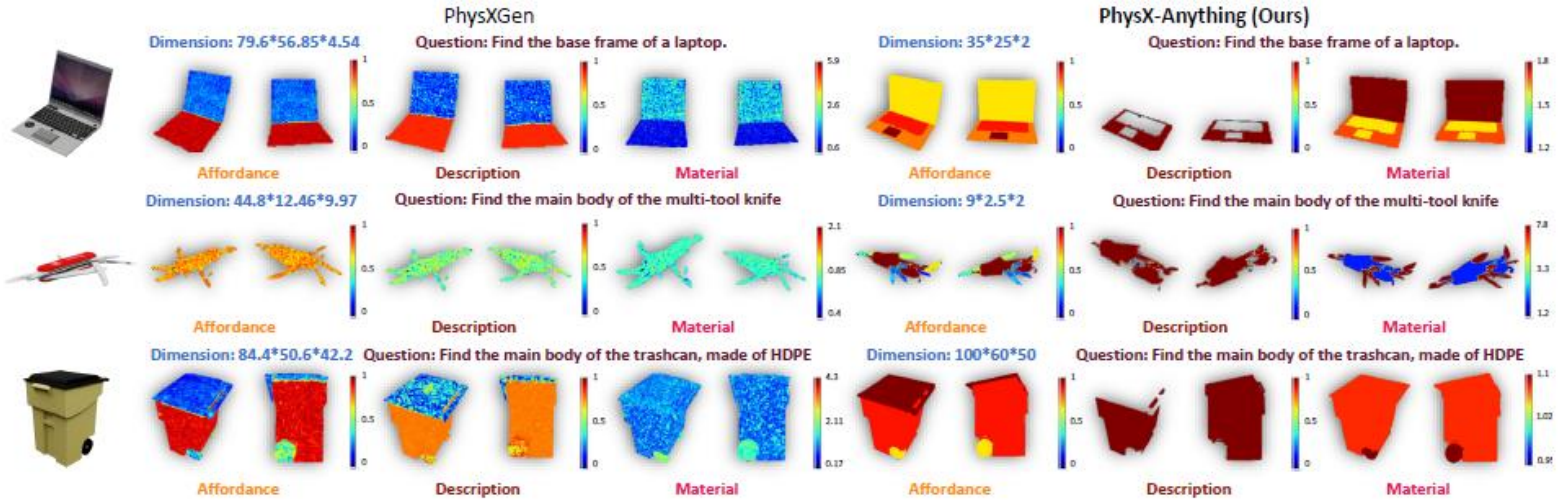
Method: Physical Representation Decoder



Results: Qualitative Comparisons



Results: Qualitative Comparisons



Results: Quantitative Comparisons

Methods	Geometry			Physical Attributes				
	PSNR ↑	CD ↓	F-score ↑	Absolute scale ↓	Material ↑	Affordance ↑	Kinematic parameters (VLM) ↑	Description ↑
URDFormer [11]	7.97	48.44	43.81	–	–	–	0.31	–
Articulate-Anything [16]	16.90	17.01	67.35	–	–	–	0.65	–
PhysXGen [3]	20.33	14.55	76.3	43.44	6.29	9.75	0.71	12.89
PhysX-Anything (Ours)	**20.35**	**14.43**	**77.50**	**0.30**	**17.52**	**14.28**	**0.83**	**19.36**

Results: Simulated Objects

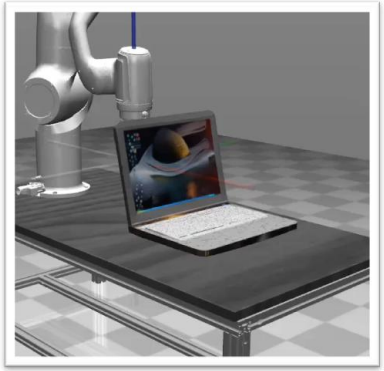


Results: Simulated Objects

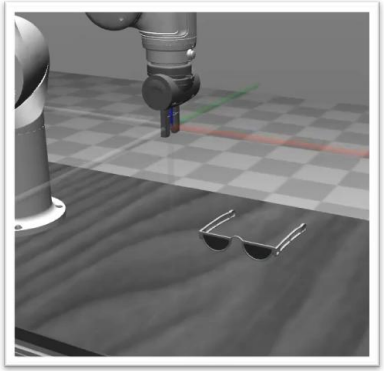


Results: Simulated Interactions

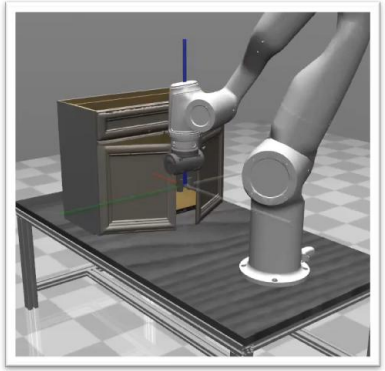
Laptop Closing



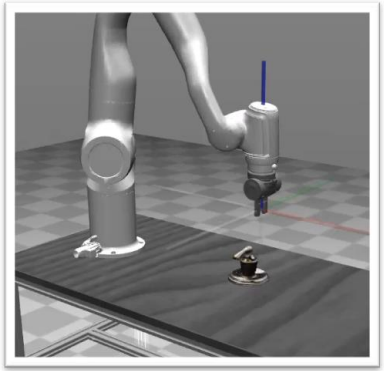
Eyeglass Temple Folding



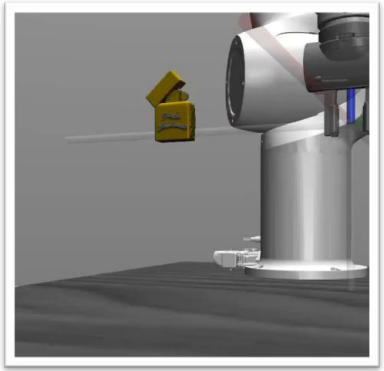
Door Opening and Closing



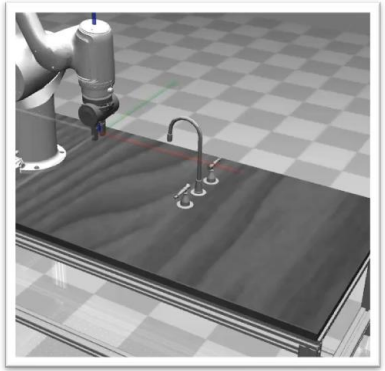
Handle Manipulation



Lighter Snapping Open



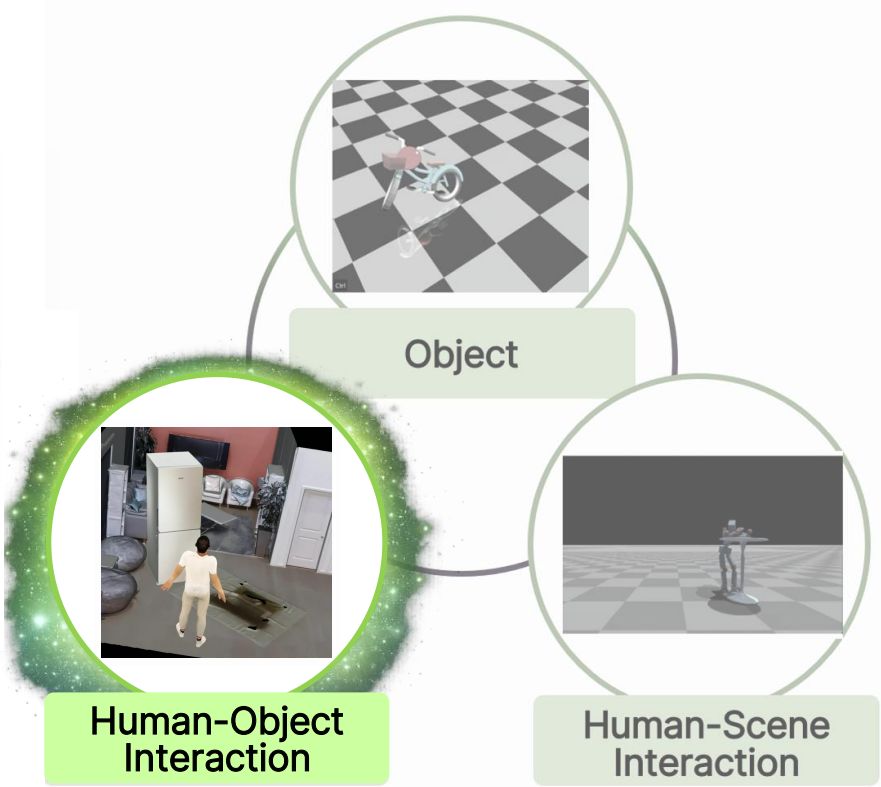
Faucet Switch Manipulation



From **Interactive Experience** to **Simulation-Ready** Modeling

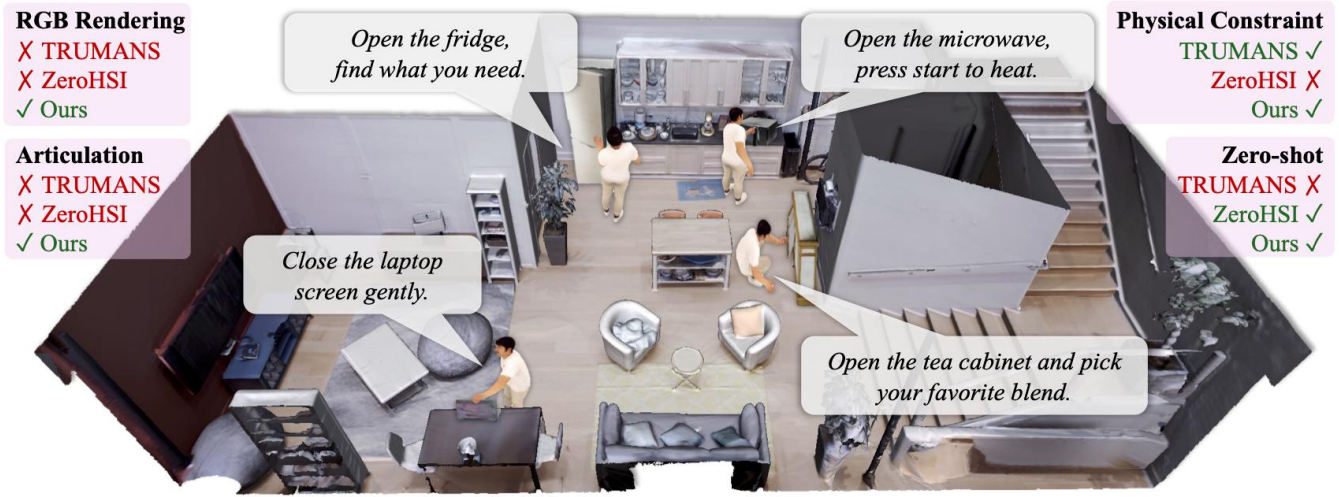


Xeperience-10M



ArtHOI: **Articulated Human-Object Interaction** Synthesis by 4D Reconstruction from Video Priors

Zihao Huang, Tianqi Liu, Zhaoxi Chen, Shaocong Xu, Saining Zhang, Lixing Xiao, Zhiguo Cao, Wei Li, Hao Zhao, Ziwei Liu



Task: Reconstruction of Articulated HOI

Input



Open the fridge, find what you need.

3D Assets + Text Prompt

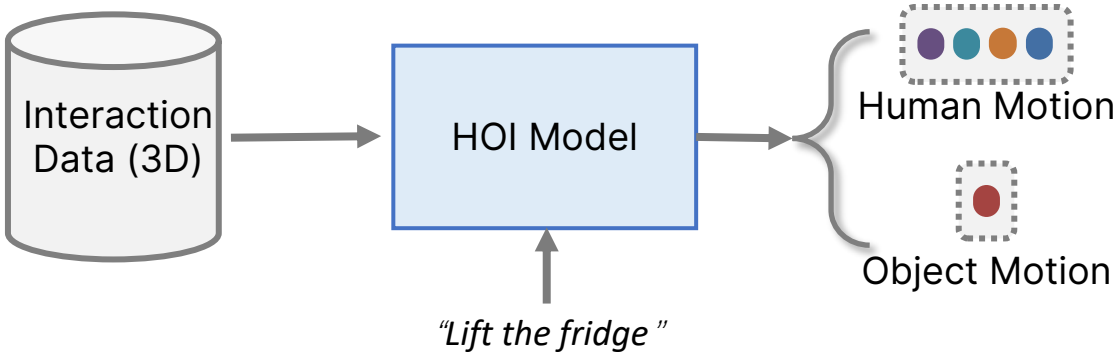
Reconstruction

Output

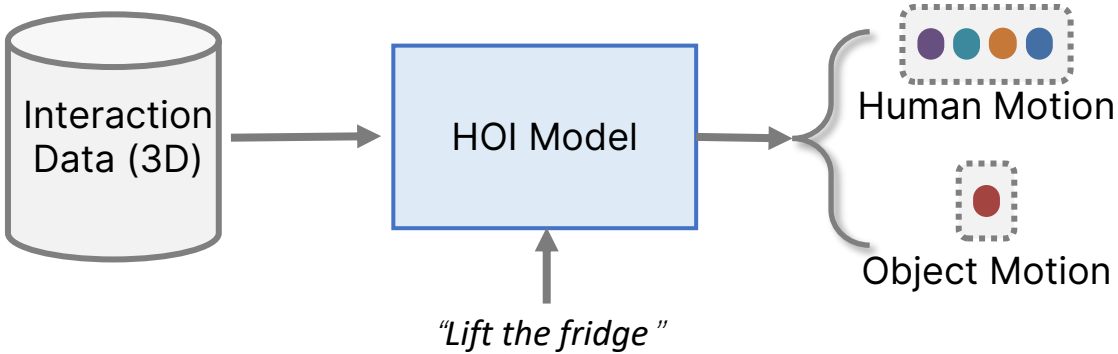


Articulated Human-object Interaction

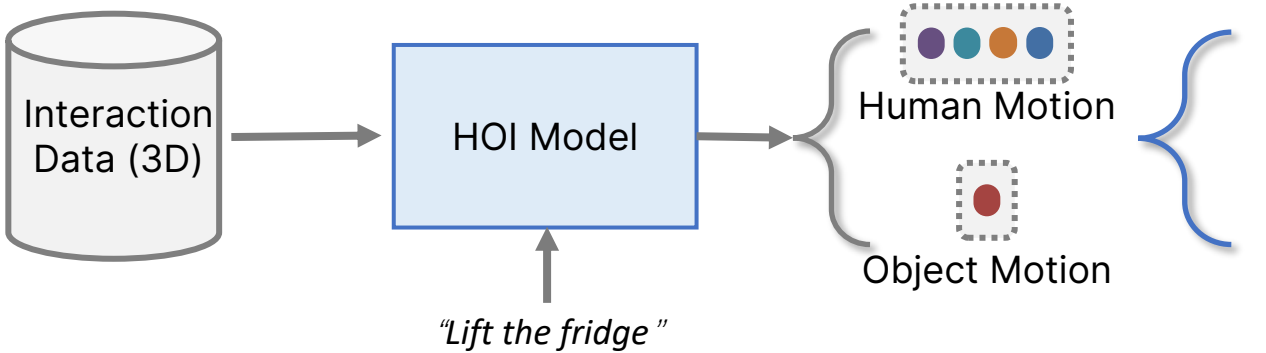
Challenge


Interaction Data (3D)
HOI Model
Human Motion
Object Motion
"Lift the fridge"

Challenge

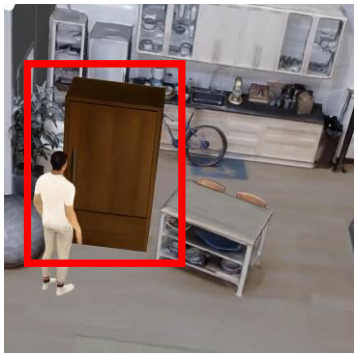

Interaction Data (3D)
HOI Model
Human Motion
Object Motion
"Lift the fridge"

Challenge



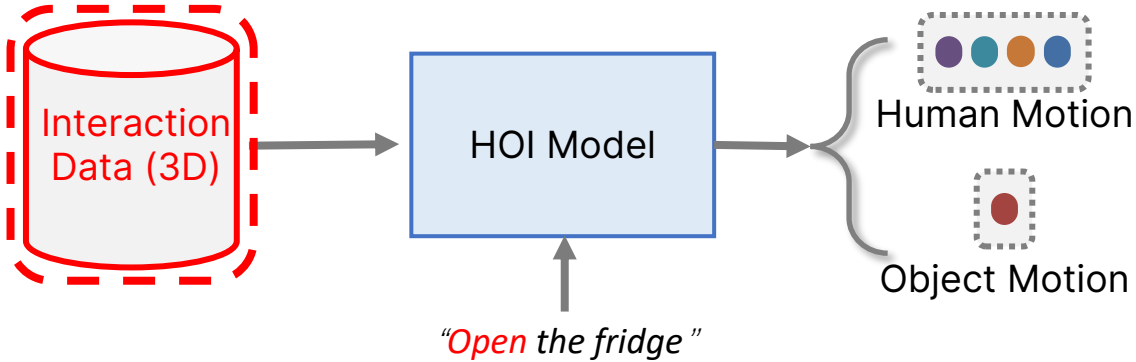
1. Lack of 3D Interaction Data
2. Limited to Rigid Objects
3. Human-object Contact Coupling





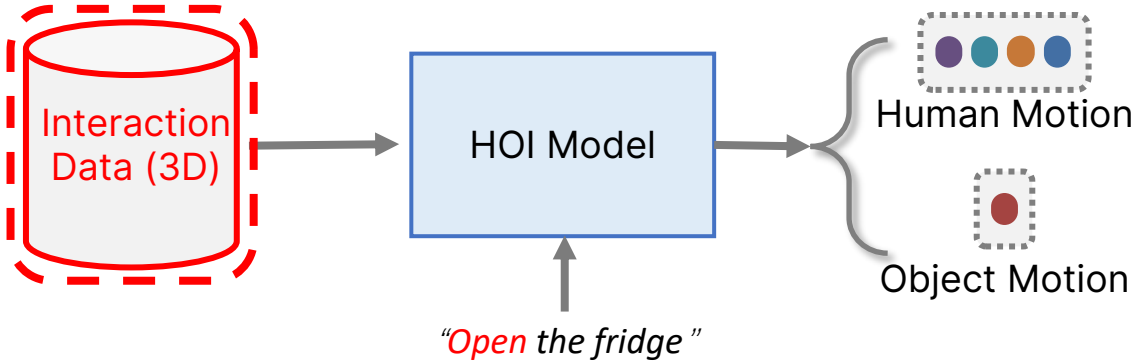


Key Idea



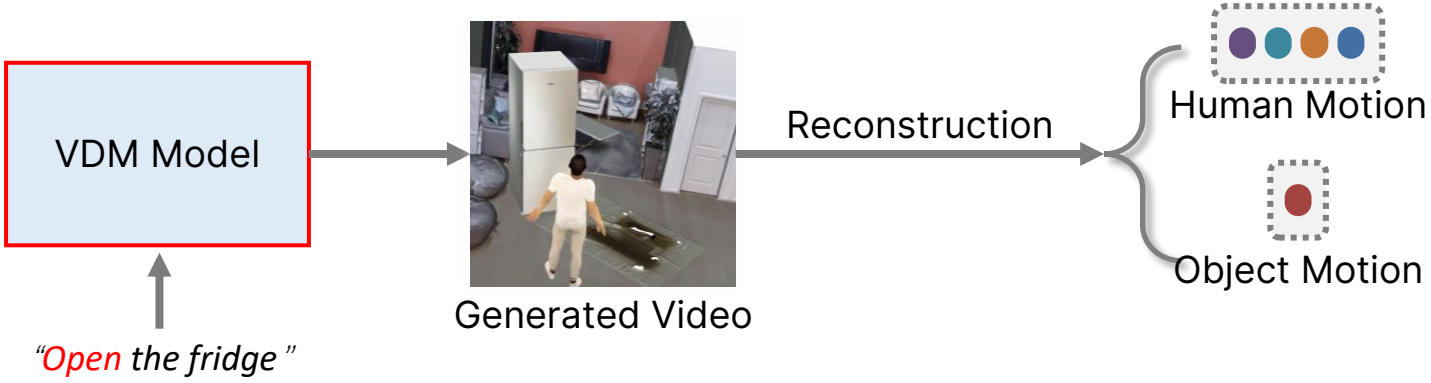
1. Lack of 3D Interaction Data
2. Limited to Rigid Objects
3. Human-object Contact Coupling

Key Idea



1. Lack of 3D Interaction Data
2. Limited to Rigid Objects
3. Human-object Contact Coupling

Key Idea

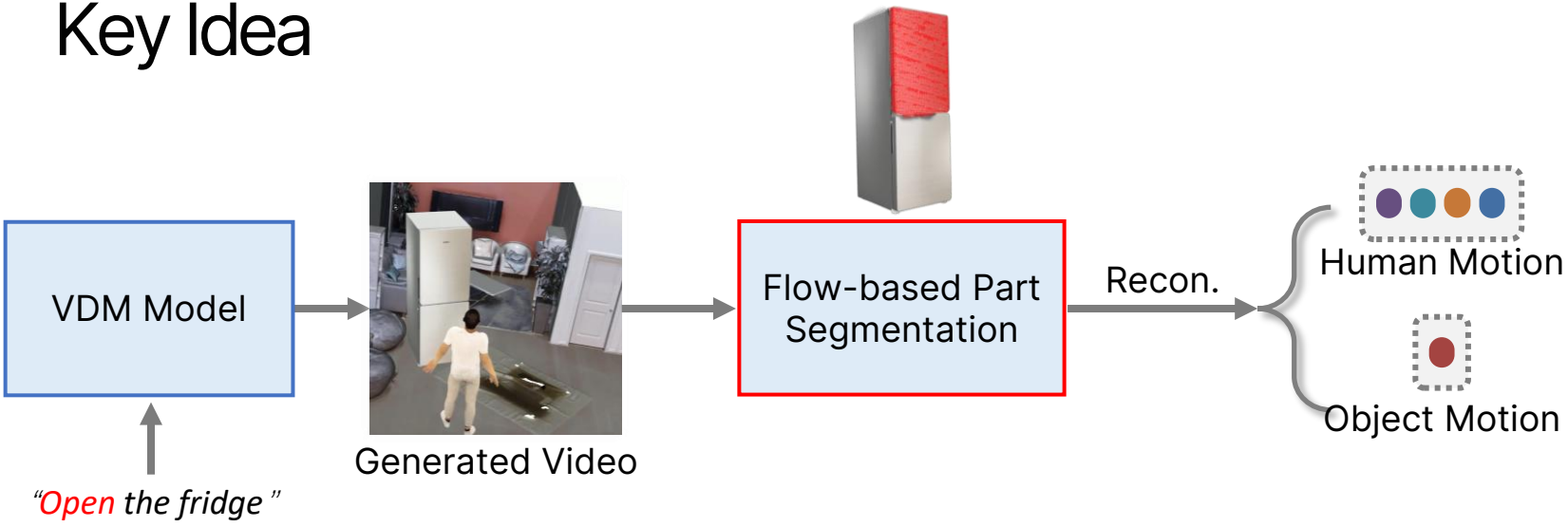


~~1. Lack of 3D Interaction Data~~ → Video Prior

2. Limited to Rigid Objects

3. Human-object Contact Coupling

Key Idea



~~1. Lack of 3D Interaction Data~~

~~2. Limited to Rigid Objects~~ → Flow-based Part Segmentation

3. Human-object Contact Coupling

Key Idea



VDM Model



Generated Video

Flow-based Part Segmentation

Recon.

Two-stage

Contact Condition

Object Motion

Human Motion

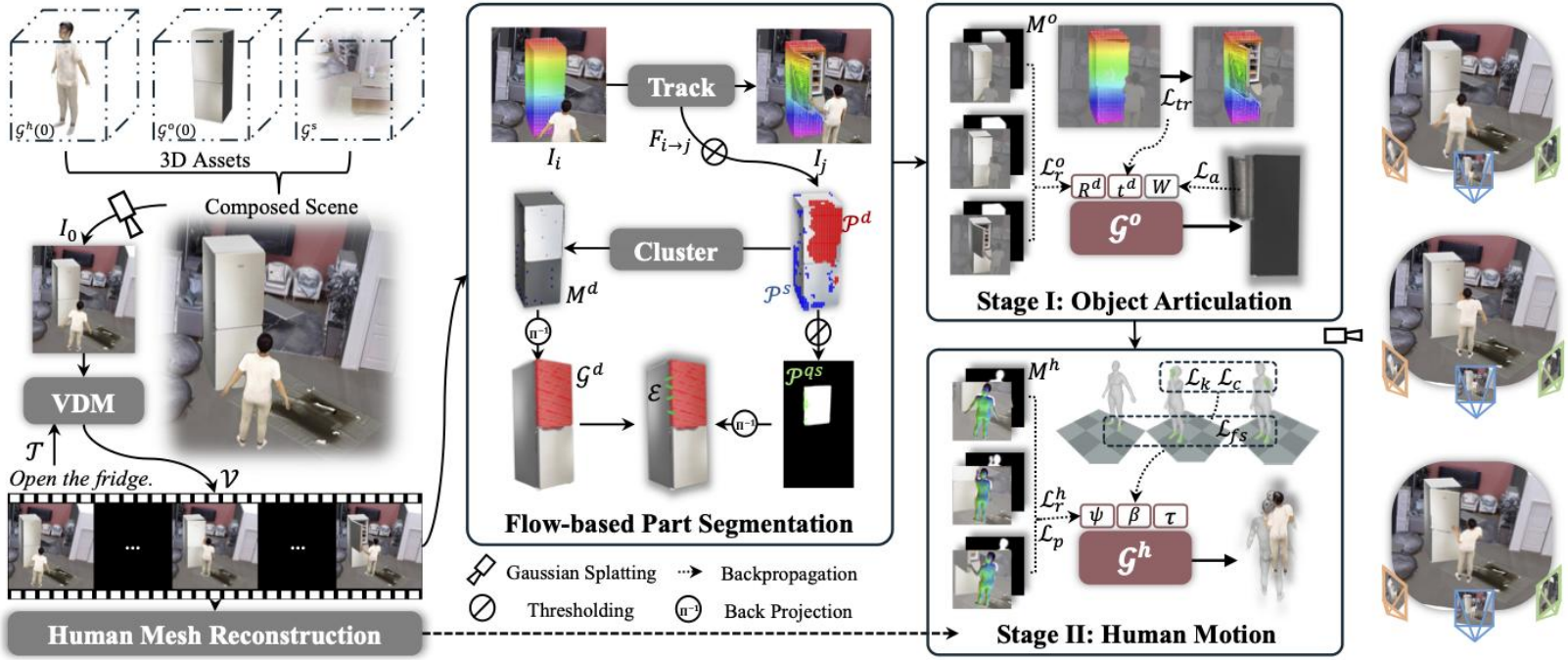
"Open the fridge"

~~1. Lack of 3D Interaction Data~~

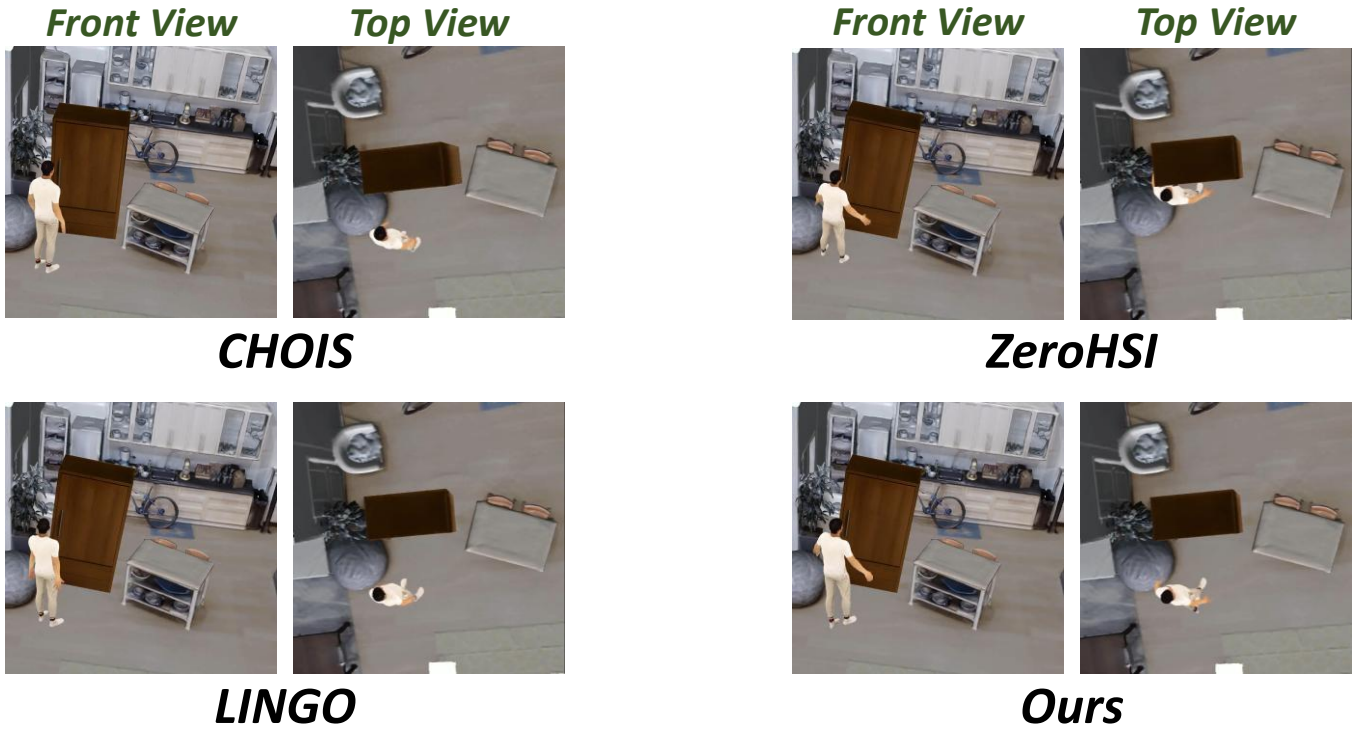
~~2. Limited to Rigid Objects~~

3. Human-object Contact Coupling → Two-stage Decoupled Reconstruction

Method

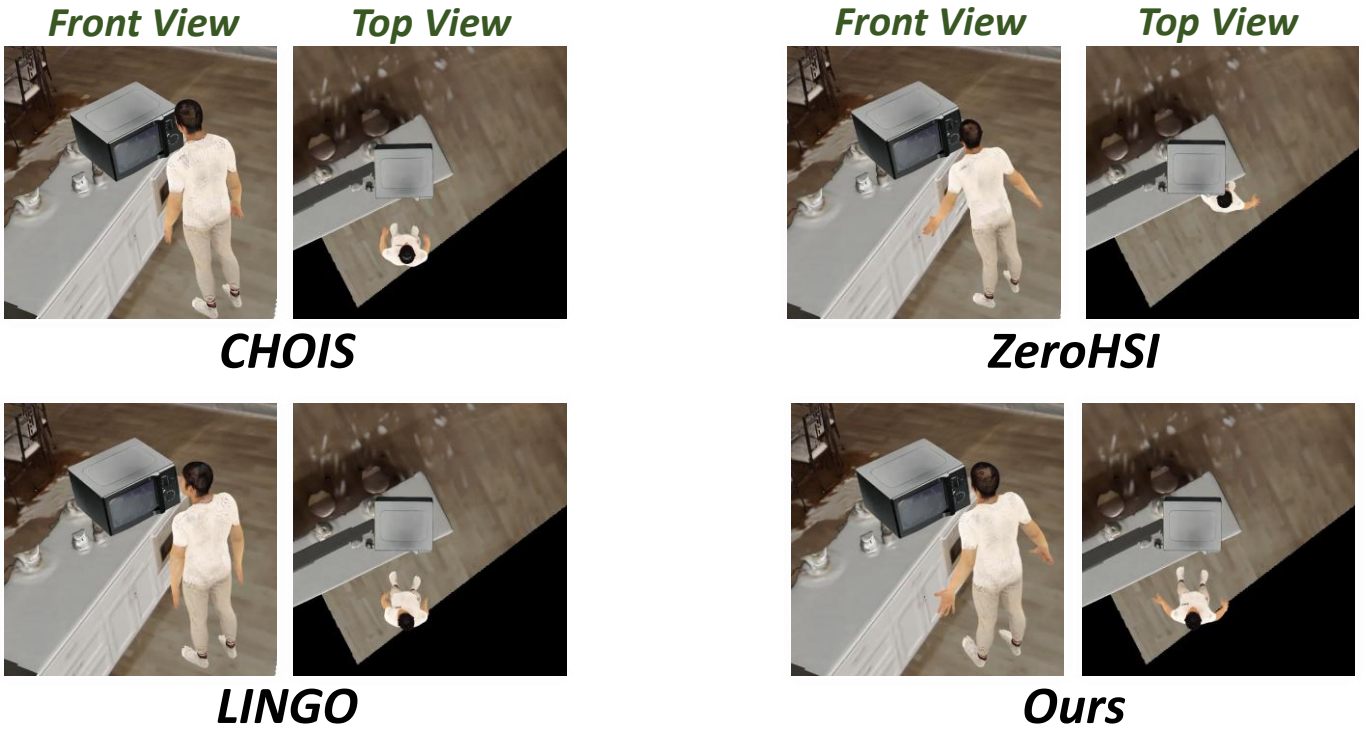

3D Assets
Composed Scene
VDM
Open the fridge.
Human Mesh Reconstruction
Track
Cluster
Flow-based Part Segmentation
Stage I: Object Articulation
Stage II: Human Motion
Gaussian Splatting
Backpropagation
Thresholding
Back Projection

Results: Qualitative Comparisons



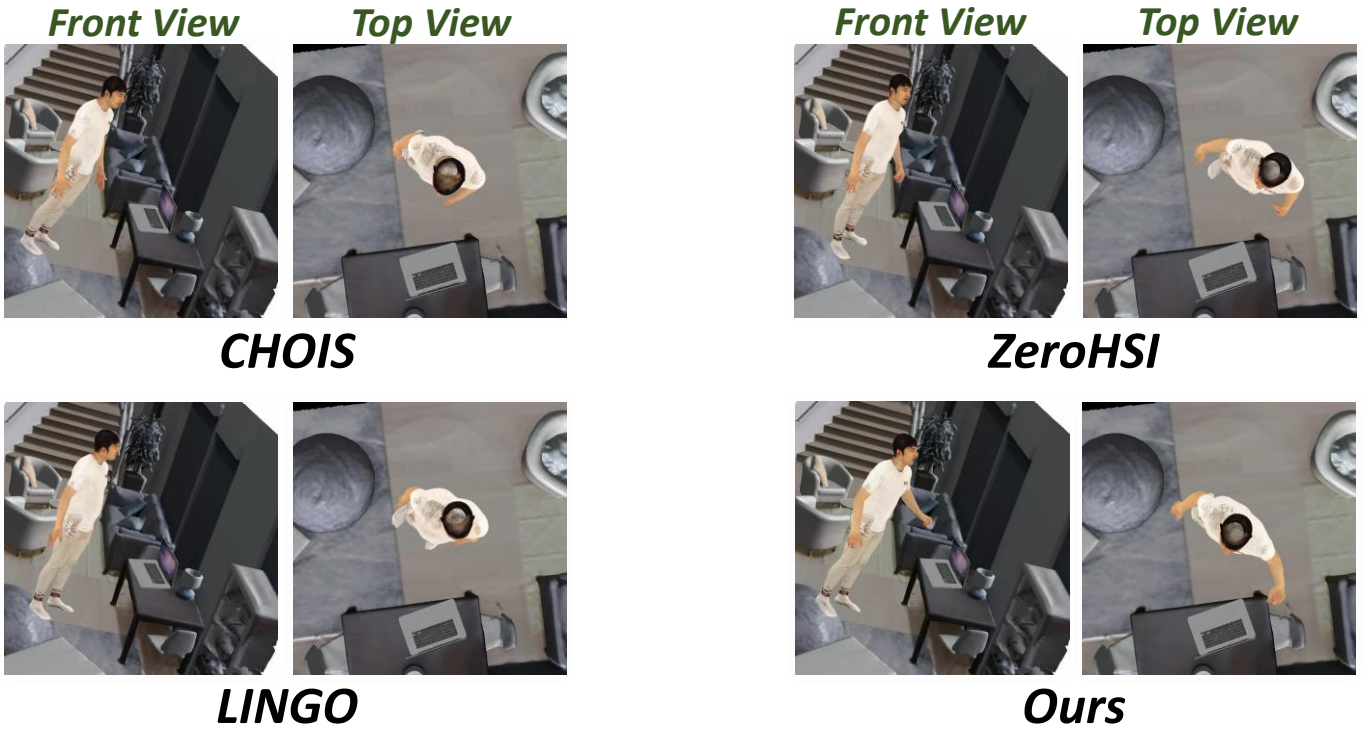
"Open the cabinet, look inside."

Results: Qualitative Comparisons



"Open the microwave"

Results: Qualitative Comparisons

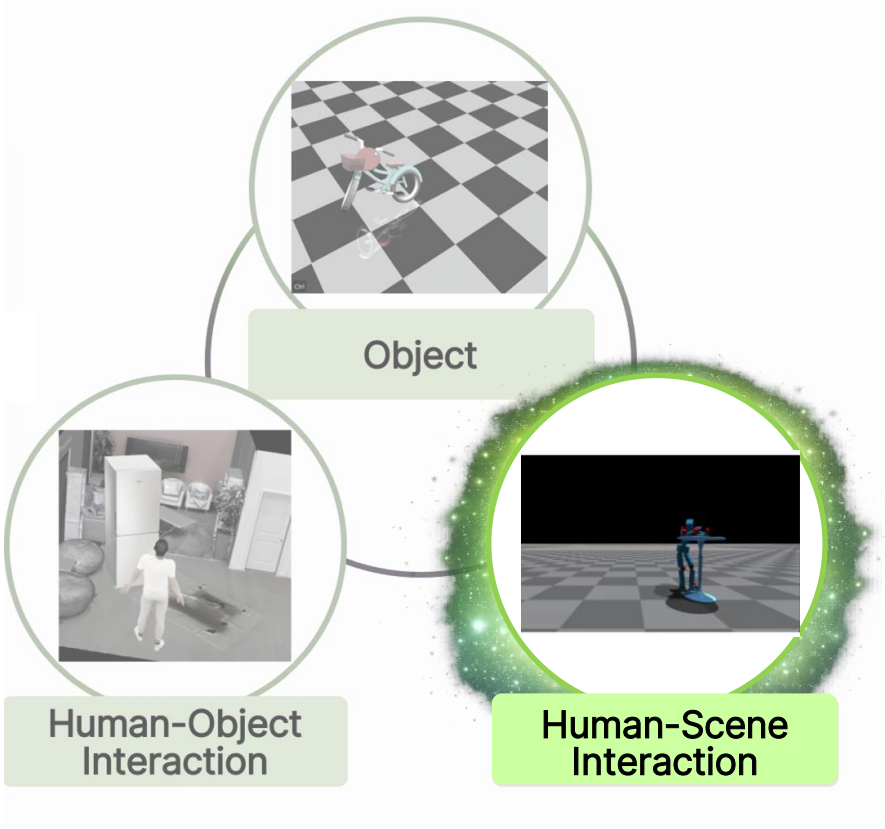


"Close the laptop screen gently."

From **Interactive Experience** to **Simulation-Ready** Modeling



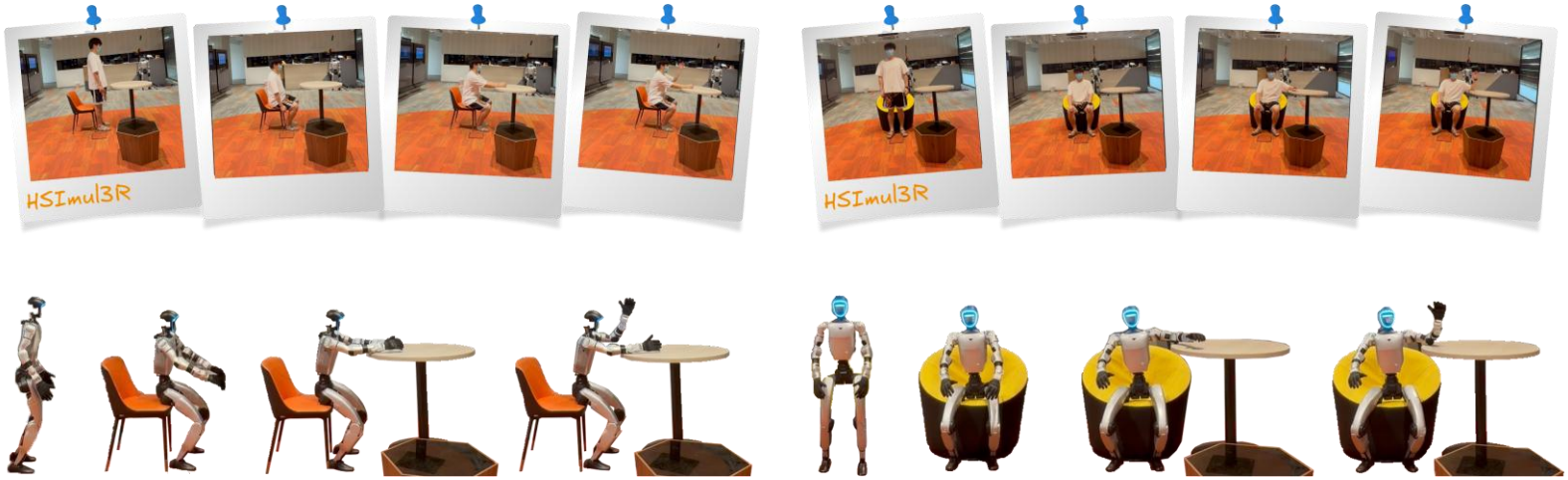
Xeperience-10M



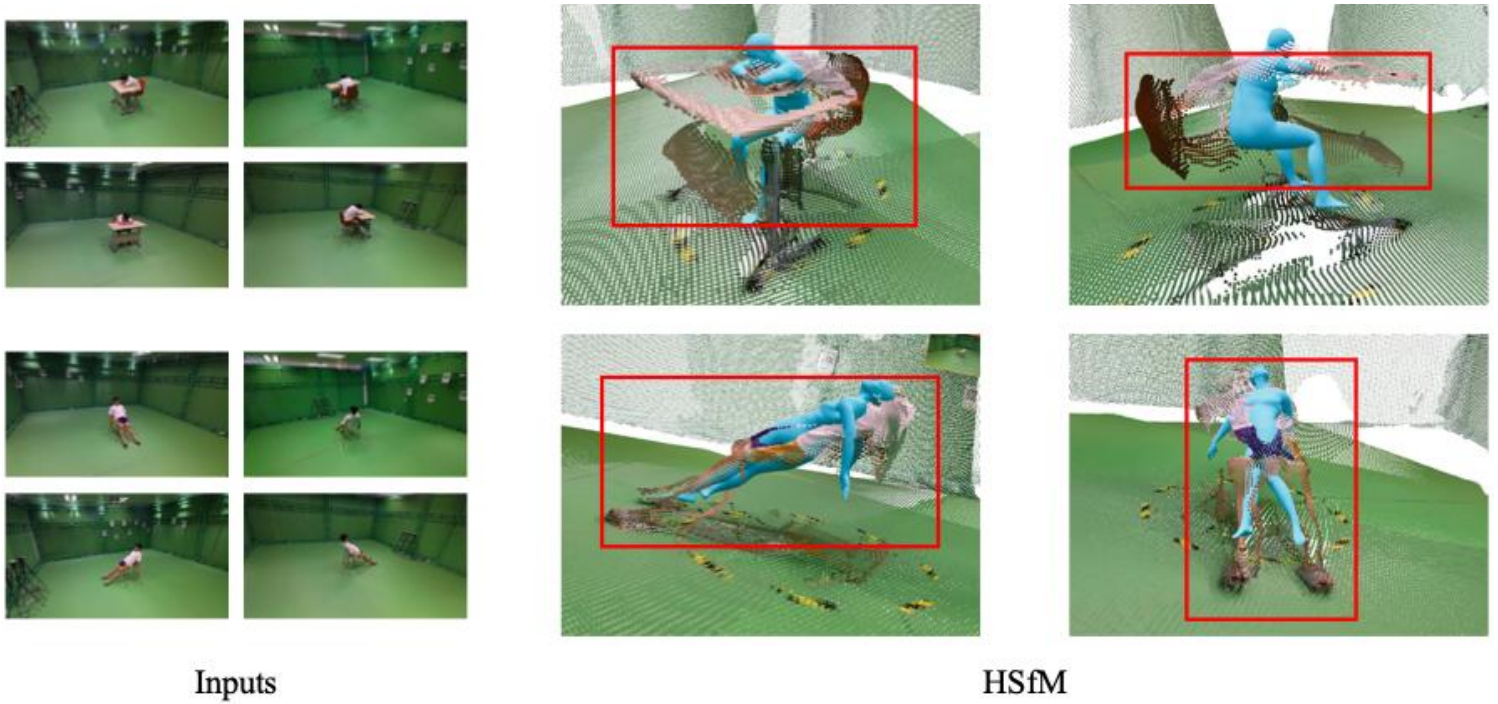
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

HSImul3R: Physics-in-the-Loop Reconstruction of **Simulation-Ready** Human–Scene Interactions

Yukang Cao, Haozhe Xie, Fangzhou Hong, Long Zhuo, Zhaoxi Chen, Liang Pan, Ziwei Liu

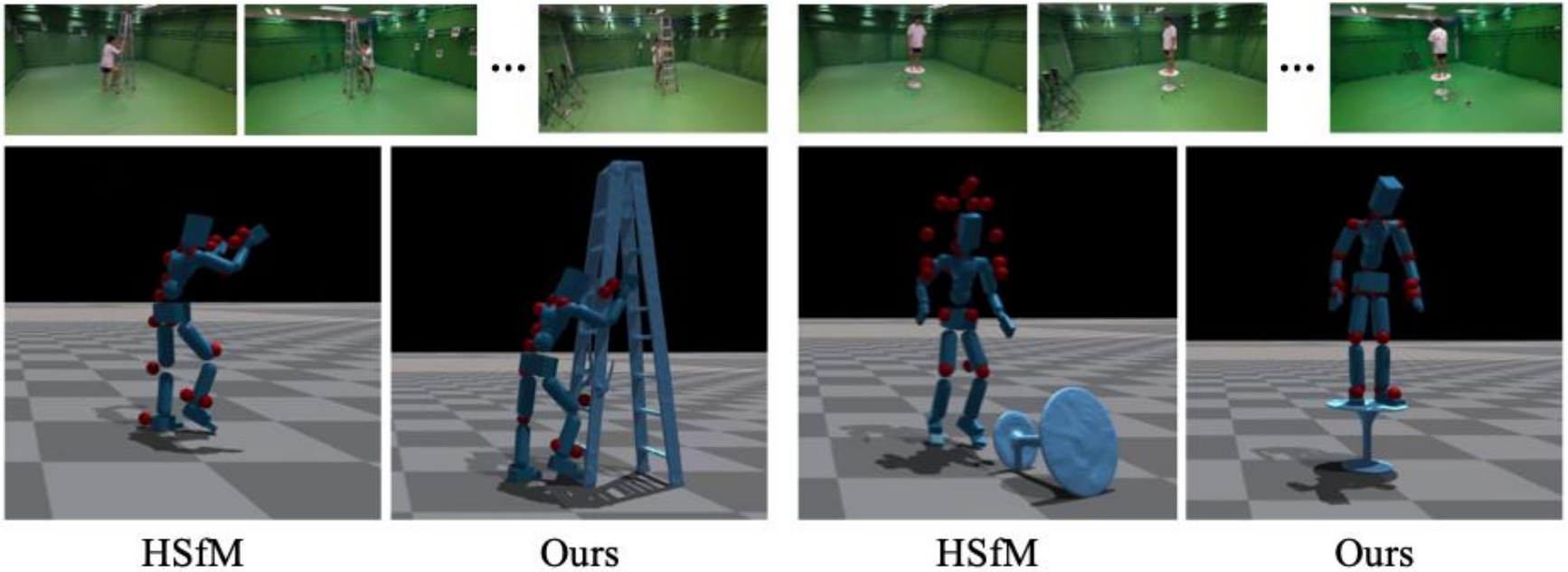


Challenge



Reconstructing People, Places, and Cameras. Lea Müller, Hongsuk Choi, et al. CVPR 2025

Challenge



Reconstructing People, Places, and Cameras. Lea Müller, Hongsuk Choi, et al. CVPR 2025

Key Idea: Leverage **Simulator Interaction Feedback** to Optimize Human Motion and Scene Geometry



Casual Capture



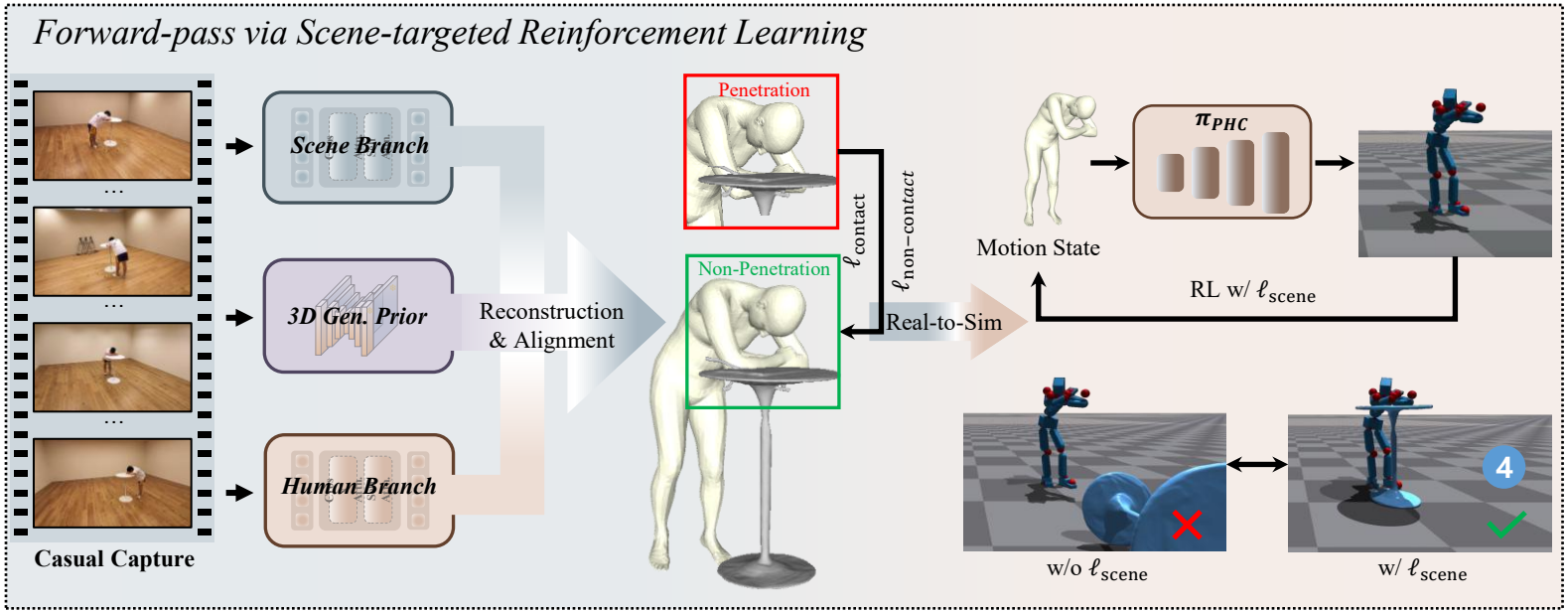


Optimize Human Motion

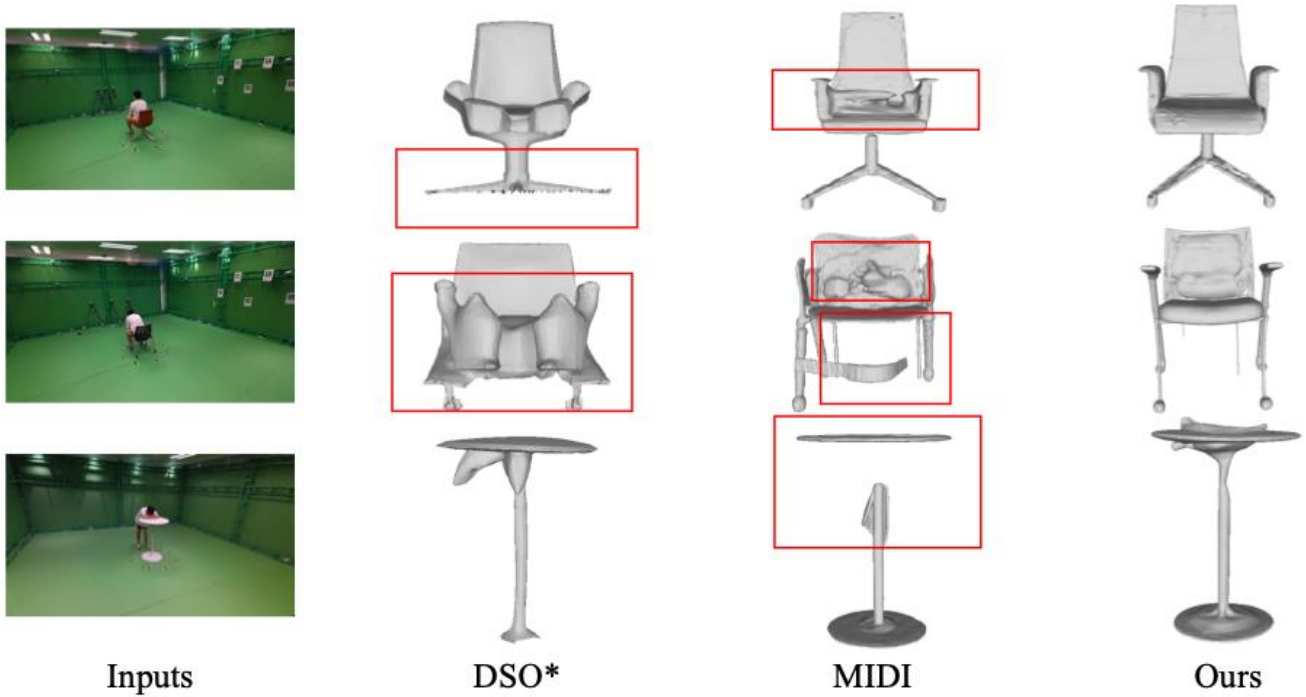


Optimize Scene Geometry

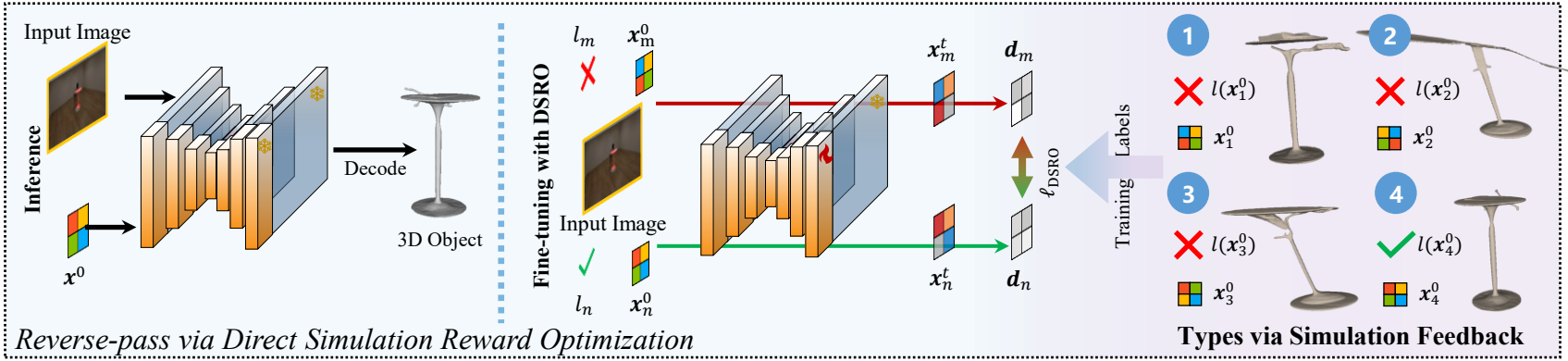
Stage I: Motion-Optimization via Scene-targeted Reinforcement Learning



Issue: Structural Suboptimality in the Reconstructed Scene



Stage II: Scene-Optimization via Simulator Feedback



Dataset: HSIBench

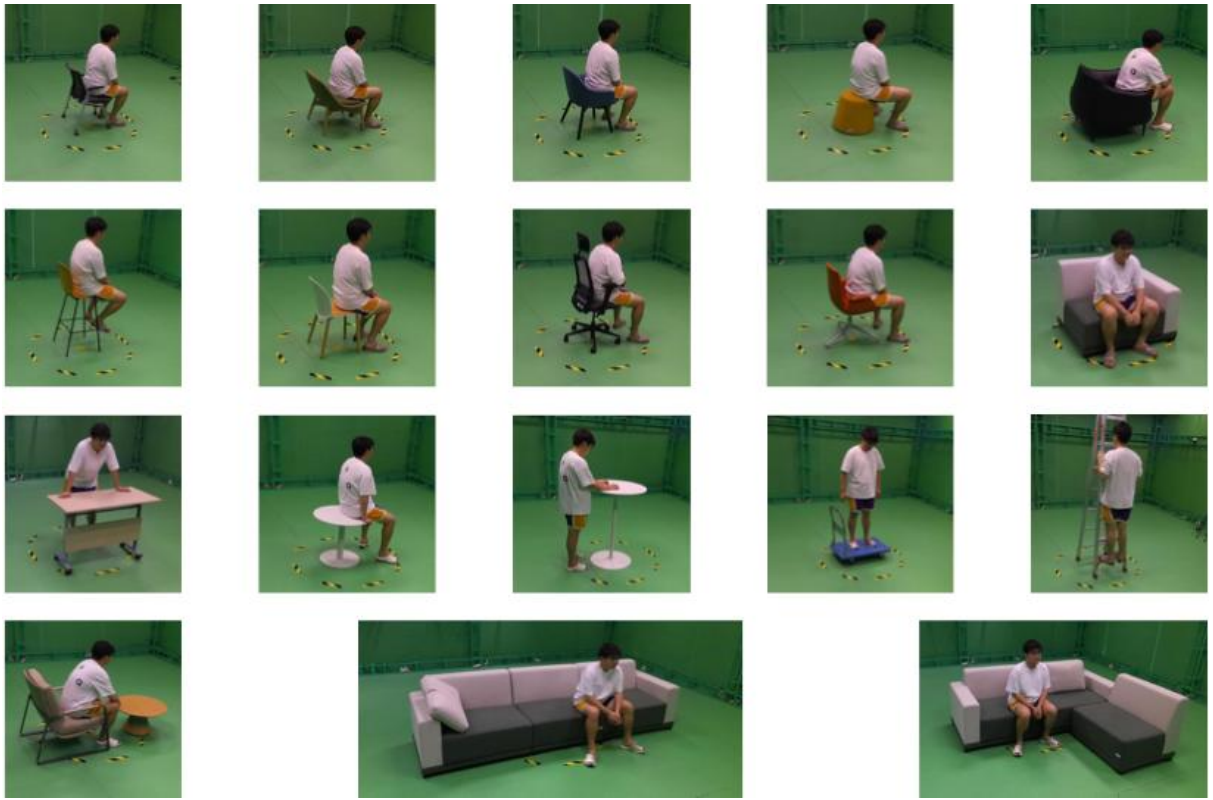
- a new dataset comprising 19 objects and over 50 motion sequences recorded by two male and one female participants, totaling 300 unique interaction instances.



Dataset: HSIBench



Dataset: HSIBench



Results: Quantitative Comparisons

Table 1: Quantitative comparison regarding human-scene-interaction reconstruction and simulation. We assess (1) post-simulation interaction stability, (2) human–scene penetration in the 3D reconstruction, and (3) changes in the human motions after simulation. Our method consistently outperforms the baseline method (HSfM) and different variants.

Method	Stability-HSI (%) ↑			Scene Penetration	Human Motion Quality	
	Easy	Medium	Hard	SP-3D (%) ↓	W-MPJPE ↓	PA-MPJPE ↓
HSfM [52]	10.52	4.50	2.66	69.51	5.02	2.79
V1	13.96	8.81	4.17	77.12	6.18	3.20
V2	39.56	22.71	7.05	-	4.91	2.71
V3	42.57	23.84	10.18	-	4.60	2.42
V4	29.56	16.62	5.17	-	4.57	2.39
Ours	**53.68**	**30.56**	**13.92**	**22.9**	**4.09**	**2.17**

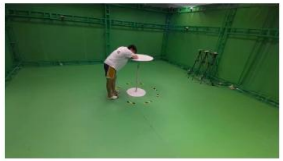
Results: Quantitative Comparisons

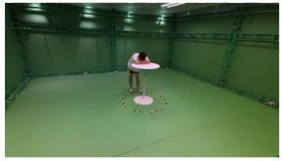
Table 2: Quantitative comparison regarding image-to-3D generation quality. Stability-HSI' denotes the ratio of simulations in which the human–scene interaction remains stable, while "Stability-Gravity"(SG) refers to placing the object alone in the simulator and evaluating whether it can stand stably under gravity. Note that we fine-tune DSO on pre-trained MIDI model for fair comparison.

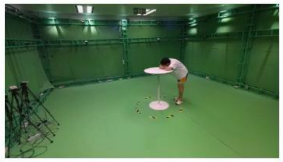
Method	Stability-HSI (%) ↑			SG (%) ↑	CD ↓	F-Score ↑
	Easy	Medium	Hard			
MIDI [23]	29.56	16.62	5.17	79.19	0.198	81.95
DSO* [31]	38.75	25.91	7.88	87.23	0.191	86.26
Ours	**53.68**	**30.56**	**13.92**	**91.50**	**0.173**	**88.25**

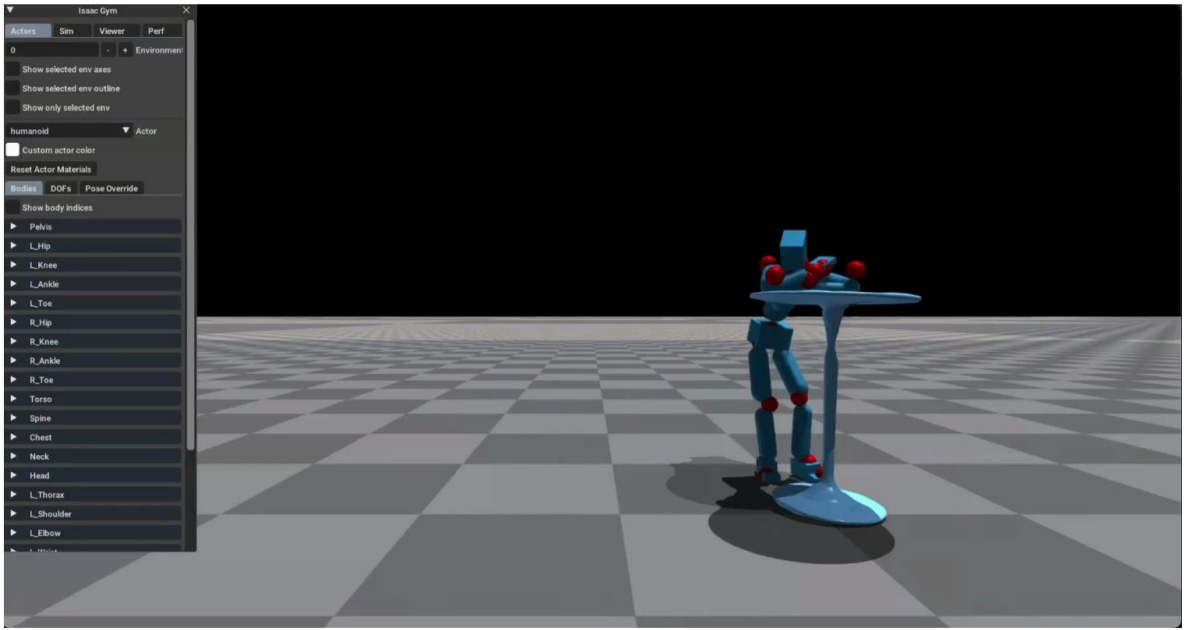
Results: Real-to-Sim



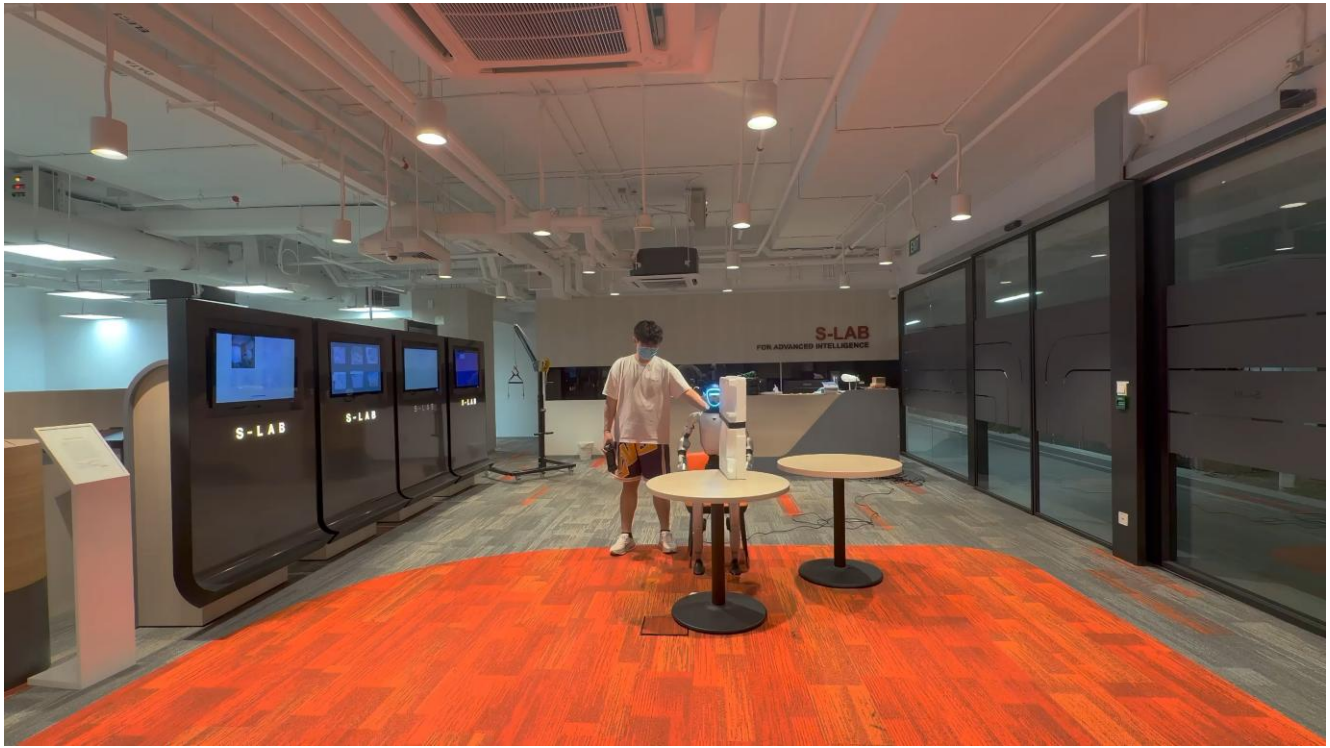








Results: Sim-to-Real



We are moving from a world that is seen,
to a world that is experienced

