UC SANTA BARBARA

TOYOTA

ICLR
International Conference On Learning Representations
2025

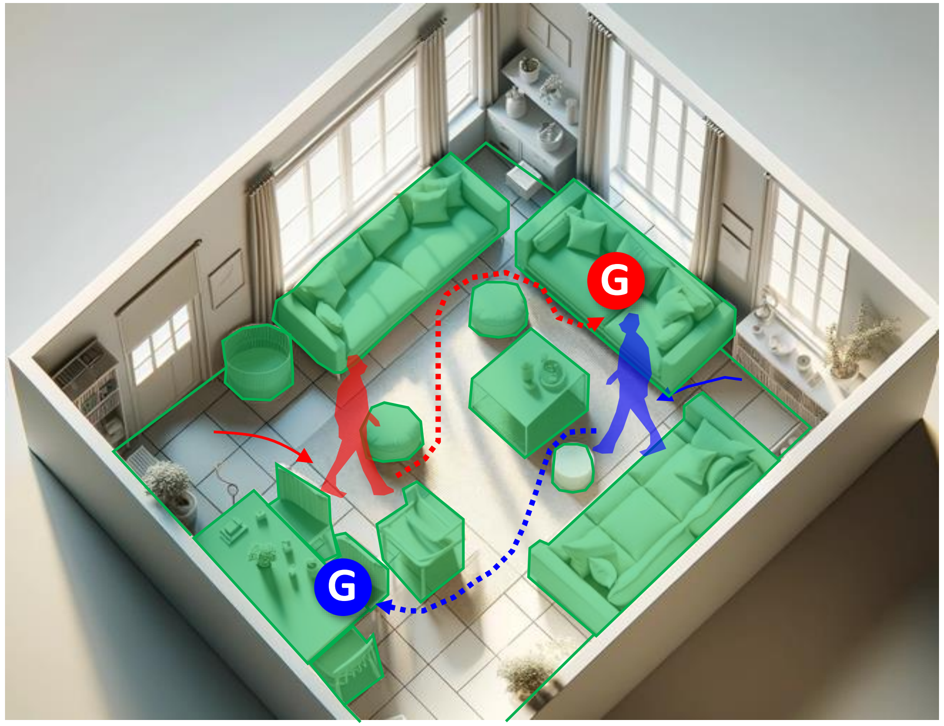
LocoVR: Multiuser Indoor Locomotion Dataset in Virtual Reality

Kojiro Takeyama[1,2], Yimeng Liu[1], Misha Sra[1]

1: University of California Santa Barbara,
2: Toyota Motor North America

1. Background

▪ Modeling multi-person trajectories in complex indoor environments is essential for various tasks. (e.g. path planning and motion prediction)



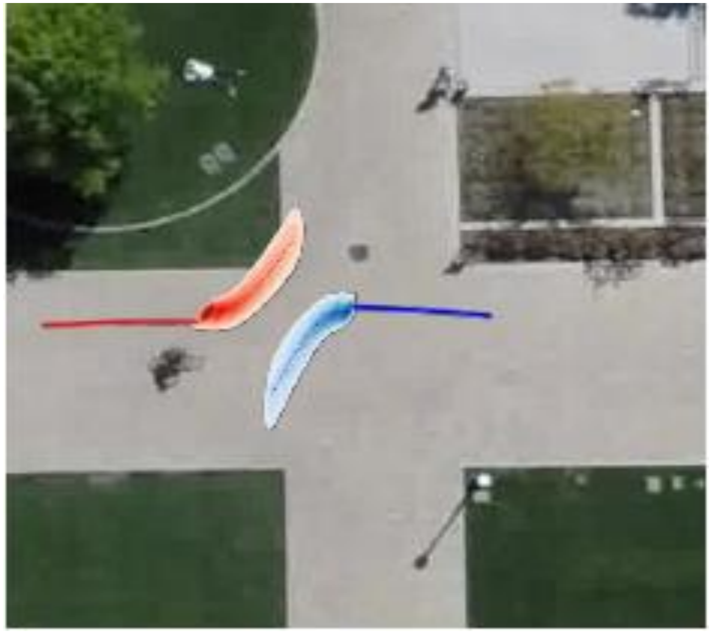
- Geometric constraints
- Social constraints

Our goal is to create a comprehensive dataset capturing multi-person trajectories across a variety of indoor environments

1. Background

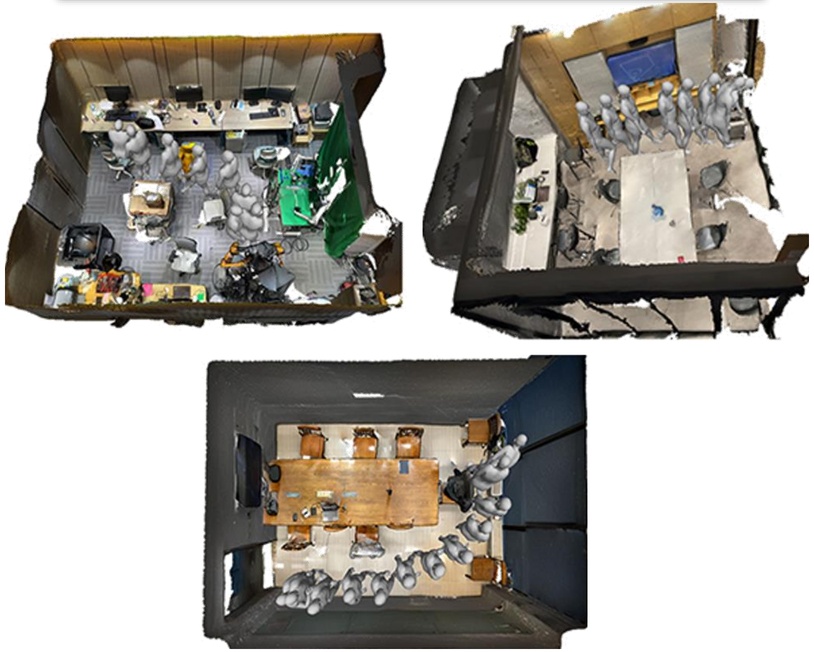
Limitations in existing datasets:

Outdoor trajectory datasets



▪ Scalability and complexity of the scene are different from indoor scenes

Indoor motion datasets



▪ Data collection efficiency is low (each scene requires system setup, scene measurement, and human-subject experiments)

2. Our approach

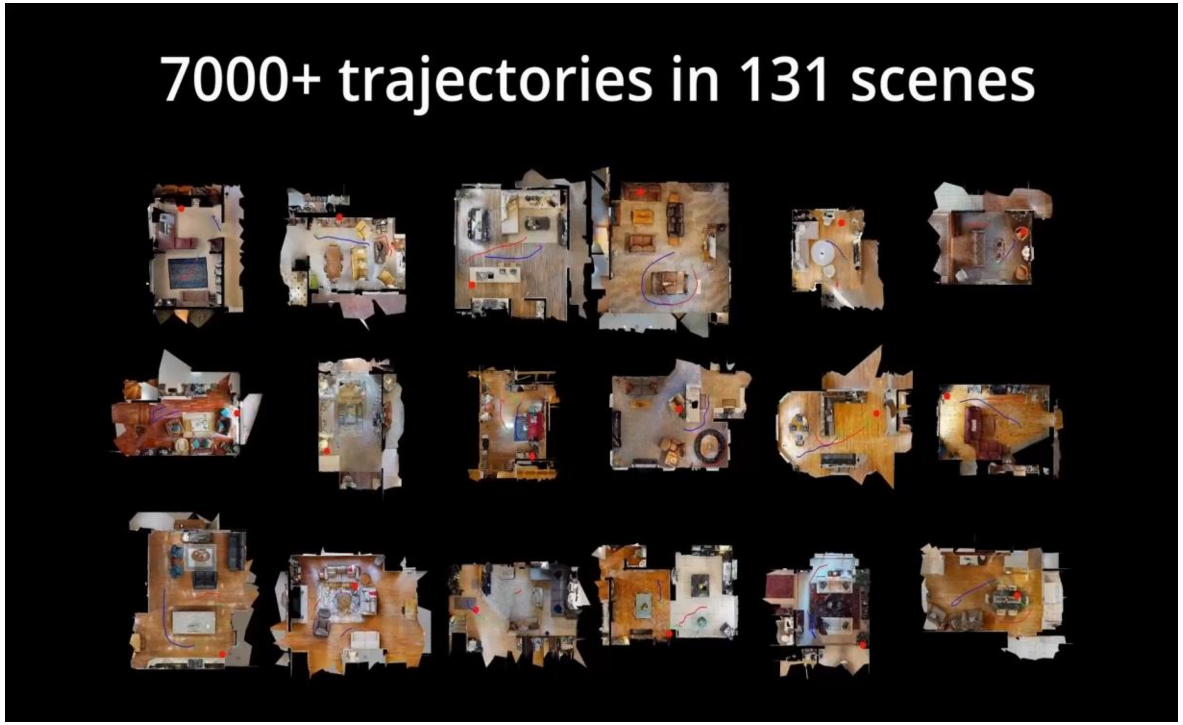
Data collection using VR:

- Efficient data collection
- Accurate and rich scene information

3. LocoVR dataset

Overview:

- Over 7000 multi-person trajectories in 131 complex indoor home environments
- Rich and accurate scene geometries and accurate body motions including head pose
- Social motion behaviors

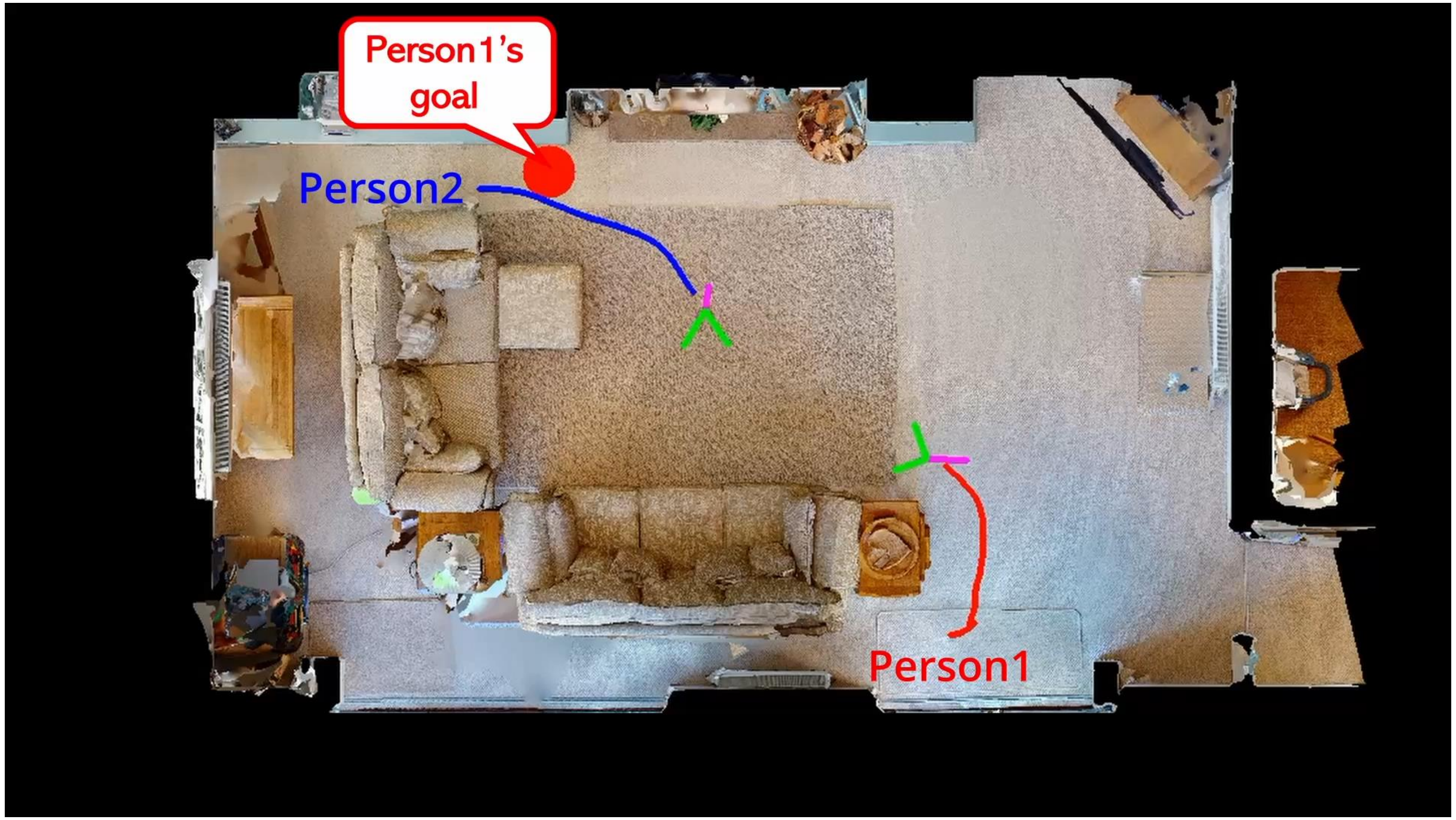


3. LocoVR dataset

ICLR 2025

Scene examples:

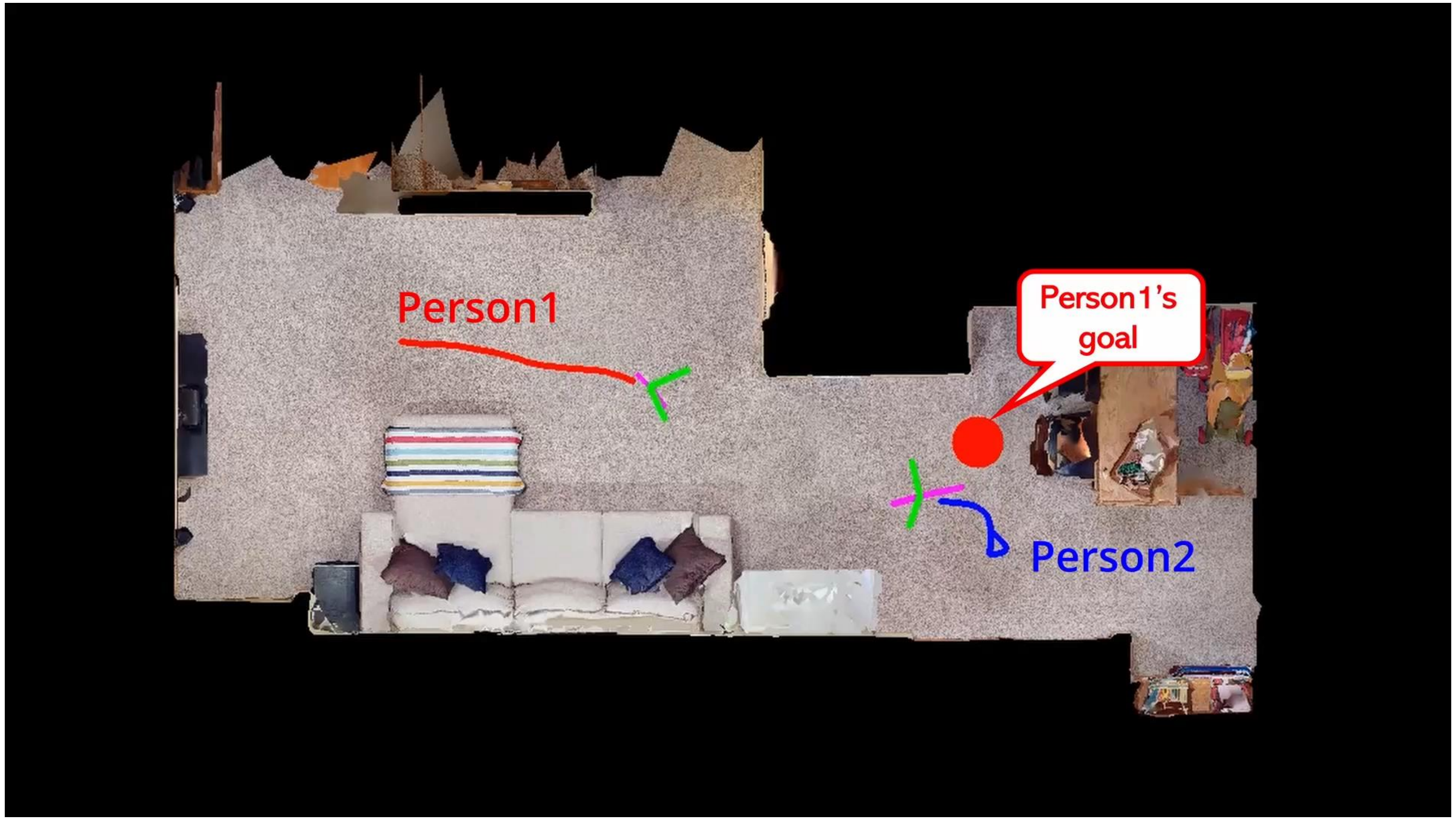
Maintaining a social distance



3. LocoVR dataset

Scene examples:

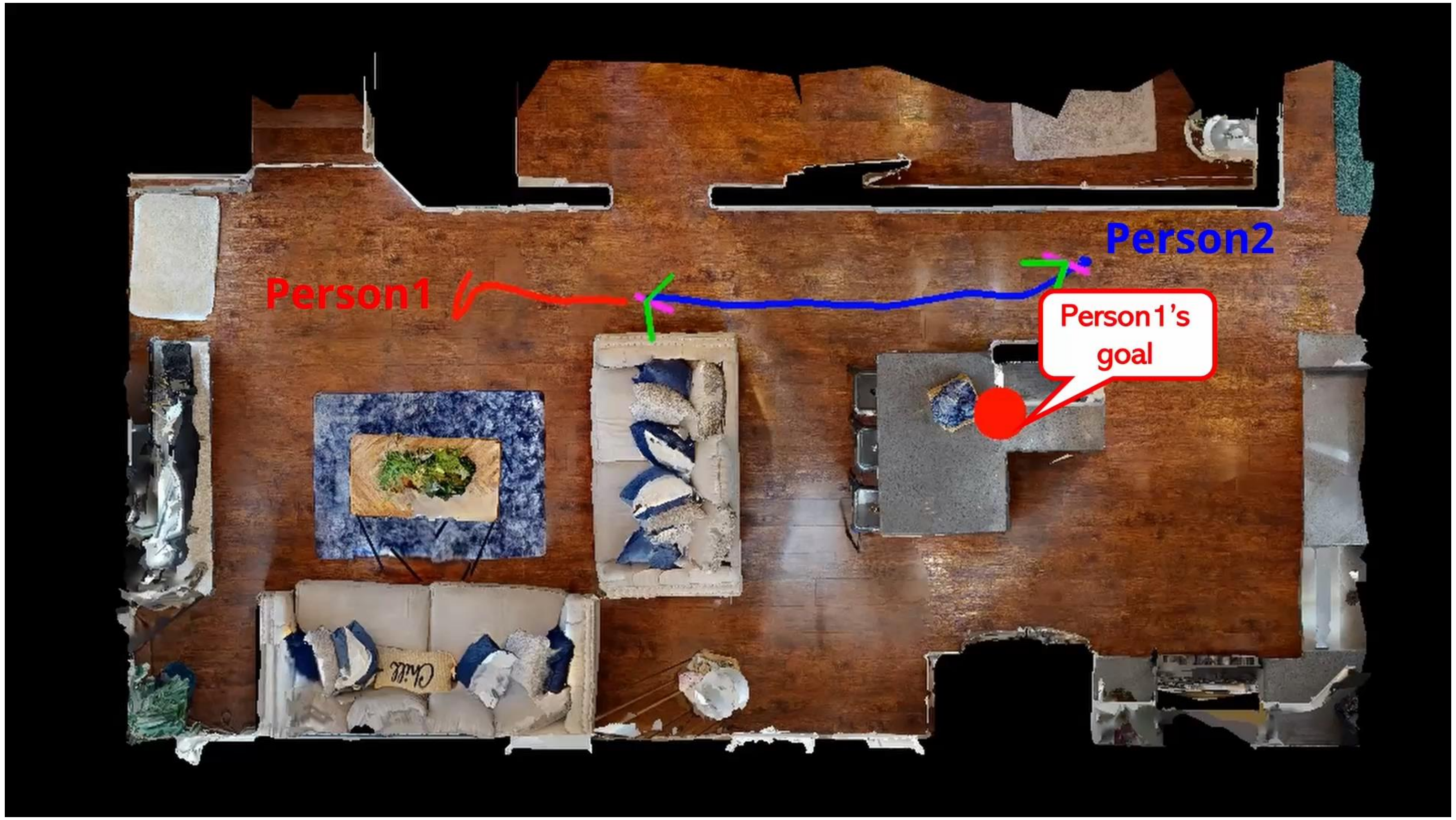
Yielding the path to allow others to pass



3. LocoVR dataset

Scene examples:

Taking a detour to avoid others



3. LocoVR dataset

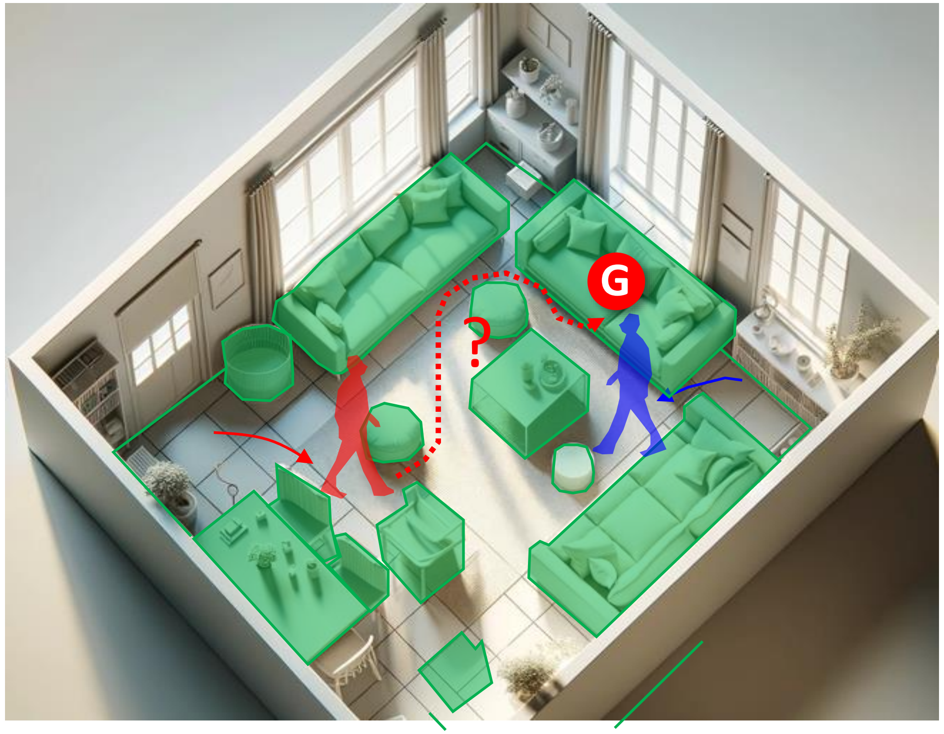
Dataset benchmark:

Dataset	*Frame*	*Scene*			*Subject*			
		Count	*Geometry*	*Location*	*Pos/Pose*	*Multi*	*Motion**	*Target – action*
HPS (Guzov et al., 2021)	300K	8	✓(3D mesh)	Out/Indoor	3D	✓	Real	Daily actions
EgoBody (Zhang et al., 2022)	153K	15	✓(3D mesh)	Indoor	3D	✓	Real	Daily actions
PROX (Hassan et al., 2019)	100K	12	✓(3D mesh)	Indoor	3D		Real	Daily actions
GIMO (Zheng et al., 2022)	129K	19	✓(3D mesh)	Indoor	3D, Gaze		Real	Daily actions
Grand Station (Zhou et al., 2012)	50K	1	✓(Aerial image)	Outdoor	2D	✓	Real	Trajectory
SDD (Robicquet et al., 2016)	929K	6	✓(Aerial image)	Outdoor	2D	✓	Real	Trajectory
ETH (Ess et al., 2007)	50K	2	✓(Aerial image)	Outdoor	2D	✓	Real	Trajectory
THOR (Rudenko et al., 2020)	360K	3	✓(3D point cloud)	Indoor	2D	✓	Real	Trajectory
JRDB (Vendrow et al., 2023)	636K	30	✓(3D point cloud)	Out/Indoor	3D	✓	Real	Trajectory
GTA-IM (Cao et al., 2020)	1000K	10	✓(3D mesh)	Indoor	3D		Synthetic	Trajectory
HUMANISE (Wang et al., 2022)	1200K	643	✓(3D mesh)	Indoor	3D		Synthetic	Daily actions
CIRCLE (Araújo et al., 2023)	4300K	9	✓(3D mesh)	Indoor	3D		Real	Daily actions
THOR-MAGNI (Schreiter et al., 2024)	1260K	4	✓(3D mesh)	Indoor	3D, Gaze	✓	Real	Trajectory, Daily actions
LocoVR (Ours)	2500K	131	✓(3D mesh)	Indoor	3D, Head	✓	Real	Trajectory, Social motion

4. Evaluation

ICLR International Conference On Learning Representations 2025

Task1: Global path prediction



Inputs
- Scene map
- Past trajectory
- Goal
- Another person's trajectory

Model
- Unet + A*
- Ynet

Output
- Future trajectory

Benchmarks

Training data	Test data
GIMO	LocoReal
THOR-MAGNI	
LocoVR	

4. Evaluation

Task1: Global path prediction

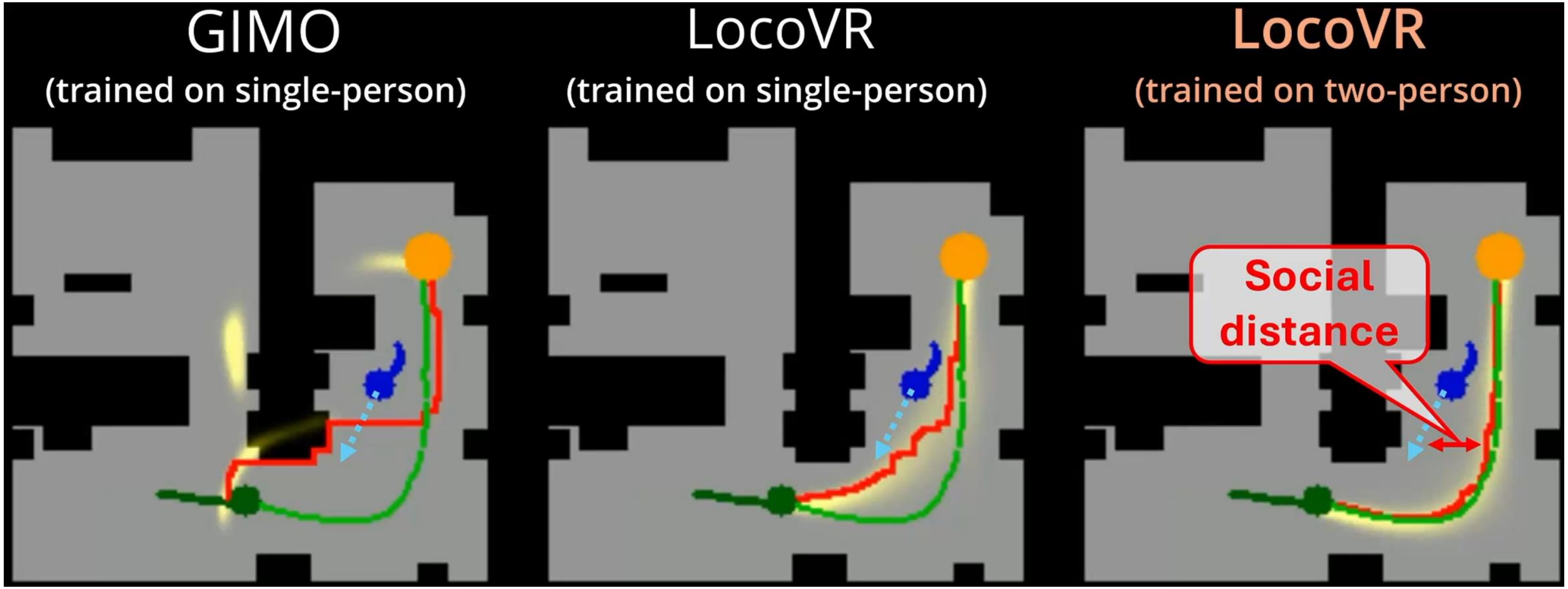
Method	*Mean*			*Max*			
	$0m \leq d \leq 3m$	$3m \leq d \leq 6m$	$6m \leq d$	$0m \leq d \leq 3m$	$3m \leq d \leq 6m$	$6m \leq d$	
Ynet (GIMO)	**0.08**$\pm$**0.003**	0.22$\pm$0.012	0.51$\pm$0.011	**0.17**$\pm$**0.003**	0.46$\pm$0.022	1.11$\pm$0.016	Baselines
Ynet (THOR-MAGNI)	0.10$\pm$0.003	0.30$\pm$0.006	0.65$\pm$0.014	0.19$\pm$0.004	0.56$\pm$0.008	1.29$\pm$0.023	
Ynet (LocoVR)	0.09$\pm$0.002	**0.18**$\pm$**0.004**	**0.42**$\pm$**0.050**	0.18$\pm$0.002	**0.37**$\pm$**0.005**	**0.92**$\pm$**0.089**	Ours
A* + MAP	0.10$\pm$0	0.27$\pm$0.000	0.40$\pm$0.000	0.22$\pm$0.000	0.58$\pm$0.000	0.89$\pm$0.000	Baselines
A* + DISTMAP	0.102$\pm$0	0.18$\pm$0.000	0.26$\pm$0.000	0.24$\pm$0.000	0.46$\pm$0.000	0.66$\pm$0.000	
A* + U-Net (GIMO)	0.09$\pm$0.002	0.23$\pm$0.006	0.36$\pm$0.011	0.20$\pm$0.004	0.53$\pm$0.013	0.84$\pm$0.024	
A* + U-Net (THOR-MAGNI)	0.07$\pm$0.001	0.21$\pm$0.007	0.30$\pm$0.005	0.17$\pm$0.001	0.45$\pm$0.014	0.71$\pm$0.015	
A* + U-Net (LocoVR)	**0.06**$\pm$**0.001**	**0.12**$\pm$**0.002**	**0.19**$\pm$**0.003**	**0.15**$\pm$**0.001**	**0.29**$\pm$**0.004**	**0.50**$\pm$**0.014**	Ours

4. Evaluation

ICLR
International Conference On Learning Representations
2025

Task1: Global path prediction

Sample1: Maintaining a social distance

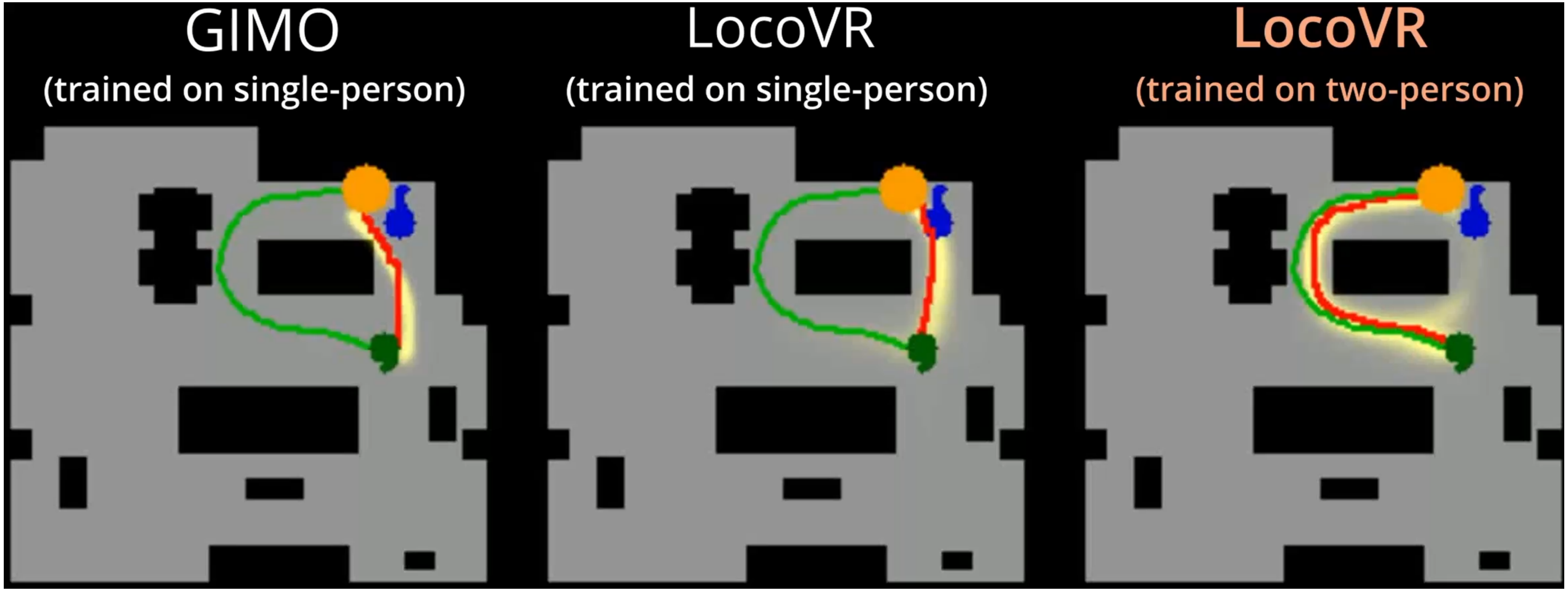


Past trajectory (p1)
Past trajectory (p2)
Predicted global path (p1)
Groundtruth global path (p1)
Predicted global path distribution
Walkable area
Non-walkable area (obstacle)

ICLR International Conference On Learning Representations 2025

Task1: Global path prediction

Sample2: Taking a detour to avoid others



5. Conclusion

- We introduced the LocoVR dataset to model geometrically and socially aware human trajectories, capturing accurate trajectories and detailed spatial information of two-person interactions across 131 indoor home environments.

- Experimental results on indoor tasks demonstrated that models trained on LocoVR significantly outperformed those trained on previous indoor datasets, highlighting its effectiveness in adapting to unseen indoor environments.

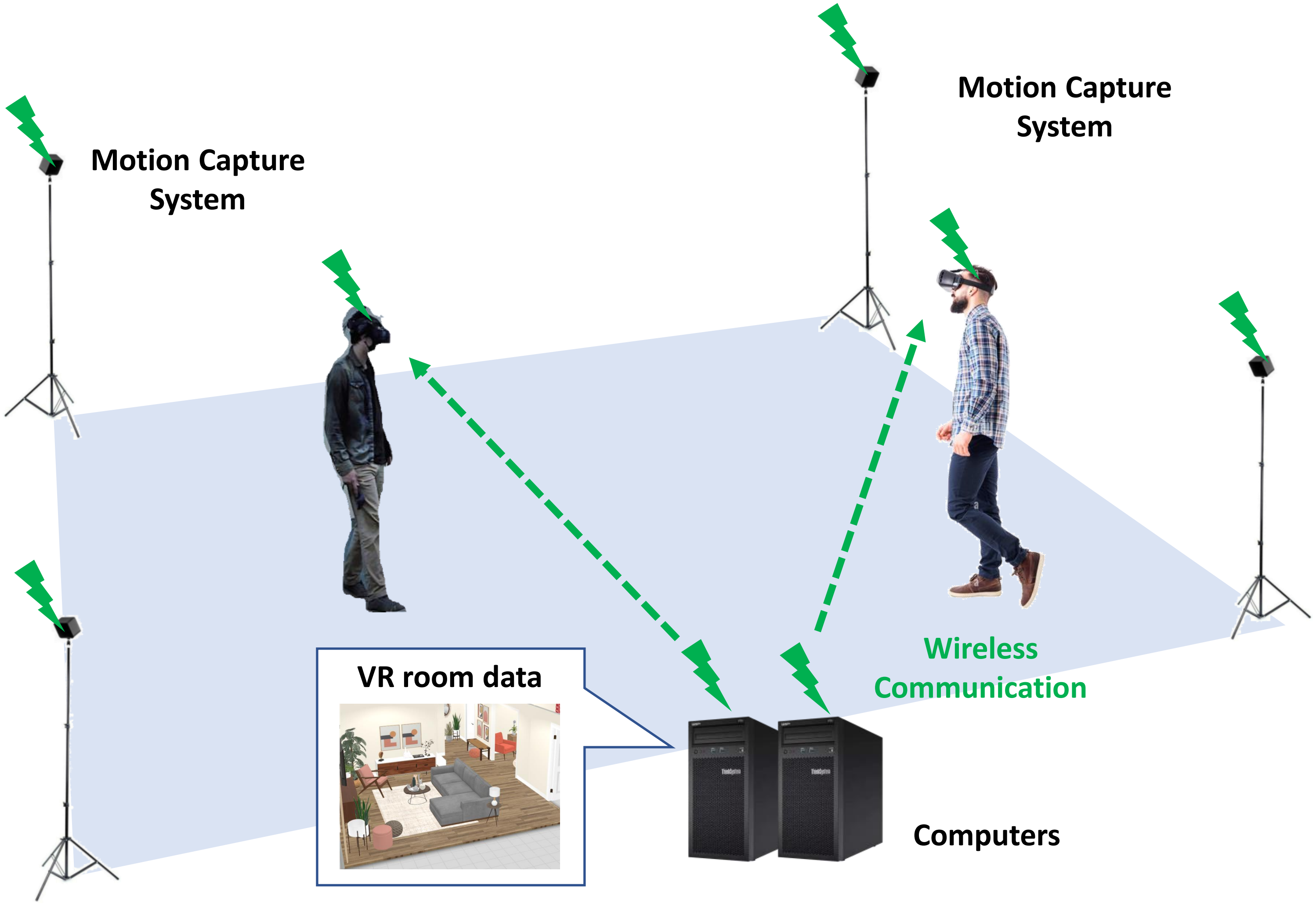
- These findings also showcase the potential of virtual environments for training generalizable models, establishing a benchmark for future indoor human motion and trajectory research.

Thank you for listening!

Visit our website for more information!

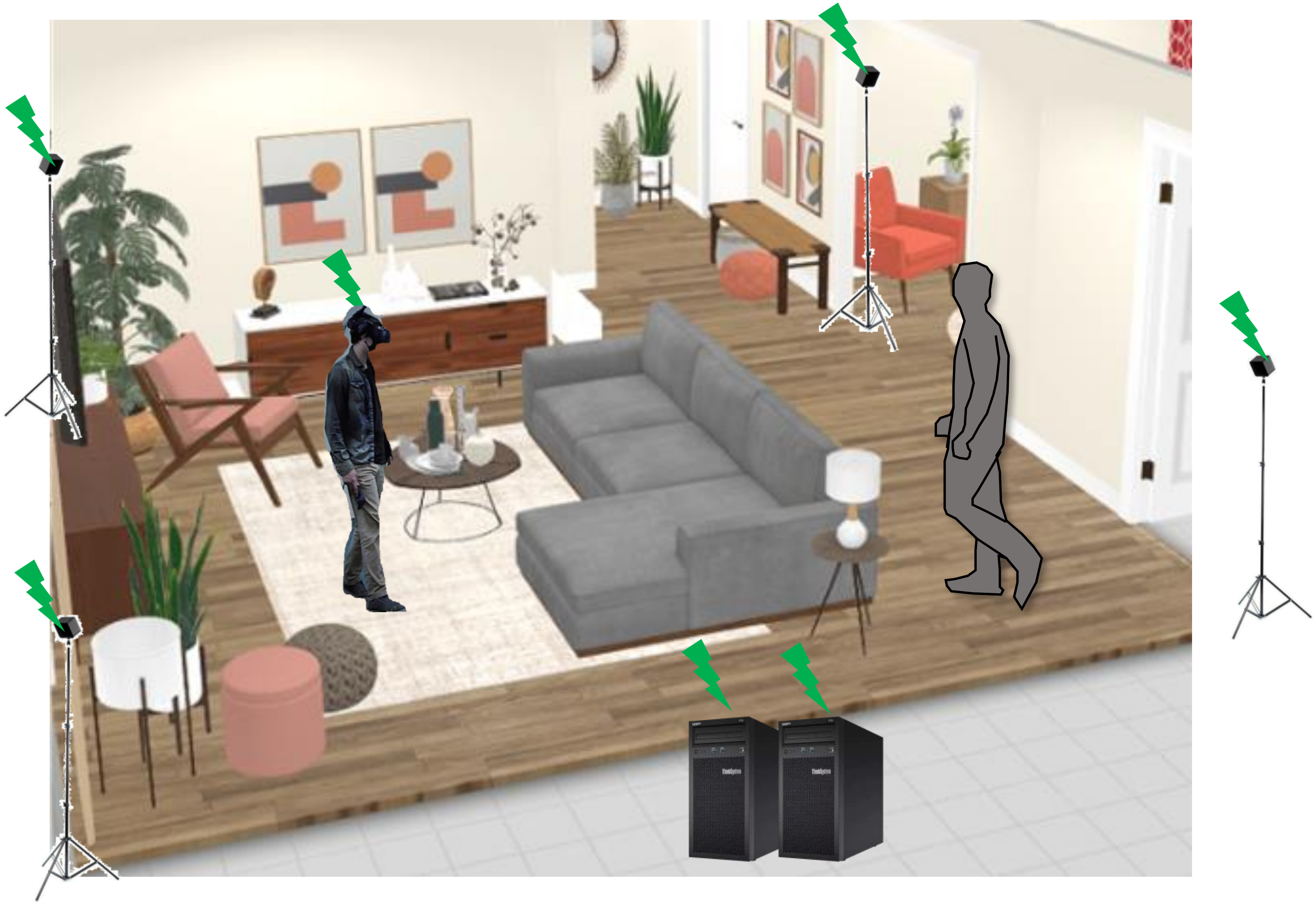
2. Our approach

System setup:



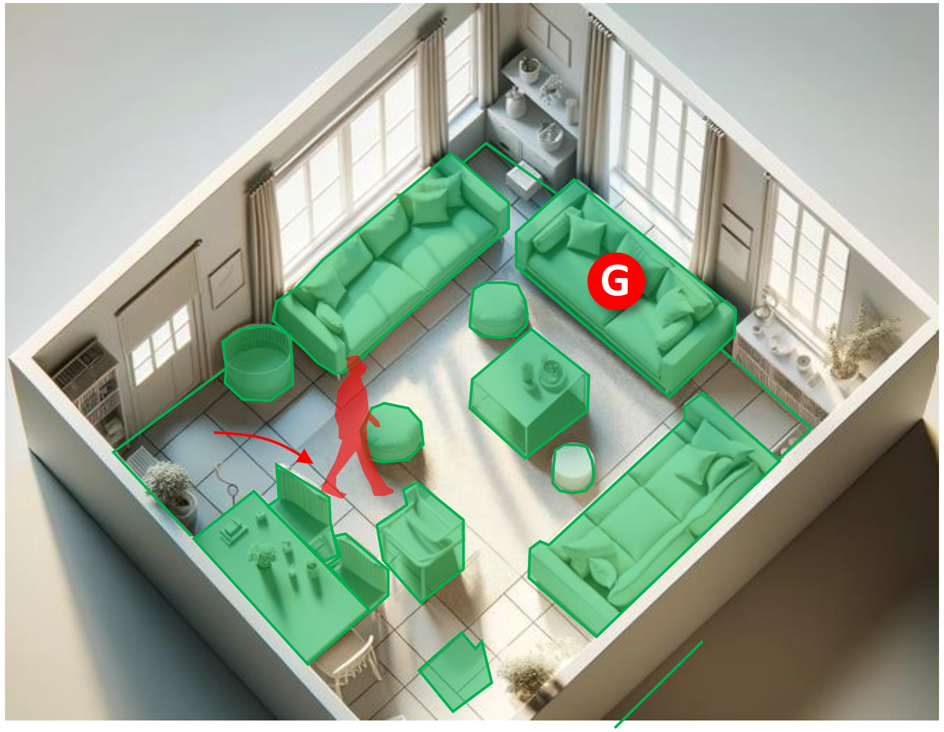
2. Our approach

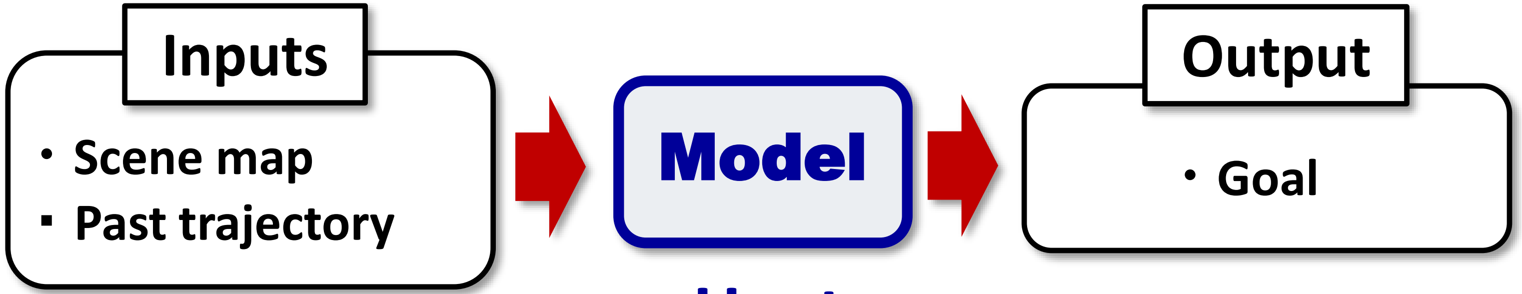
System setup:



4. Evaluation

Task2: Goal prediction





▪ Unet

Benchmarks

Training data	Test data
GIMO	LocoReal
THOR-MAGNI	
LocoVR	

4. Evaluation

Task2: Goal prediction

Method	*Goal position error*			*Object prediction accuracy*		
	$0m \le d \le 3m$	$3m \le d \le 6m$	$6m \le d$	$0m \le d \le 3m$	$3m \le d \le 6m$	$6m \le d$
RANDOM	3.70$\pm$0.02	3.75$\pm$0.02	3.76$\pm$0.03	15.5$\pm$1.0	16.1$\pm$0.5	**15.3$\pm$1.2**
NEAREST	1.76$\pm$0.00	3.89$\pm$0.00	4.73$\pm$0.00	42.7$\pm$0.0	0.5$\pm$0.0	0.0$\pm$0.0
U-Net (GIMO)	1.58$\pm$0.32	2.47$\pm$0.06	**3.35$\pm$0.23**	49.2$\pm$6.7	17.8$\pm$2.0	3.9$\pm$0.8
U-Net (THOR-MAGNI)	1.82$\pm$0.04	3.29$\pm$0.04	4.23$\pm$0.09	40.1$\pm$1.3	18.9$\pm$0.6	9.5$\pm$1.6
U-Net (LocoVR)	**0.83$\pm$0.03**	**1.89$\pm$0.02**	3.45$\pm$0.04	**72.2$\pm$2.6**	**40.1$\pm$2.0**	13.5$\pm$2.7

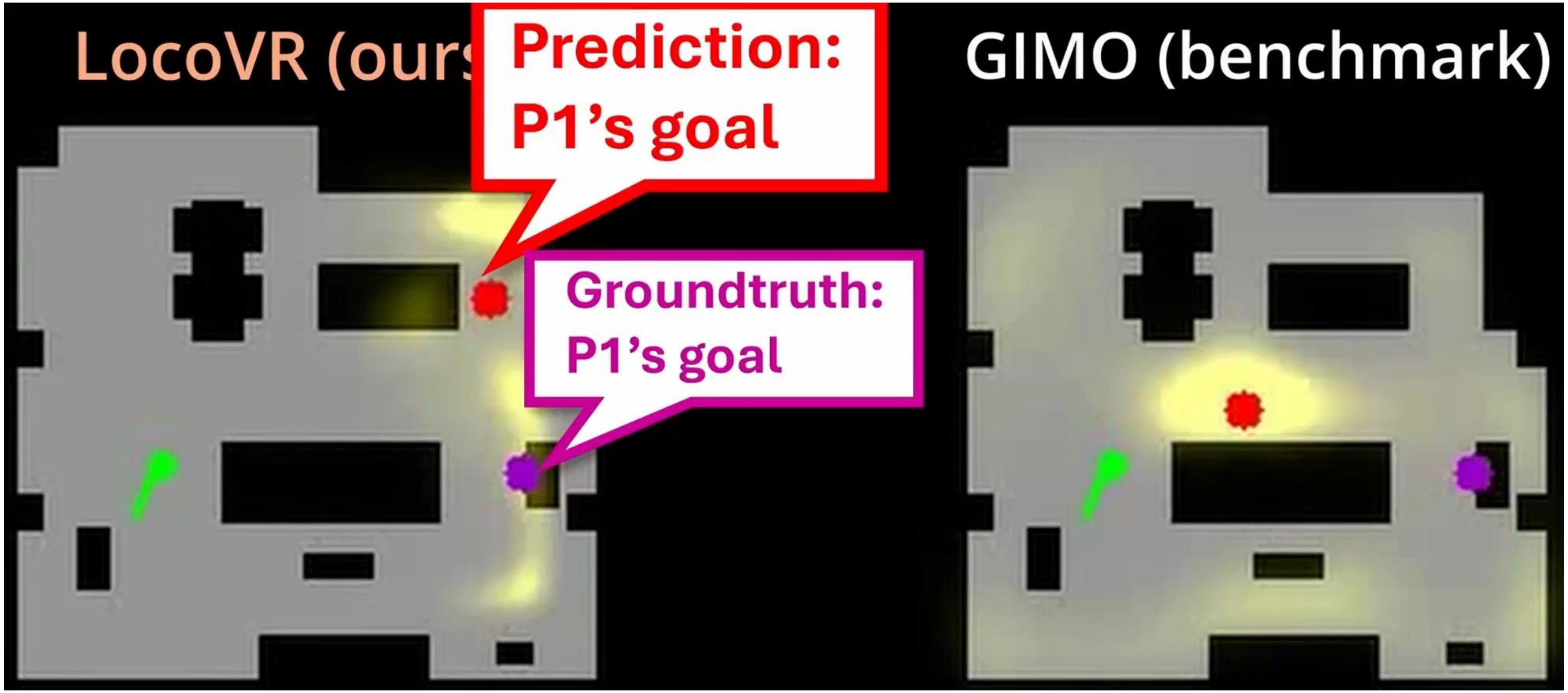
Baselines

Ours

4. Evaluation

ICLR International Conference On Learning Representations 2025

Task2: Goal prediction



Past trajectory
Predicted distribution of goal position
Walkable area
Groundtruth goal position
Predicted goal position
Non-walkable area (obstacle)