SiT Dataset:

Socially interactive Pedestrian Trajectory Dataset for Social Navigation Robots

Jong Wook Bae*, Jungho Kim*, Junyong Yun*, Changwon Kang*,
Junho Lee, Jeongseon Choi, Chanhyeok Kim, Jungwook Choi, Jun Won Choi†

*Equal contribution. †Corresponding author.

NEURAL INFORMATION PROCESSING SYSTEMS

HANYANG UNIVERSITY

SPA LAB
Signal Processing & Artificial-intelligence Laboratory

1. Challenges

- The Advent of diverse driving robots
  - Explosive growth of service robot market
- Insufficiency of comprehensive datasets for autonomous mobile robots
  - Necessity for socially interactive robots
  - Human Perception in 3D and movement prediction for safe and agile navigation

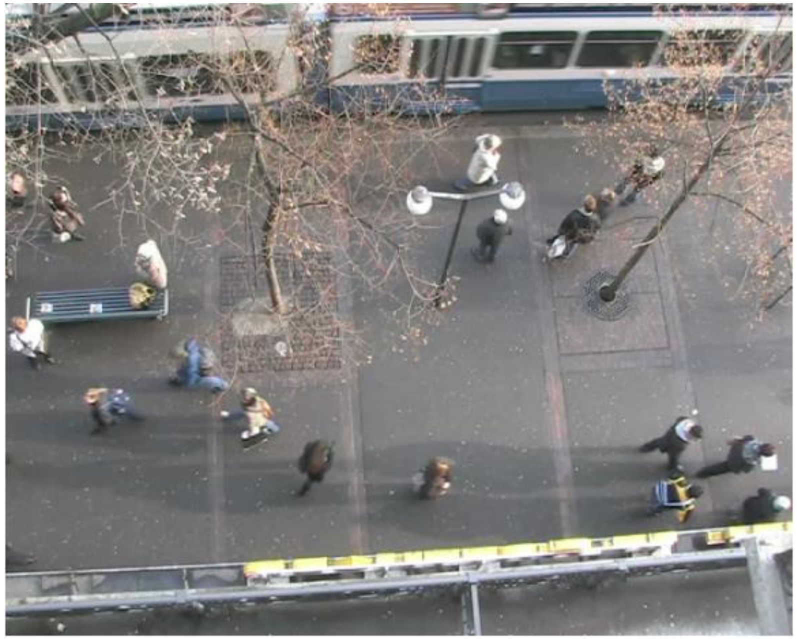
(a) Food delivering robot[1]

(b) Serving robot [2]

(c) Guide robot [3]

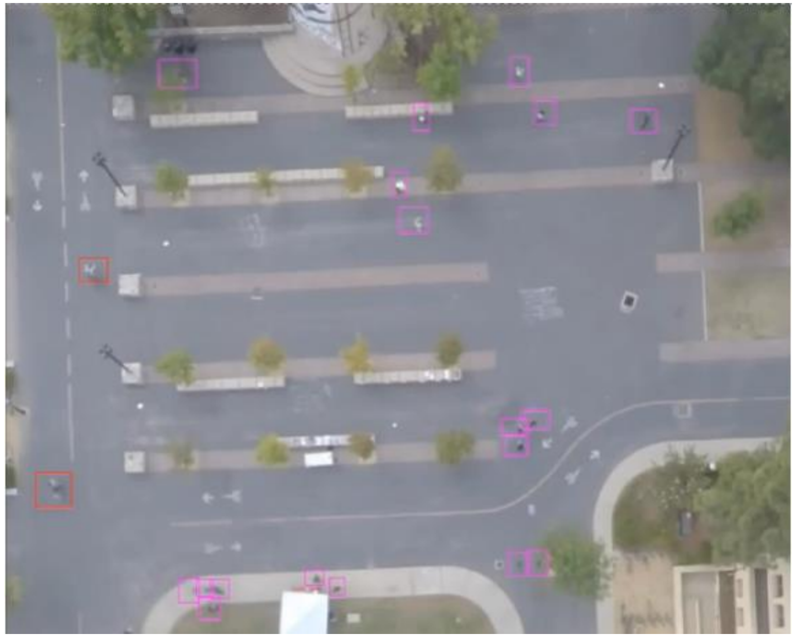
1. Challenges: Pedestrian Trajectory Datasets

- Data collected from fixed positions, potentially restricting the range of data variability
- Hard to reflect *Human-Robot Interaction* (HRI)
- Mostly consisting of camera images and data on pedestrian location



(a) ETH-Hotel [4]

(b) UCY-Zara [5]

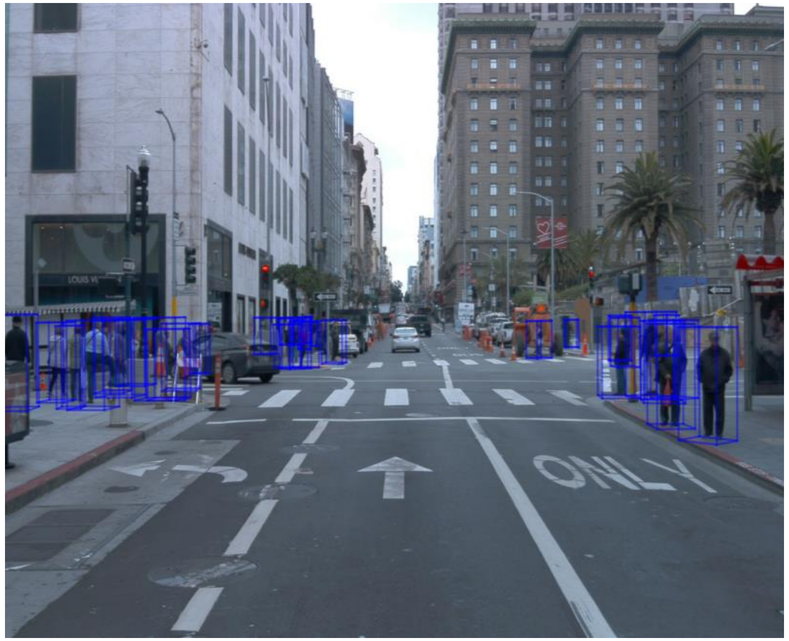


(c) SDD [6]

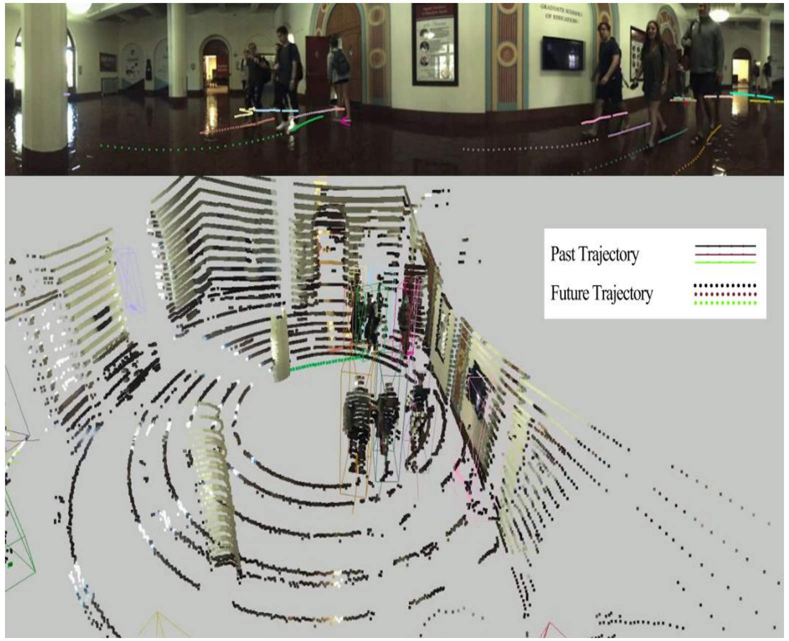
1. Challenges: Autonomous Driving Datasets

- Vehicle-centric on autonomous driving datasets rather than pedestrian-centric
  - A Shortage of vehicle-pedestrian interaction in autonomous driving datasets
  - Gaps in real-world robot and pedestrian interaction behavior
- Asynchronous multi-sensor data in robot-based datasets

(a) nuScenes[7] *(vehicle-based)*



(b) Waymo Open[8] *(vehicle-based)*



(c) JRDB [9] *(robot-based)*

1. Challenges: Comparison with other datasets

- Pedestrian trajectory datasets:
  - Data collected from fixed positions, potentially restricting the range of data variability
- Autonomous driving datasets:
  - Vehicle-centric on autonomous driving datasets rather than pedestrian-centric
  - Asynchronous multi-sensor data in robot-centric datasets

*†: Multi – layered map

Dataset	Platform	Task	Sync.	Map	E2E	Location
UCY	Fixed	Tracking, Prediction	-			Outdoor
ETH	Fixed	Tracking, Prediction	-			Outdoor
SDD	Fixed	Tracking, Prediction	-			Outdoor
nuScenes	Vehicle	Detection, Tracking, Prediction	✓	✓†		Outdoor
Waymo Open	Vehicle	Detection, Tracking, Prediction	✓	✓		Outdoor
Argoverse	Vehicle	Detection, Tracking, Prediction	✓	✓†	✓	Outdoor
JRDB	Robot	Detection, Tracking				Indoor & Outdoor
SiT(Ours)	**Robot**	**Detection, Tracking, Prediction**	✓	✓†	✓	**Indoor & Outdoor**

2. SiT Dataset: Real-world Context

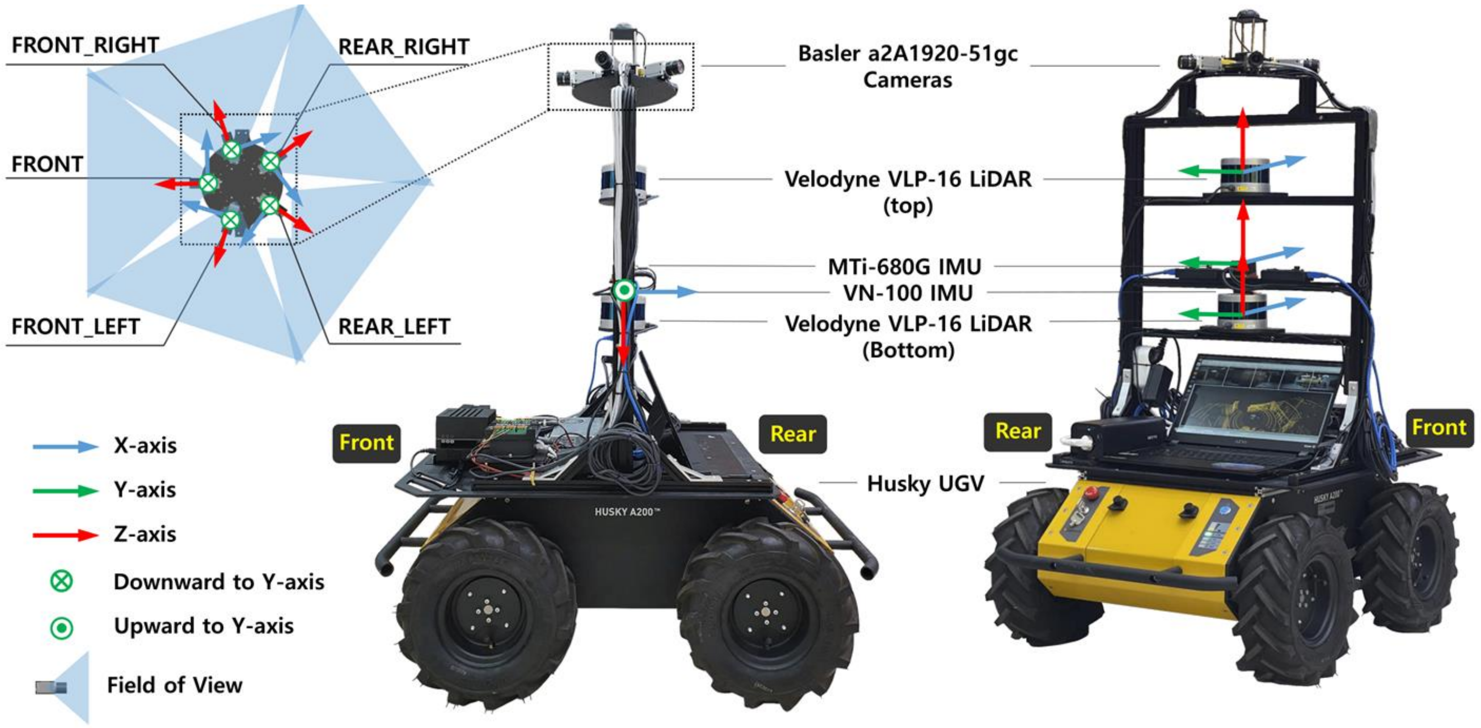
- Collected data from dense areas like campuses and public roads
- Authentic Human-Robot Interactions in real-world settings
  - Capturing data without any actors or pre-arranged scenarios

(a) Outdoor scene *(Crosswalk)*

(b) Indoor scene *(Hallway)*

2. SiT Dataset: Diverse Data Collection

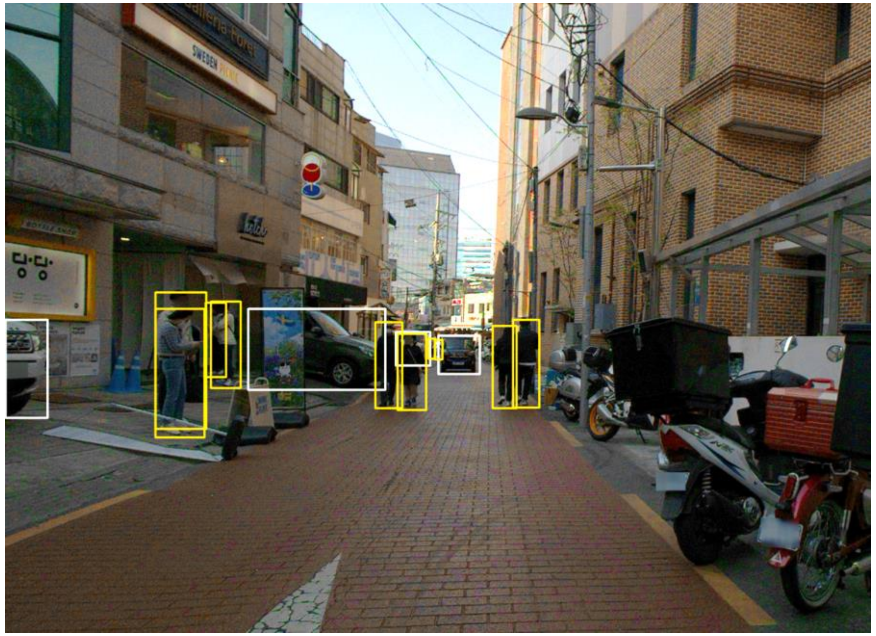
- **Sequential raw data from various sensors**
  - **60 scenes with 60K images and 12K point cloud frames at 10 Hz**
- 2D and 3D bounding boxes for 6 classes
  - Car, bus, truck, pedestrian, cyclist, motorcyclist



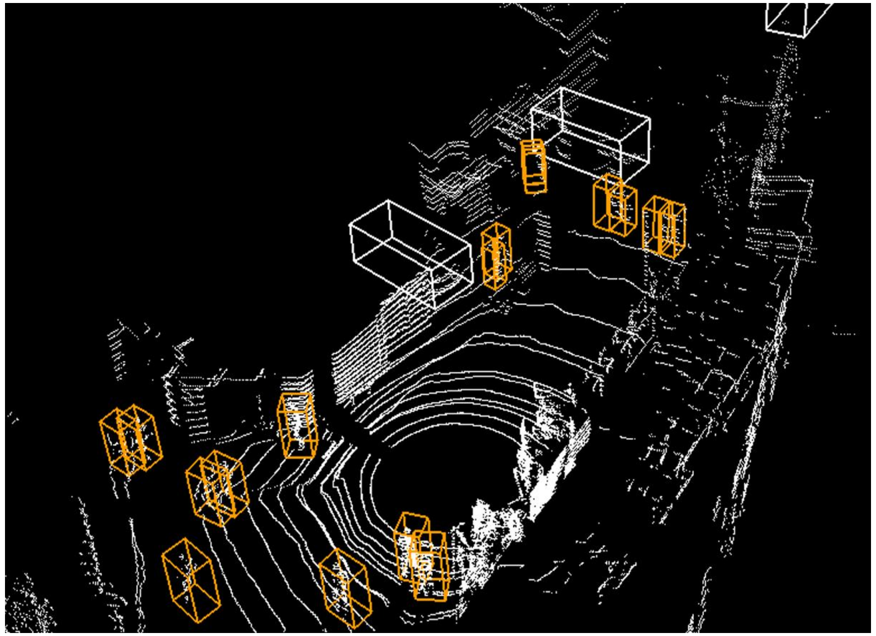
Husky UGV platform equipped with various sensors

2. SiT Dataset: Diverse Data Collection

- Sequential raw data from various sensors
  - 60 scenes with 60K images and 12K point cloud frames at 10 Hz
- **2D and 3D bounding boxes for 6 classes**
  - **Car, bus, truck, pedestrian, cyclist, motorcyclist**



(a) 2D Bounding Boxes labeled on Image



(b) 3D Cuboid labeled on Point Clouds

2. SiT Dataset: Unique Features

- **Precise multi-sensor synchronization**
- Multi-layered indoor & outdoor semantic maps from SLAM
- Cover tasks from 3D detection to motion forecasting (End-to-end)
- Emphasis on *Human-Robot Interactions* (HRI)

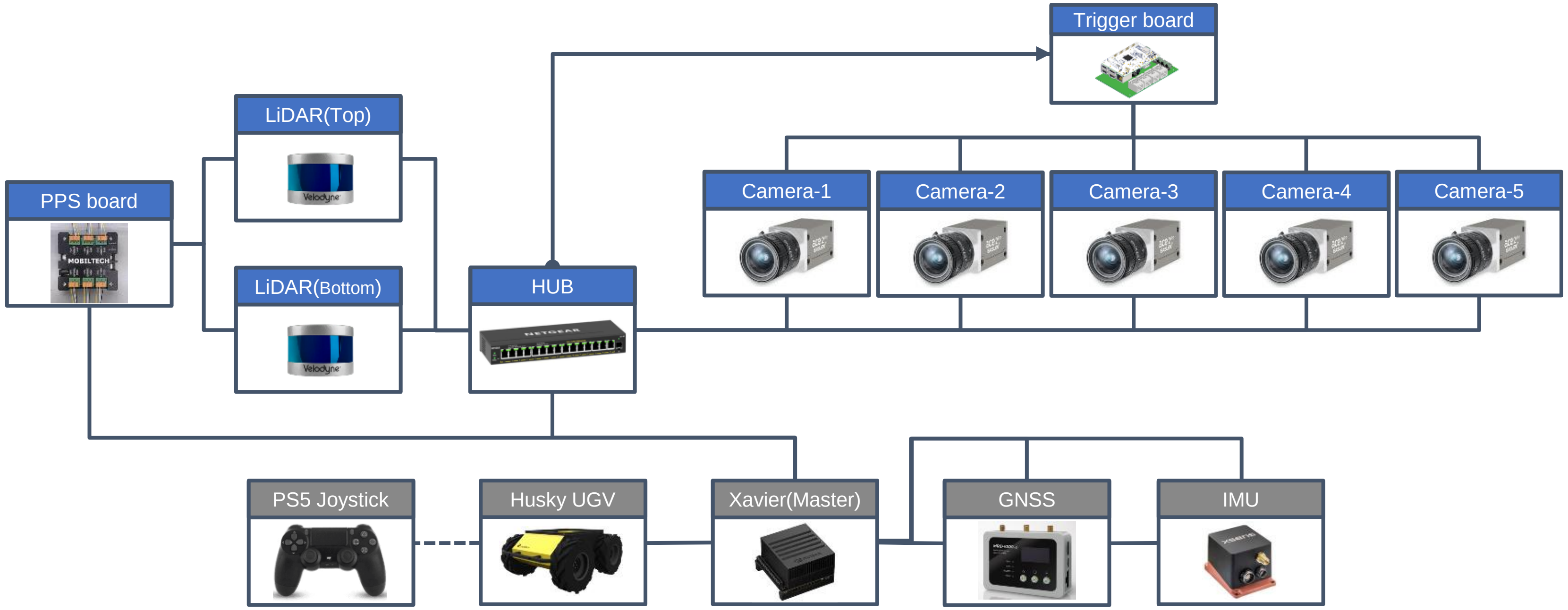
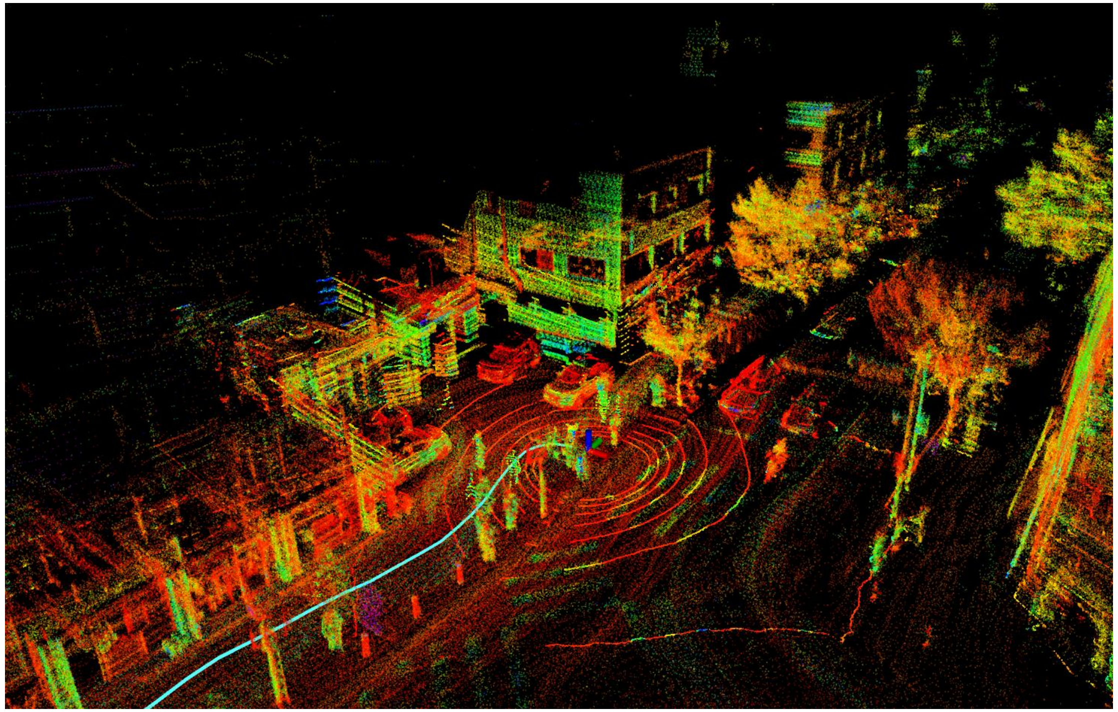


Diagram for multi-sensor synchronization

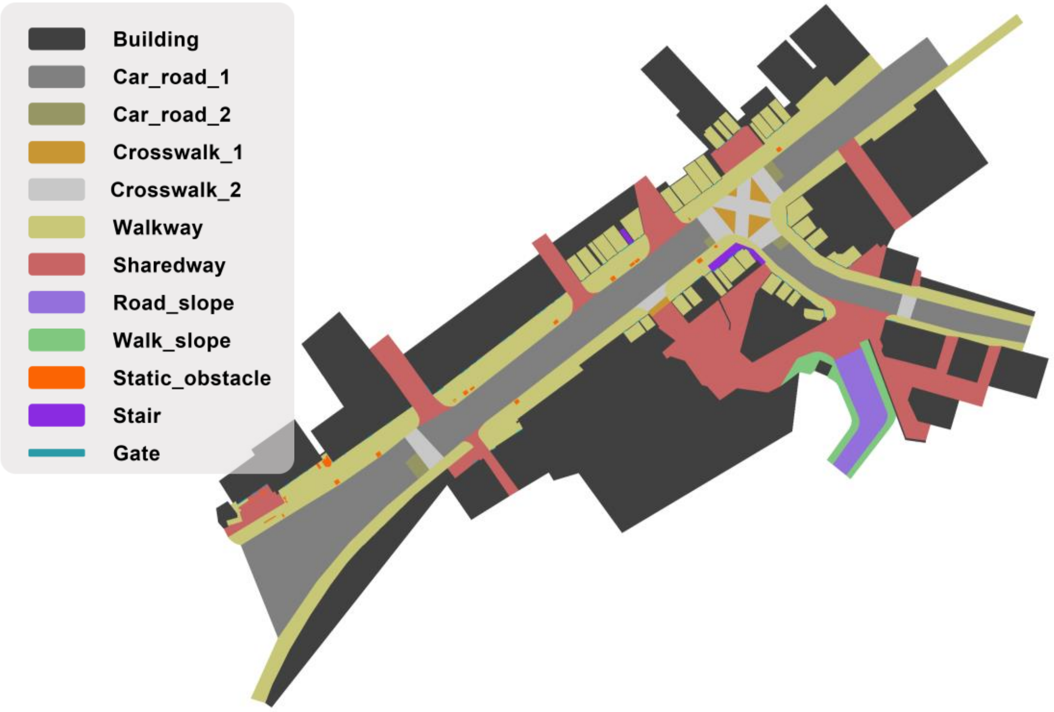
2. SiT Dataset: Unique Features

- Precise multi-sensor synchronization
- **Multi-layered indoor & outdoor semantic maps from SLAM**
- Cover tasks from 3D detection to motion forecasting (End-to-end)
- Emphasis on *Human-Robot Interactions* (HRI)

SLAM: *Simultaneous Localization And Mapping*



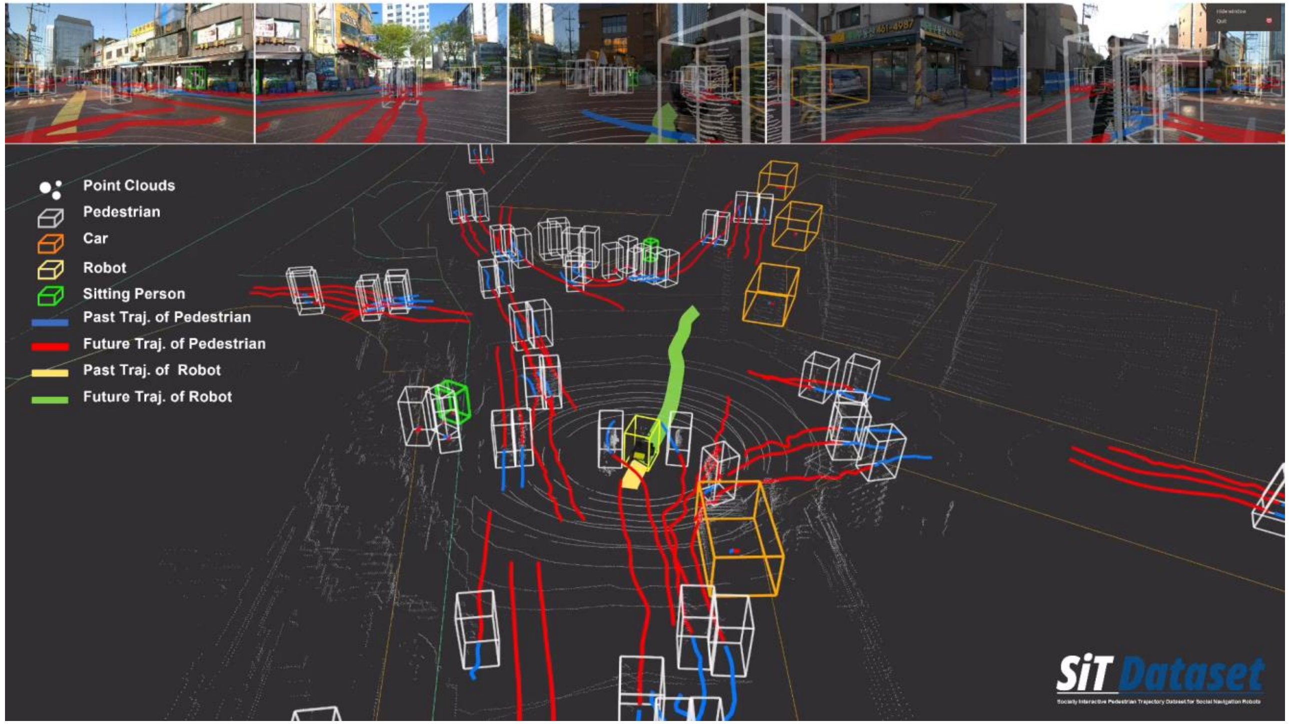
(a) SLAM-based 3D Point Cloud map



(b) 12-layered semantic map of outdoor scene

2. SiT Dataset: Unique Features

- Precise multi-sensor synchronization
- Multi-layered indoor & outdoor semantic maps from SLAM
- **Cover tasks from 3D detection to motion forecasting (End-to-end)**
- Emphasis on *Human-Robot Interactions* (HRI)

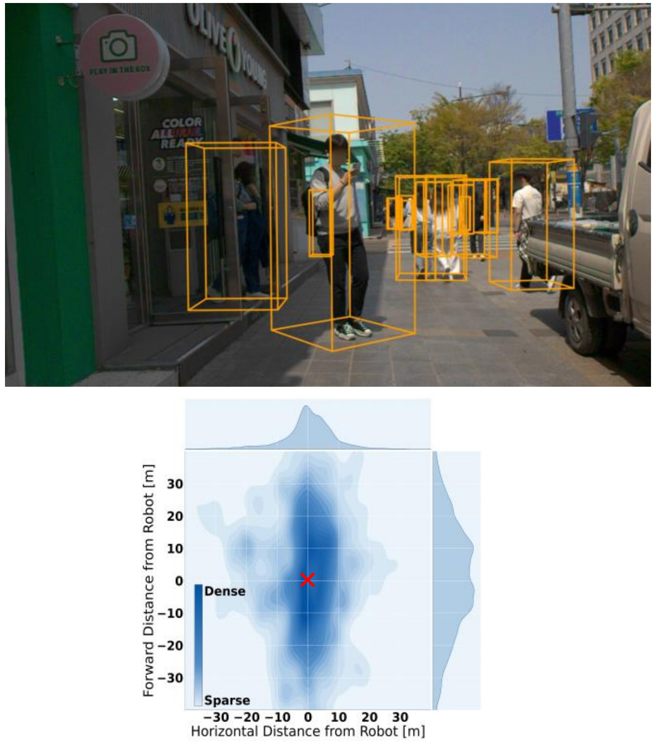


Visualization of SiT dataset of outdoor scene *(Cafe_Street_3)*

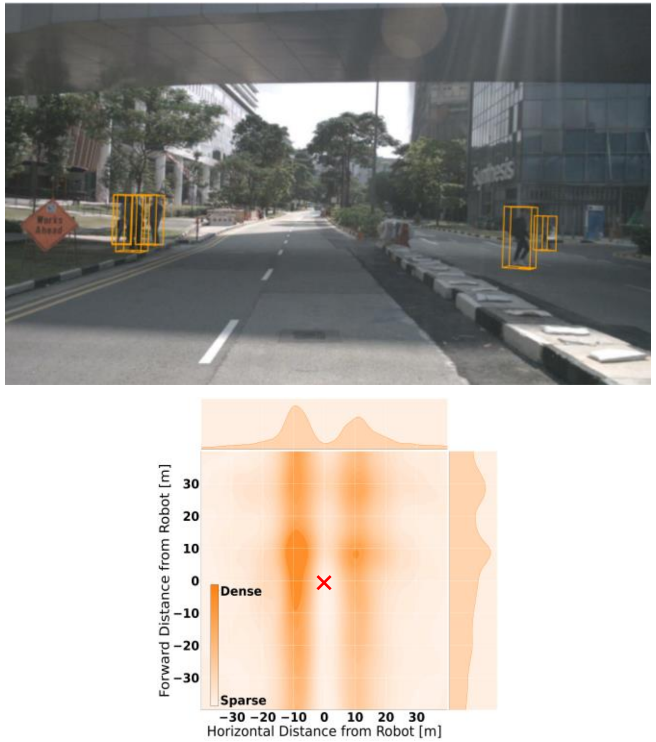
2. SiT Dataset: Unique Features

- Precise multi-sensor synchronization
- Multi-layered indoor & outdoor semantic maps from SLAM
- Cover tasks from 3D detection to motion forecasting (End-to-end)
- **Emphasis on *Human-Robot Interactions* (HRI)**

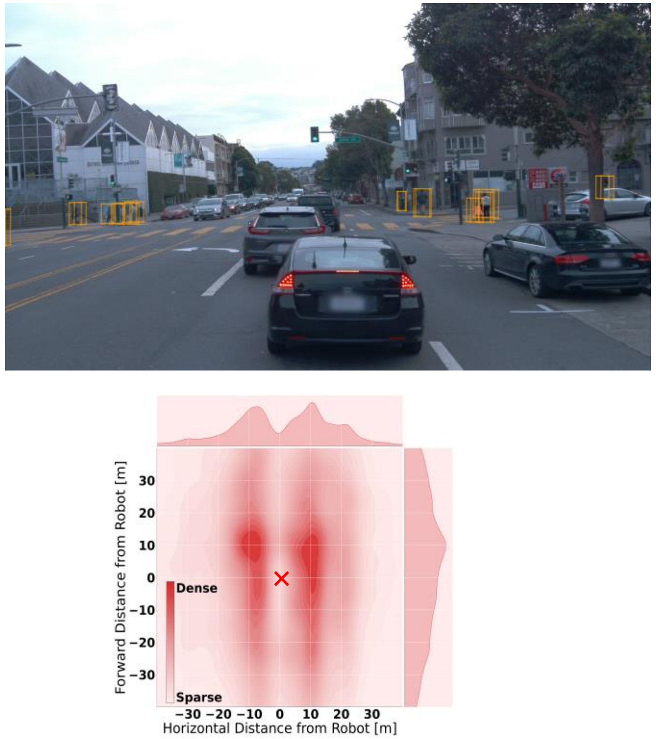
: 3D Cuboid of Pedestrian
× : Position of ego vehicle



(a) SiT (Ours)



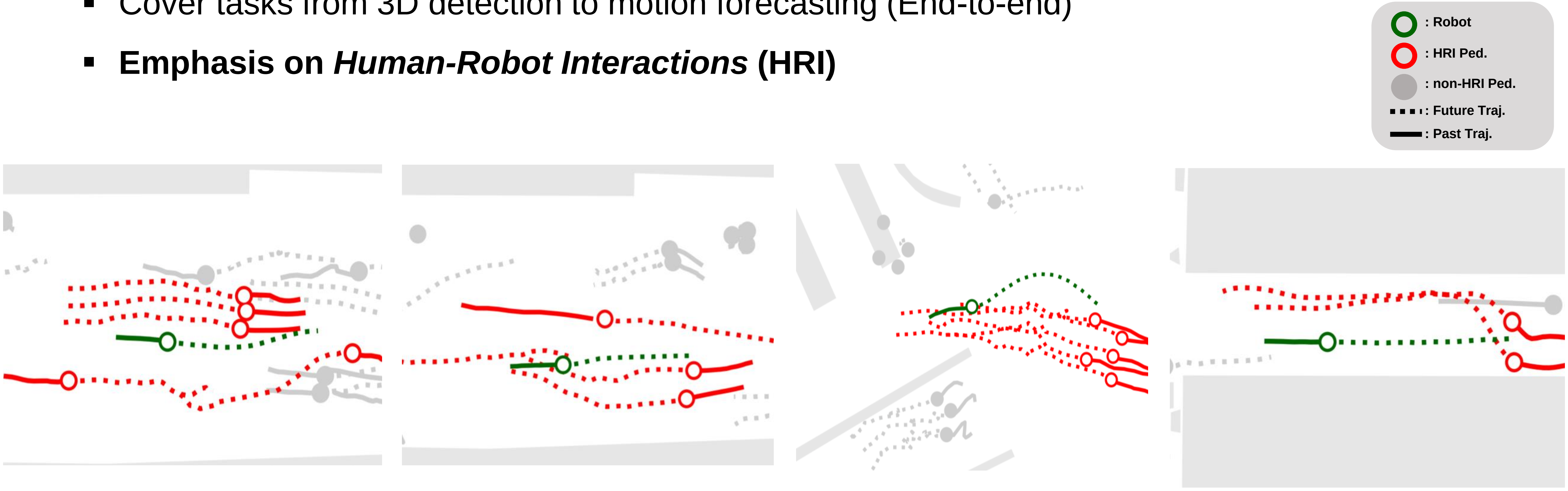
(b) nuScenes[7]



(c) Waymo Open[8]

2. SiT Dataset: Unique Features

- Precise multi-sensor synchronization
- Multi-layered indoor & outdoor semantic maps from SLAM
- Cover tasks from 3D detection to motion forecasting (End-to-end)
- **Emphasis on *Human-Robot Interactions* (HRI)**



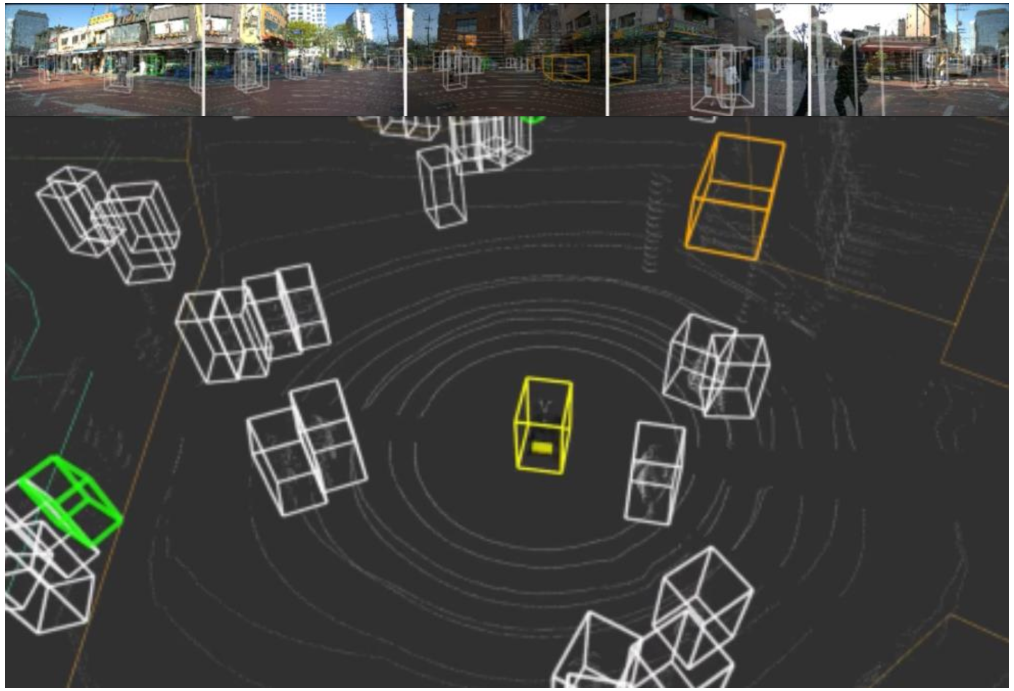
(a) Approach **(b) Followed by Pedestrians** **(c) Avoidance by Robot** **(d) Avoidance by Pedestrians**

3. Benchmarks & Challenges

- **3D object detection based on image and point clouds**
- 3D multi-object tracking
- Trajectory prediction
- End-to-end 3D detection to motion forecasting
- Challenges open on Eval.AI *(Feb. 2024)*

Methods	Modality	mAP ↑	AP(0.25) ↑	AP(0.5) ↑	AP(1.0) ↑	AP(2.0) ↑
FCOS3D [33]	Camera	0.244	0.024	0.159	0.329	0.463
PointPillars [15]	LiDAR	0.351	0.260	0.354	0.374	0.418
Centerpoint-P [39]	LiDAR	0.414	0.300	0.424	0.446	0.486
Centerpoint-V [39]	LiDAR	0.518	**0.397**	0.531	0.553	0.592
TransFusion-P [2]	LiDAR+Camera	0.390	0.248	0.371	0.437	0.507
TransFusion-V [2]	LiDAR+Camera	**0.531**	0.318	**0.536**	**0.607**	**0.665**

Evaluation of 3D pedestrian detection baselines.



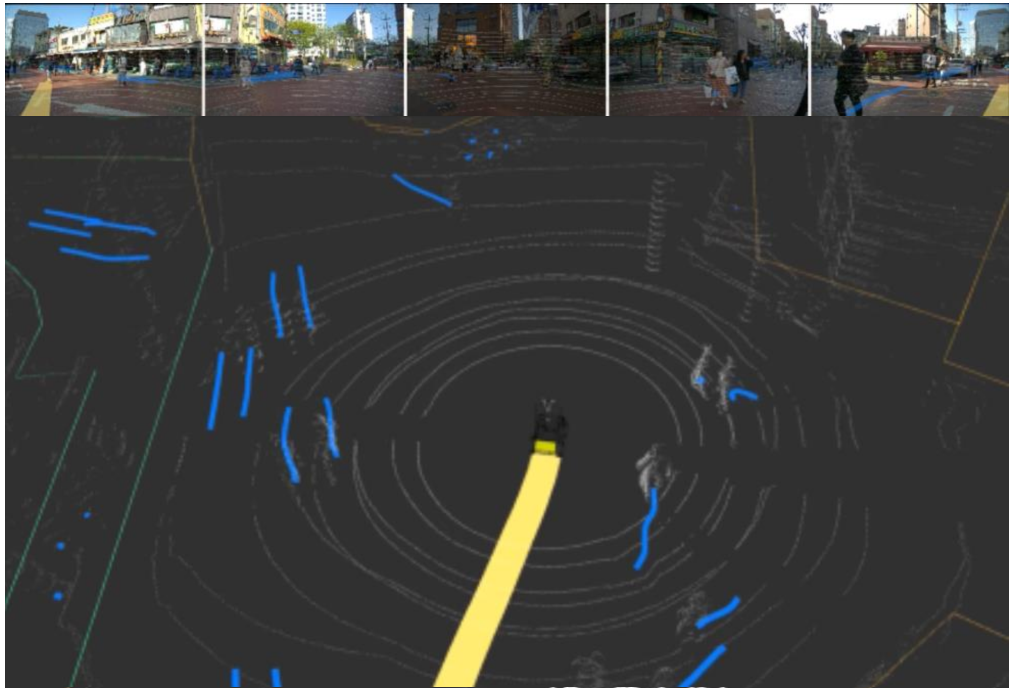
3D cuboids on each object

3. Benchmarks & Challenges

- 3D object detection based on image and point clouds
- **3D multi-object tracking**
- Trajectory prediction
- End-to-end 3D detection to motion forecasting
- Challenges open on Eval.AI *(Feb. 2024)*

Methods	sAMOTA ↑	AMOTA ↑	AMOTP(m) ↓	MOTA ↑	MOTP(m) ↓	IDS ↓
PointPillars [15] + AB3DMOT [34]	0.4110	0.1047	0.3580	0.4086	1.0277	1048
Centerpoint Detector [39] + AB3DMOT [34]	0.4841	0.1398	0.3958	0.4586	0.9836	**554**
Centerpoint Tracker [39]	**0.6070**	**0.2007**	**0.2679**	**0.4760**	**0.5140**	1136

Evaluation of 3D pedestrian tracking baselines.



Past trajectories of each object

3. Benchmarks & Challenges

- 3D object detection based on image and point clouds
- 3D multi-object tracking
- **Trajectory prediction**
- End-to-end 3D detection to motion forecasting
- Challenges open on Eval.AI *(Feb. 2024)*

Methods	Map	$\text{ADE}_5 \downarrow$	$\text{FDE}_5 \downarrow$	$\text{ADE}_{20} \downarrow$	$\text{FDE}_{20} \downarrow$
Social-LSTM [1]		1.638	3.121	1.630	3.103
Y-Net [22]		1.527	2.802	0.836	1.878
Y-Net [22]	✓	1.361	2.624	0.675	1.547
NSP-SFM [41]		1.346	2.261	0.634	1.087
NSP-SFM [41]	✓	**1.061**	**1.818**	**0.517**	**0.925**

Evaluation of pedestrian trajectory prediction baselines

Past and future trajectories of each objects

3. Benchmarks & Challenges

- 3D object detection based on image and point clouds
- 3D multi-object tracking
- Trajectory prediction
- **End-to-end 3D detection to motion forecasting**
- Challenges open on Eval.AI *(Feb. 2024)*

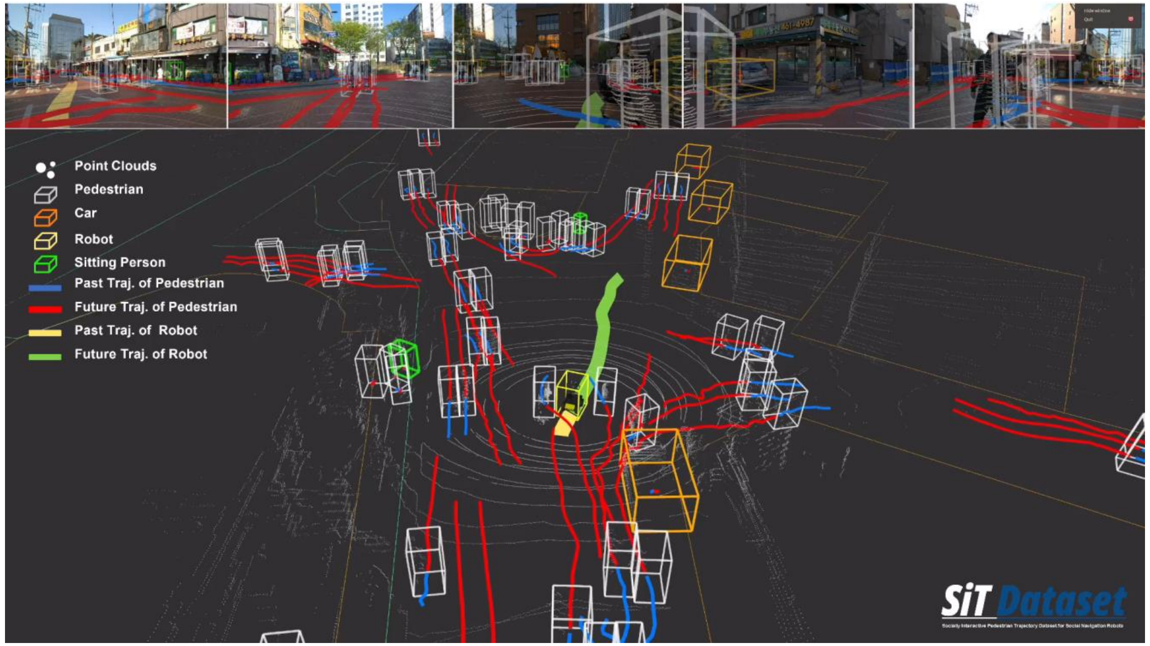
Methods	mAP ↑	$\mathrm{mAP_f}$ ↑	$\mathrm{ADE_5}$ ↓	$\mathrm{FDE_5}$ ↓
FaF [21]	**0.490**	**0.079**	**1.915**	**3.273**
FutureDet-P [26]	0.209	0.037	2.532	4.537
FutureDet-V [26]	0.408	0.053	2.416	4.409

Evaluation of end-to-end motion prediction baselines.

3D cuboids, past and future trajectories of each object

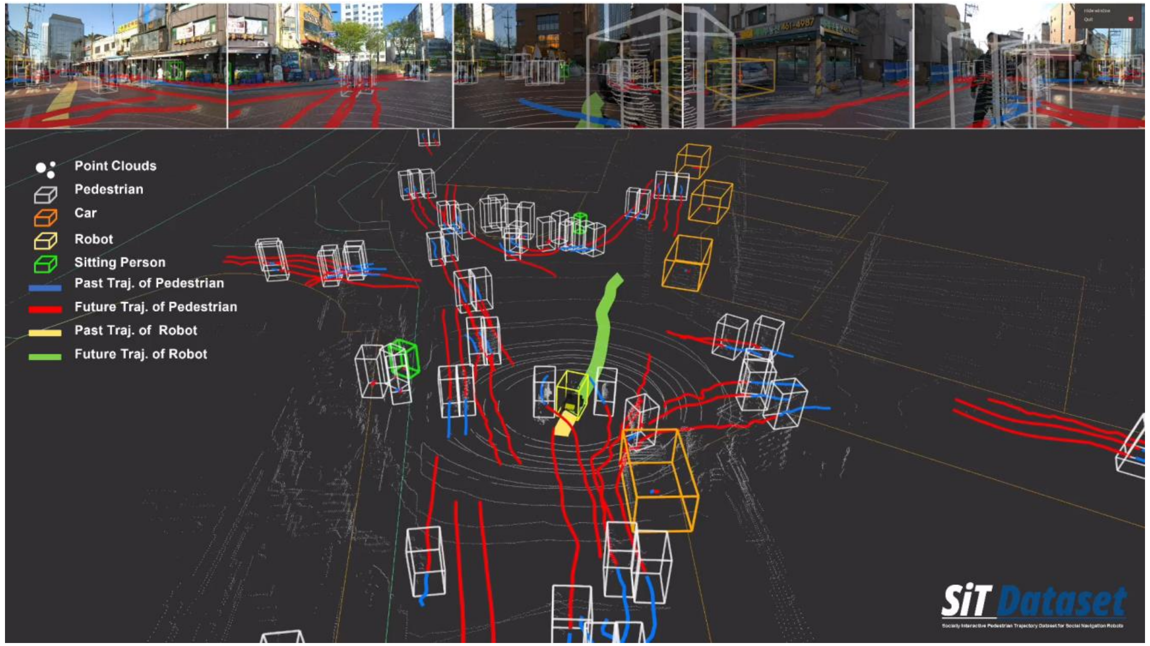
3. Benchmarks & Challenges

- 3D object detection based on image and point clouds
- 3D multi-object tracking
- Trajectory prediction
- End-to-end 3D detection to motion forecasting
- **Challenges open on Eval.AI *(Feb. 2024)***



4. Conclusion

- **SiT Dataset**: Socially interactive Pedestrian Trajectory Dataset for Social Navigation Robots
  - Include diverse pedestrian trajectories captured in human-robot interactive scenarios
  - High-quality 2D and 3D annotations for various perception tasks
  - 12-layered semantic maps covering a wide range of scene information
  - Facilitate design of end-to-end motion prediction models



References

- [1] Starship: food delivering robot, https://en.wikipedia.org/wiki/Delivery_robot
- [2] Dadawan: serving robot, https://www.businessinsider.com/robot-serves-food-takes-temperatures-covid-19-in-the-netherlands-2020-6#next-the-human-staff-still-have-to-actually-take-orders-from-a-safe-distance-6
- [3] Airstar: guide robot, https://m.hankookilbo.com/News/Read/201807111499787598
- [4] ETH-Hotel, https://icu.ee.ethz.ch/research/datsets.html
- [5] UCY-Zara, https://www.youtube.com/watch?v=jyKO4rGDn0o
- [6] SDD, https://www.youtube.com/watch?v=c6xQ6iz6wH8
- [7] nuScenes, https://www.nuscenes.org/
- [8] Waymo Open, https://waymo.com/open/
- [9] JRDB, https://jrdb.erc.monash.edu/

End of Document