Mobile Robot Oriented Large-Scale Indoor Dataset for Dynamic Scene Understanding

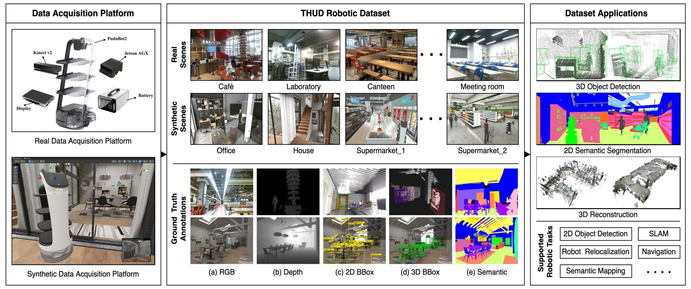
Yi-Fan Tang[1], Cong Tai[1], Fang-Xing Chen[1], Wan-Ting Zhang[1], Tao Zhang[2], Xue-Ping Liu[1], Yong-Jin Liu[1] and Long Zeng[1]
[1]Tsinghua University; [2]Pudu Robotics

ICRA2024 YOKOHAMA | JAPAN

Introduction

Large-Scale Dynamic Indoor Dataset

- Most robotic datasets capture static scene data and thus are limited in evaluating robots' dynamic performance. We present a mobile robot oriented large-scale indoor dataset, denoted as THUD (Tsinghua University Dynamic) robotic dataset for robots' dynamic scene understanding.

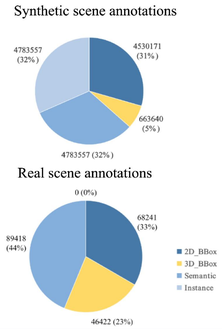


Contributions

- Data annotated with dynamic instances for large-scale indoor scenes, closer to robots' real working environment.
- Supports training and testing for various robotic scene understanding tasks
- Contains both real and synthetic annotated data, and satisfy the testing of mobile robot in different scenes.

Statistics

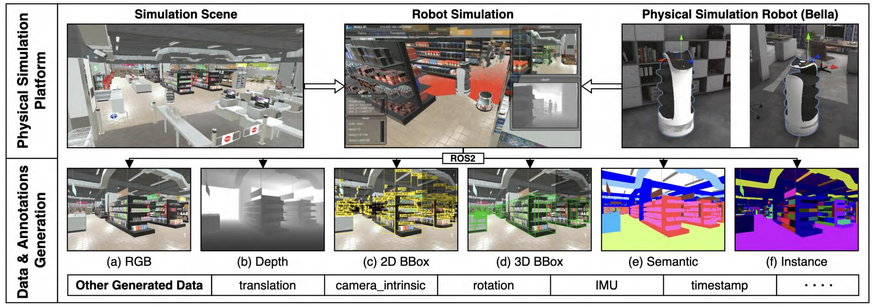
- Consisting of 84,984 frames from synthetic data collection and 5,191 frames from real-world data collection.
- In a total of over 20M labels, with over 1.2M labels for dynamic objects each frame contains 176 data labels.
- The dataset is still continuously expanding.



Method

Synthetic data acquisition

- The platform is designed to simulate the realistic working environment for mobile robots, allowing for the collection of virtual sensor data such as RGB images, depth maps, robot pose, and IMU data.



- Different levels of dynamic complexity: dynamic scenarios with moving obstacles and special scenarios that pose potential hazards for robots in the reality.

Real data acquisition

- Data collection platform

- Different levels of dynamic complexity

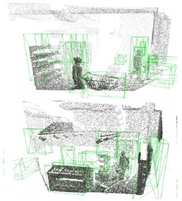
Ground Truth Annotation

- Combination of automated and semi-automated methods to enhance efficiency and accuracy in data annotation.

Experiments

- 3D objects detection on dynamic and static objects in scene with different dynamic complexity

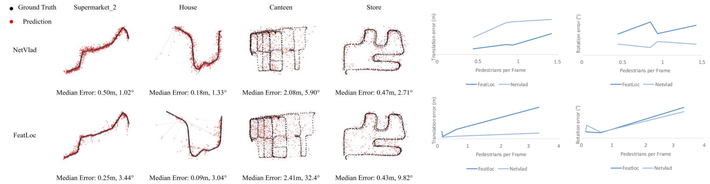
Scene	Method	Dynamic Objs (mAP)	Static Objs (mAP)
Supermarket	F-PointNet [24]	7.89	8.92
	ImVoteNet [23]	17.49	17.29
	DeMF [25]	34.51	38.24
Canteen	F-PointNet [24]	18.16	36.67
	ImVoteNet [23]	26.72	43.37
	DeMF [25]	28.56	45.43



- Semantic Segmentation in scene with different dynamic complexity

Scene	Method	Backbone	MIoU(%)
Supermarket	ACNet [29]	3×R50	74.83
	RedNet [30]	2×R34	76.92
	ESANet [31]	2×R34	78.42
	SA-Gate [32]	2×R101	83.19
Canteen	ACNet [29]	3×R50	51.85
	RedNet [30]	2×R34	59.83
	ESANet [31]	2×R34	65.97
	SA-Gate [32]	2×R101	58.34

- Robot relocalization in synthetic and real-world dynamic scene



- 3D scene reconstruction in scene with different dynamic complexity

- **Up Coming:** Pedestrian trajectory prediction