

# Technique Report



Figure 1. A demonstration of a navigation process, in which a robot move to several places to accomplish a task.

In this report, we are going to introduce the technical details our work [2], including the annotation pipeline, the dataset, and evaluation. During our implementation, we find challenges that hinders us to build a robust navigation agent. We will discuss the shortcoming of our work and welcome the community to solve these problems.

## 1. Annotation Framework

We annotate our FAO dataset by four steps: 1) annotate the bounding boxes of an object from all viewpoints that can reach it; 2) annotate instructions to describe annotated objects and local scenes, which enable an agent to reach any target object from anywhere in this house; 3) merge the bounding boxes and descriptions that belongs to the same object manually; 4) post-processing, including orientation alignment and trajectory generation.

**Data Noise in Annotation** As shown in Fig 1, we annotate an bounding box for an object. We also can navigate to other viewpoints by clicking blue boxes. However, this process introduces noise by two means. First, humans sometimes do not label all of the viewpoints of an object. This problem harms the navigation performance when an agent navigation to reach the target correctly but get punished. Second, humans sometimes do not label the object at the center of the window. An labeled bounding box has four parameters, the windows orientation in a panoramic space, including heading and eleva-

tion, and the 2D offset of an object bounding box in the plan of this window. However, it is challenging to align the 2D offset with the panoramic window orientation. We develop an automatic program to rotate windows so that the bounding boxes is eventually located in the center of the window.

**Data Noise in Panoramic Space** Another reason of introducing data noise is the application of panoramic space. In the panoramic space,  $0^\circ = 360^\circ$ . Different from linear space where  $c - a > c - b$  is always holds if  $a < b < c$ , the optimize the prediction module in the panoramic space is much more challenging. For example, if the ground truth is  $1^\circ$  while the prediction is  $350^\circ$ , encouraging the agent to predict a larger number is closer to the ground truth than punishing the agent to predict a smaller number.

## 2. Baseline Performance

The performance of our paper [2] is reported with the previous version of dataset, which data split does not follow the standard Matterport split [1]. In our challenge, we clean our training dataset, adding more testing data, and re-split our dataset following the standard Matterport split. Here we provide the baseline performance in the new SOON dataset.

## 3. Dataset

**Visualization Interface** We also provide a visualization interface at <https://label-auto.dm-ai.com/admin.html>. Here you can navigation within our 90 houses, see the bounding box annotations and its corresponding natural language instructions.

### 3.1. Instruction Examples

In Fig. 2, we list more instructions in our dataset. We annotate different semantic parts use different colors. The blue color stands for the object name; the green color stands for the object attributes; the

Splits	Unseen House (Val)					Unseen House (Test)				
Metrics	LE	OSR $\uparrow$	SR $\uparrow$	SPL $\uparrow$	SFPL $\uparrow$	LE	OSR $\uparrow$	SR $\uparrow$	SPL $\uparrow$	SFPL $\uparrow$
GBE (Ours)	28.96	28.54	19.52	13.34	1.16	27.88	21.45	12.90	9.23	0.45

Table 1. The results of the GBE in the new dataset).

red color stands for the room where it sets; the purple color stands for the nearby regions.

### 3.2. Data Formats

```

1 {"path": [
2   ["ef5c95eaff9c487d8ae05b2f8d0ea532",
3    "caf815b583284834a7a07526e93de4e5",
4    ...
5   ],
6   ...
7  ],
8  "instruction": [
9    "cabinet",
10   "rectangular, black and brown",
11   "between two ovens, under a wash basin",
12   "It is in the big living room.",
13   "The living room is on the first floor
14     and ...",
15   "Find a rectangular cabinet ..."
16  ],
17  "bboxes": [
18    {
19      "scan": "B6ByNegPMKs",
20      "image_id": "433e85ac1f964f0fa890b355
21        ba051fc5",
22      "camera_pose_matrix": [...],
23      "heading": 2.127971314578369,
24      "elevation": -0.5920715901805768
25    },
26    ...
27  ]
28 }
```

The data format is shown above. The path is annotated by a sequence of point id (use hash id to ensure it is unique across the dataset). The instruction includes different granularities. In "bboxes", there are several bounding boxes indicating all the positions that is able to reach the target object.

### References

- [1] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niebner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. In

2017 International Conference on 3D Vision (3DV), pages 667–676, 2017. 1

- [2] Fengda Zhu, Xiwen Liang, Yi Zhu, Qizhi Yu, Xiaojun Chang, and Xiaodan Liang. Soon: Scenario oriented object navigation with graph-based exploration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12689–12699, 2021. 1

---

---

I want to find a big, transparent and glass glass, which is set in the spacious living room. The glass is on the cabinet, in front of the picture and on the right of the lamp. The living room is on the first floor and next to the dining room.

---

Find a rectangular, white, cloth chair which is next to a table, in front of a bed. The chair is settled in the corner of a big bedroom. The bedroom is on the second floor and there is a balcony outside.

---

There is an old brown wooden table on the carpet, by the table which is situated at the center of the room. Behind the chair there is a tall cupboard and the chair is the nearest to the downstairs.

---

Please help me find a middle size, woolen, brown shelf, which is next to a door and fixed on wall. It is in a storage room on the second floor. The storage room is in a bedroom which is near a corridor and a living room.

---

At the corner of a living room, there is a medium size, wooden and brown cabinet settled under a silver ship, opposite to a beige floriated sofa and near six pictures. The cabinet is near a study and a corridor on the second floor.

---

Figure 2. Examples of instructions in the dataset.