Seeing the Un-Scene: Learning Amodal Semantic Maps for Room Navigation

Anonymous ECCV submission

Paper ID 3023

1 Overview

In this supplementary section, we describe in more detail the data collection process, statistics of houses in Matterport3d, and design choices made by us while creating the Room Navigation dataset. We also include qualitative results in the form of videos.

2 MatterPort3D Room Analysis

Scene Selection. For our Room Navigation dataset, we select a subset of the scenes in Matterport3D [1] which are houses. A scene is considered to be a house if it contains a bedroom. Matterport3D scenes are equipped with room type and bounding box annotations which we use to select houses and for generating RoomNav episodes. The bounding boxes of rooms on different floors would overlap at times, *e.g.* a room on the first floor would have a overlapping z axis with the room right above it on the second floor. Also, the floors are uneven, making it difficult to distinguish between the different levels of the house. For these reasons, we only choose houses which have at least one bedroom on the first floor of the house. We likewise construct all the episodes on the first floor. In total, we have 32/4/10 houses in train/val/test splits respectively.

Room Statistics. Table. 1 contains room count and shape statistics in the Matterport3D houses. In general, bathrooms are the smallest in size and living rooms are the largest. There are also more bathrooms compared to other room types in the whole dataset as typically each house consists of more than one bathroom.

Ground truth target point. Each episode in the Room Navigation dataset consists of a starting location and a target room ID. To select the ground truth target point in the room which is closest to the source position, we first sample all the points at 0.1m intervals which lie 0.2m inside the target room bounding box. Often, the room bounding boxes extend beyond the walls of the room and choosing points 0.2m inside the bounding box ensures this point will definitely lie inside the room. From the sampled points, the one which has the smallest geodesic distance along the shortest path to the source point is chosen as the ground truth target point. Fig. 5 in the main paper illustrates this procedure.

3 Room Navigation Dataset Statistics

In this section, we provide statistics of our Room Navigation dataset. Fig. 1 is a visualization of the distribution of episodes on all three splits over (i) geodesic

Table 1: **Room Statistics for Matterport3D.** Statistics of room count and size in the houses we use in our experiments.

	Bathroom	**Bedroom**	**Dining Room**	**Kitchen**	**Living Room**
Count	143	90	48	53	47
Average Size ($l \times w$ in meters)	2.90×2.84	4.97×4.47	5.83×5.65	5.40×5.09	6.08×6.11

Table 2: Model hyperparameters for Map Generation and Point Prediction.

Hyperparamter	**Map Generation**	**Point Prediction**
Batch size	256	256
Epochs	200	200
Learning rate	2.5e-5	2.0e-5
Optimizer	Adam	Adam
Weight decay	0.01	0.00
Dropout	0.1	0.1
LSTM Layers	2	-
Generated map size	26	-
Frame sequence length	20	-

distance along the shortest path from start position to the ground truth target point in target room, (ii) Euclidean distances between the start position and ground truth target point in the target room (iii) the ratio of (i) and (ii). The ratio is an approximate measure of the complexity of an episode. A higher value indicates that the geodesic distance along shortest path is greater than the euclidean distance, thus suggesting the route is more complex that a simple straight line between source and target. The average geodesic distance on train, val, and test are 11.48m, 10.25m, and 11.21m respectively.

"Nearest" Target Room. The target for each Room Navigation episode is a target room type, *e.g.* "Kitchen". As described in Sec. 3 of the main paper, if there exist multiple rooms of the same type, the agent needs to navigate to the one closest to the source position. Here, we analyze how often there exist multiple rooms of the same type and also report the average RoomNav SPL and Success metrics on these episodes.

There are 1.2 million episodes in train, 89 in validation and 284 in test where there are multiple target rooms of the same type. The average RoomNav SPL and Success of our best perfoming model (Map Generation + Point Prediction + PointNav + Fine-tune) on the 89 episodes in validation are 0.34 and 0.36 and on the 284 episodes in test are 0.30 and 0.33.

4 Random Oracle Baseline

Here, we show results on the validation set of a random agent with an oracle stopping policy. The agent moves similar to the random agent described in Sec. 6 of the main paper, *i.e.* it takes action randomly among `move_forward`, `turn_left`, and `turn_right` with uniform distribution. It calls `stop` as soon as it reaches the

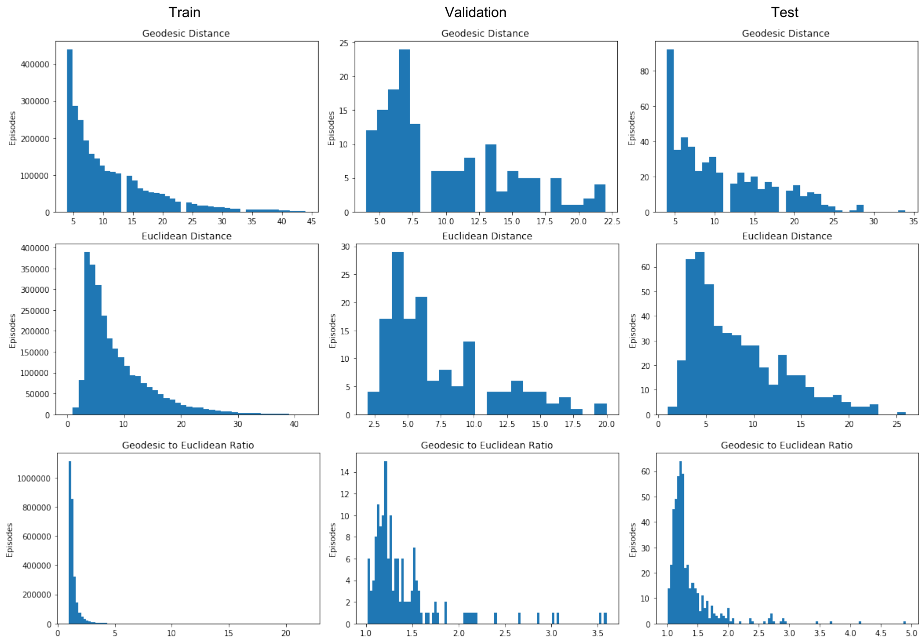

Fig. 1: **Statistics of Room Navigation Episodes.** Distribution over geodesic distance along shortest path between start and goal, distribution over Euclidean distance between start and goal, and distribution over the ratio of geodesic to Euclidean distance. Typically, the larger the ratio, the more complex the trajectory.

target room. We implement this by checking if the agent's current position lies 0.2m within the bounds of the target room. This agent achieves an RoomNav SPL of 0.008 and Success of 0.013.

Table 3: Hyperparameters for Fine-tuning Point Navigation.

Hyperparamter	**Value**
Batch size	256
Epochs	200
Learning rate	1e-4
Weight decay	0.0
Dropout	0.1
Learning rate decay	Linear

5 Hyperparameters

We describe our hyperparameter choices for each of the components in our model in Table. 3. For training the point navigation model, we use the same settings as described in [2]. For fine tuning the point navigation model on the predicted points, we use the hyperparameters specified in Table. 3.

6 Videos

We attach 3 videos showing qualitative results of our model. The first video is the same as Fig. 6 in the main paper. The second and third video include more examples showing success and failure cases of our method.

References

1. Chang, A., Dai, A., Funkhouser, T., Halber, M., Nießner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3d: Learning from rgb-d data in indoor environments. arXiv preprint arXiv:1709.06158 (2017) 1
2. Wijmans, E., Kadian, A., Morcos, A., Lee, S., Essa, I., Parikh, D., Savva, M., Batra, D.: Decentralized distributed ppo: Solving pointgoal navigation (2019) 4