

Supplementary Material

1. Experiments

1.1. Auto-Labeling

In Fig. 1, we visualize the results from the baseline methods MaskFormer and MaskFormer+SAM alongside our approach, MaskFormer+SAM+Detic+MV, when applied to the AVD dataset.

In Fig. 2, we show the results on the ADE20K dataset. The multi-view verification stage is omitted because of the lack of information to associate different images in the dataset, so only MaskFormer and MaskFormer+SAM results are visualized.

1.2. Object Part Discovery

We additionally annotate 'bottle cap' via the object part discovery approach detailed in Sec.5.1.

Fig. 3 shows the annotated bottle caps.

1.3. Object Goal Navigation

Table. 1 provides comprehensive details about the testing episodes, such as the testing environment, the target object, and the navigation success rates for both the VL-Map and GT-Map approaches.

To gain further insight into the performance of our navigation episodes, Fig. 4, Fig. 5, Fig. 6 and Fig. 7 present qualitative results across various AVD environments.

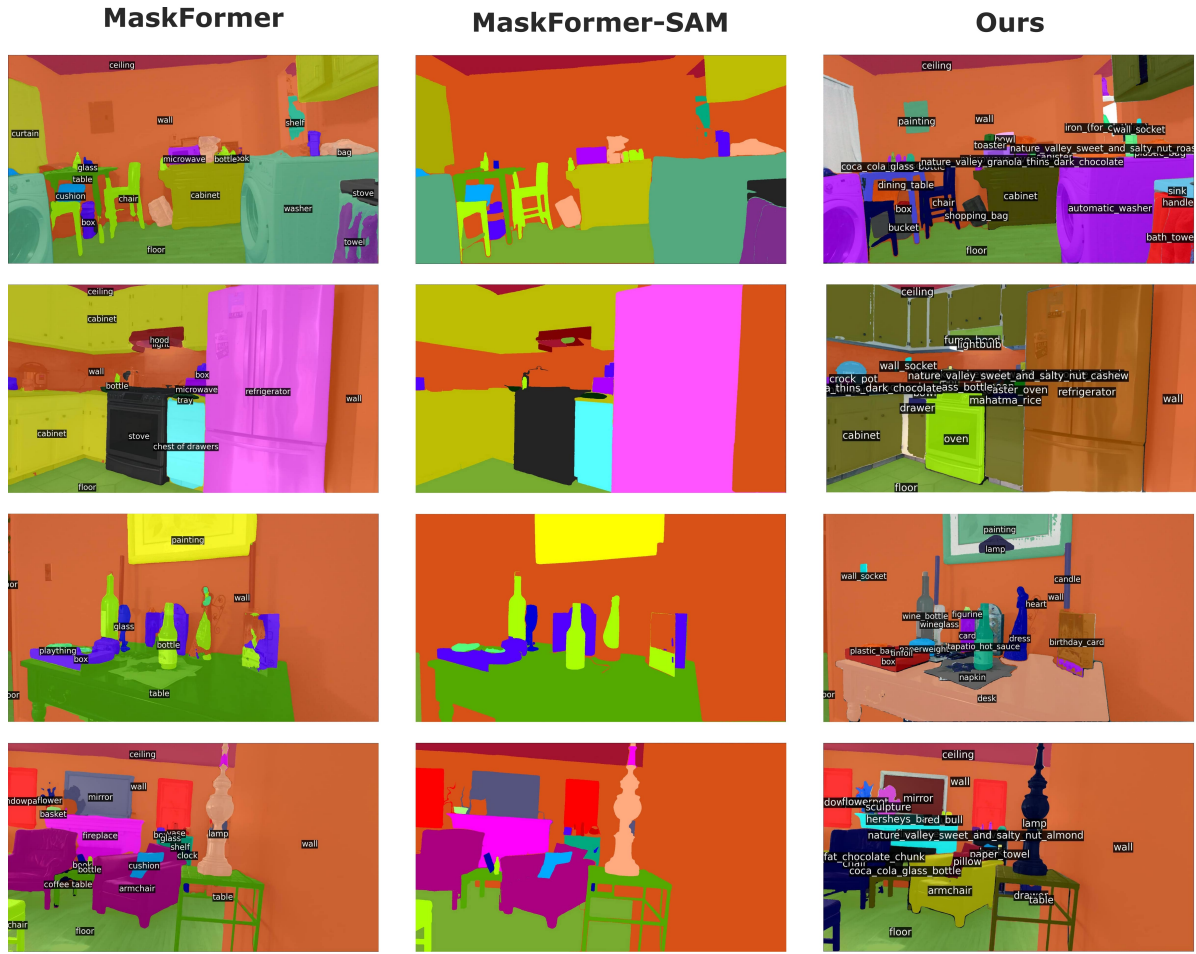


Figure 1. **Comparison of auto-labeling results on MaskFormer and MaskFormer+SAM, and MaskFormer+SAM+Detic+MV (ours) on AVD images.** Notably, MaskFormer+SAM demonstrates improved boundary delineation between distinct objects. Our approach also utilizes Detic and existing bounding boxes from AVD for getting more object masks, and the multi-view verification stage improves over single-view results.



Figure 2. A comparison between MaskFormer and MaskFormer-SAM on the ADE20k-indoor dataset. In certain examples, MaskFormer-SAM showcases notable improvements, surpassing MaskFormer by 10% in mIoU and 12% in mIoU-small.

bottle_cap annotation examples



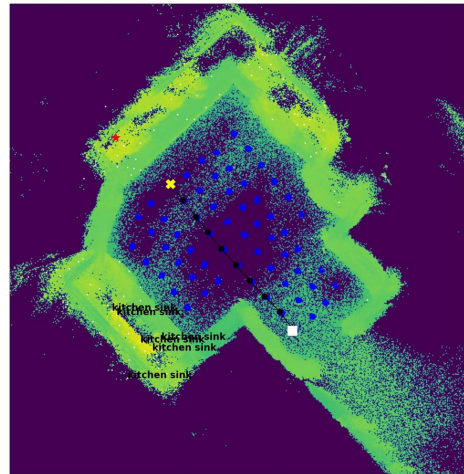
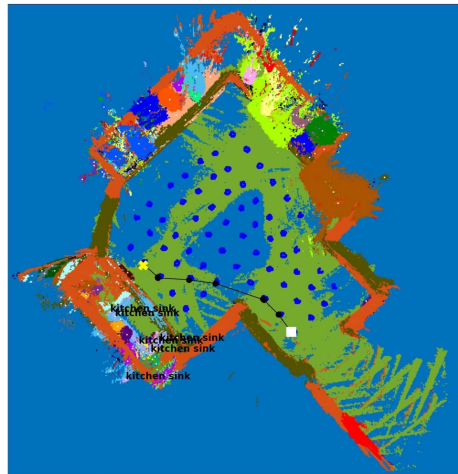
Figure 3. This figure gives examples of bottle caps marked as red rectangles. For enhanced clarity, the labeled bottle caps have been enlarged to improve visualization.

Instruction: Go to *'Kitchen Sink'*

GT Map (Semantic Map)

VL-Map (Heatmap)

Home_005_1



Last View in the Navigation Episode

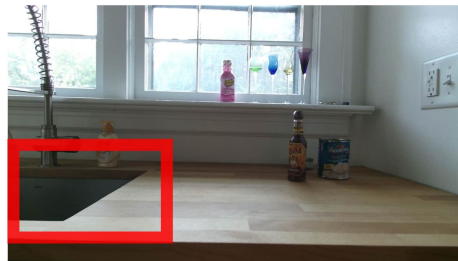


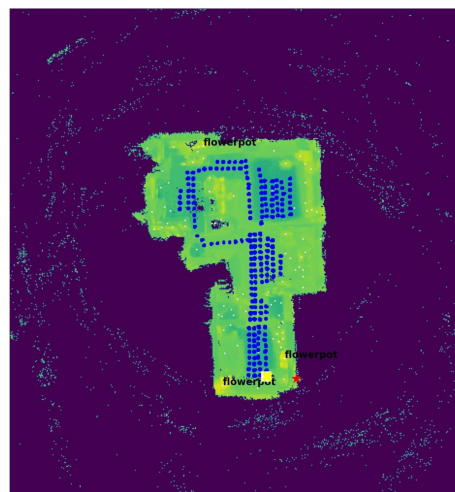
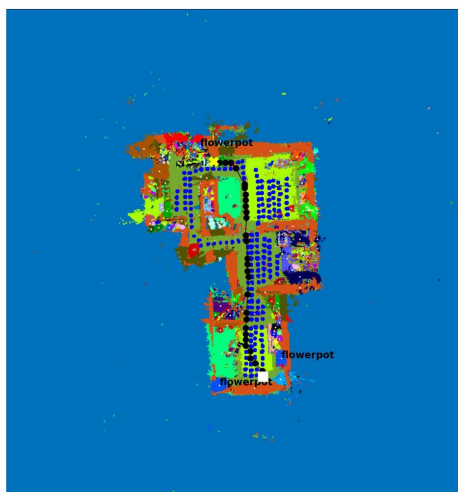
Figure 4. A qualitative result of an episode for navigating to a **kitchen sink** in **Home-005**. Our approach successfully drives the agent to the sink, while the VL-Map approach fails to approach the target object.

Instruction: Go to 'Flowerpot'

GT Map (Semantic Map)

VL-Map (Heatmap)

Home_006_1



Last View in the Navigation Episode

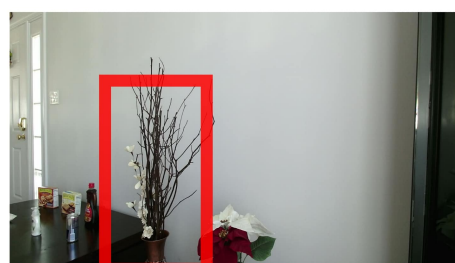
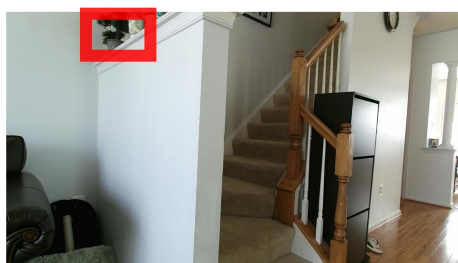


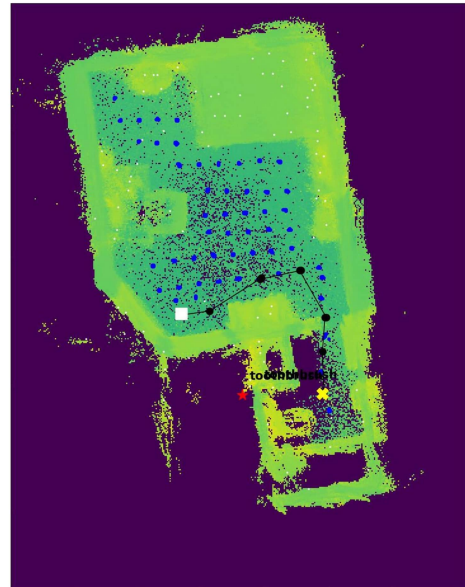
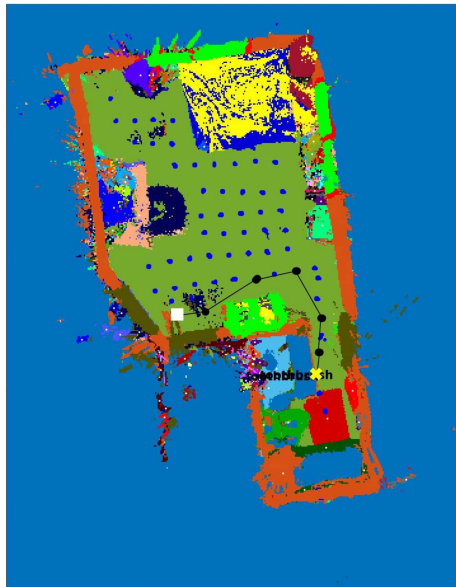
Figure 5. A qualitative result of an episode for navigating to a **flowerpot** in **Home-006**. Both approaches successfully reach the target object.

Instruction: Go to *'Toothbrush'*

GT Map (Semantic Map)

VL-Map (Heatmap)

Home_014_1



Last View in the Navigation Episode

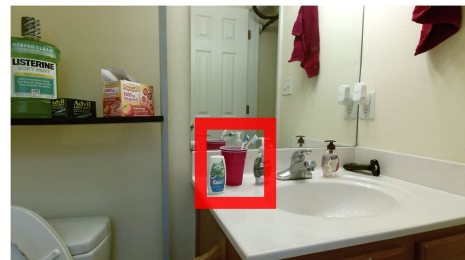
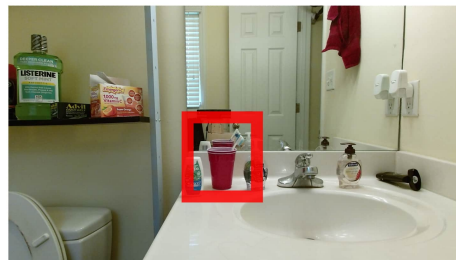


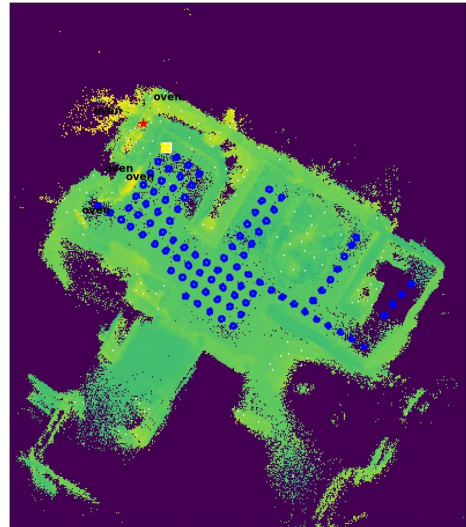
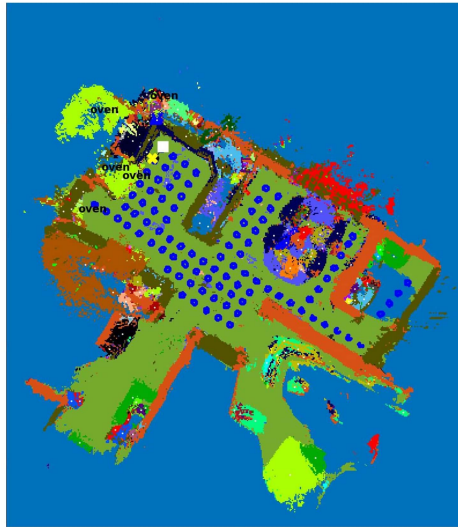
Figure 6. A qualitative result of an episode for navigating to a **toothbrush** in **Home-014**. Both approaches successfully reach the target object.

Instruction: Go to 'Oven'

GT Map (Semantic Map)

VL-Map (Heatmap)

Home_016_1



Last View in the Navigation Episode

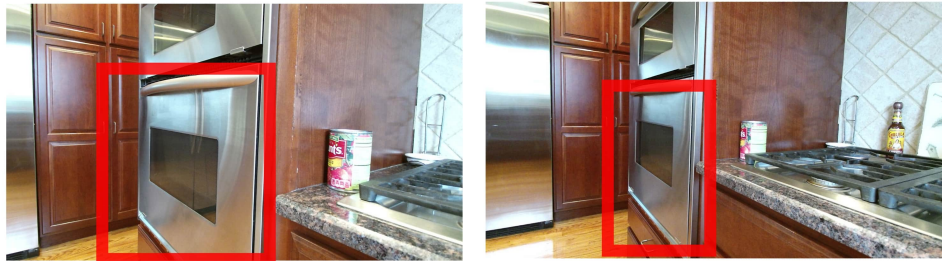


Figure 7. A qualitative result of an episode for navigating to an **oven** in **Home-016**. Both approaches successfully reach the target object.

Table 1. Detailed results for Object Goal Navigation experiment

Run #	R1				R2				R3			
	Target	VLMaps	GT-Maps	Target	VLMaps	GT-Maps	Target	VLMaps	GT-Maps	Target	VLMaps	GT-Maps
Home-001-1	remote control	1	1	sink	1	1	sofa	1	1	sofa	1	1
Home-001-2	sponge	0	1	sofa bed	1	1	cabinet	1	1	cabinet	1	1
Home-002-1	sofa	1	1	handbag	0	0	sofa	1	0	sofa	1	1
Home-003-1	dining table	1	1	toilet	1	0	fireplace	1	0	fireplace	1	1
Home-003-2	cabinet	1	1	bath towel	1	1	cabinet	1	1	cabinet	1	1
Home-004-1	refrigerator	1	1	dining table	1	1	blanket	1	1	blanket	0	1
Home-004-2	sofa	1	1	crystal hot sauce	1	1	plastic bag	1	1	plastic bag	1	1
Home-005-1	kitchen sink	0	1	refrigerator	1	1	coffee maker	1	1	coffee maker	1	1
Home-005-2	bottle	1	1	kitchen sink	1	1	faucet	0	1	faucet	0	1
Home-006-1	bath mat	1	1	box	1	1	flowerpot	1	1	flowerpot	1	1
Home-007-1	chair	1	1	chair	1	1	sofa	1	1	sofa	1	1
Home-008-1	coca cola glass bottle	0	1	cushion	1	1	refrigerator	0	1	refrigerator	0	0
Home-010-1	sofa bed	1	1	jacket	0	1	box	1	1	box	1	1
Home-011-1	hand towel	1	0	oven	1	1	nature valley sweet and salty nut cashew	1	1	nature valley sweet and salty nut cashew	1	1
Home-013-1	pringles bbq	1	1	box	0	1	refrigerator	1	1	refrigerator	1	1
Home-014-1	toothbrush	1	1	book	1	1	chair	1	1	chair	1	1
Home-014-2	desk	1	1	suitcase	0	0	drawer	0	0	drawer	0	1
Home-015-1	bottle	1	1	easel	0	0	dresser	1	0	dresser	1	1
Home-016-1	cabinet	1	1	oven	1	1	teakettle	1	1	teakettle	1	1
Average	-	84.2	94.7	-	73.6	78.9	-	78.9	78.9	-	78.9	94.7