

ManipulaTHOR: A Framework for Visual Object Manipulation

Supplementary Materials

Appendix

A. Data generation

Our data generation process starts by finding all the possible locations in each room per object. Furthermore, we need to separate out the initial object locations for which there exists a valid agent pose that can reach it. This is to ensure the objects in each task of our dataset are reachable by our agent, and there is at least a solution for picking up/dropping off the object in that location. Note that the object in the aforementioned target location can be visible or hidden from the agent’s point of view (e.g., on a table vs. in the closed fridge). In this paper we only consider the problem for visible objects. Since there is a path between any two agent locations in the room, any pair of possible locations can be considered as a task (which may or may not require navigating the agent in the scene).

B. More Dataset Stats

In Section 4, we discussed details and some statistical aspects of the dataset. In Figure 3 in the main paper, the training data split is described as a large streaming pool of tasks that can be sampled from. In Figure 1, we provide the number of tasks (a pair of possible object locations described in the previous section) per training object. Due to the physical space constraints, the number of possible object locations (and as a result the number of possible tasks) for smaller objects is higher than the bigger ones (e.g., Apple vs. Pot).

In Table 1, we illustrate the number of possible object locations per object for each data split. This emphasizes on the diversity of our dataset.

Figure 2 shows the co-occurrences of objects and their source and target receptacles. As seen, objects may lie in a diverse set of locations, requiring the agent to be able to reach for and place objects in a variety of situations.

	Train	Test	Val	Total
Apple	12523	2858	2012	17393
Bread	8003	1612	1817	11432
Tomato	12158	2641	2424	17223
Lettuce	7942	1267	1454	10663
Pot	4298	918	937	6153
Mug	6289	1555	1586	9430

Table 1: **Possible location of objects.** The distribution of possible object locations per object per data split. In total, there are 72k object locations in our dataset.

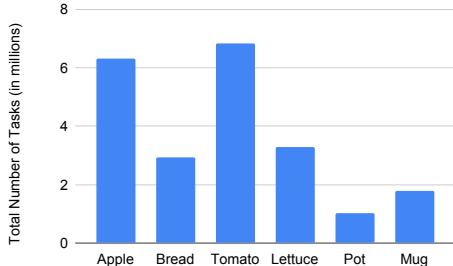


Figure 1: **Number of possible tasks.**

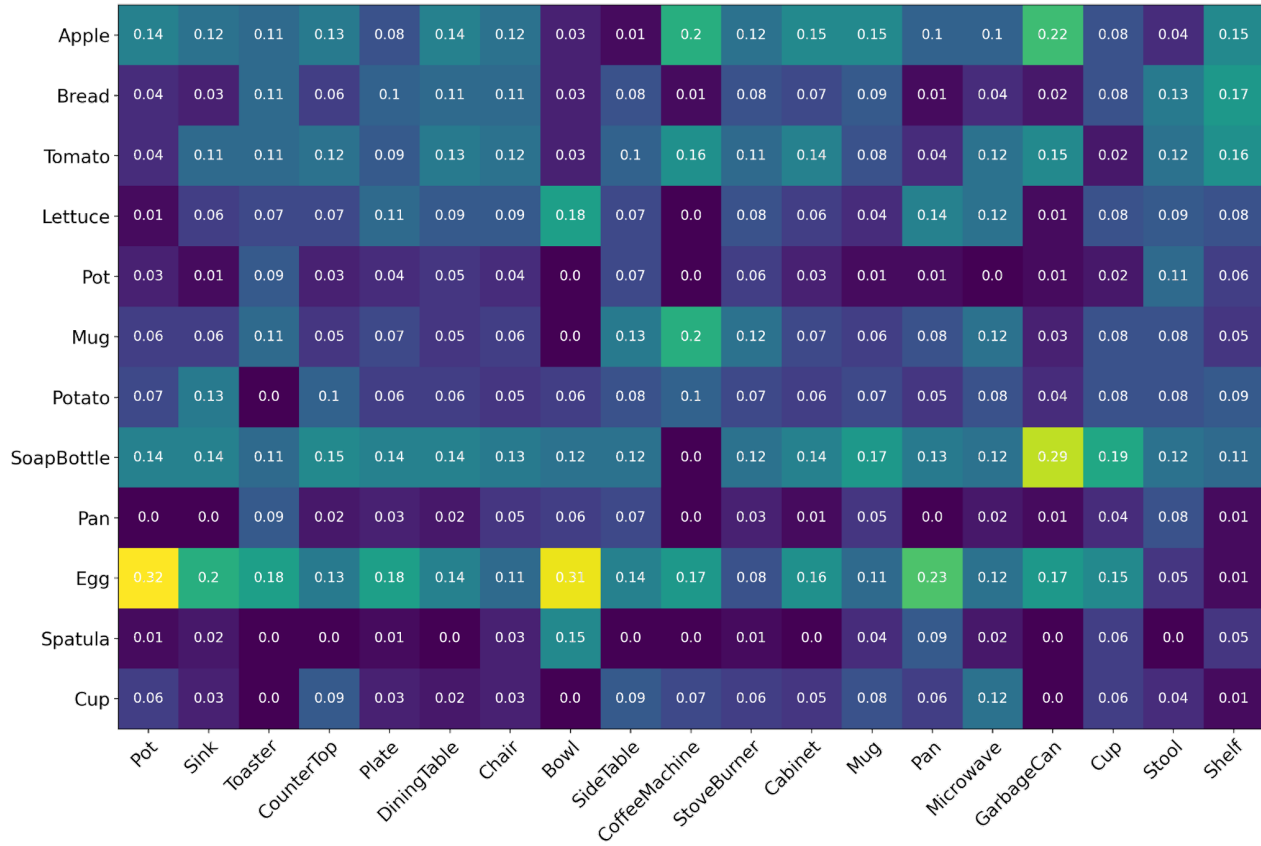


Figure 2: **Co-occurrences of objects and receptacles.** The heatmap provides the distribution of object locations on each receptacles and illustrates the diversity of our dataset.