Habitat- A Platform for Embodied AI Research: Paper Review

The authors introduce a new embodied AI research platform- Habitat, for training embodied agents in rich realistic simulators which can then be transferred to the reality. The platform consists of 2 major components. Habitat-Sim, which is a flexible high performance 3D simulator with configurable agents, multiple sensors and generic 3D dataset handling. Habitat-API, which is a modular high level library which involves defining embodied AI tasks such as navigation, question answering, or training embodied agents via Reinforcement, imitation learning or classic SLAM, and benchmarking using standard metrics. While most of the existing simulators have limitations in terms of experimentation with multiple tasks and datasets, less flexibility for large scale learning because of low frames/second, and not being able to modify the environment programmatically, Habitat provides a unifying platform which is modular, fast and is flexible. This enables to run experiments based on different paradigm and generalization of agents across datasets. The authors validate the modularity and flexiblity of this platform with cross-dataset generalization experiments with multiple sensor capabilities. They further shows comparison of visual navigation using SLAM and learning based approach and makebetween learning

The strengths of the paper lies mostly in their detailed description of the architecture of Habitat, such as software stack for multi-layer abstraction. The novelty is in introducing a new platform with unified embodied agent stack for various tasks, simulators and datasets, which allows evaluation of a broad set of tasks and experiments. The unique features of this platform are clearly demonstrated with relevant experiments, such as speed evaluation on Intel Xeon and Nvidia Titan processors, and comparison of a visual navigation tasks with a certain algorithm and various sensors. Further, the platform supports few realistic models such as noisy actuations and collision model which allows partial or no progress along the intended direction. The results and experiments are very well presented. The baselines are good enough for evaluating the performance simulator. The platform in general offers many advantages for an embodied AI research and can be used as baseline platform for many tasks.

There are few weaknesses in their method of evaluation and experiments. Firstly, finding the evidence that *learning outperforms SLAM when provided with large amount of training experiences*, might not be considered as one of the contributions of this paper, on the basis of experiments on just one simulator (*Habitat*). Few experiments of similar navigation tasks with same algorithms and same dataset in another suitable simulator could be useful in supporting this conclusion. Also, to validate the unique features of the simulator, few more experiments on various tasks (other than visual navigation) could be useful, which could actually demonstrates the *Embodied AI* aspect of the platform and show its unique features, flexibility and modularity. Further, the results demonstrated in Fig. 6 are not explained as to why navigation with just depth sensor performs better than RGBD. Further, there are no enough citations and background literature survey in the **Related work** section, while discussing the shortcomings of existing simulators. The shortcoming are explained in a very holistic way for all the existing simulators, but does not indicate existing which simulator has certain specific shortcomings.

Overall, this is a great step towards building a unified platform for an Embodied AI as it provides a single platform for testing various algorithms and tasks with different combinations of datasets, which other existing platforms do not provide. One possible research direction could be dynamics modelling of real robots (like turtlebot, husky, or few other standard robots used in the labs) and transferring the learning agent from simulation to real world. This could help building the platform more advanced and realistic and could be useful for actually transferring from sim to real with better accuracy.
I would be curious to use Habitat for my own created real-time dataset of streets, for visual navigation, and would be curious to see its performance, and where it could fail.