<u>Semi- Paramteric Topological Memory For Navigation: Review-</u>

The author introduces a hybrid approach using deep learning and classical planning algorithm, to address navigation in previously unseen environment, inspired by landmark-based navigation. The proposed architecture is a semi-parametric topological memory which consists of a non-parametric memory graph, encoding the internal representation of the environment, and the parametric deep network, called retrieval network, which estimates the similarity of two observations to predict whether they are temporally close or not. The network takes a current observation and the goal observation and generates a waypoint observation. The current observation and the waypoint observation is then provided to the locomotion network for providing short range navigation towards the goal. The internal graphical representation of the environment results in a topological graph where edges represent the similar view points or points corresponding to same time steps.
The retrieval and locomotion network are trained using self-supervised learning, where the agent first explores randomly to generate training data. The approach has been tested on DOOM environment on a goal directed navigation tasks. The approach seems to outperform all existing approaches which requires the knowledge of ego-motion or ground truth depth maps, by achieving 100% success on these navigation trials.

The strengths of the paper lies mostly in their approach where they are able to address the navigation task in unseen environment without need or access to depth map or ego-motion information, and is based only on the observed image sequence. Also, the idea of storing experiences in a graph and in using landmark similarity rather than metric embeddings is interesting. Secondly, the paper has been very clearly written, all the approaches are pictorially presented well, with decent clarity of their approach and the figures clearly represent how the network performs. The experimental results are clearly demonstrated and the results are well presented and clearly shows the advantage of this approach over existing approaches. Further, the background literature survey includes sufficient material and the introduction is motivating enough which excites the readers to read this paper.

However, there are few weaknesses in their method of evaluation and experiments. Firstly, the evaluation has been done on a single environment for a single task (goal directed navigation), which is not sufficient to analyze the performance of the architecture. There should be a broad range of navigation tasks on different domains where the images might not have enough textures (features) as present in Doom environment. Also, if the environment is full of aliased features, the architecture might not be able to encode the internal topological representation of the environment might be very inappropriate . Further, the individual evaluation of the performance of various components of the architecture like retrieval and locomotion network might be useful to demonstrate which part of the network is actually responsible for this performance in DOOM. Further, it is clear from the experiments in Fig. 6, that the agent is not taking the best route towards the goal for Track 2 . Therefore, one baseline could be a goal directed navigation given the map of the environment, for large scale maps. However, the approach does not seem to scale well to large scale maps, therefore, the analysis for different scales of maps could strengthen the analysis of the results.

Overall, this is a novel approach towards addressing navigation task by building a topological graph, which creates a good benchmark for other future work in similar areas. One possible research direction could be extending this to real robots a sim-to-real fashion. I would be curious to use apply to this more realistic images, which have aliasing, lighting differences, different types of textures, etc.