The Role of Deep Learning in SLAM

Irem ARPAG

April 11, 2021

SLAM that is defined shortly as the problem of constructing a global representation from local measurement makes great progress in the last ten years with the incremental role of deep learning in the research area. While traditional SLAM methods simply build geometric maps of static worlds and localize a localize a robot, modern SLAM methods is expected to synthesize data from heterogeneous sensors to infer object categories, operate in dynamic environments by building richer maps and also support planning. When the success of the machine learning methods for perception tasks especially in computer vision that is becoming more effective with the usage of large-scale labelled datasets are combined with the increasing efficiency of classical geometric estimation, it is begun to be awaited from SLAM methods to perform special scene understanding tasks and concurrently carry out robust, reliable, and vivid operations equipped with various sensors. Such kind of improvements motivated research in object level representation and specifically in semantic information to result to call to whole system's interaction with the environment as spatial AI.

Some significant research areas including navigation in dynamic and semi-static environments, abstractions, and hierarchy, and learned representation are standing out with the unification of semantic and geometric information enriched by deep learning that leads to SLAM system to go beyond points and planes algorithms and thus targets such as active or interactive perception, intelligent exploration, manipulation, and learning as a task in itself are on the agenda. SLAM methods integrated with semantic understanding have the ability of incorporating objects and places with a more expressive map representations and afford higher level autonomy. Semantic maps can be coded with smaller memory by reducing the aggregation of errors caused by geometric computations and also sustain human operators to engage with autonomous systems. With the extension of deep learning methods, learned perception models for object detections, recognition and pose estimation, that is not only estimating the class and position of objects but also their orientation and extend, such as through learned object descriptors which captures the shape of objects as well as the pose of them, supersede the previous object models. Modelling semi static environments in which environmental change happens on a timescale is another fundamental challenging subject of SLAM community that need to apply deep learning approaches for building updatable maps that is extremely important for autonomous vehicle systems which rely on high-resolution 3D maps for navigation. The research area of spatial-semantic hierarchy deals with the construction and maintenance of cognitive maps of the environment with the assistance of several learning-aided planning tasks such as neural network-based encoder on topological and scene graph mapping. Combining spatial and temporal information to improve learning is the domain of the learning-centered approaches and further interest is required for development.

Although great progress is tackled in SLAM systems within the last ten years for more comprehensive understanding of the world through end-to-end learning, there is still some key issues to be evaluated as feature research area like long-term autonomy or lifelong map learning, both of which are the tasks need to be dominated by the implementations of deep learning approaches to achieve the target of versatile and intelligent operations.