Late breaking results on Graph Reasoning on Situational Graphs for higher-concepts generation

Jose Andres Millan-Romera[1], Hriday Bavle[1], Holger Voos[1], and Jose Luis Sanchez-Lopez[1]

I. Motivation

To enhance Simultaneous Localization and Mapping (SLAM) in complex environments, our framework S-Graphs [1] presented a novel factor graph compounded by robot keyframes associated with a scene graph that includes higher-level semantic-relational concepts. However, it only includes simple and rigid *room* structures and lacks other useful entities such as *walls* (as a *wall surface* to *wall surface* relationship). To overcome this, we propose to generate new factorized subgraphs using the relational and generalisation power of Graph Neural Networks (GNN) [2] by:

- A novel *wall surface* clustering approach leveraging inference of pair-wise relations.
- New *room* and *wall* node feature definition from already computed *wall surface* node embeddings.

II. Problem Statement and Methodology

We aim to define new *room* and *wall* nodes from a set of *wall surface* nodes defined as 2D finite lines.

The solution proposed is inspired in [2]. In the preliminary approach, no edges embedding are used and the initial nodes embeddings, v_i^0 are defined as $[c_i, length_i, n_i]$ where c_i is the centroid of i_{th} node and n_i, the normal from the observed side. As a final approach, v_i is defined as $[length_i, n_i]$ and initial edge embedding, e_{ij}^0 as, $[c_j - ci, cd_{ij}, n_j - n_i]$ where cd_{ij} is the closest distance. The embeddings are updated through the GNN iterations following Eq. 1 and Eq. 2. Where $g_v()$ and $g_e()$ are multi-layer perceptrons, N(i) are the neighbours of i_{th} node and $NN()$ is SAGEConv for the preliminary approach and Feature-wise Attention for the final approach.

$$v_i^{l+1} = g_v([v_i^l, \max_{j \in \mathcal{N}(i)}(NN(v_i^l, e_{ij}^l, v_j^l))] \tag{1}$$

$$e_{ij}^{l+1} = g_e([v_i^l, e_{ij}^l, v_j^l]) \tag{2}$$

As no prior graph structure can be leveraged, new edges are artificially created in the message-passing graph by node proximity. Ground truth edges contain node relations indicating nodes which belong to the same *room* or *wall*, label the message-passing graph for training of GNN convolutions which update the embeddings of nodes and edges. After the GNN convolutions, the edge embeddings in the last layer are passed to a linear layer to classify if each edge corresponds to one of the desired relations. Subsequently, the set of predicted relations is used to cluster nodes into *rooms* and *walls* by chained edges of the same type. Once clustered, new *room* and *wall* nodes and *belong to same room* or *belong to same wall* edges are generated. Their features are obtained by employing the same GNN architecture as in Eq. 2 to generate new embeddings. Finally, the node embeddings from Eq. 1 are fed to a linear layer to fit the node feature dimension.

III. Preliminary Results

The preliminary results for *same room* relations obtained with the described preliminary approach are presented in Fig. 1. The data employed has been generated from a new synthetic dataset to randomize squared rooms. The results demonstrate the network successfully learns to correctly classify at training time. However, it fails to generalize when exposed to new validation data. It motivates the evolution to the final approach as future work.

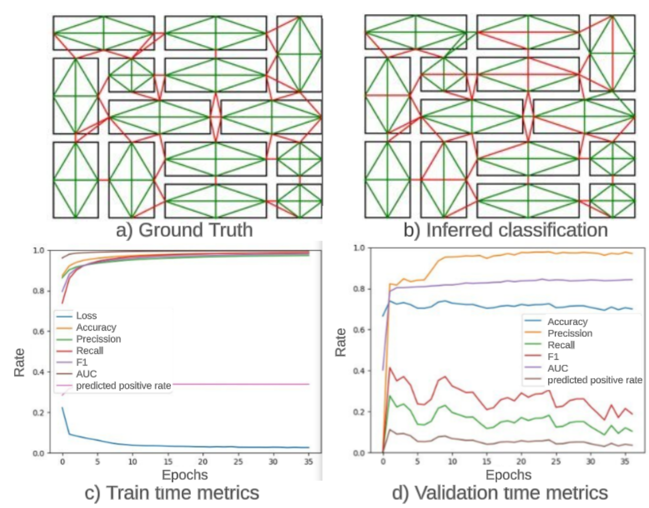

Fig. 1: A), B) black lines: wall surfaces; green/red lines: positive/negative edges. D) Validation after each epoch.

[1]Authors are with the Automation and Robotics Research Group, Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg (UL). Holger Voos is also associated with the Faculty of Science, Technology and Medicine, University of Luxembourg, Luxembourg. {jose.millan, hriday.bavle, holger.voos, joseluis.sanchezlopez}@uni.lu

*This work was partially funded by the Fonds National de la Recherche of Luxembourg (FNR) under the project 17097684/RoboSAUR.

References

[1] H. Bavle, J. L. Sanchez-Lopez, M. Shaheer, J. Civera, and H. Voos, "S-graphs+: Real-time localization and mapping leveraging hierarchical representations," *IEEE Robotics and Automation Letters*, 2023.

[2] S.-C. Wu, J. Wald, K. Tateno, N. Navab, and F. Tombari, "Scenegraph-fusion: Incremental 3d scene graph prediction from rgb-d sequences," in *IEEEConference on Computer Vision and Pattern Recognition*, 2021.