Evaluation of Visual Semantic Navigation Models in Real Robots

C. Gutiérrez-Álvarez[1*], P. Ríos-Navarro[1*], R. Flor-Rodríguez-Rabadán[1], F. J. Acevedo-Rodríguez[1] and R. J. López-Sastre[1]

I. Motivation

Visual Semantic Navigation (VSN) is the ability of a robot to learn visual semantic information for navigating in unseen environments. Visual representations are learnt to reduce the exploration time and better generalize to unseen scenes and objects. These VSN models are typically tested in those virtual environments where they are trained, hence, we do not yet have an in-depth analysis of how these models behave in the real world. In this work we propose, for the first time, a detailed study of how several state-of-the-art VSN models behave when embedded in real robotic platforms. We release a novel ROS-based framework[2] for VSN, so that any VSN-model can be easily deployed in any ROS-compatible robot and tested in the real world. Fig. 1 depicts how these VSN models work when deployed on our robotic platforms.

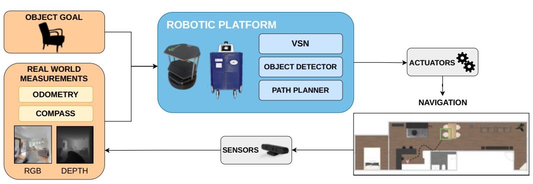

Fig. 1. ROS VSN framework working scheme.

II. ROS Visual Semantic Navigation

The ROS framework we present consists of two packages: `ros_visual_semantic_navigation` (R_{VSN}) and `ros_discrete_move` (R_{DM}). The R_{VSN} package embeds the VSN models in a ROS node that predicts discrete actions from visual inputs taken from robot sensors. The ROS nodes and services provided by the R_DM package convert the discrete actions predicted by the model to continuous input velocities. These velocities are sent to the robot motors to produce the movement associated with the predicted action.

III. Main Results

We have performed an experimental evaluation of two state-of-the-art VSN models (VLV [1] and PIRLNav [2]) in two robotics platforms, by using our novel ROS-based software architecture. The robots invited to our experiments have been Turtlebot 2 and LOLA2 [3]. We have designed a navigation experiment that mimics the evaluation performed in the ObjectNav [4] task. First, we have defined 15 starting positions in an apartment of 75 m^2 (Fig 2 shows 2 of these positions and the trajectories followed by the robot). From these positions, the robot is asked to navigate towards the different object categories that every VSN model has been trained to reach. We perform navigation experiments for every target and measure the success of the episode if the robot reaches the selected object category in less than 150 steps, without collisions. We report the success rate (SR) of the VSN models as the percentage of episodes in which navigation was successful. Table I compares the SR obtained by these VSN models in virtual environments and in our real-world experimental setup.

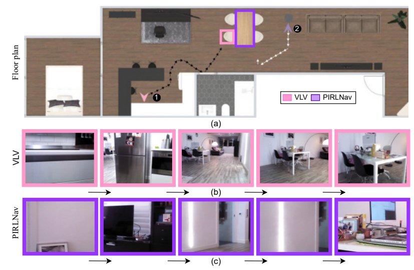

Fig. 2. VSN results. a) Shows the floor plan where the experiments where performed, indicating two of the starting positions used. b) and c) show the egocentric view of the robot while navigating using VLV and PIRLNav models, respectively.

Models	SR (Real world)	SR (virtual environment)
VLV [1]	29.33%	79%
PIRLNav [2]	21.11%	65%

TABLE I
Real world success rate against simulation.

There is an important gap that we hope can be studied and reduced thanks to the use of our VSN ROS library. More results are shown in this video[3]. Finally, our VSN system is shown to robustly operate through long-duration experiments. With the LOLA2 and Turtlebot2 our VSN ROS architecture was running for more than 38 hours, and the robots traveled a cumulative distance of 5.22 km.

*Equal contribution.

[1]University of Alcalá, Department of Signal Theory and Communications, Spain. Email: carlos.gutierrezalva@uah.es

[2]Code will be released.

[3] Link: https://youtu.be/s1rem3c6fw8.

References

[1] M. Chang, A. Gupta, and S. Gupta, "Semantic Visual Navigation by Watching Youtube Videos," in *NeurIPS*, 2020.
[2] R. Ramrakhya, D. Batra, E. Wijmans, and A. Das, "PIRLNav: Pretraining with Imitation and RL Finetuning for ObjectNav," in *CVPR*, 2023.
[3] N. Nasri, R. J. López-Sastre, S. P. da Costa, I. Fernández-Munilla, C. Gutiérrez-Álvarez, T. Pousada-García, F. J. Acevedo-Rodríguez, and S. Maldonado-Bascón, "Assistive robot with an ai-based application for the reinforcement of activities of daily living: Technical validation with users affected by neurodevelopmental disorders," *Applied Sciences*, 2022.
[4] D. Batra, A. Gokaslan, A. Kembhavi, O. Maksymets, R. Mottaghi, M. Savva, A. Toshev, and E. Wijmans, "ObjectNav Revisited: On Evaluation of Embodied Agents Navigating to Objects," in *arXiv:2006.13171*, 2020.