BEyond observation: an approach for ObjectNav

Daniel V. Ruiz and Eduardo Todt
Department of Informatics, Federal University of Paraná, Brazil

UFPR

Visão Robótica Imagem

CVPR VIRTUAL JUNE 19-25

In this work, we present our exploratory research of how sensor data fusion and state-of-the-art machine learning algorithms can perform the Embodied Artificial Intelligence (E-AI) task called Visual Semantic Navigation. This task, a.k.a Object-Goal Navigation (ObjectNav) consists of autonomous navigation using egocentric visual observations to reach an object belonging to the target semantic class without prior knowledge of the environment. Our method reached fourth place on the Habitat Challenge 2021 ObjectNav on the Minival phase and the Test-Standard Phase.

The main aspects of our method are:

- Use of semantic segmentation to distill visual information
- Projection to a top-down semantic 2D grid to ease spatial awareness
- Decoupled global and local planner so modules can be replaced
- Modified reward function combining reinforcement and imitation learning to improve convergence.

Two variants were proposed Cartesian and Polar.

Cartesian:

- Policy prediction x,y represents the global point goal. It is converted to integers and fed to A* to compute the closest non-obstructed local point goal.
- The local point goal is transformed into episodic polar coordinates and fed to the local planner along with the RGB-D observation.

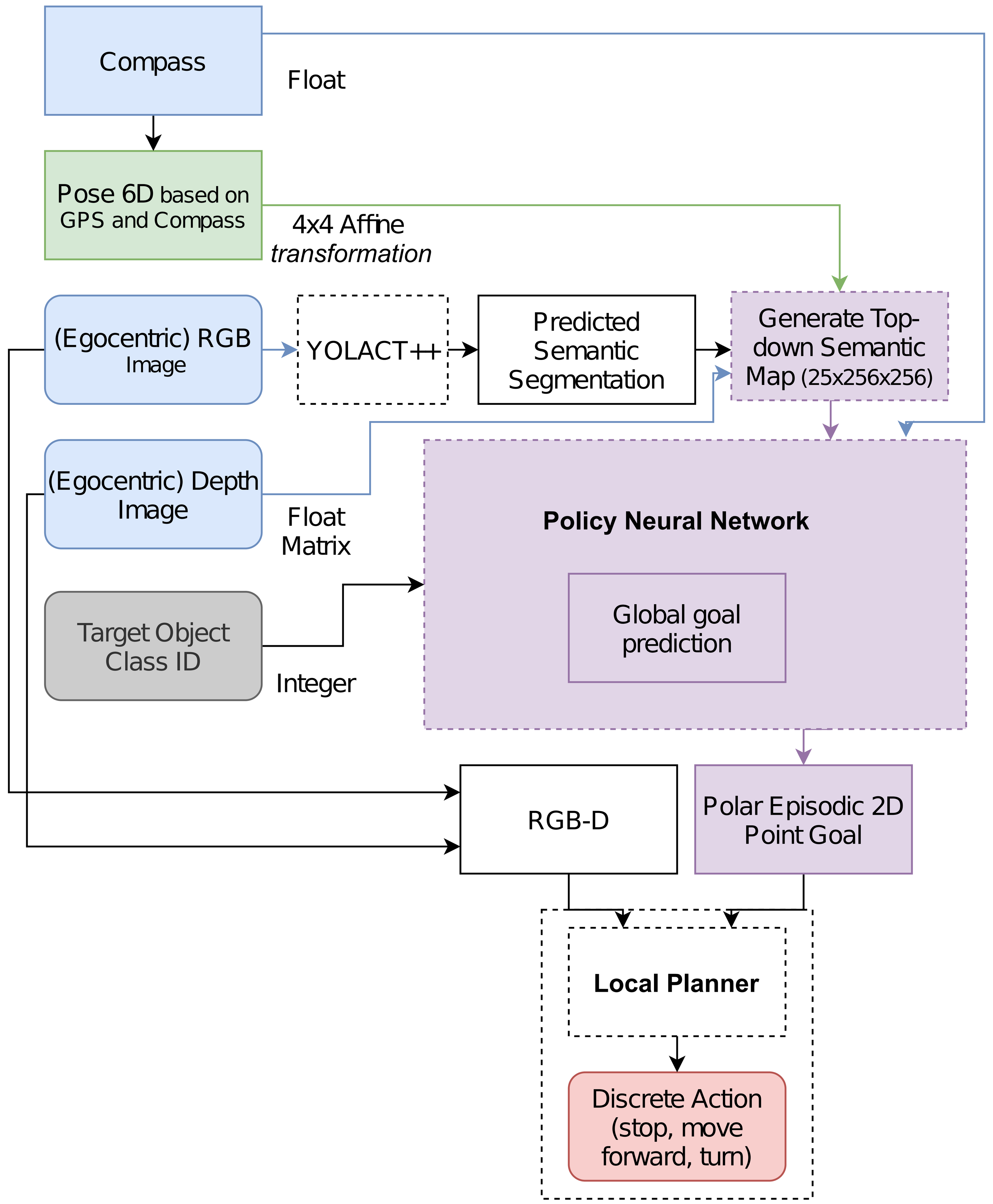
Polar:

- Policy prediction x,y is converted to ρ in [0,1] range, φ in [-π,+π] range and already represents the local point goal.
- The local point goal is fed to the local planner along with the RGB-D observation like in the Cartesian variant.

Reward Function:

$$f(a,d,s) = -10^{-4} + f'(a) + (d_{t-1} - d_t) + 2.5 s_t$$

BEyond flowchart. In green is the resulting pose using GPS and Compass data provided by the simulation. In blue are the Compass value, RGB, and depth images observed at step *t*. In gray the target class ID provided by the episode's settings. In purple are the module that creates the top-down semantic map, our policy neural network that performs the global goal prediction and the subsequent transformation to episodic polar coordinates. In red is the output containing the predicted action.



Habitat Challenge 2021: ObjectNav Leaderboard Minival Phase

Rank	Team	SPL	SoftSPL	Distance to goal	Success rate
1	TreasureHunt	0.15	0.25	3.41	0.27
2	Alstar (RL)	0.09	0.16	3.32	0.23
3	Clueless-Wanderers (Peter)	0.03	0.10	4.41	0.13
4	**BEyond-VRI-UFPR**	**0.00**	**0.14**	**5.71**	**0.00**
5	See through pixels (init)	0.00	0.00	6.39	0.00
6	Black Swan	0.00	0.01	6.38	0.00

Habitat Challenge 2021: ObjectNav Leaderboard Test-Standard Phase

Rank	Team	SPL	SoftSPL	Distance to goal	Success rate
1	TreasureHunt	0.09	0.17	9.23	0.21
2	Alstar (RL)	0.03	0.11	9.41	0.10
3	Clueless-Wanderers (Peter)	0.02	0.10	9.07	0.07
4	**BEyond-VRI-UFPR**	**0.00**	**0.08**	**10.18**	**0.00**
5	Habitat Team (RGBD+DD-PPO)	0.00	0.01	10.33	0.00

We achieved fourth place on the Minival phase and Test-Standard phase, marking our method as a promising approach for dealing with ObjectNav, but with a large room for future improvements.

Contact: dvruiz@inf.ufpr.br, todt@inf.ufpr.br

Public code at: https://github.com/VRI-UFPR/BeyondSight/tree/beyond_habitat_challenge_2021