Uncertainty-Informed Active Perception for Open Vocabulary Object Goal Navigation

Utkarsh Bajpai, Julius Rückin, Marija Popović and Cyrill Stachniss

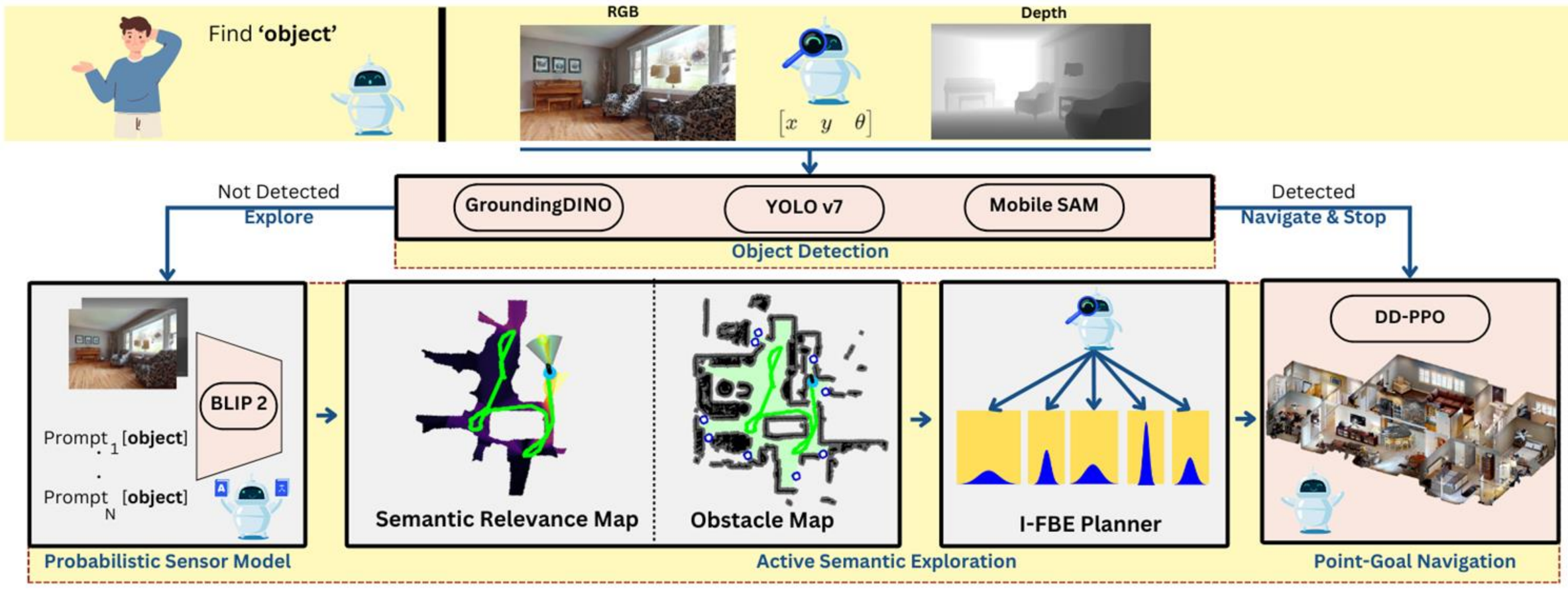
PHENOROB

UNIVERSITÄT BONN

Abstract

- Robots use vision-language models to find named objects in images.
- Existing methods rely on prompt engineering but ignore semantic uncertainty in prompt phrasing.
- We model semantic uncertainty probabilistically and integrate it into a geometric-semantic map.
- Our uncertainty-informed planner performs Object Goal Navigation without prompt engineering.

Method

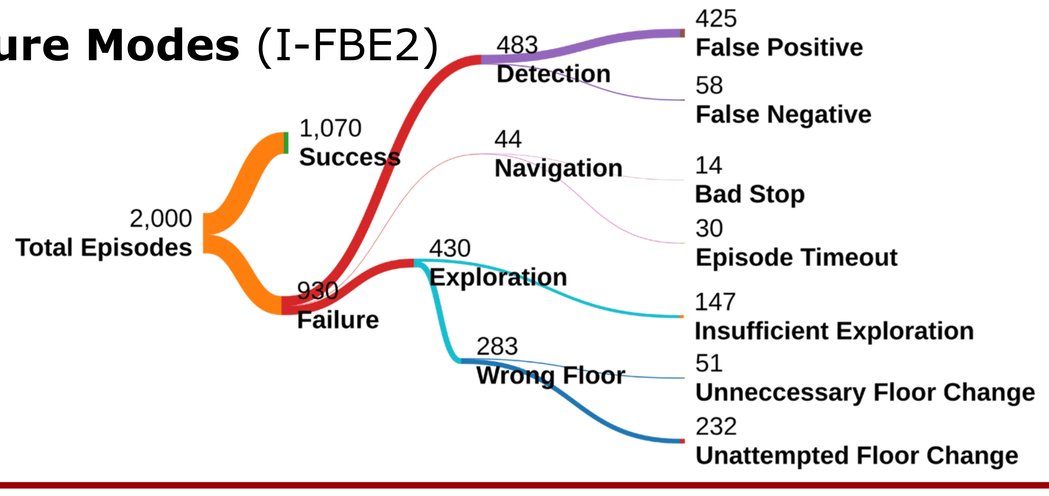


Experiments and Results

Prompt choice directly **influences** ObjectNav **performance**

Prompt	SR ↑	SPL ↑
Seems like there is a `target_object` ahead	52.60	30.42
A place where `target_object` can be found	51.00	29.71
A `target_object` can be in the vicinity	53.20	31.20
Seems like a `target_object` is ahead	53.20	30.50
A `target_object` is in the vicinity	51.65	28.67
`target_object` likely ahead	52.45	29.86
`target_object`	50.60	28.28
Ours (I-FBE1)	52.25	28.96
Ours (I-FBE2)	53.50	27.31

Failure Modes (I-FBE2)



Our uncertainty informed planners **do not require** hand engineered prompts and **perform comparably** to SOTA ObjectNav approaches

Approach	HM3D		MP3D	
	SR↑	SPL↑	SR↑	SPL↑
Closest-FBE	11.80	9.34	-	-
Random-FBE	37.30	23.32	-	-
ZSON	25.50	12.60	15.30	4.80
CoW	-	-	7.40	3.70
ESC	39.20	22.30	28.70	14.20
VLFM	52.60	**30.40**	**36.40**	**17.50**
Ours (I-FBE1)	52.25	28.96	35.26	16.47
Ours (I-FBE2)	**53.50**	27.31	35.63	16.52