Navigating to Objects in the Real World


Theophile Gervet [1]


Soumith Chintala [4]


Dhruv Batra [3,4]


Jitendra Malik [2,4]


Devendra Chaplot [4]

1 Carnegie Mellon University

2 Berkeley UNIVERSITY OF CALIFORNIA

3 Georgia Tech

4 Meta AI


Unseen environment: No experience, No map
Inputs
Toilet
Goal Category
Observation (RGBD)
(x, y, θ)
Pose Sensor
Output
Action


Spatial Scene Understanding
Navigable Space Detection


Spatial Scene Understanding
Navigable Space Detection
Semantic Scene Understanding
Object Detection
Chair
Oven

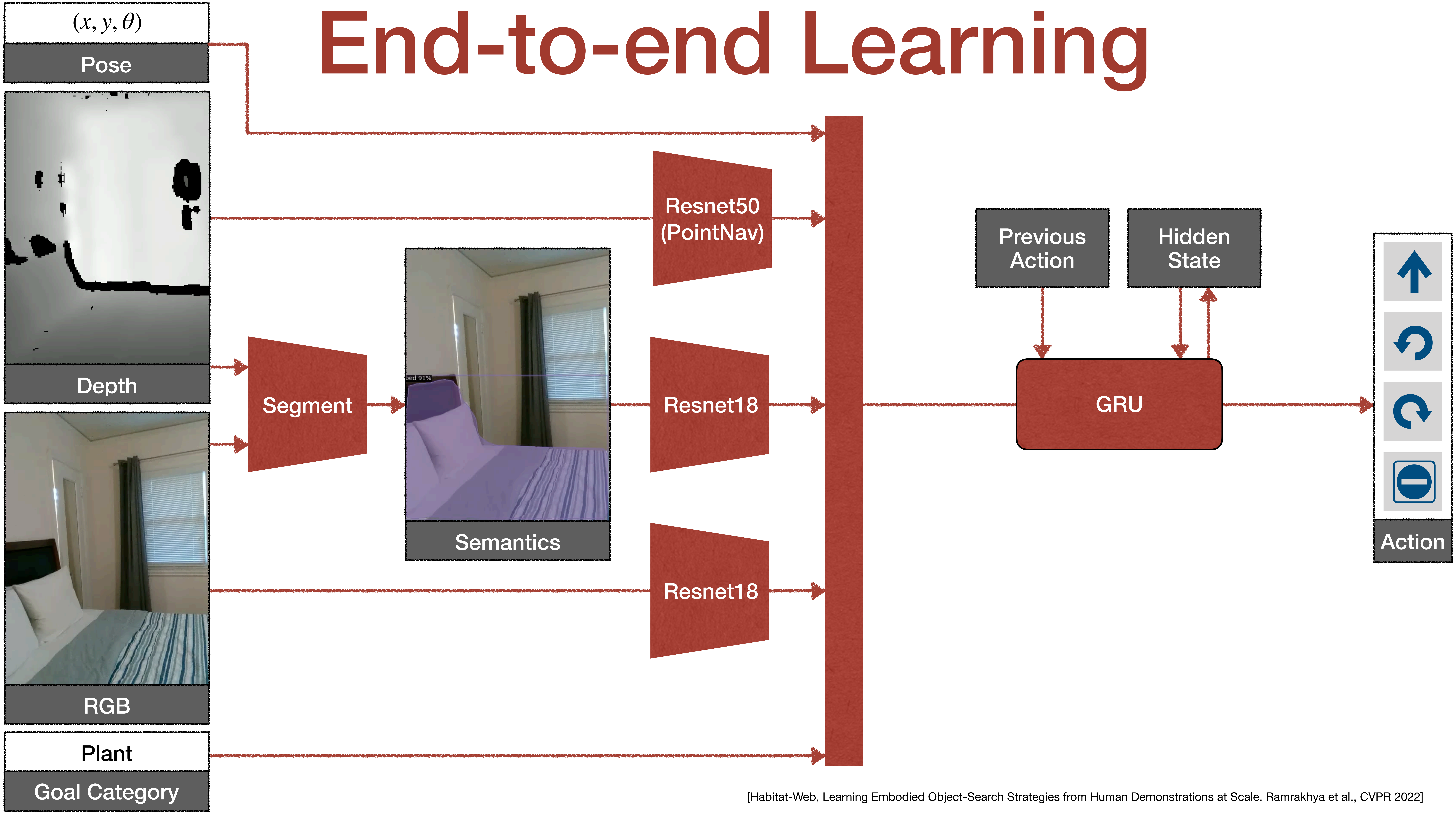
Spatial Scene Understanding
Navigable Space Detection
Semantic Scene Understanding
Object Detection
Semantic Exploration Priors
Where is a toilet more likely to be found?

Spatial Scene Understanding
Navigable Space Detection
Semantic Scene Understanding
Object Detection
Semantic Exploration Priors
Where is a toilet more likely to be found?
Episodic Memory
Keep track of explored and unexplored areas
Unexplored
Explored

Classical Navigation

[A Frontier-based Approach for Autonomous Exploration. Yamauchi, *CIRA 1997*]

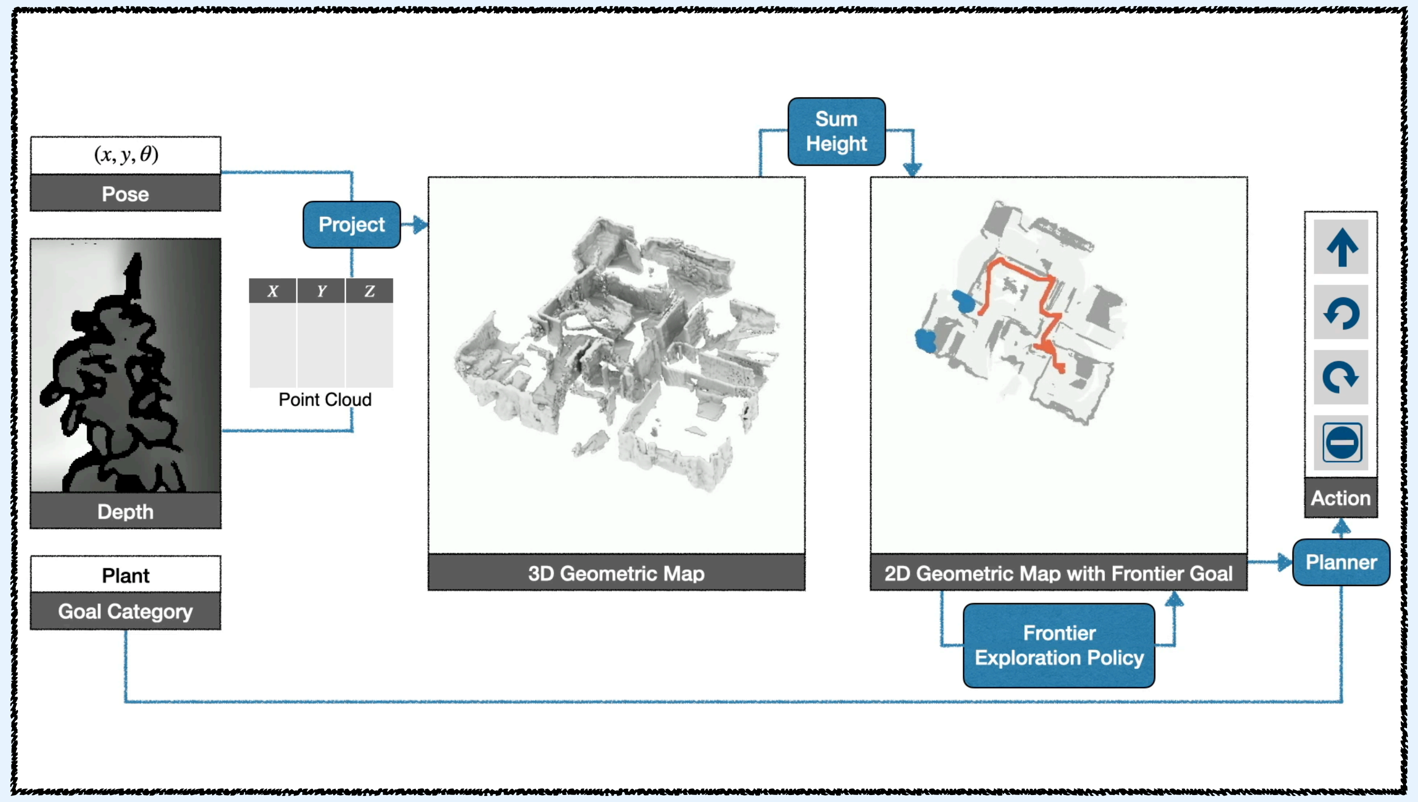
End-to-end Learning



[Habitat-Web, Learning Embodied Object-Search Strategies from Human Demonstrations at Scale. Ramrakhya et al., CVPR 2022]

Modular Learning

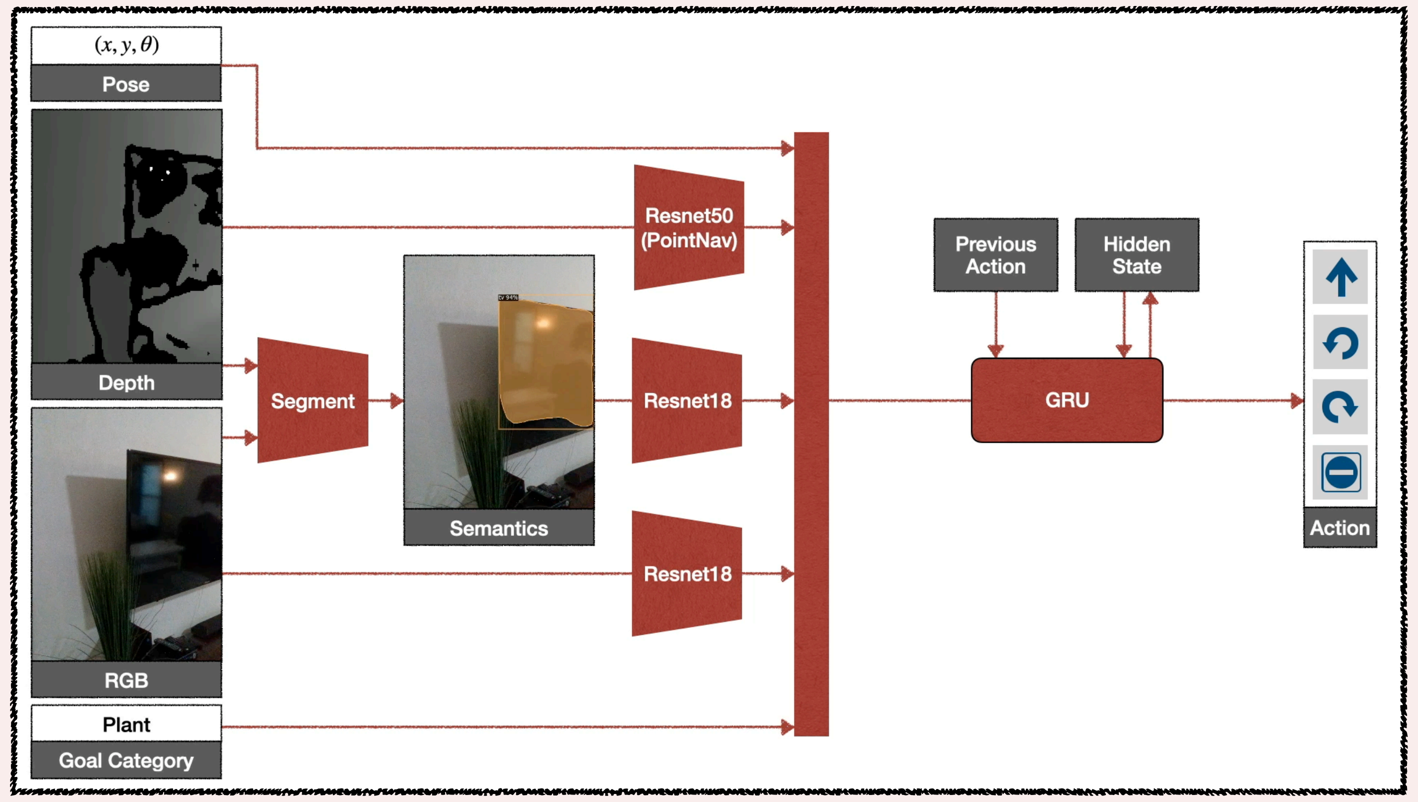
Classical



- Modular
- Explicit Memory/Maps & Planning
- Heuristic Policy

- ➕ Long-term Memory and Planning
- ➖ Semantic Exploration Priors

End-to-end Learning



- End-to-end
- Implicit Memory & Planning
- Learned Policy

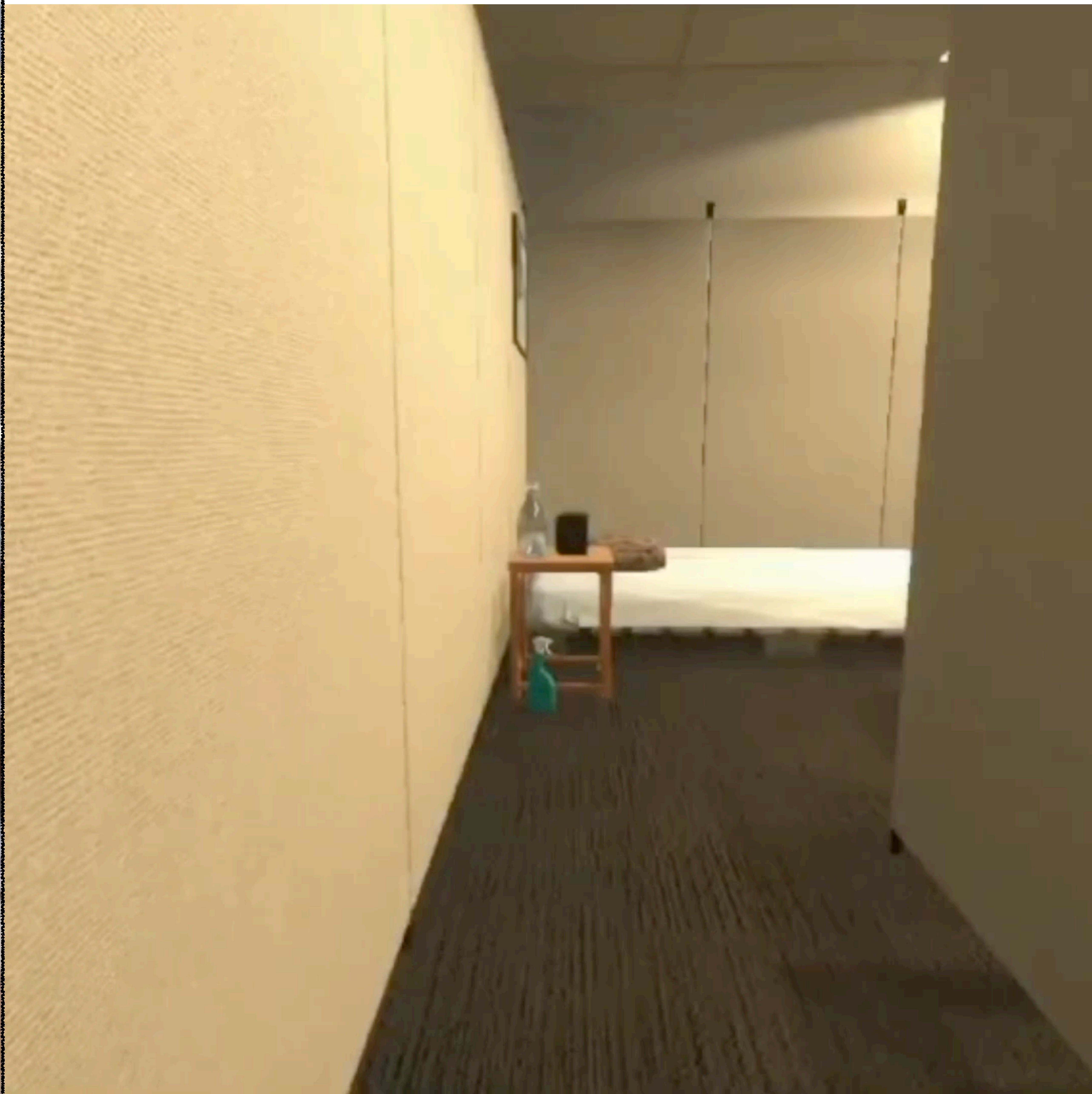
- ➖ Long-term Memory and Planning
- ➕ Semantic Exploration Priors

Modular Learning

[Object Goal Navigation Using Goal-Oriented Semantic Exploration. Chaplot et al., NeurIPS 2020]

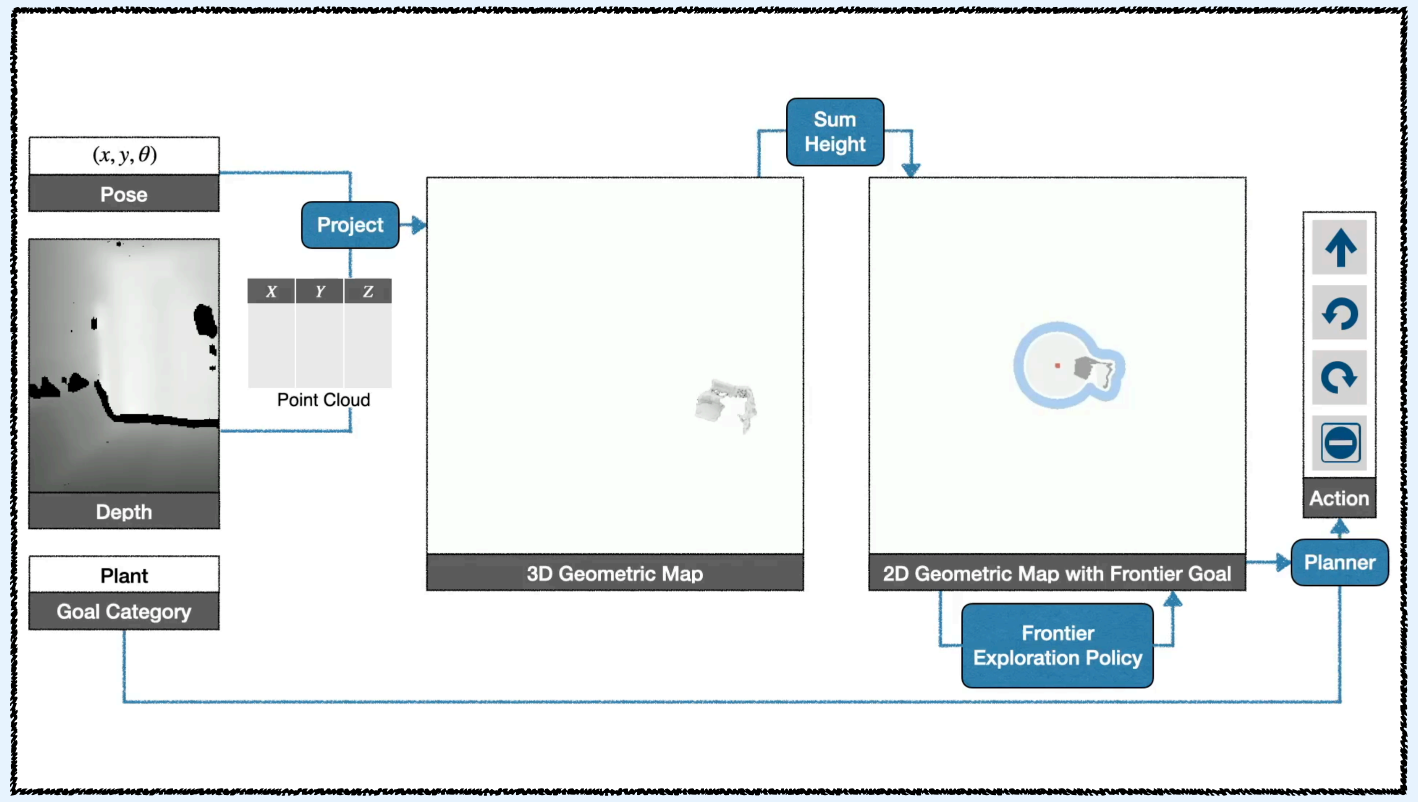
Habitat

AI2-Thor



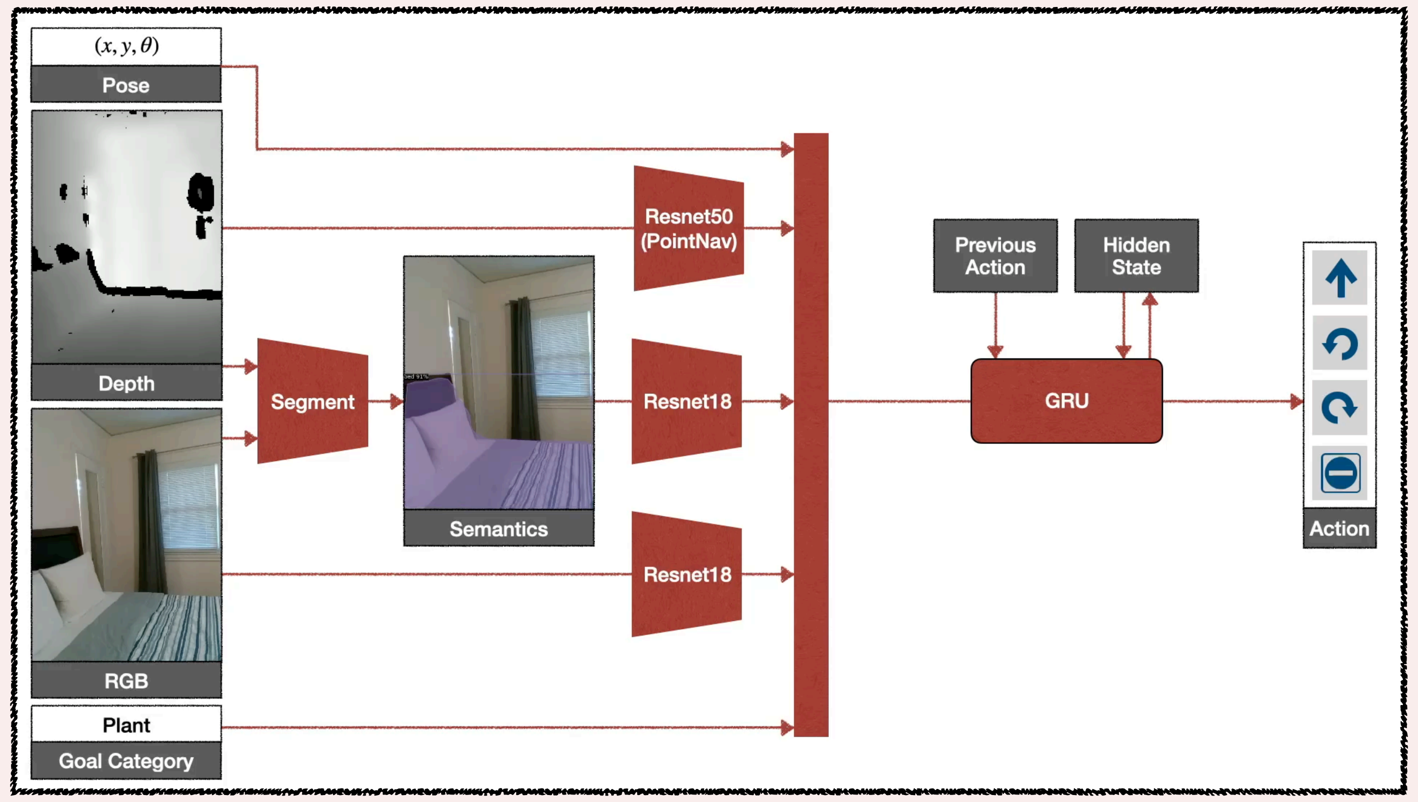
Methods

Classical



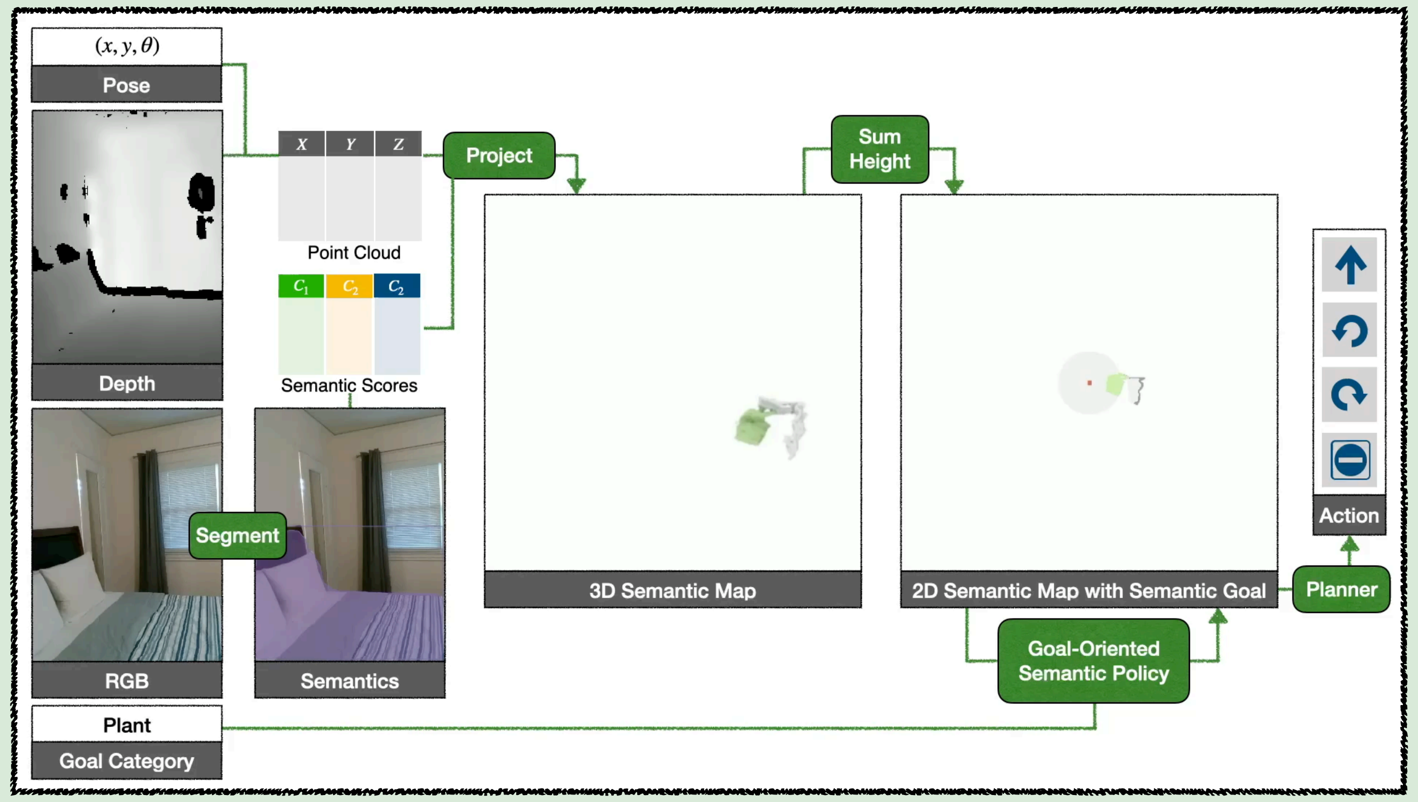
- Geometric Map
- Heuristic Exploration
  - No Training

End-to-end Learning



- End-to-end
- Large-scale IL + RL fine-tuning
  - 77,000 human trajectories
  - 200M frames of RL

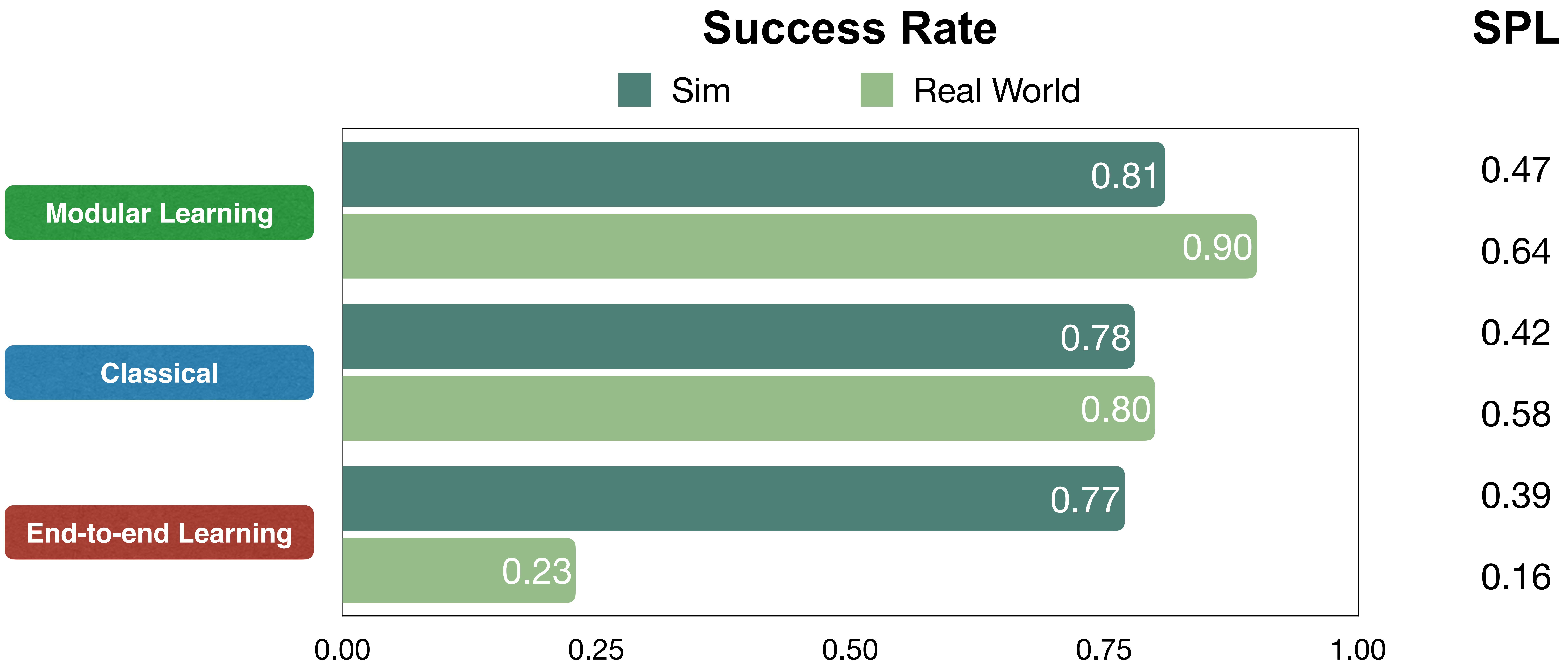
Modular Learning



- Semantic Map
- Goal-Oriented Exploration
  - 10M frames of RL


Empirical Evaluation
3 Approaches
6 Unseen Homes
6 Goal Object Categories

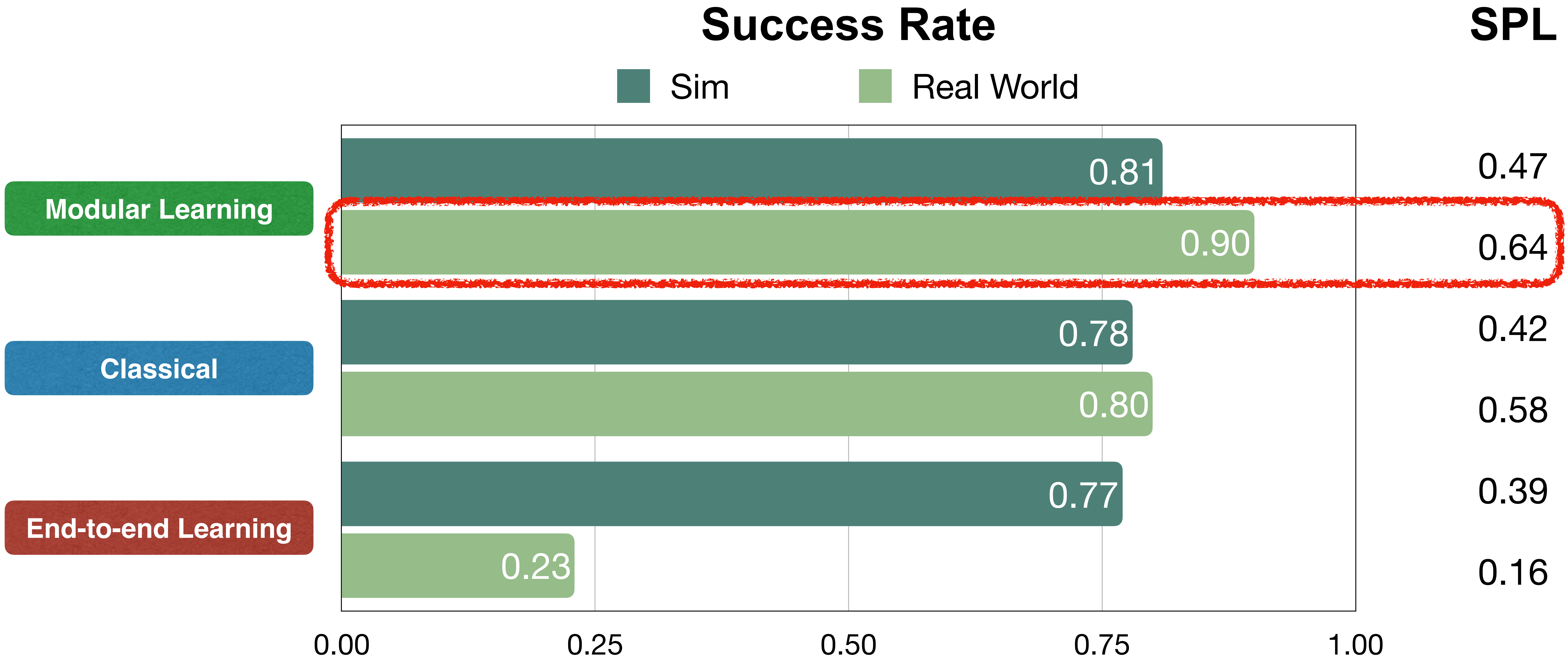
Results



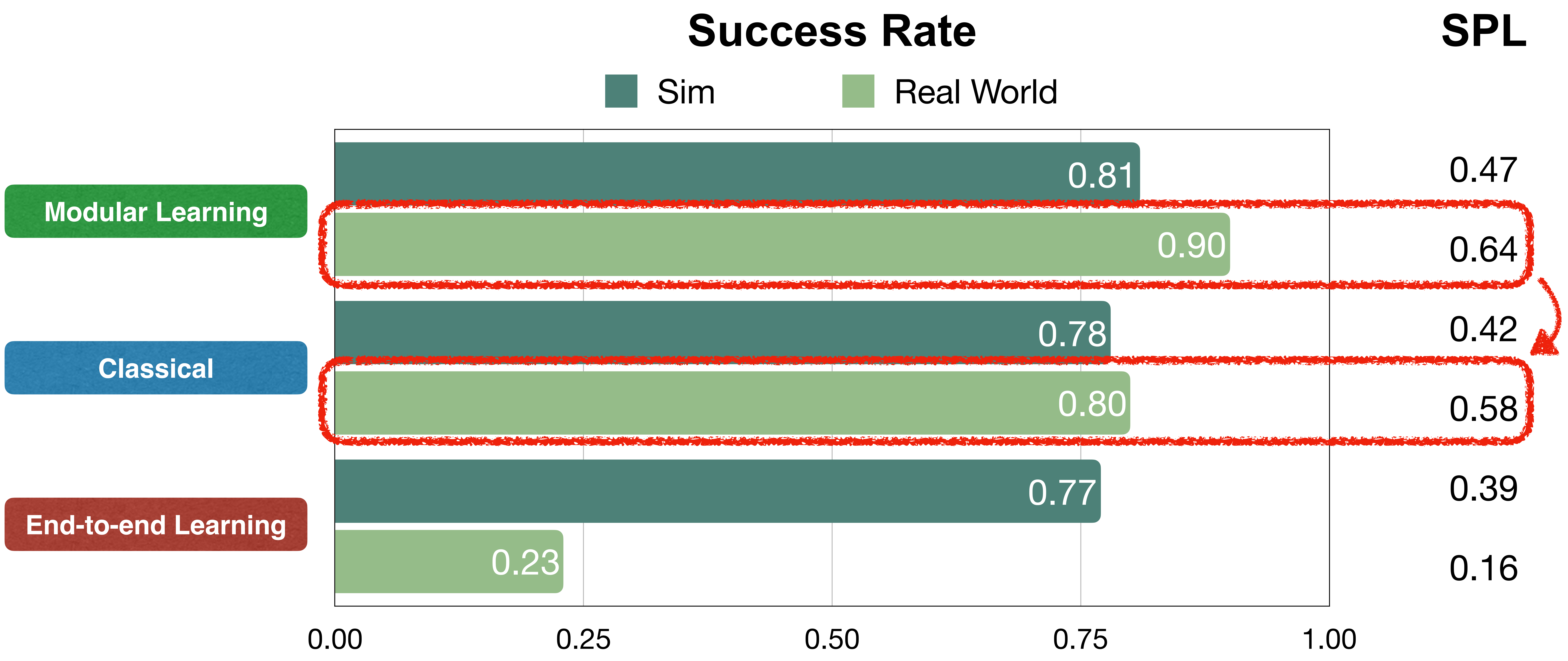
Success Rate
SPL
Sim
Real World
Modular Learning
Classical
End-to-end Learning
0.81
0.90
0.78
0.80
0.77
0.23
0.47
0.64
0.42
0.58
0.39
0.16
0.00
0.25
0.50
0.75
1.00

Goal: *couch*
SPL: 0.74, 78 steps
SPL: 0.0, 121 steps
SPL: 0.33, 181 steps
Modular
End-to-End
Classical
Third-person view
Third-person view
Third-person view
Success
Failure
Success
Observation
Predicted Semantic Map
Observation
Predicted Semantic Map

Modular Learning is Reliable



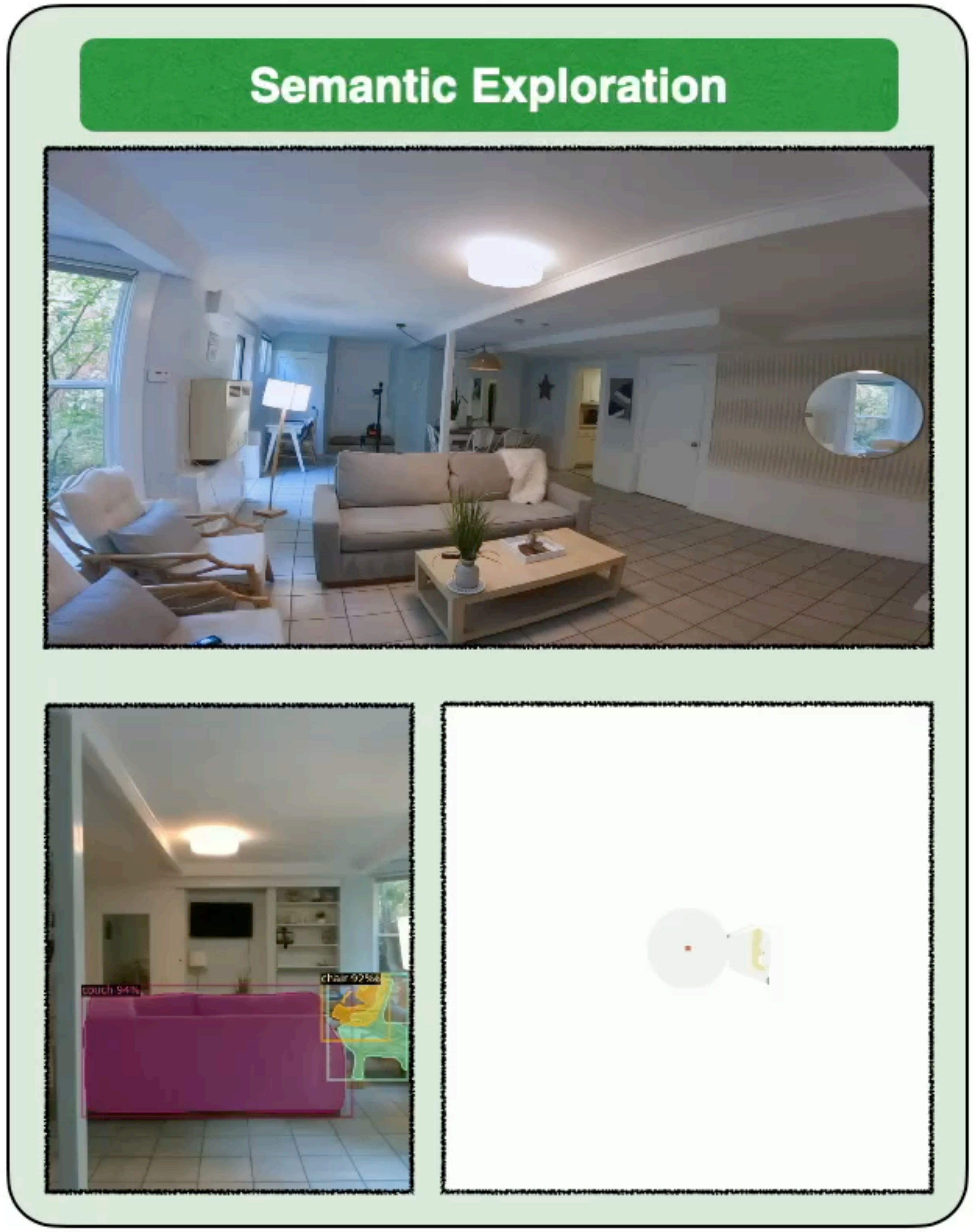
Classical vs Modular Learning



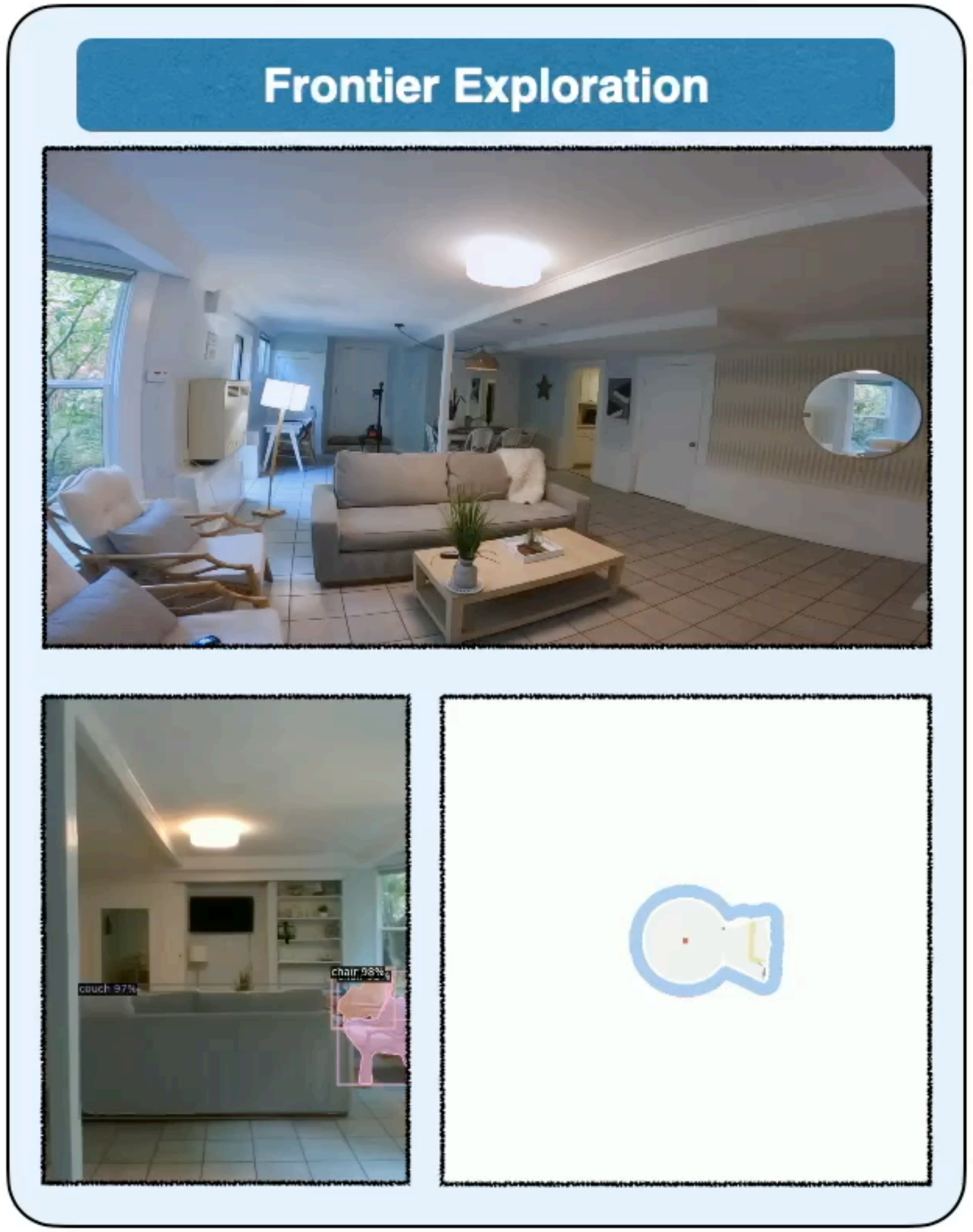
Classical vs Modular Learning

Goal: *bed*

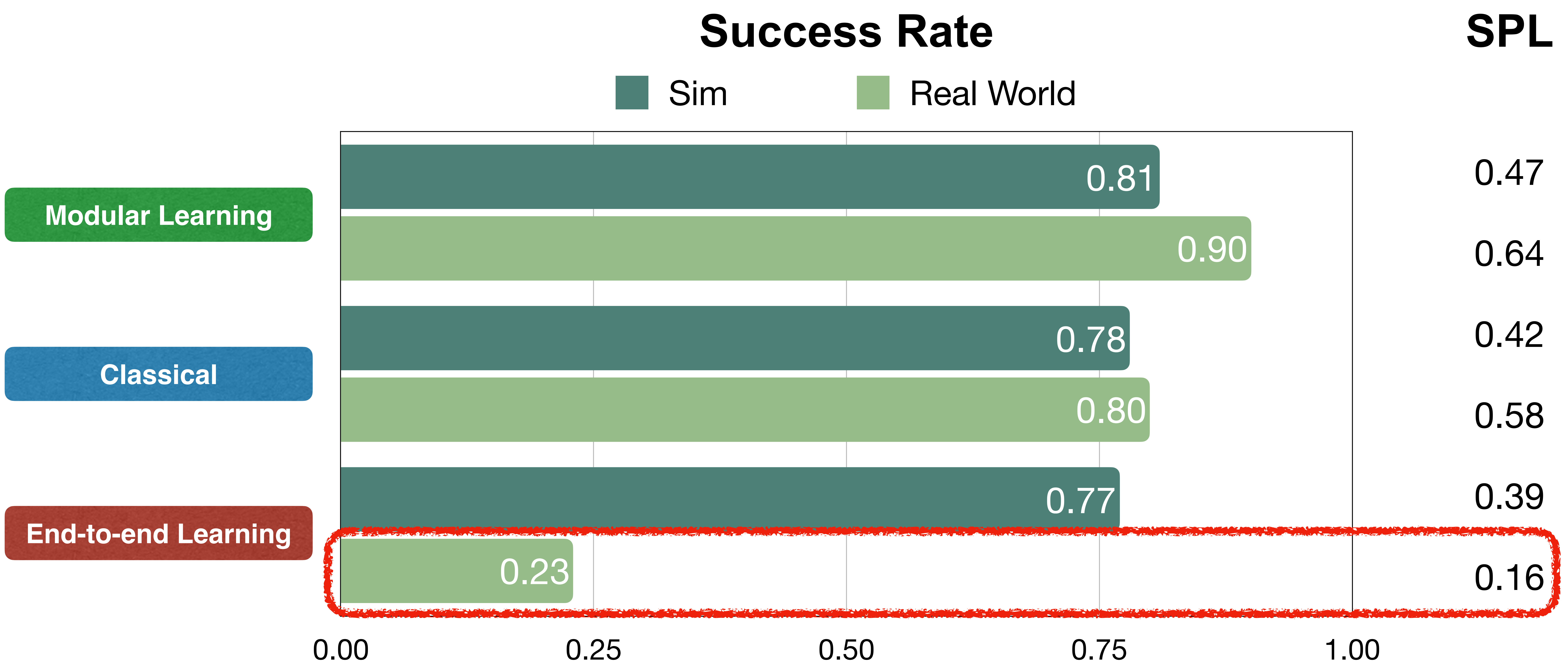
SPL: 0.90, 98 steps



SPL: 0.52, 152 steps

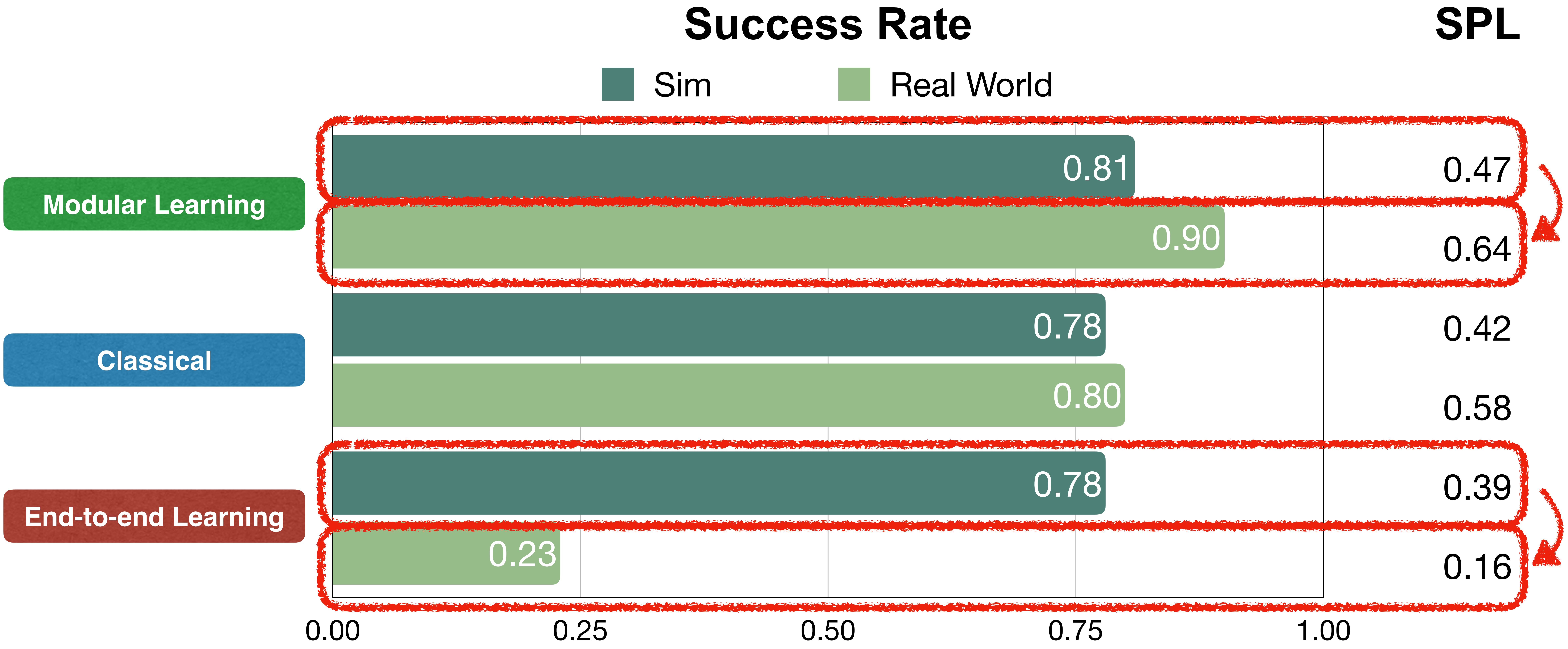


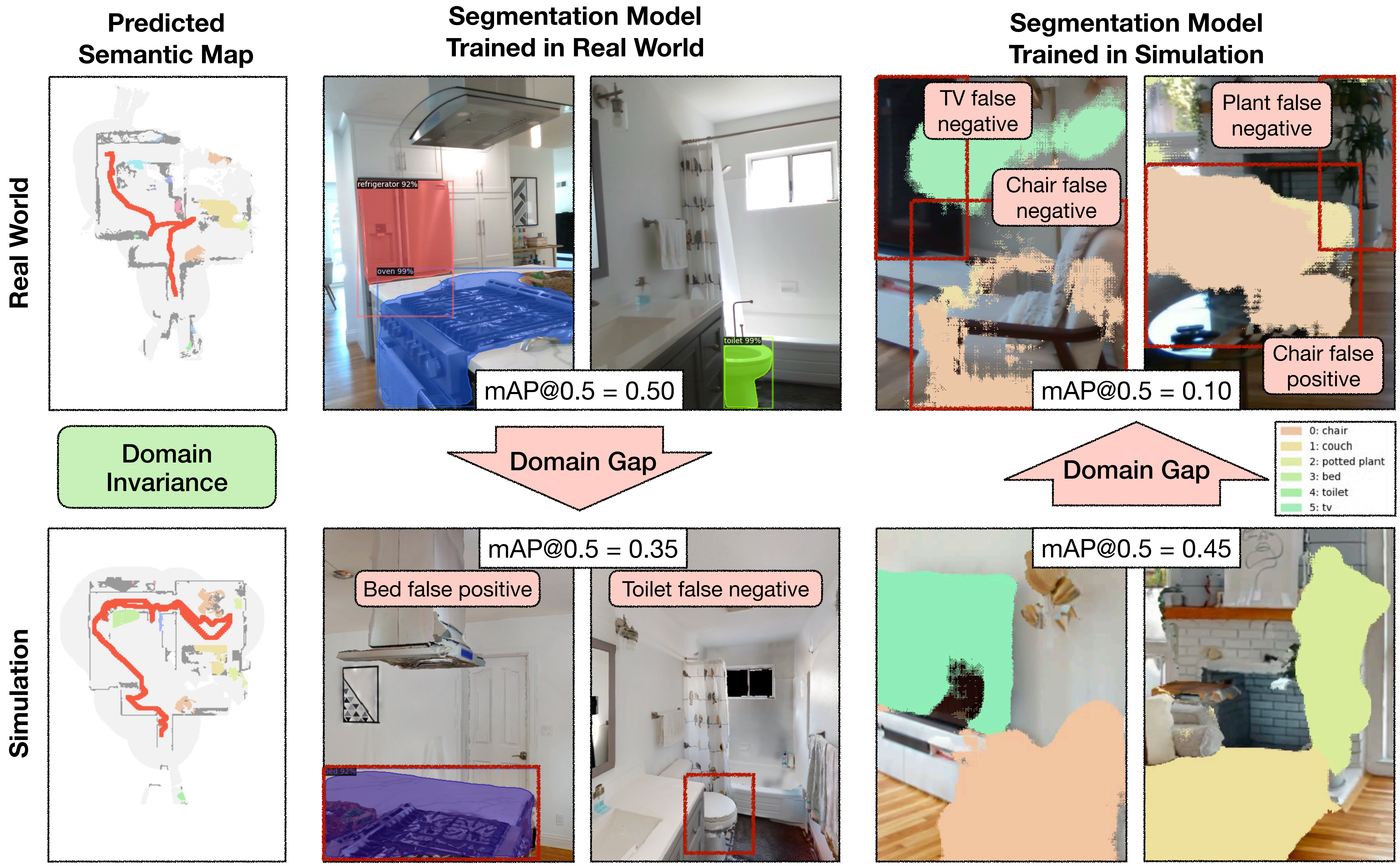
End-to-end fails to Transfer



End-to-end Failures

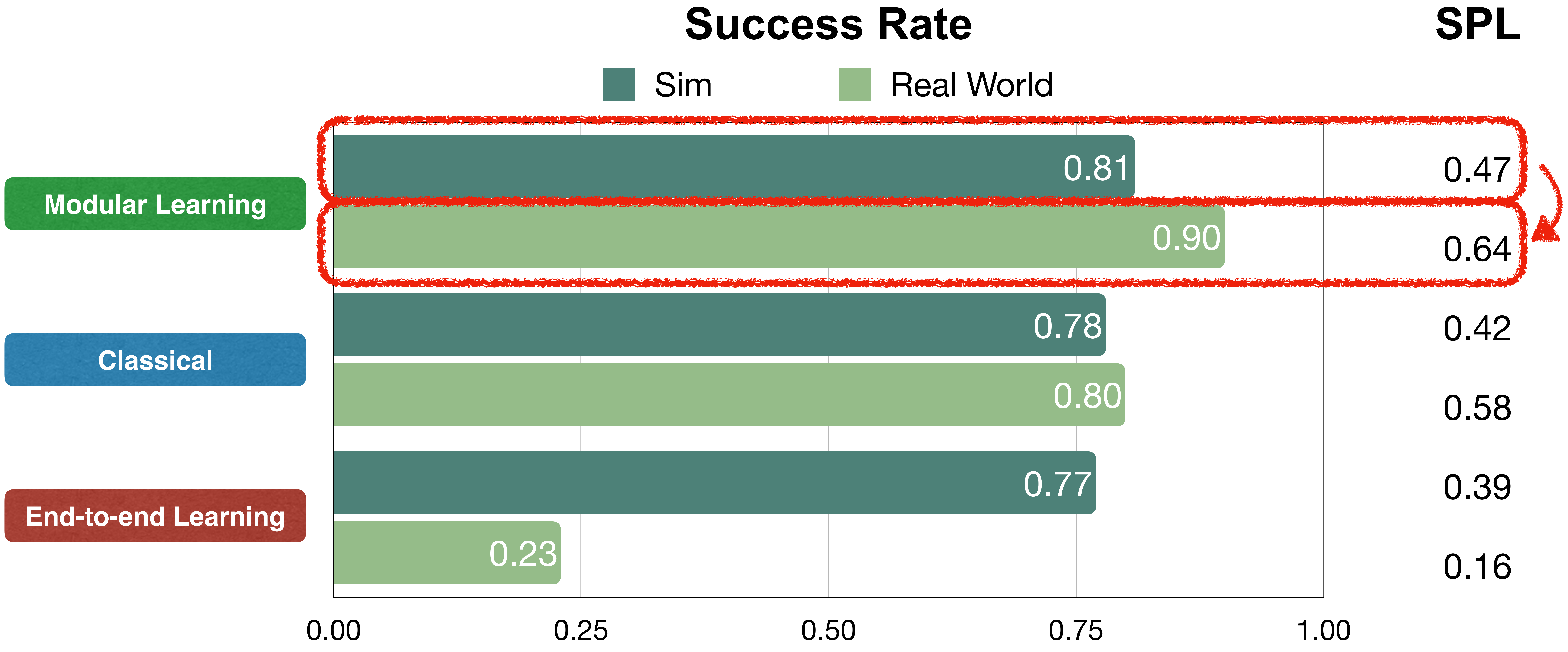
Modular vs End-to-end Transfer



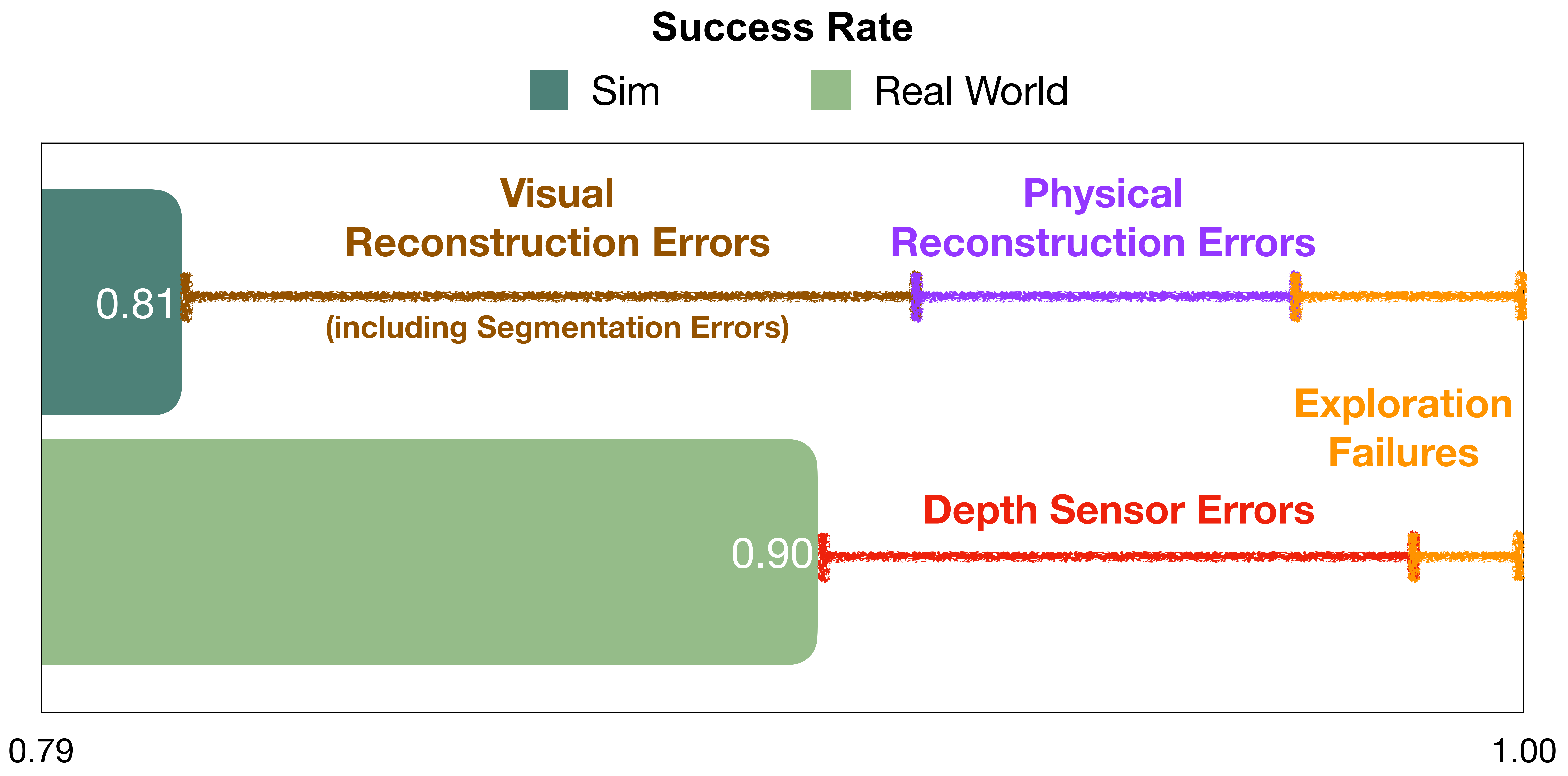


Predicted Semantic Map
Segmentation Model Trained in Real World
Segmentation Model Trained in Simulation
Real World
Simulation
refrigerator 92%
oven 99%
toilet 99%
mAP@0.5 = 0.50
TV false negative
Plant false negative
Chair false negative
Chair false positive
mAP@0.5 = 0.10
Domain Invariance
Domain Gap
Domain Gap
0: chair
1: couch
2: potted plant
3: bed
4: toilet
5: tv
mAP@0.5 = 0.35
Bed false positive
Toilet false negative
mAP@0.5 = 0.45

Modular Learning Sim vs Real



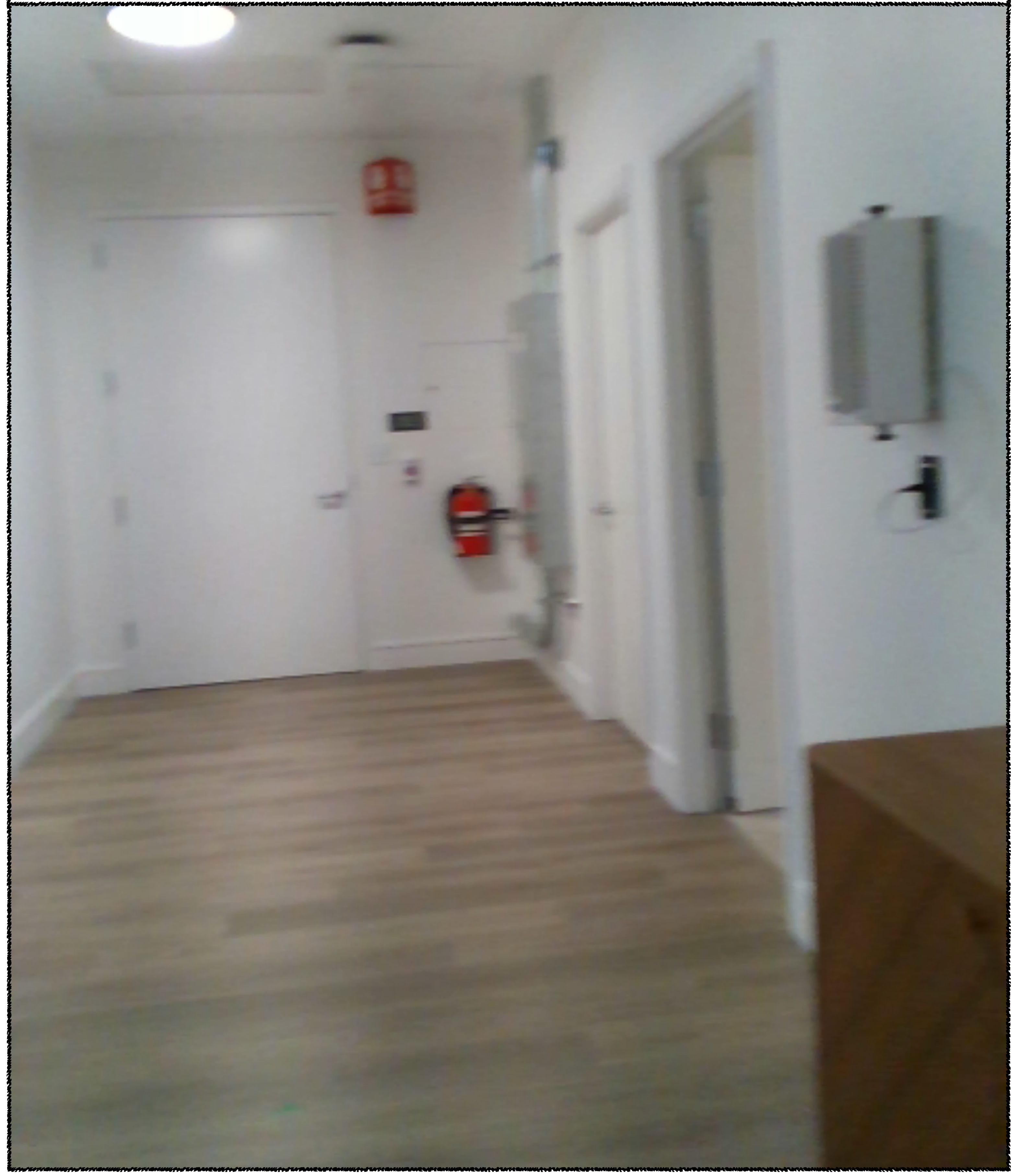
Modular Learning Sim vs Real

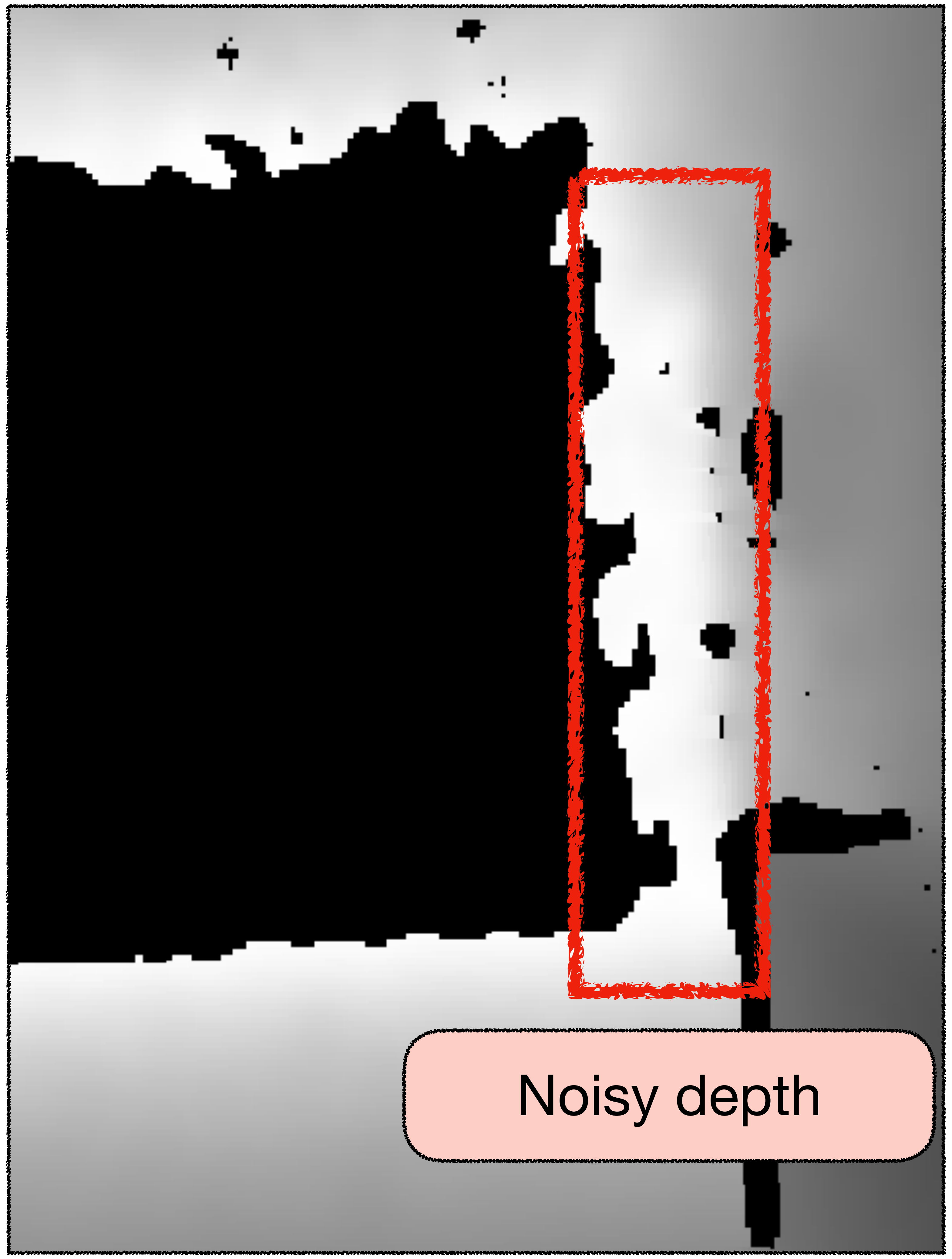


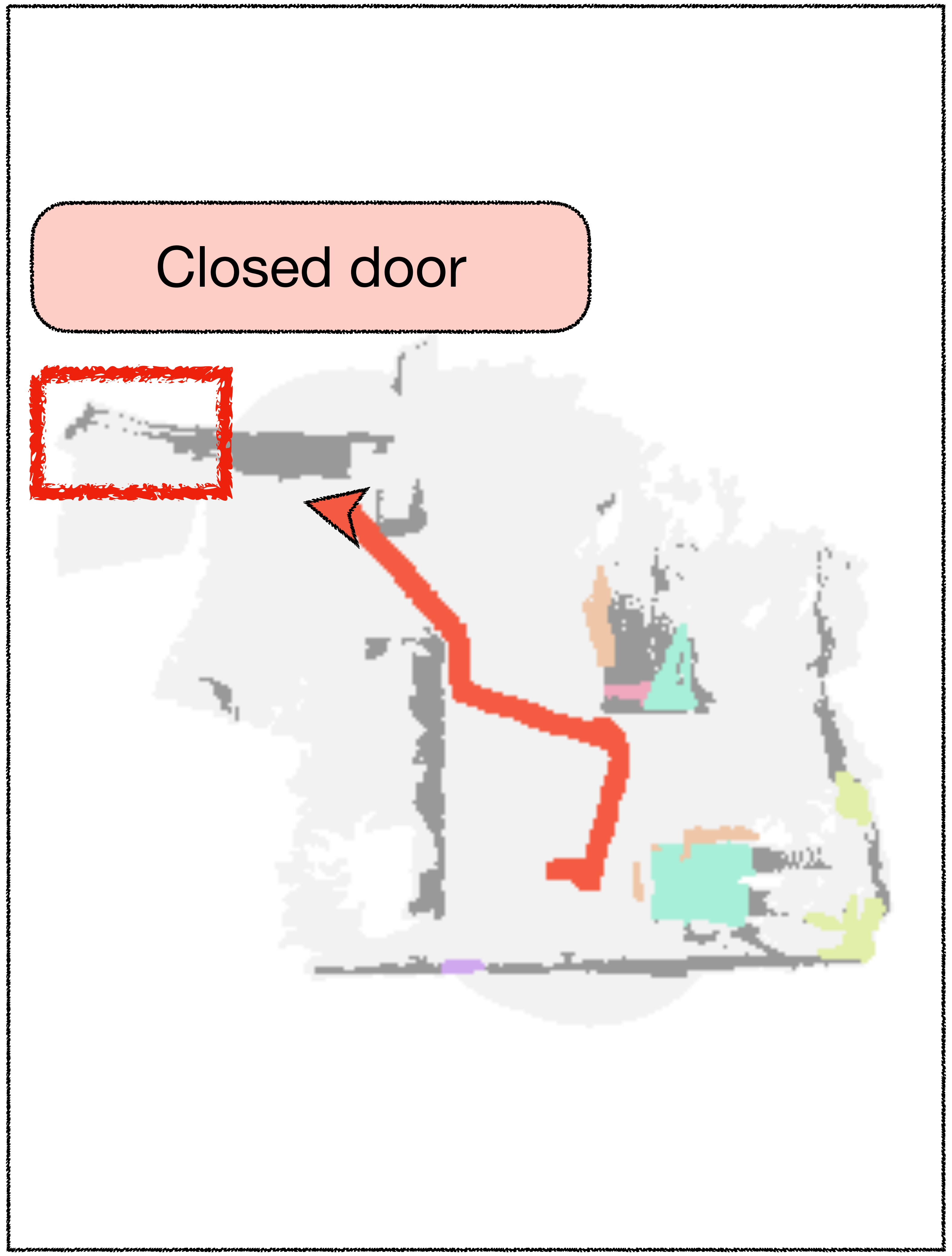

bed|0.93

Real-world Depth Sensor Errors

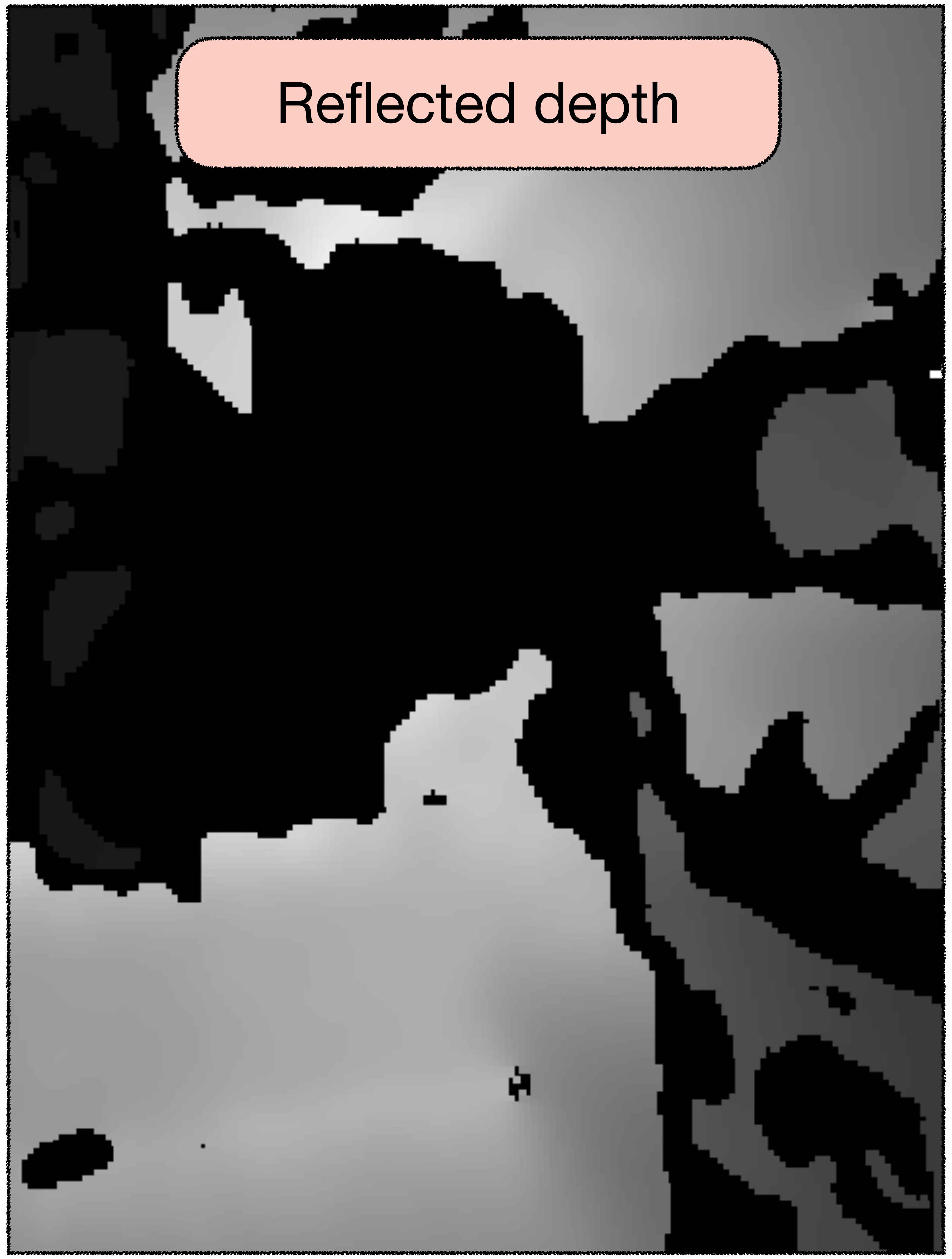
Door approach at an angle

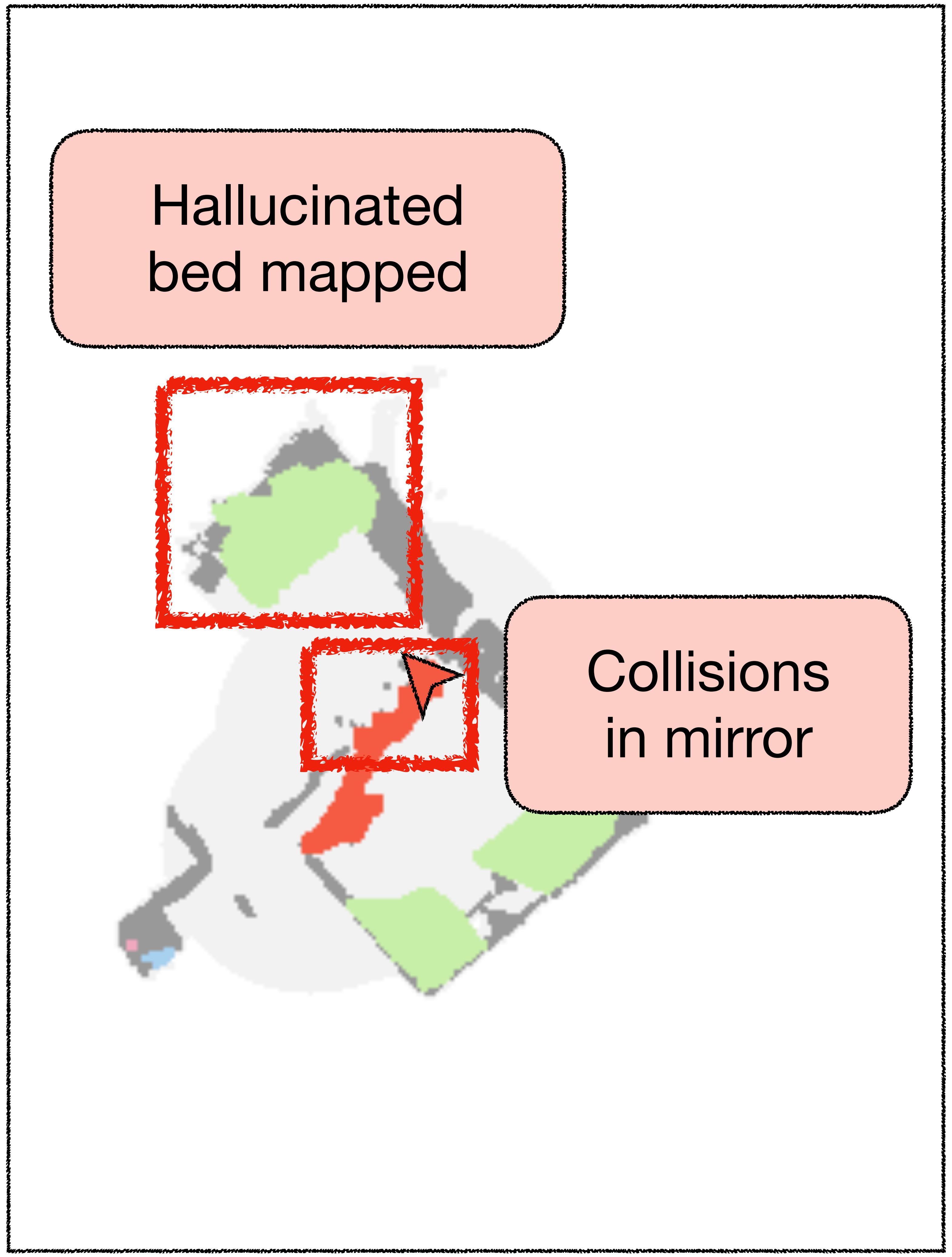






Real-world Depth Sensor Errors





Takeaways

For practitioners:

- Modular learning can reliably navigate to objects with 90% success

For researchers:

- Models relying on RGB images are hard to transfer from sim to real ➔ *leverage modularity and abstraction in policies*
- Disconnect between sim and real error modes ➔ *evaluate semantic navigation on real robots*

Thank you!

Webpage: https://theophilegervet.github.io/projects/real-world-object-navigation