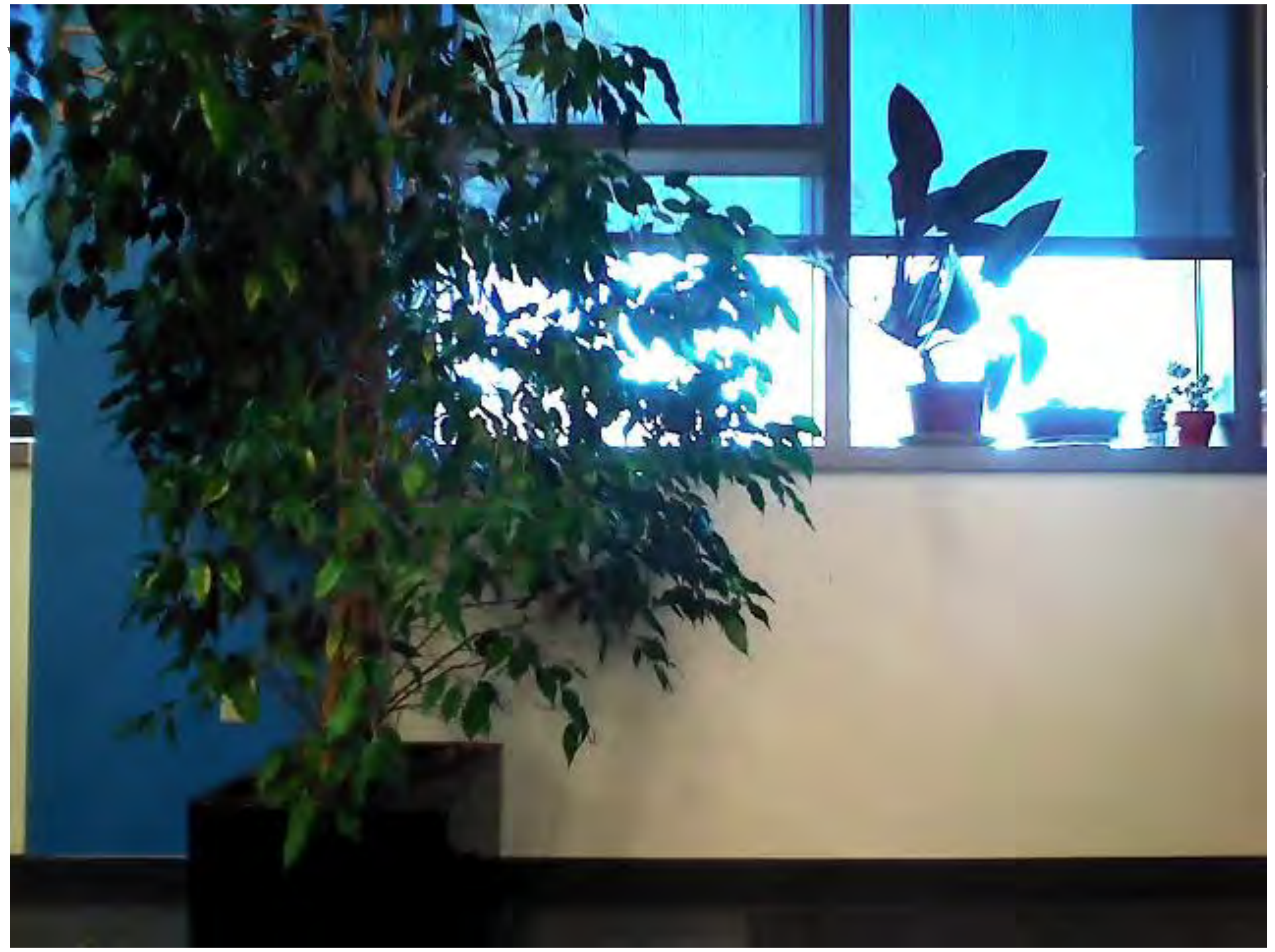
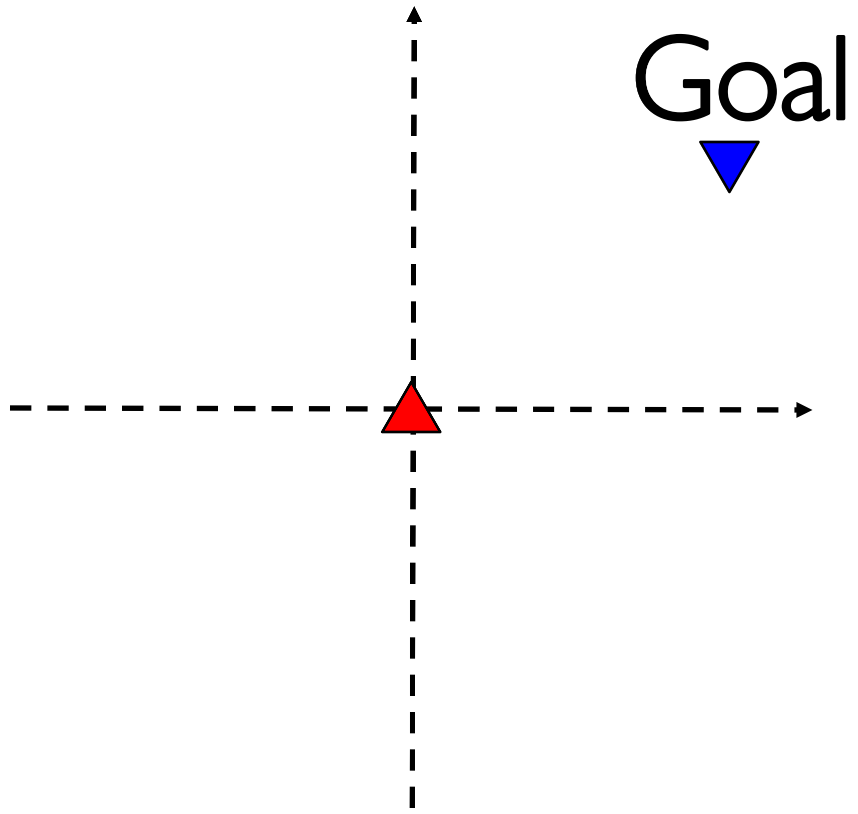
Robot with a first person camera



Dropped into a novel environment

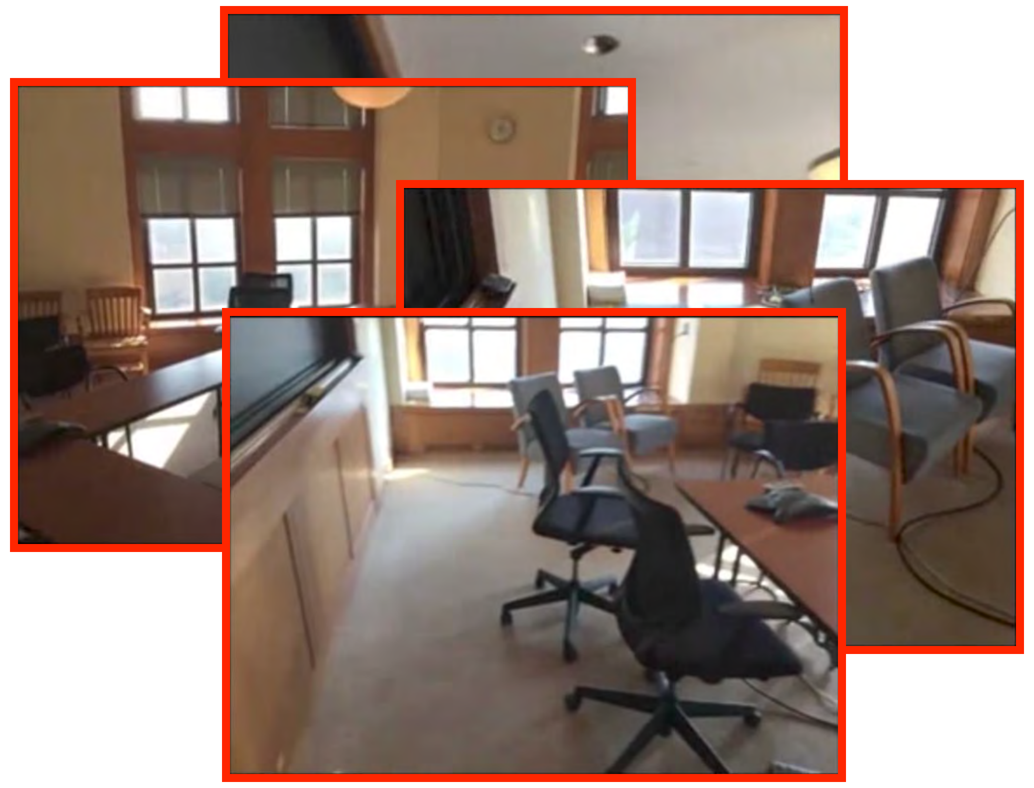


“Go 300 feet North, 400 feet East”

“Go Find a Chair”

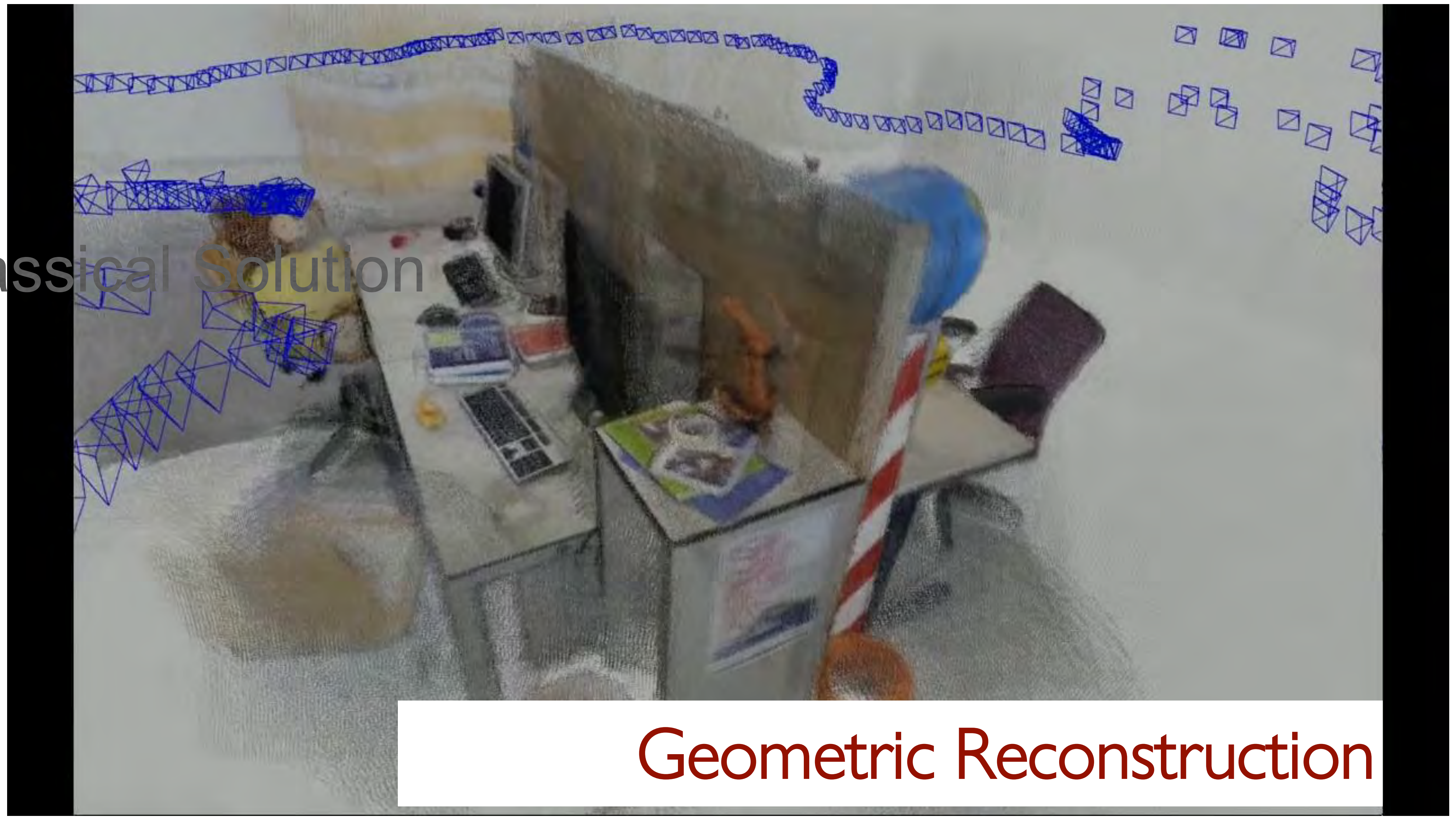
Navigate around

Classical Solution

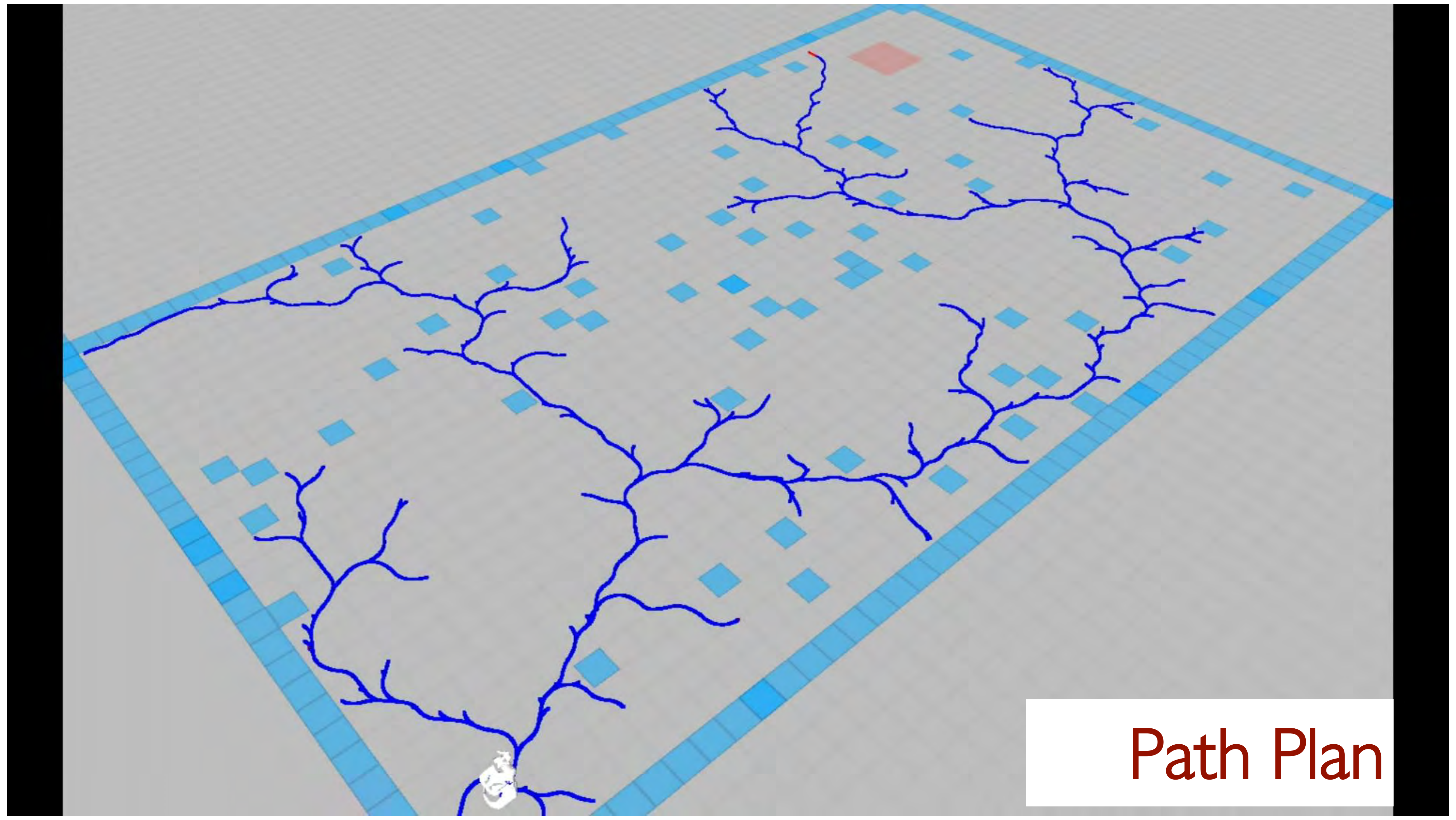


Observed Images

Mapping



Planning

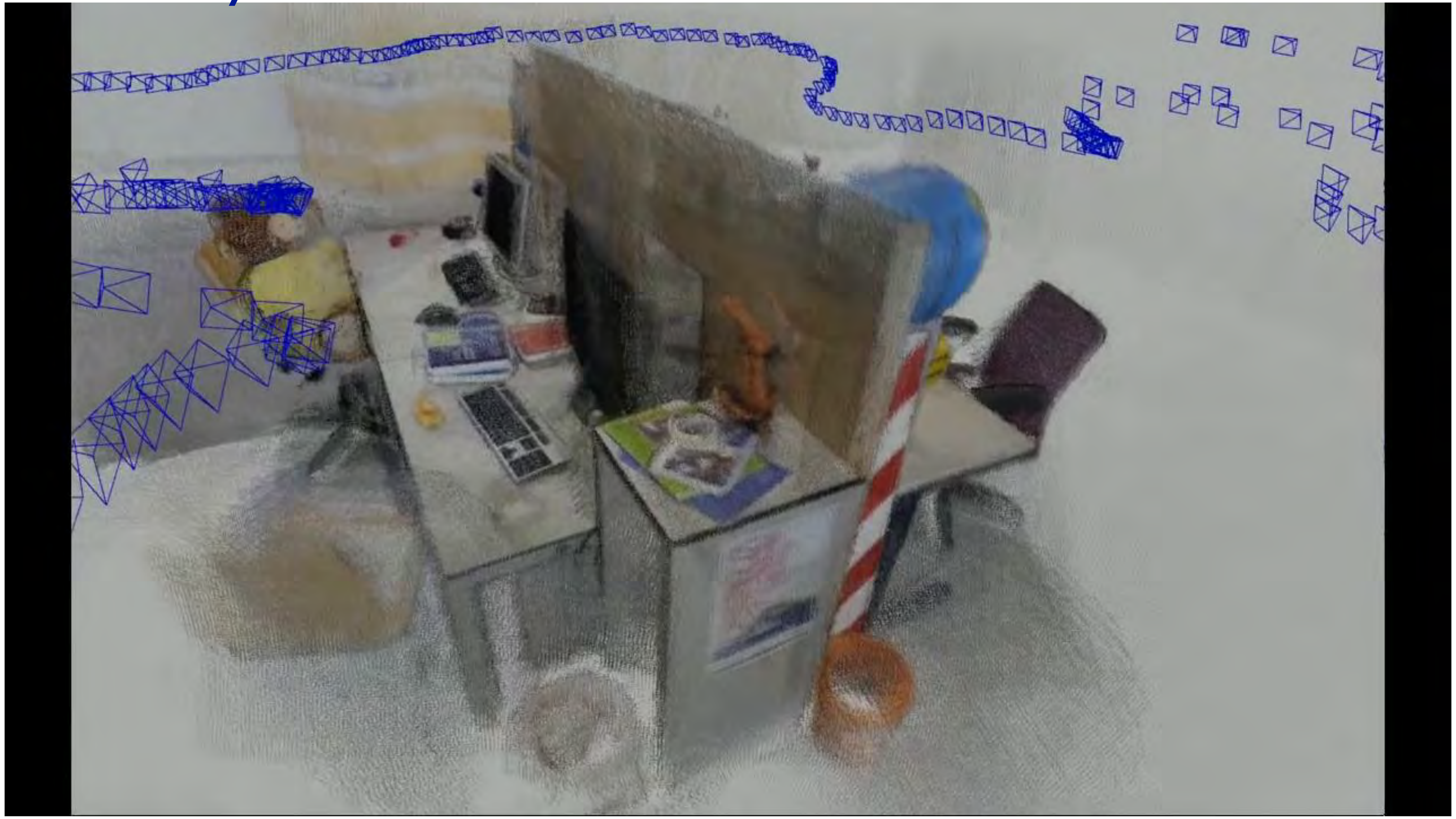


Hartley and Zisserman. 2000. Multiple View Geometry in Computer Vision
Thrun, Burgard, Fox. 2005. Probabilistic Robotics

Canny. 1988. The complexity of robot motion planning.
Kavraki et al. RA1996. Probabilistic roadmaps for path planning in high-dimensional configuration spaces.
Lavalle and Kuffner. 2000. Rapidly-exploring random trees: Progress and prospects.

Video Credits: Mur-Artal et al., Palmieri et al.

Unnecessary

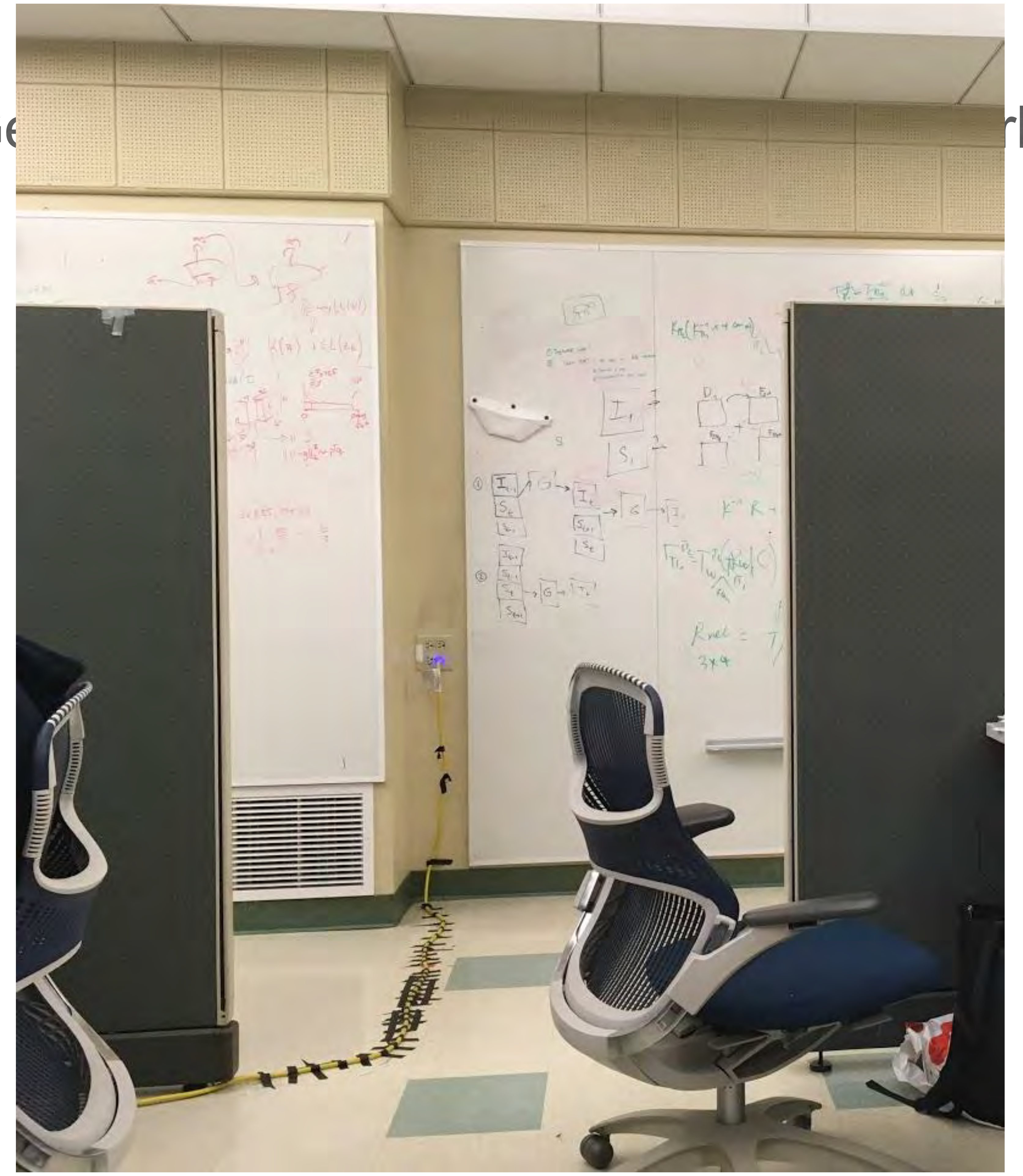


Do we need to tediously reconstruct everything on this table?

Video Credit: Mur-Artal and Tardos, TRobotics 2016. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras.

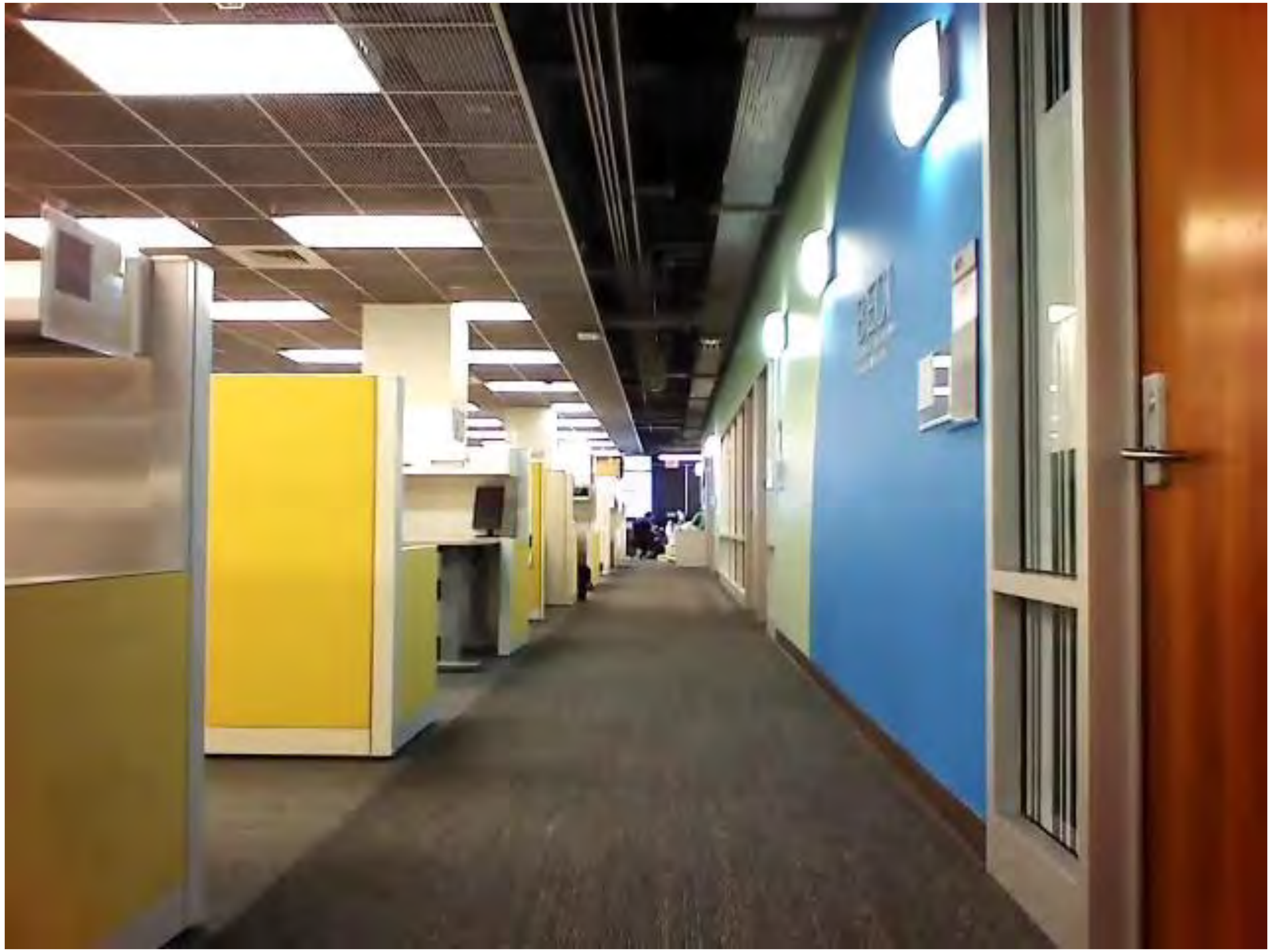
Insufficient

Ge … rld



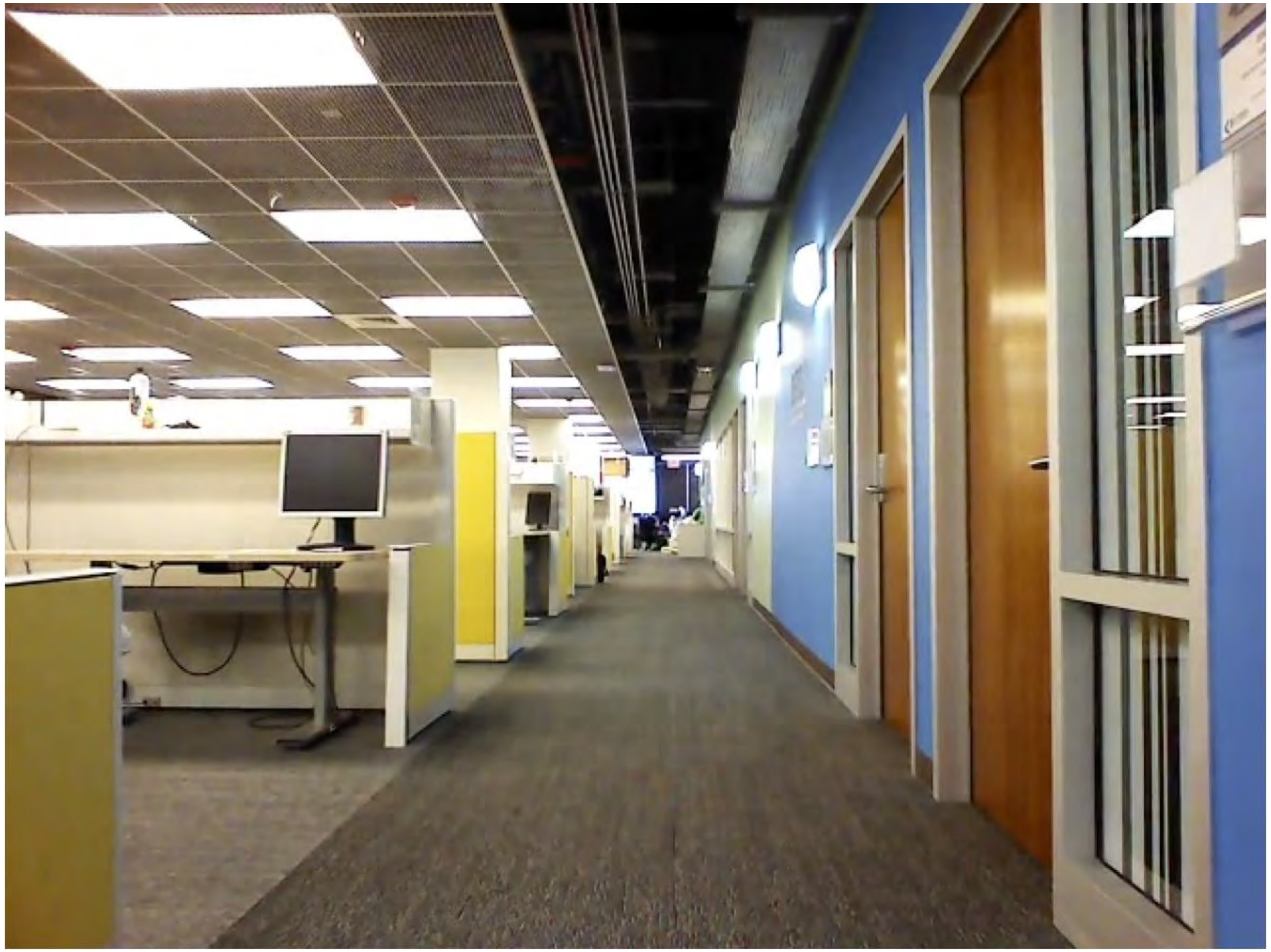
Can’t speculate about space not directly observed.

Insufficient



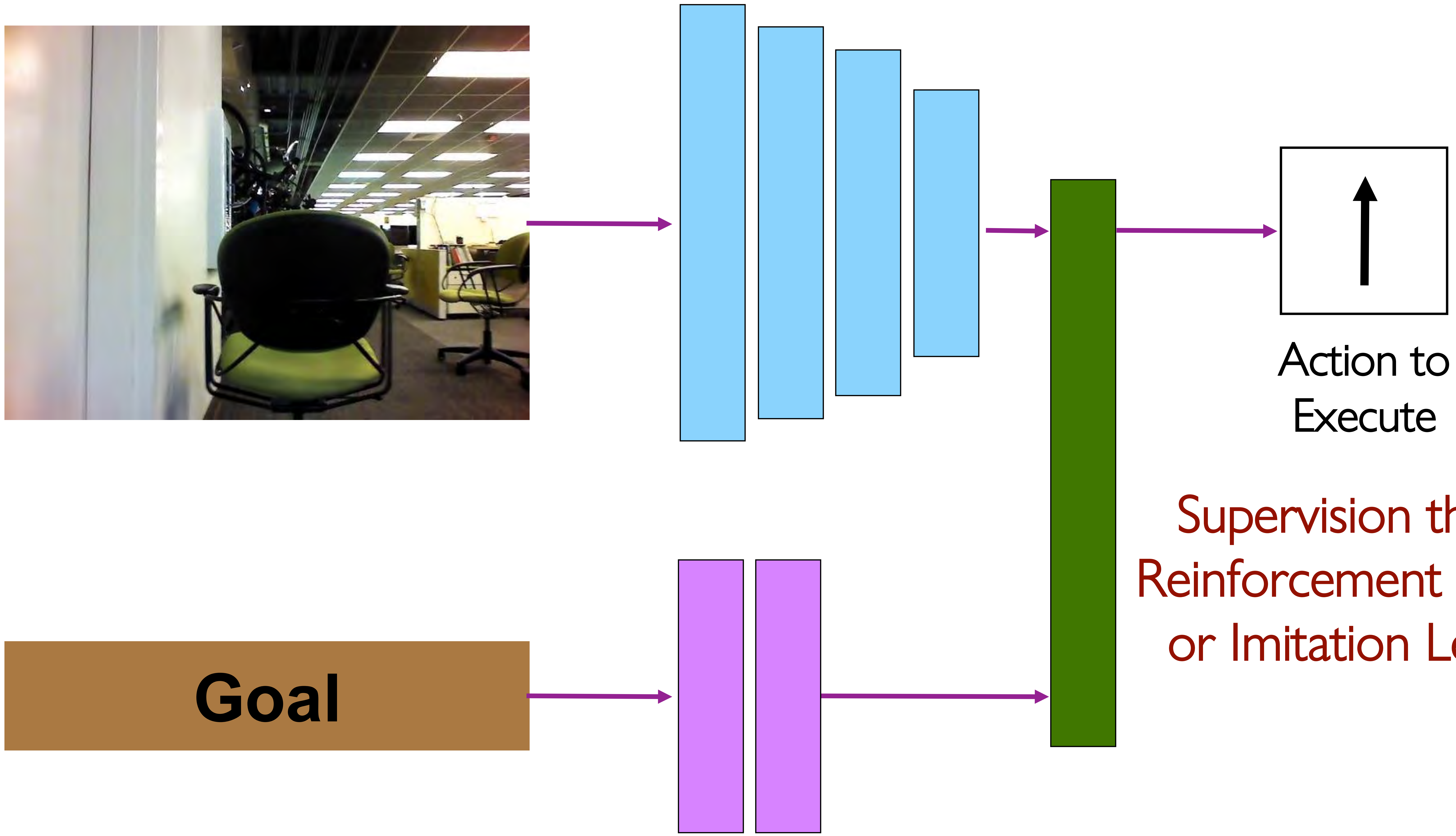
Can't exploit patterns in layout of indoor spaces.

Insufficient



Ignore navigational affordances.

Modern End-to-End Learned Navigation



Supervision through Reinforcement Learning or Imitation Learning

Mnih et al., Nature 2014. Human-level control through deep reinforcement learning.

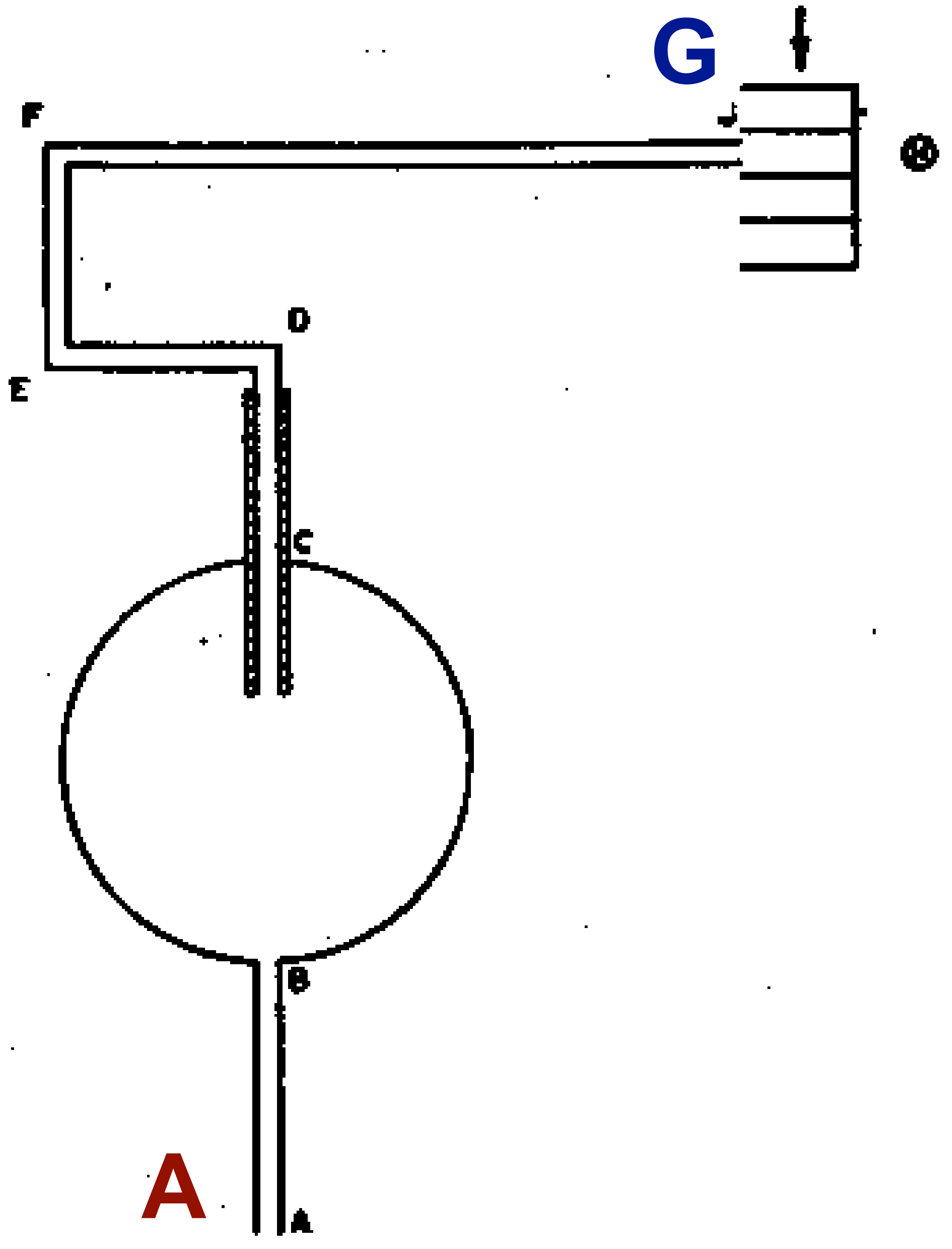
Levine et al., JMLR 2015. End-to-End Training of Deep Visuomotor Policies.

Zhu et al., ICRA 2017. Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning.

COGNITIVE MAPS IN RATS AND MEN[1]

BY EDWARD C. TOLMAN

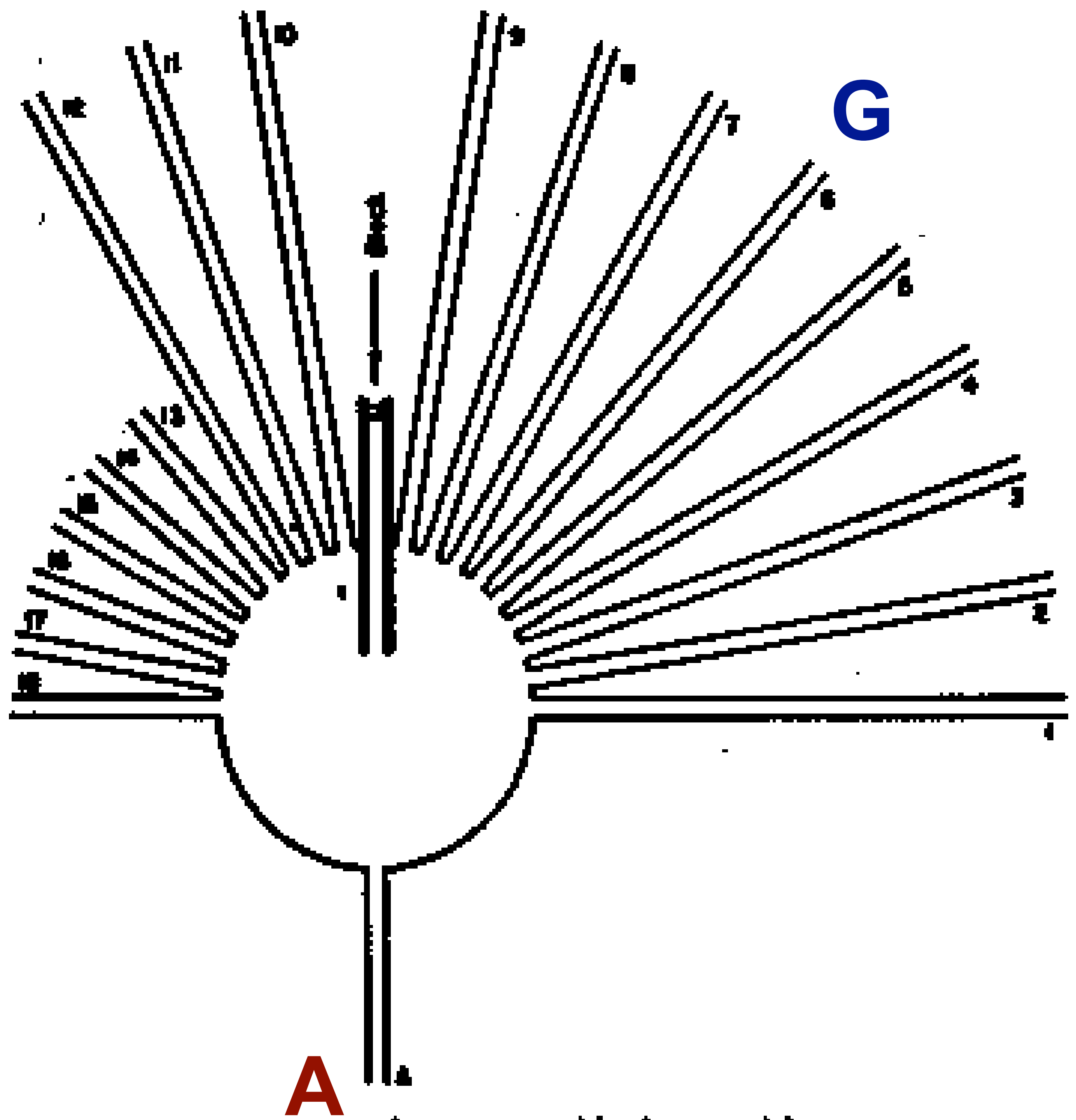
University of California



Apparatus used in preliminary training

Fig. 15

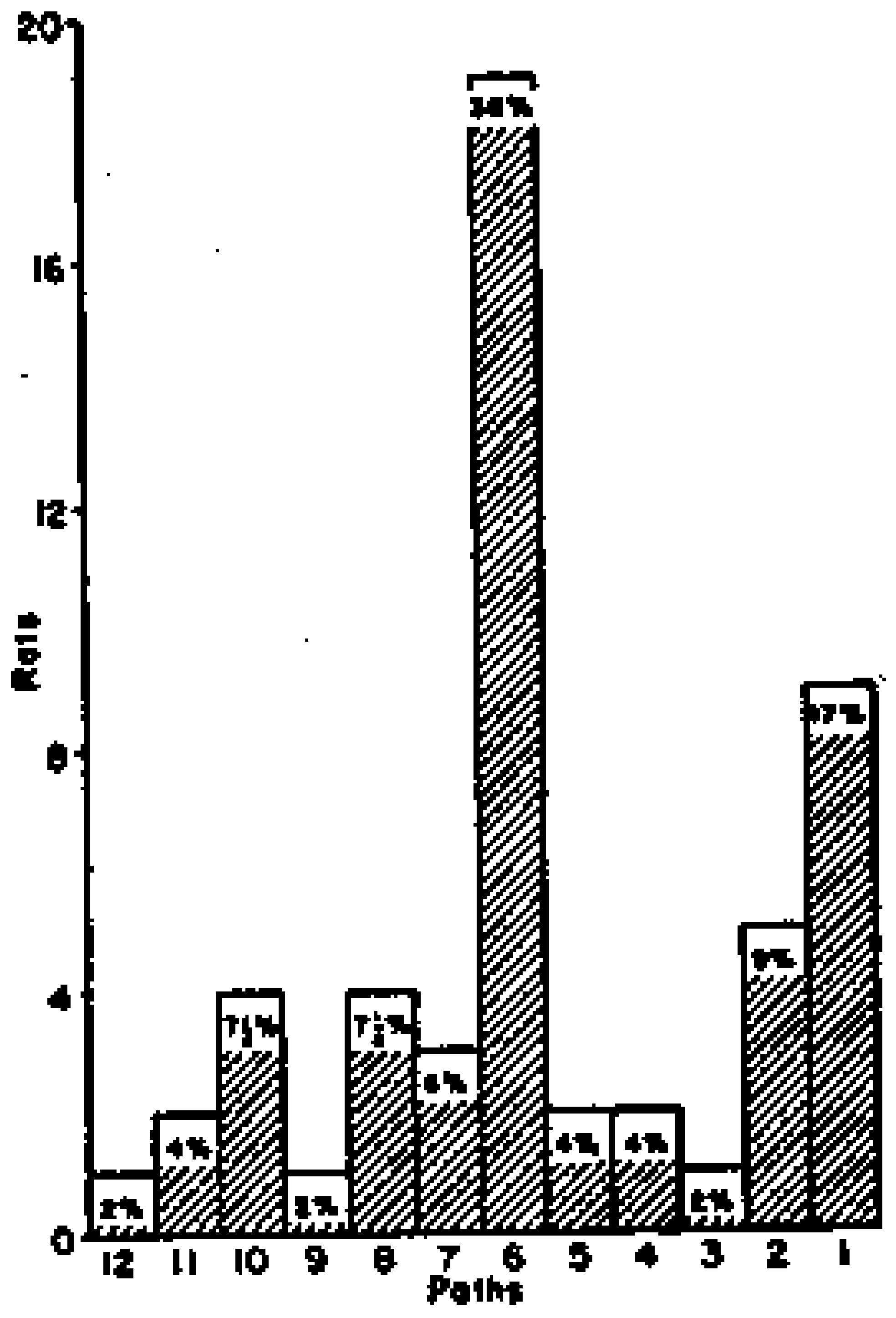
(From E. C. Tolman, B. F. Ritchie and D. Kalish, Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, 36, p. 16.)



Apparatus used in the test trial

Fig. 16

(From E. C. Tolman, B. F. Ritchie and D. Kalish, Studies in spatial learning. I. Orientation and short-cut. *J. exp. Psychol.*, 1946, 36, p. 17.)



Numbers of rats which chose each of the paths

Fig. 17

(From E. C. Tolman, B. F. Ritchie and D. Kalish, Studies in spatial learning. I. Orientation and the short-cut. *J. exp. Psychol.*, 1946, 36, p. 19.)

Tolman. Psychological Review 1948. Cognitive Maps in Rats and Men

Navigating to Objects in the Real World


Theophile Gervet [1]


Soumith Chintala [4]


Dhruv Batra [3,4]


Jitendra Malik [2,4]


Devendra Chaplot [4]

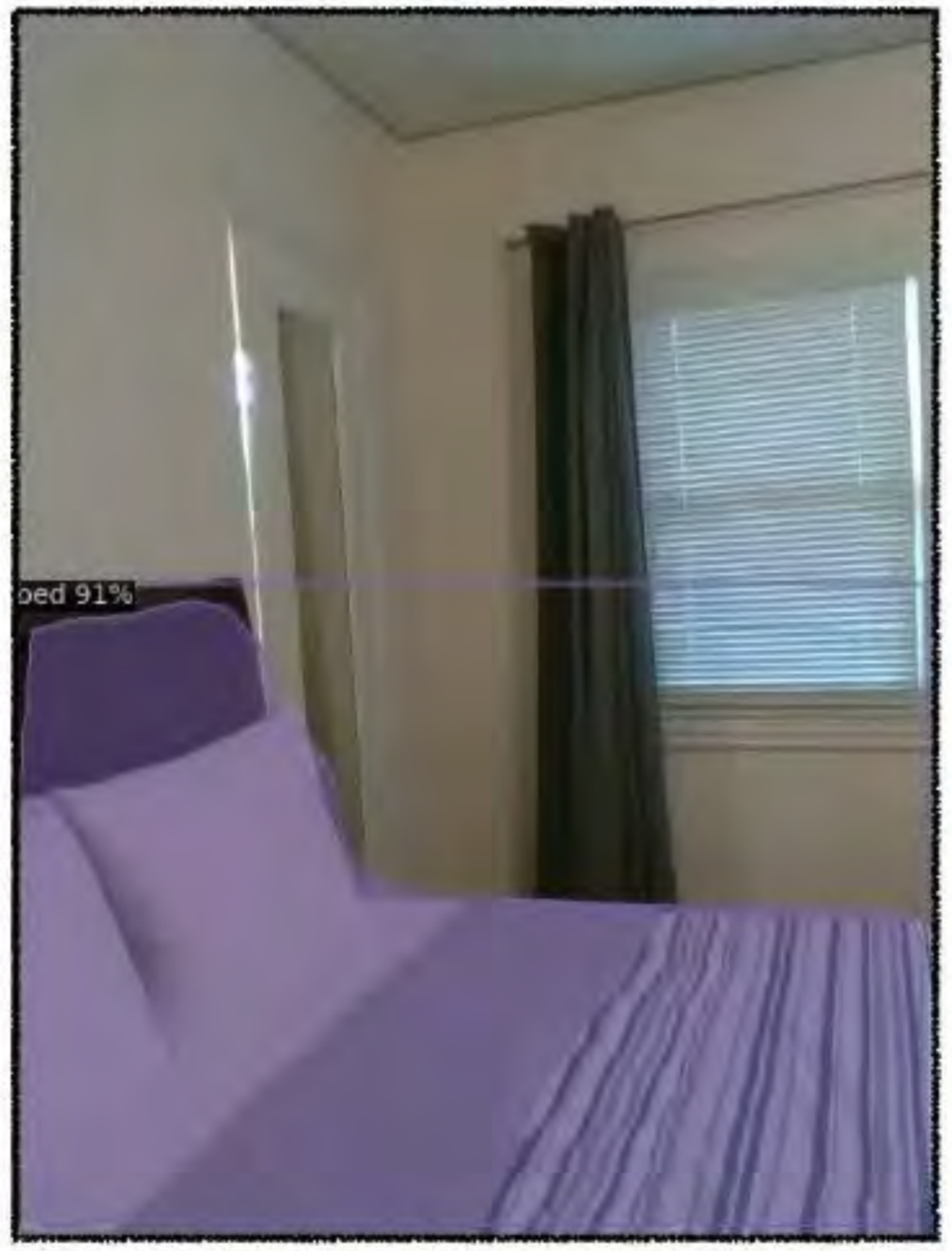
1 Carnegie Mellon University

2 Berkeley UNIVERSITY OF CALIFORNIA

3 Georgia Tech

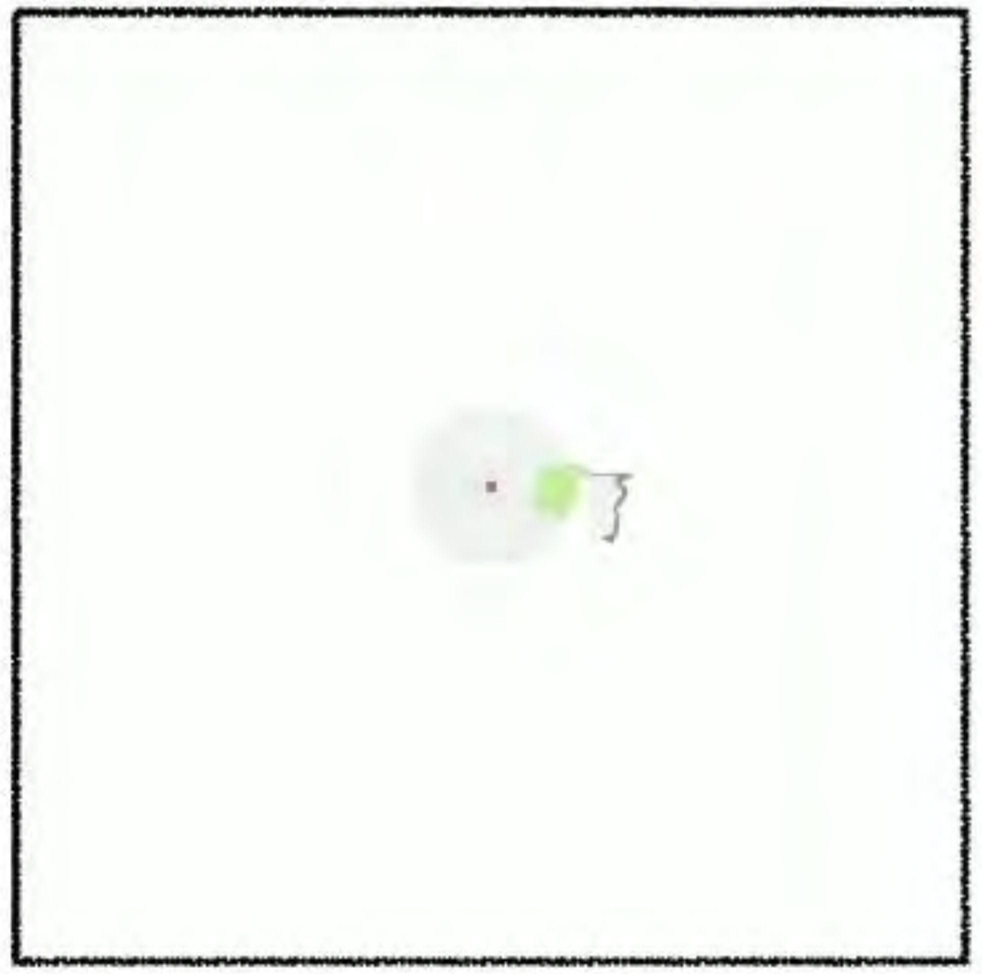
4 Meta AI

Observation



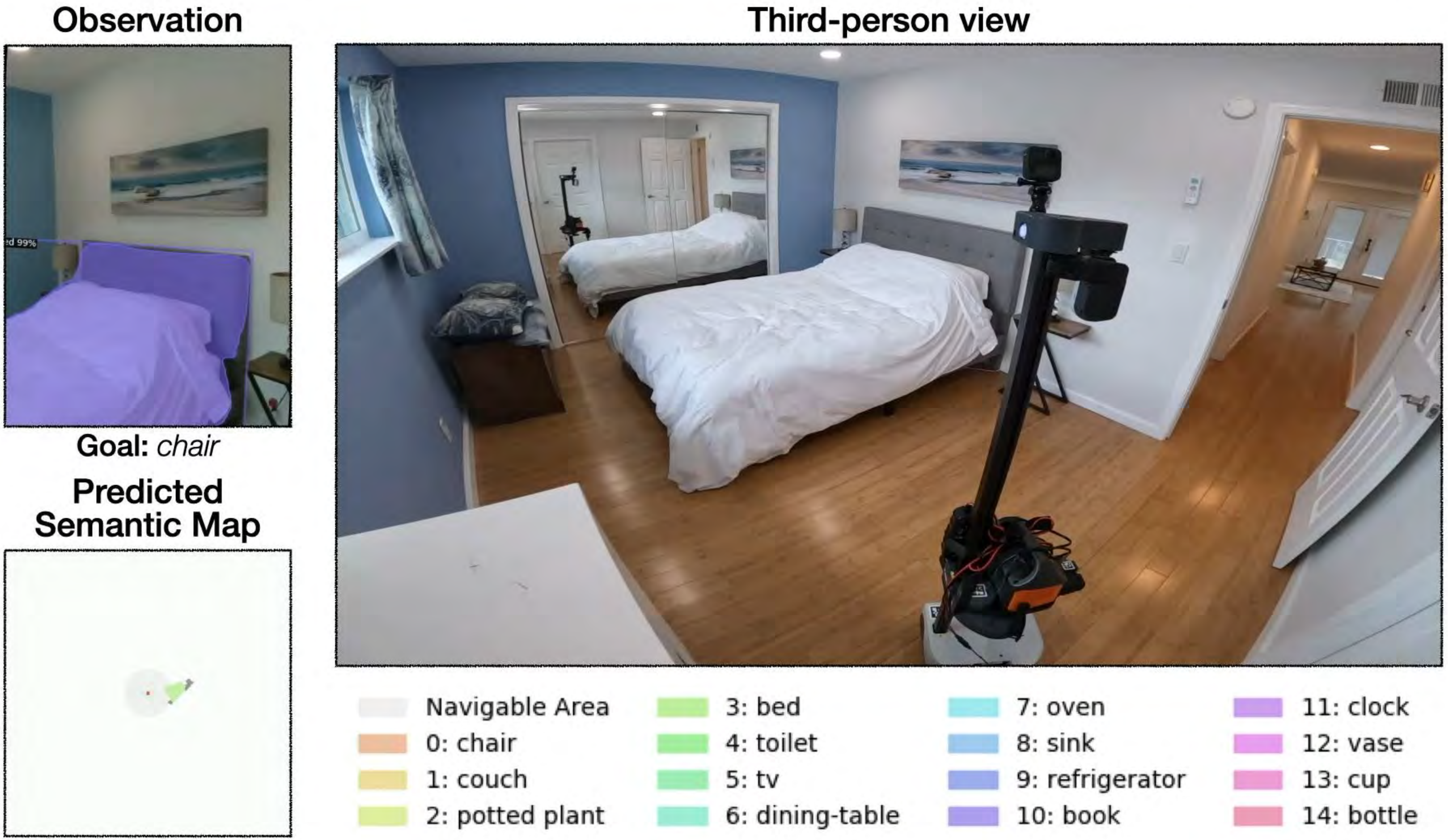
Goal: *plant*

Predicted Semantic Map



Third-person view

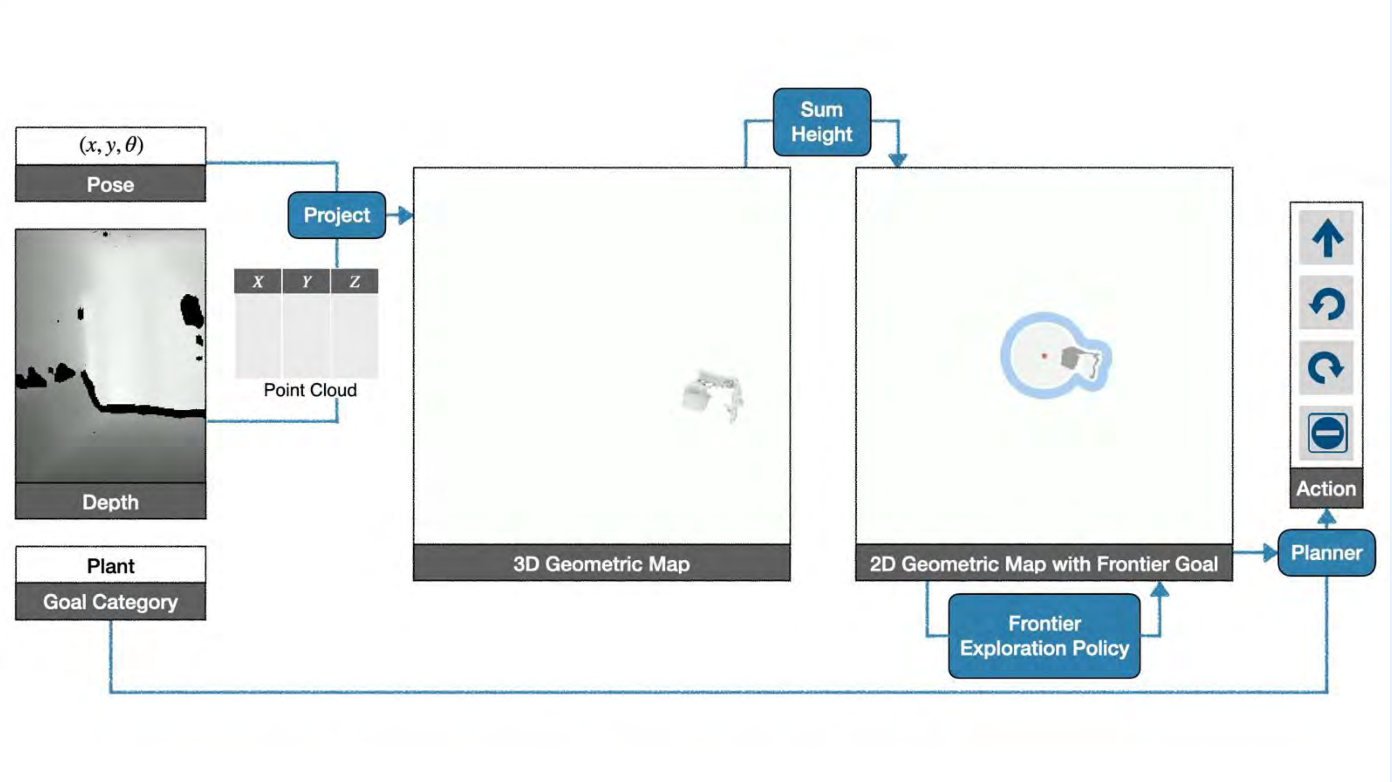
Navigable Area
0: chair
1: couch
2: potted plant
3: bed
4: toilet
5: tv
6: dining-table
7: oven
8: sink
9: refrigerator
10: book
11: clock
12: vase
13: cup
14: bottle


Observation
Third-person view
Goal: chair
Predicted Semantic Map
Navigable Area
0: chair
1: couch
2: potted plant
3: bed
4: toilet
5: tv
6: dining-table
7: oven
8: sink
9: refrigerator
10: book
11: clock
12: vase
13: cup
14: bottle


Empirical Evaluation
3 Approaches
6 Unseen Homes
6 Goal Object Categories

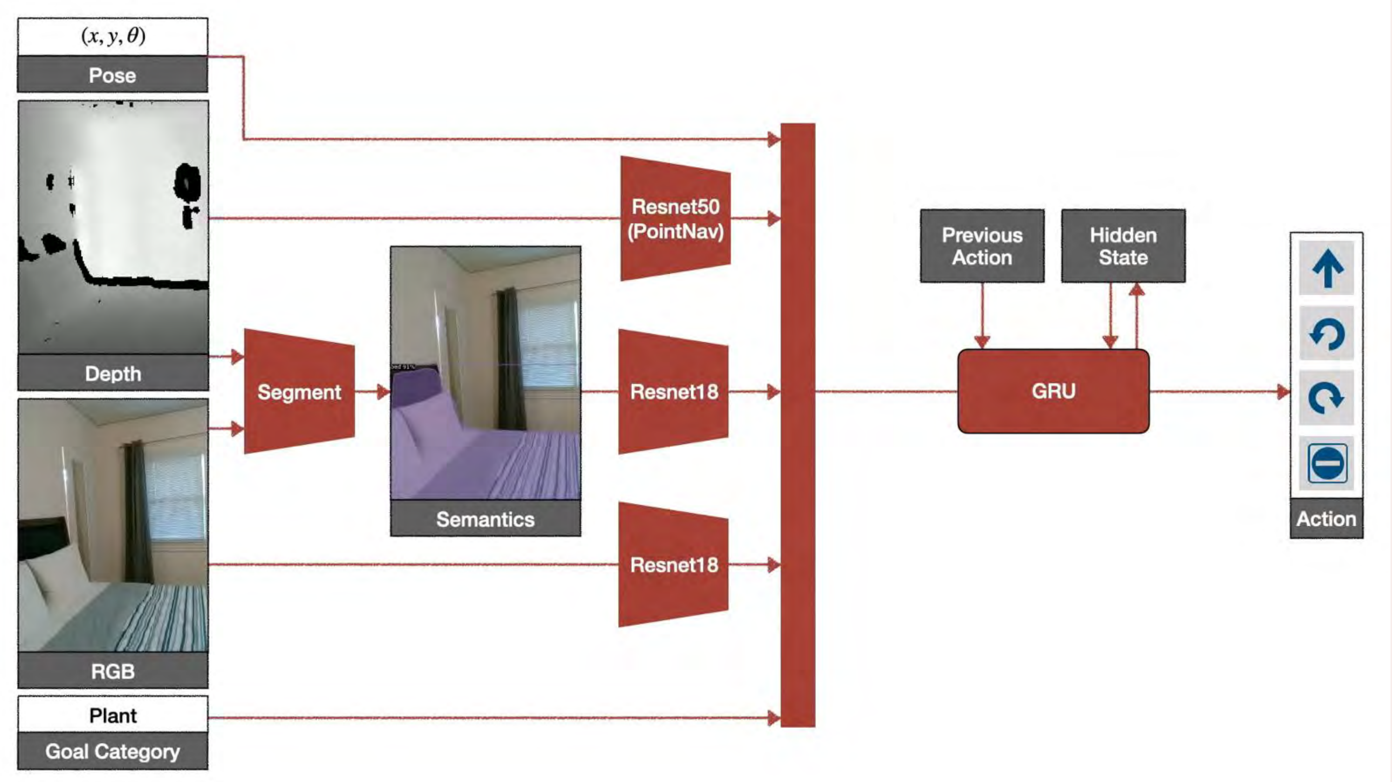
Methods

Classical



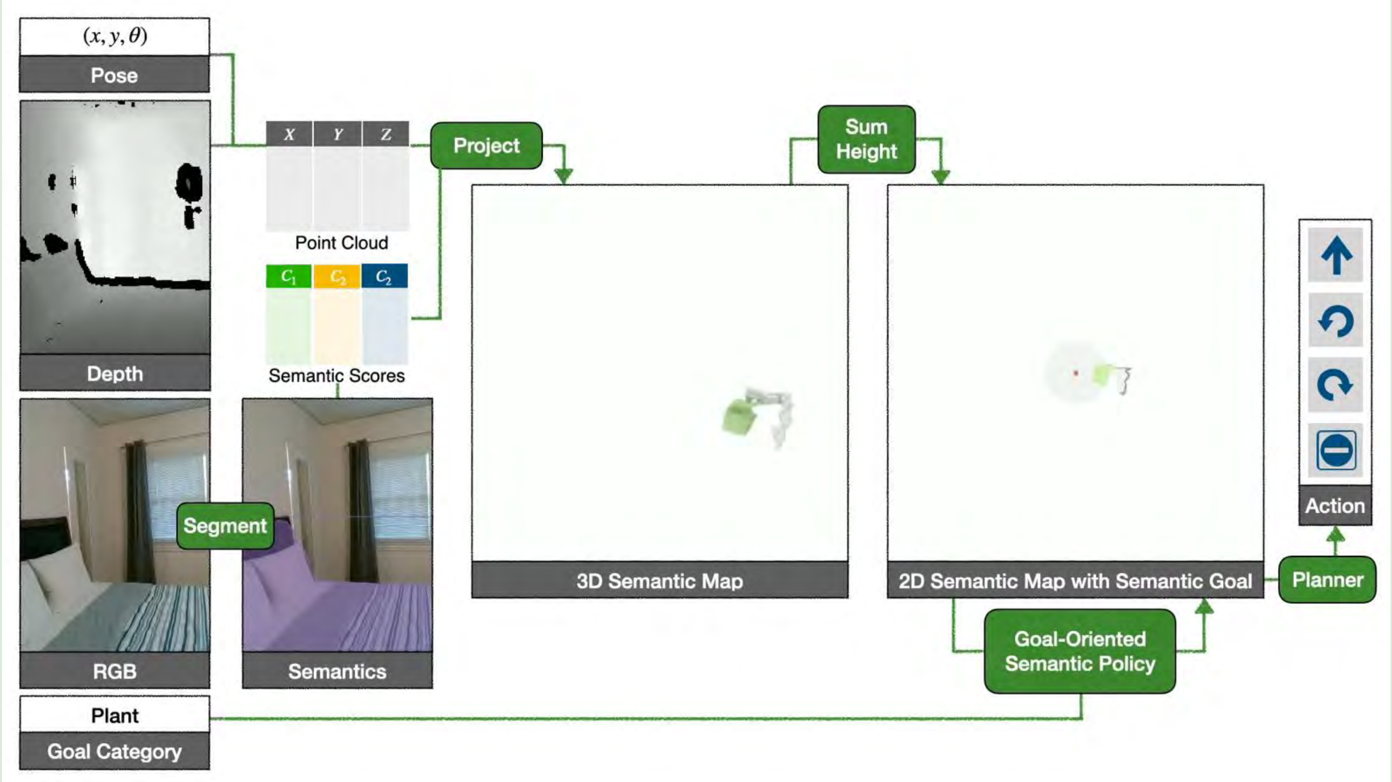
- Geometric Map
- Heuristic Exploration
  - No Training

End-to-end Learning



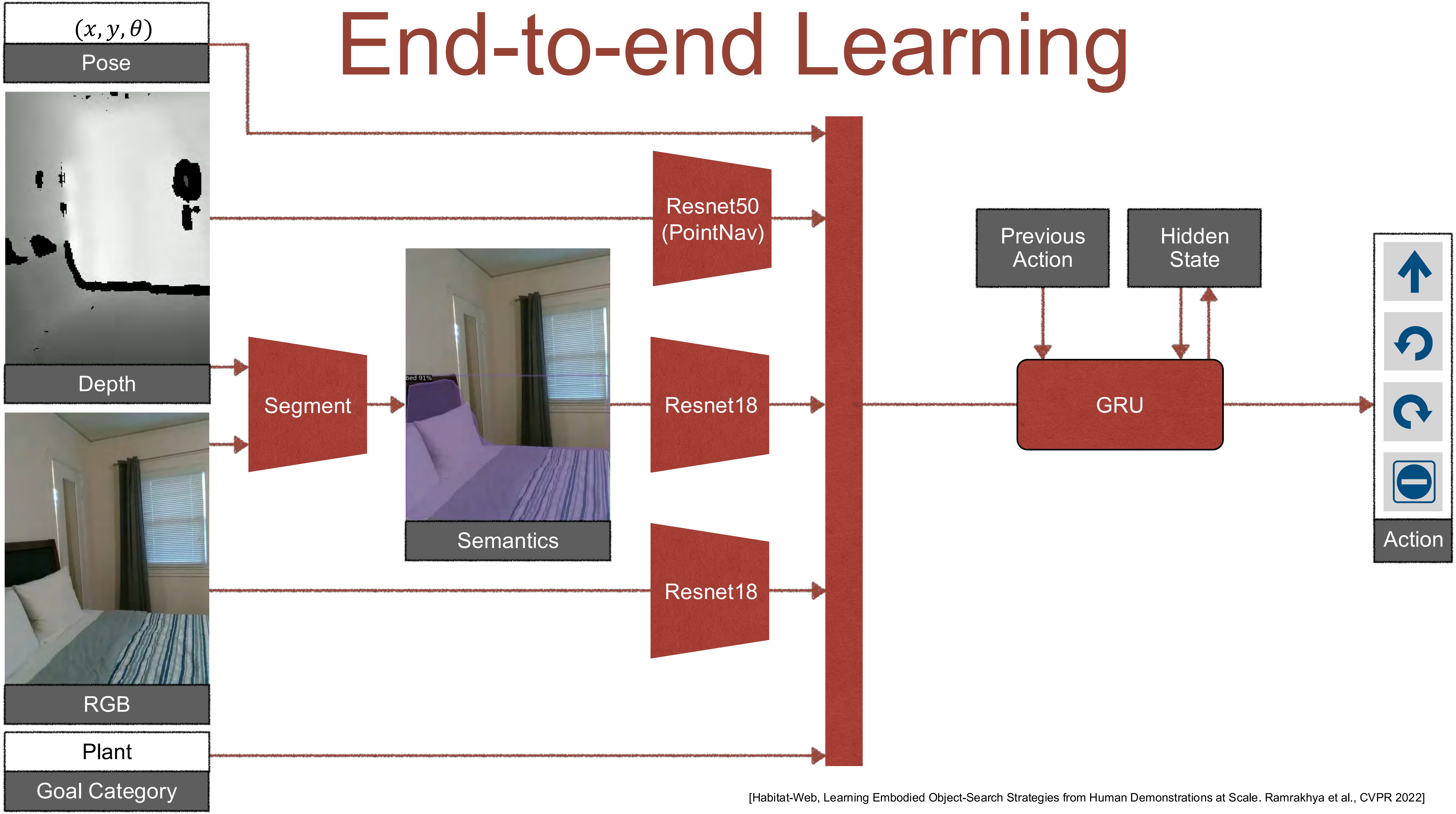
- End-to-end
- Large-scale IL + RL fine-tuning
  - 77,000 human trajectories
  - 200M frames of RL

Modular Learning



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

End-to-end Learning



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

Modular Learning

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

Habitat

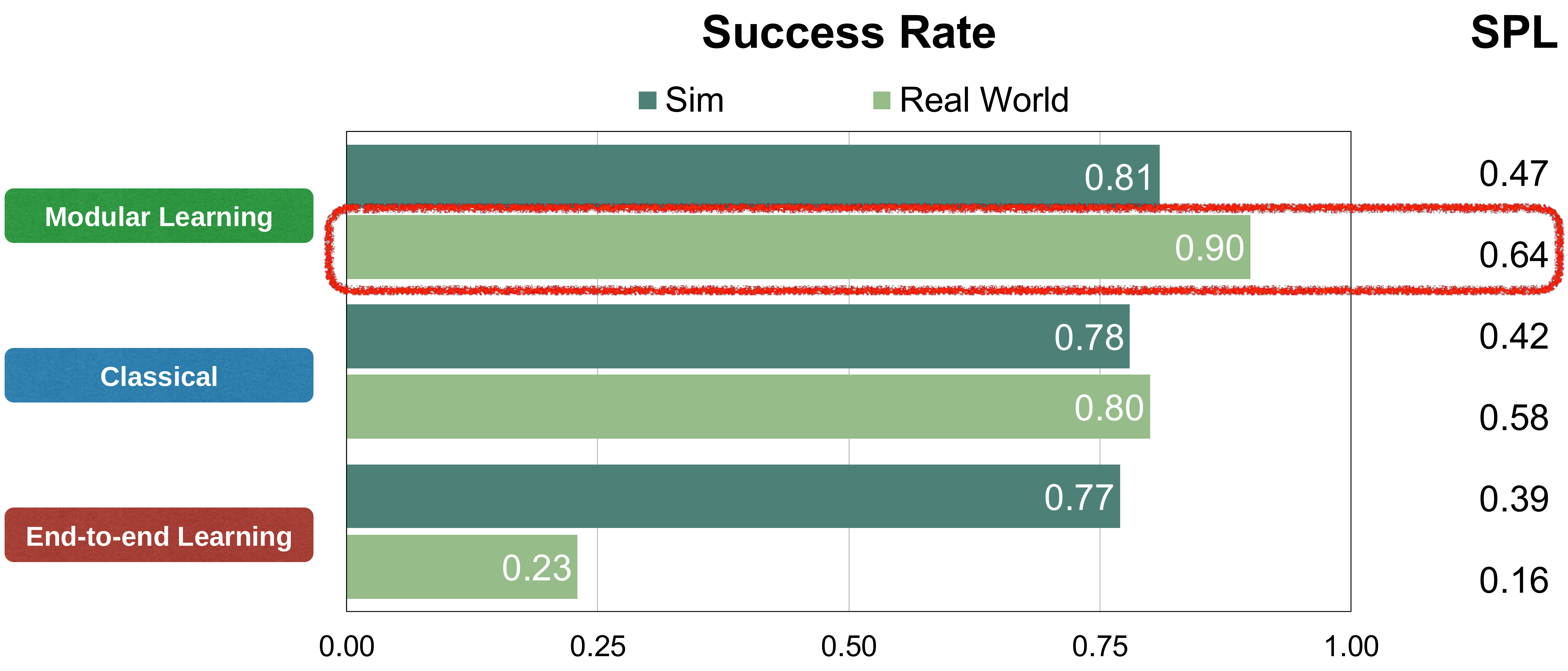
AI2-Thor

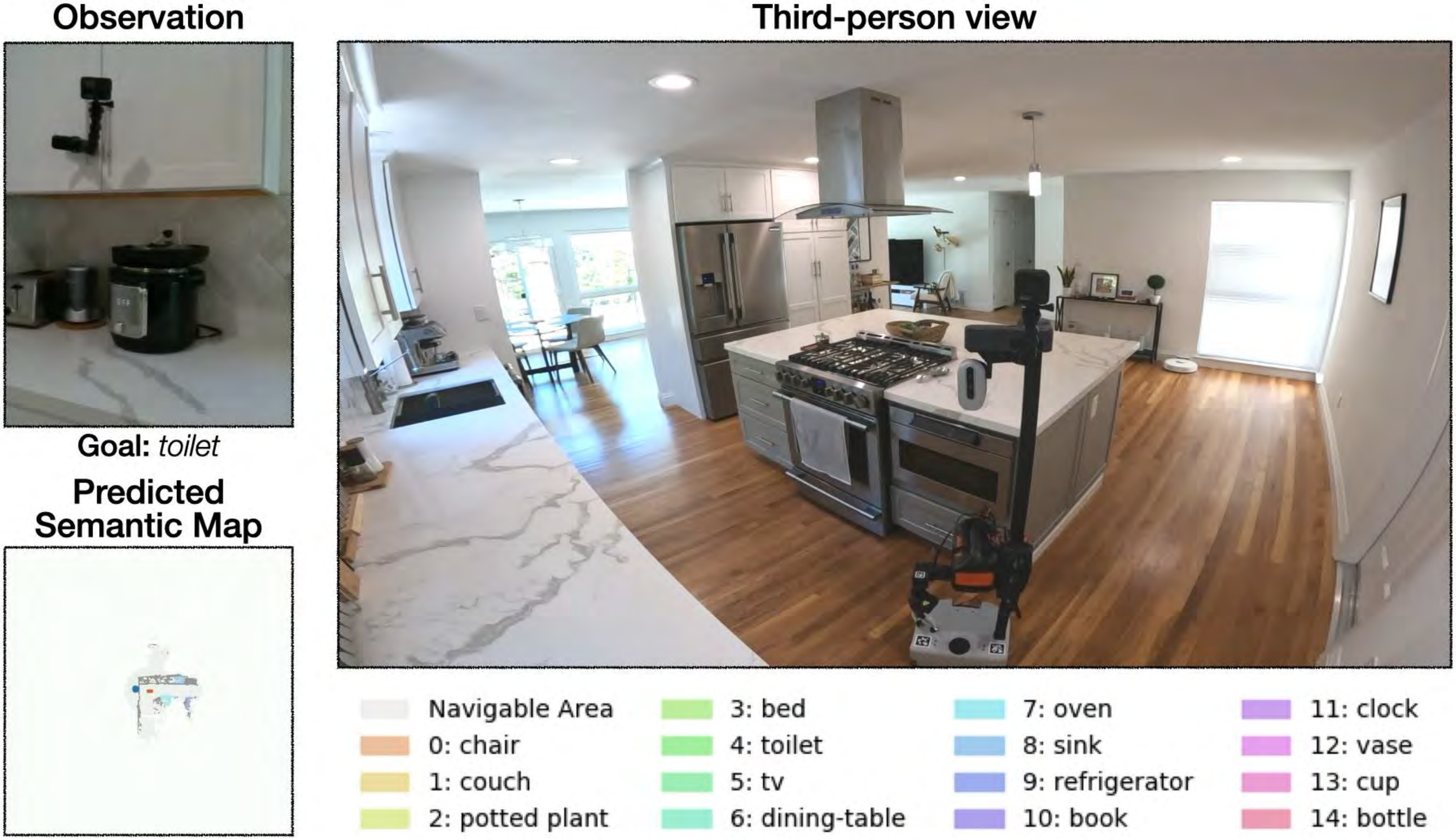
Results

Success Rate
SPL
Sim
Real World
Modular Learning
Classical
End-to-end Learning
0.00
0.25
0.50
0.75
1.00

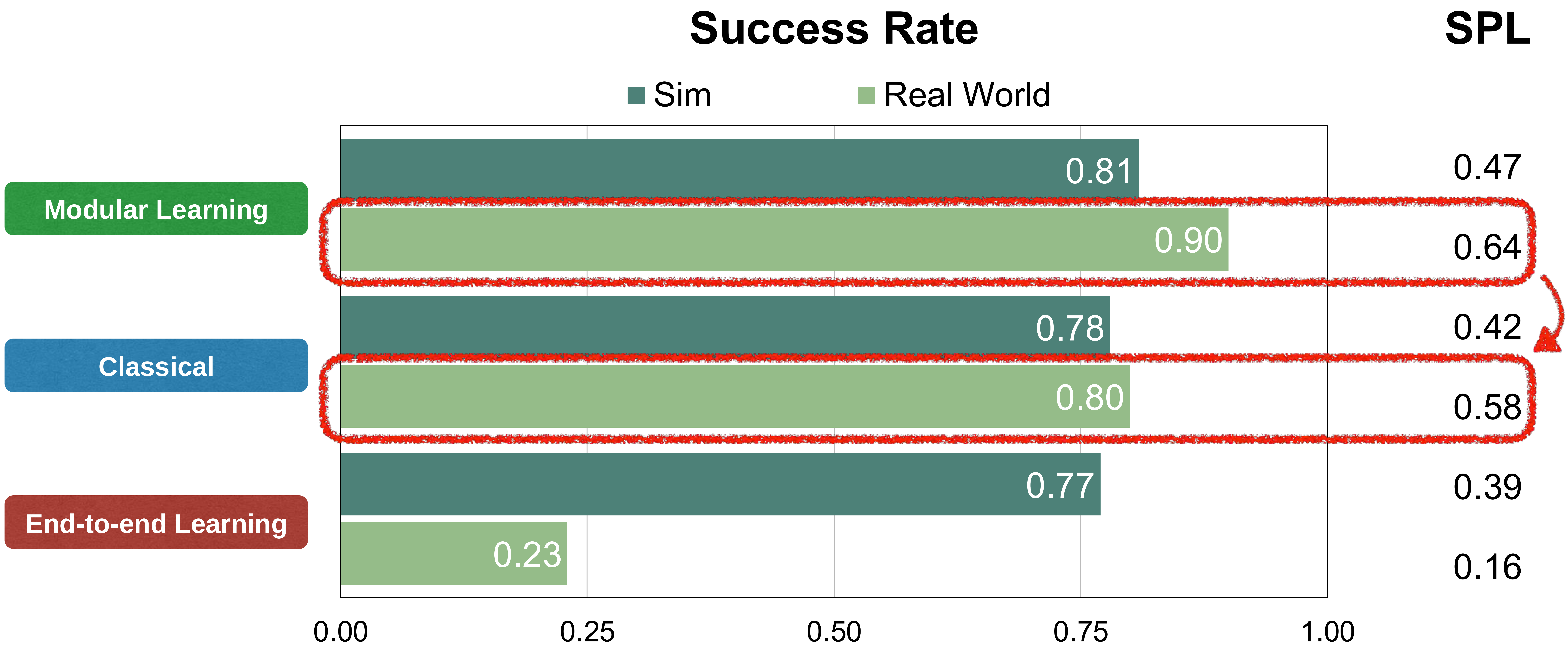

Goal: couch
SPL: 0.74, 78 steps
SPL: 0.0, 121 steps
SPL: 0.33, 181 steps
Modular
End-to-End
Classical
Third-person view
Third-person view
Third-person view
Success
Failure
Success
Observation
Predicted Semantic Map
Observation
Predicted Semantic Map

Modular Learning is Reliable




Observation
Third-person view
Goal: *toilet*
Predicted Semantic Map
Navigable Area
0: chair
1: couch
2: potted plant
3: bed
4: toilet
5: tv
6: dining-table
7: oven
8: sink
9: refrigerator
10: book
11: clock
12: vase
13: cup
14: bottle

Classical vs Modular Learning



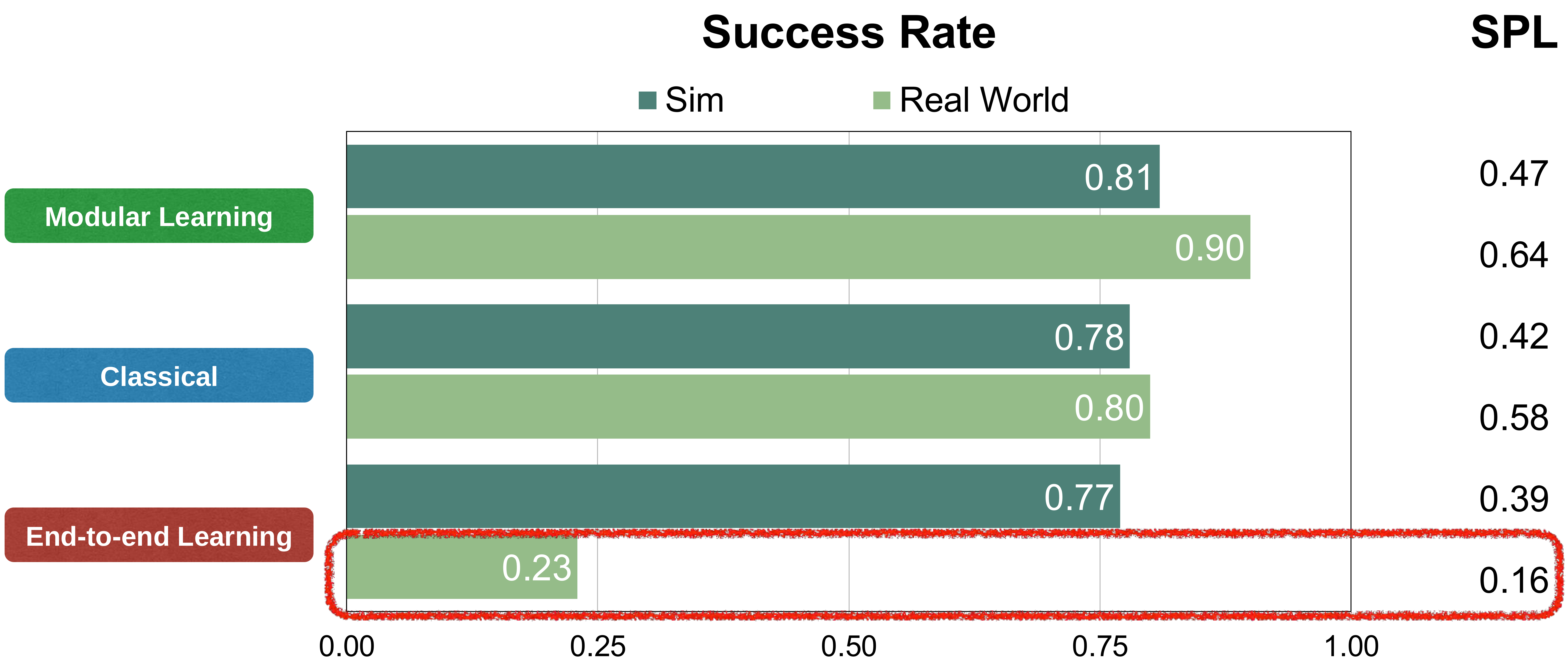
Classical vs Modular Learning

Goal: *bed*

SPL: 0.90, 98 steps

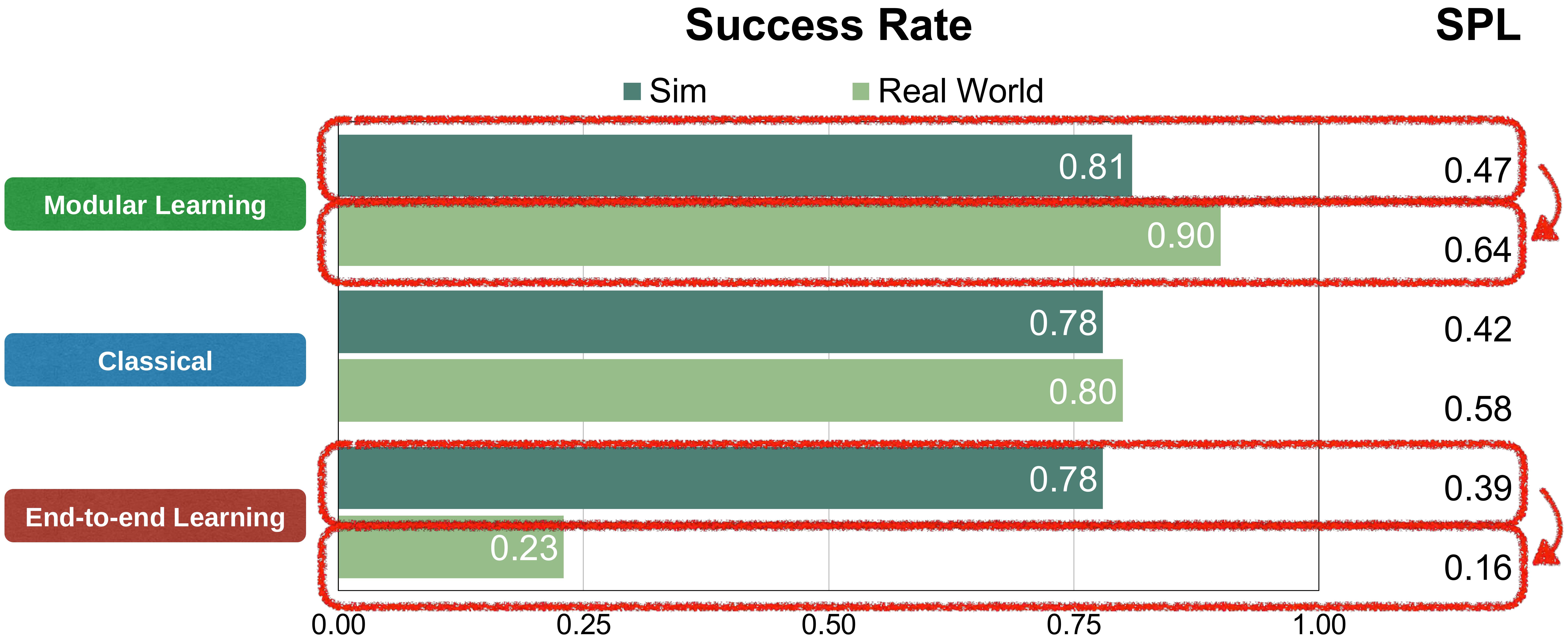
SPL: 0.52, 152 steps

End-to-end fails to Transfer

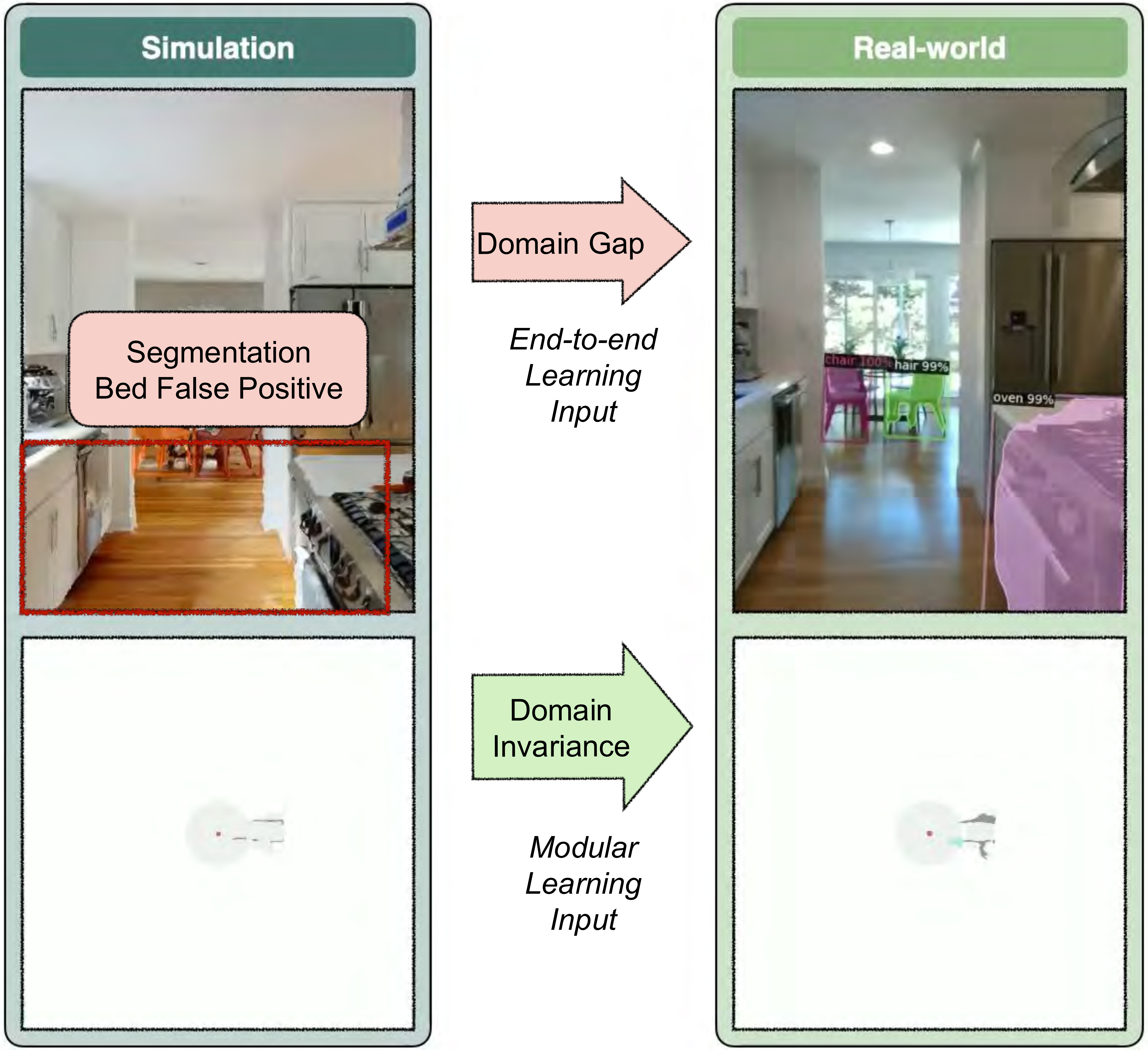


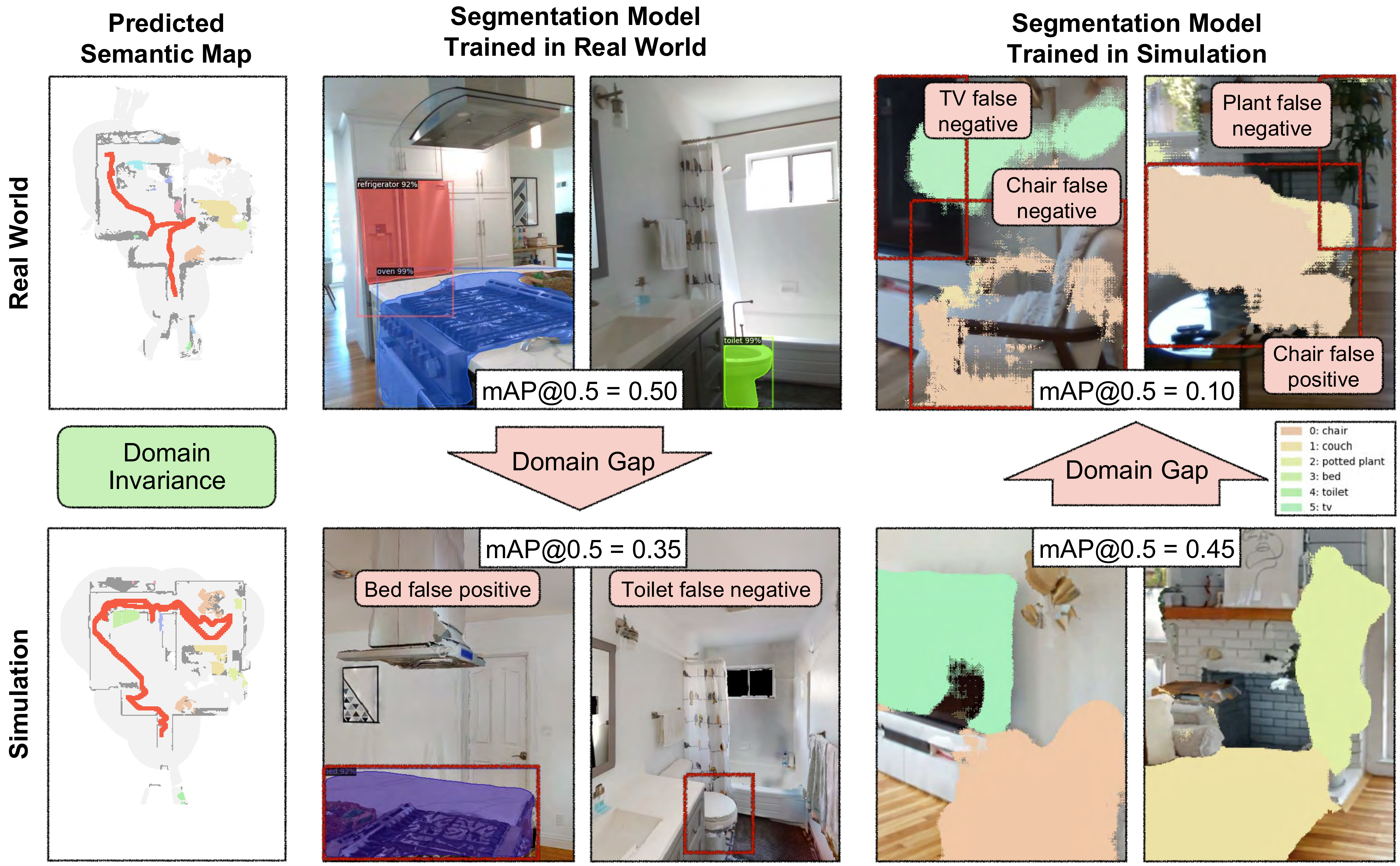
End-to-end Failures

Modular vs End-to-end Transfer

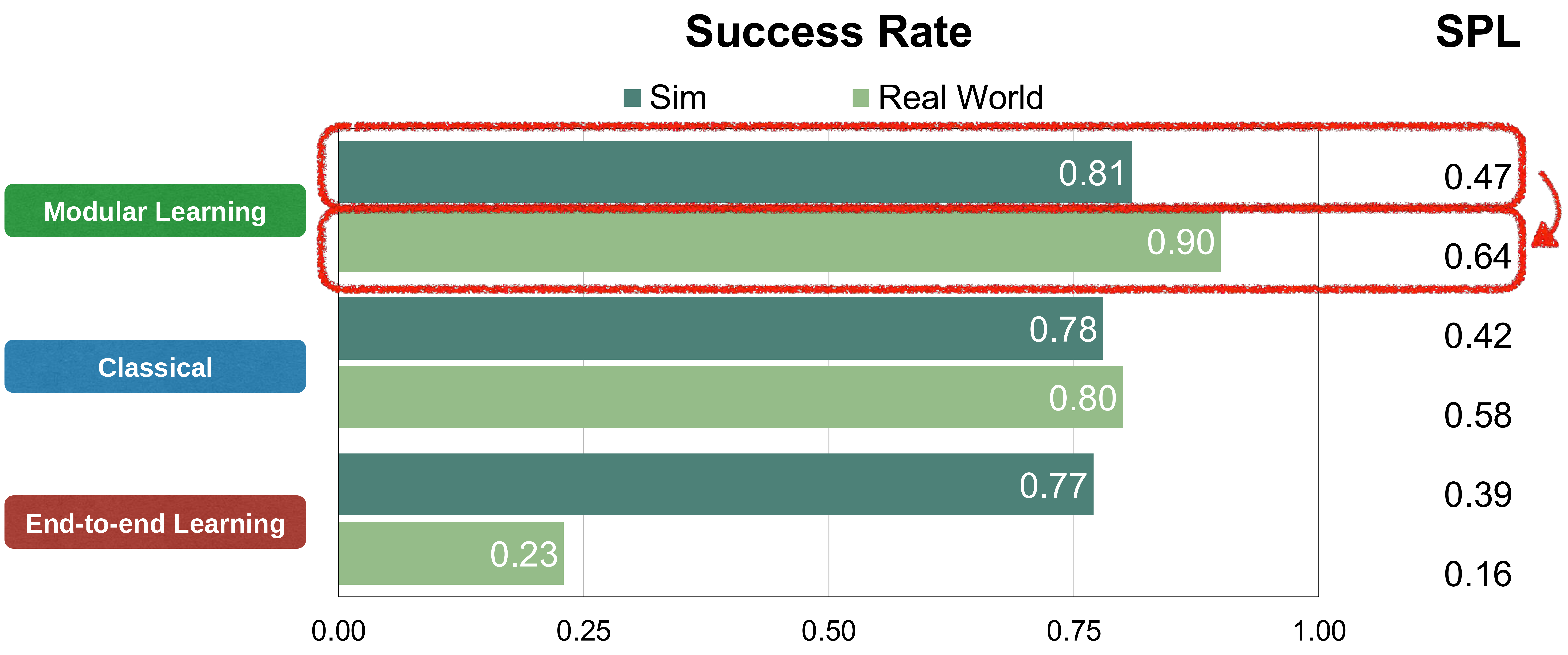


Simulation
vs
Reality




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



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*

GOAT: GO to Any Thing

Matthew Chang*, Theophile Gervet*, Mukul Khanna*, Sriram Yenamandra*, Dhruv Shah, Tiffany Min, Kavit Shah, Chris Paxton, Saurabh Gupta, Dhruv Batra, Roozbeh Mottaghi, Jitendra Malik*, Devendra Singh Chaplot*

Third-person view

1. GOAT Problem
2. GOAT System Architecture
3. Results
4. Applications
   - Pick & Place
   - Social Navigation
   - Platform Agnostic


Unknown Environment
Explore
Perception
Detect and Localize Objects
Lifelong Memory
Remember Object Locations
Control
Navigate to / Pick & Place Objects

Multimodal:

Reach Any Object Specified in Any Way

Image

Language

Find ***the fruit basket on the kitchen counter***

Category

Bring me ***a CUP***

Lifelong:

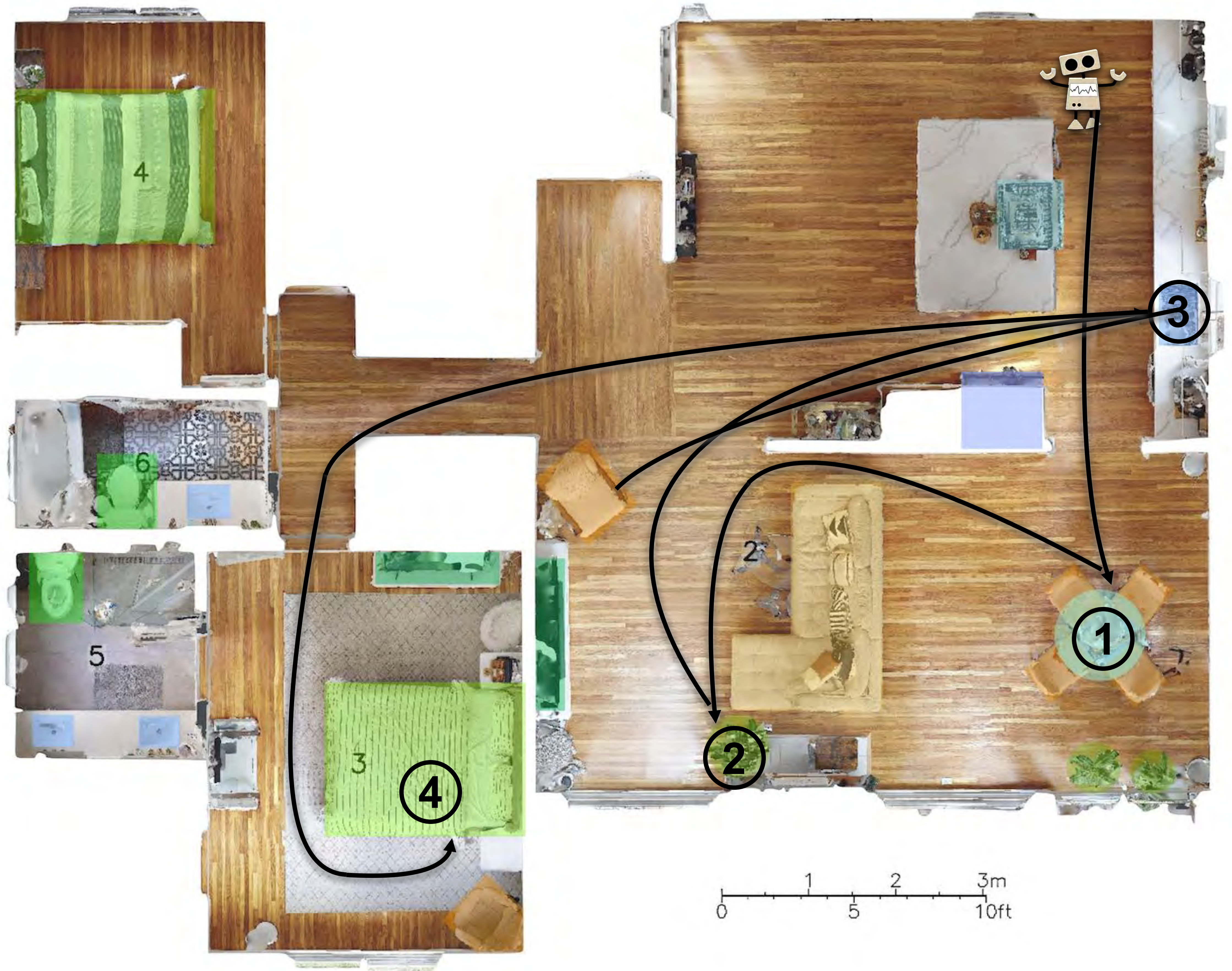
Remember Object Locations

1

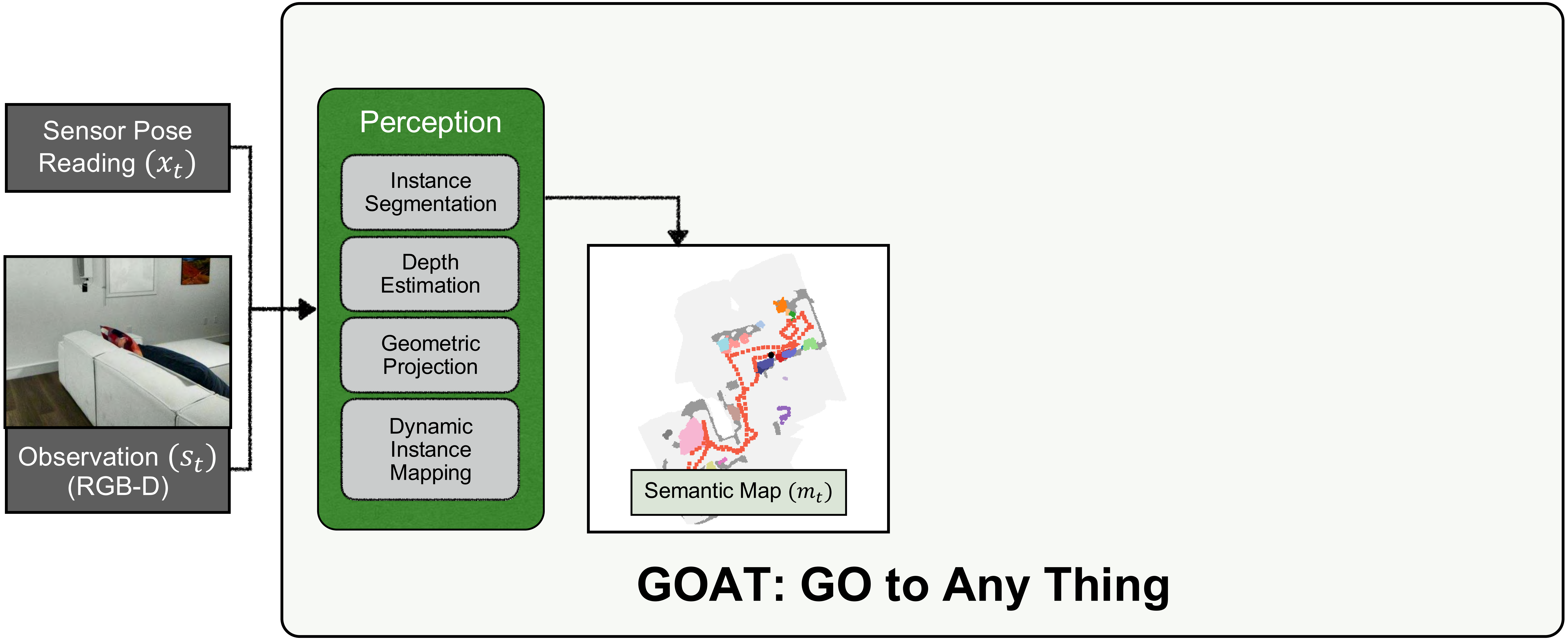
2 *Go to the potted plant next to the couch*

3 *Go to a SINK*

4 *Go to the black and white striped bed*



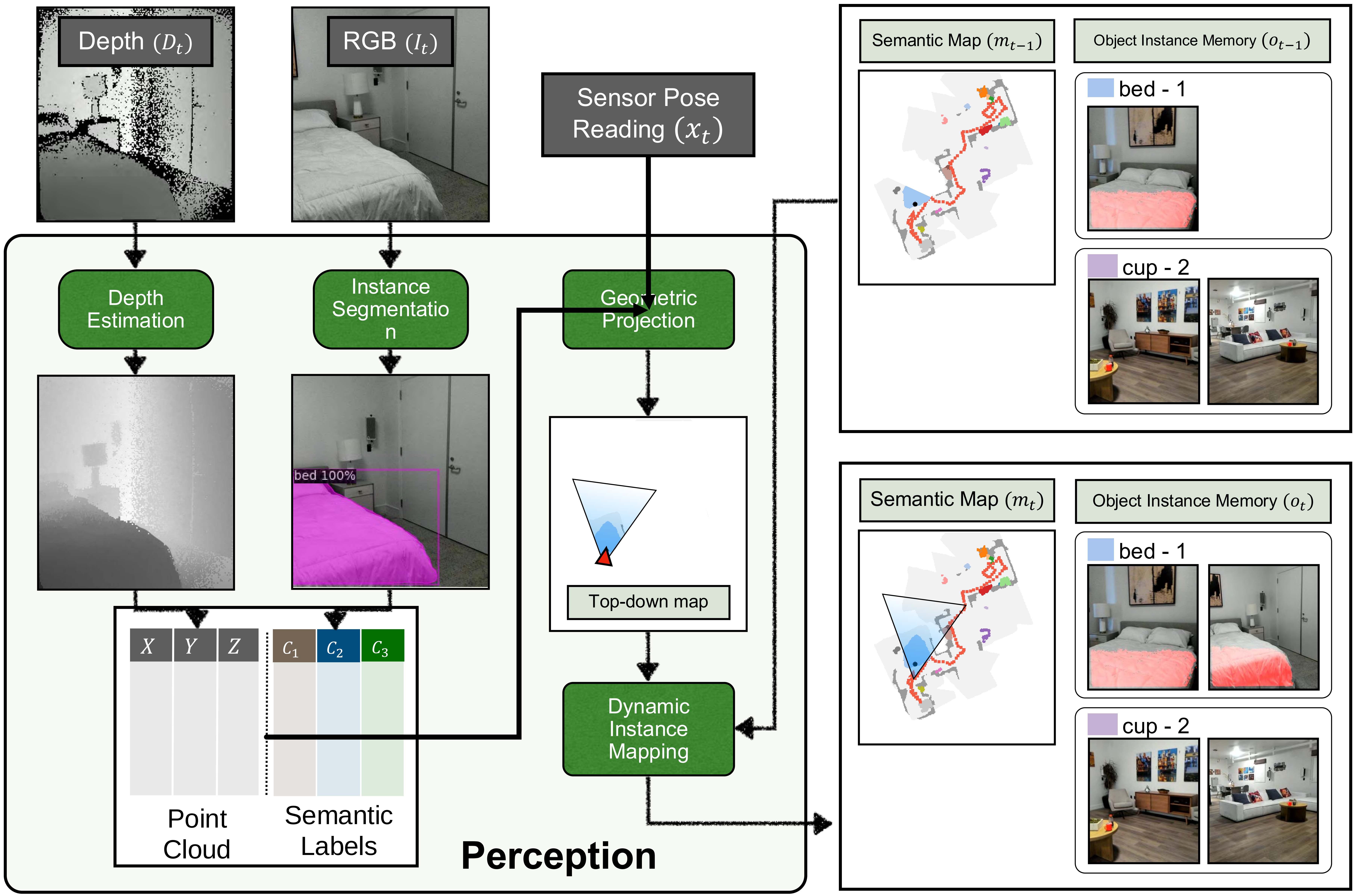
1. GOAT Problem
2. GOAT System Architecture
3. Results
4. Applications
   - Pick & Place
   - Social Navigation
   - Platform Agnostic


Sensor Pose Reading (x_t)
Observation (s_t) (RGB-D)
Perception
Instance Segmentation
Depth Estimation
Geometric Projection
Dynamic Instance Mapping
Semantic Map (m_t)
GOAT: GO to Any Thing

Language Goal
Brown office chair
Bed with white sheets
Cup on the kitchen counter
Image Goal
Category Goal
CHAIR
BOOK
CUP
Sensor Pose Reading (x_t)
Observation (s_t) (RGB-D)
Perception
Instance Segmentation
Depth Estimation
Geometric Projection
Dynamic Instance Mapping
π_G Global Policy
π_L Local Policy
Long-term goal (g_t)
chair - 3
couch - 1
Semantic Map (m_t)
Object Instance Memory (o_t)
Action (a_t)
GOAT: GO to Any Thing

System Architecture
Language Goal
Brown office chair
Bed with white sheets
Cup on the kitchen counter
Image Goal
Category Goal
CHAIR
BOOK
CUP
Sensor Pose Reading (x_t)
Observation (s_t) (RGB-D)
Perception
Instance Segmentation
Depth Estimation
Geometric Projection
Dynamic Instance Mapping
Global Policy π_G
Local Policy π_L
Long-term goal (g_t)
chair - 3
couch - 1
Semantic Map (m_t)
Object Instance Memory (o_t)
Action (a_t)
GOAT: GO to Any Thing

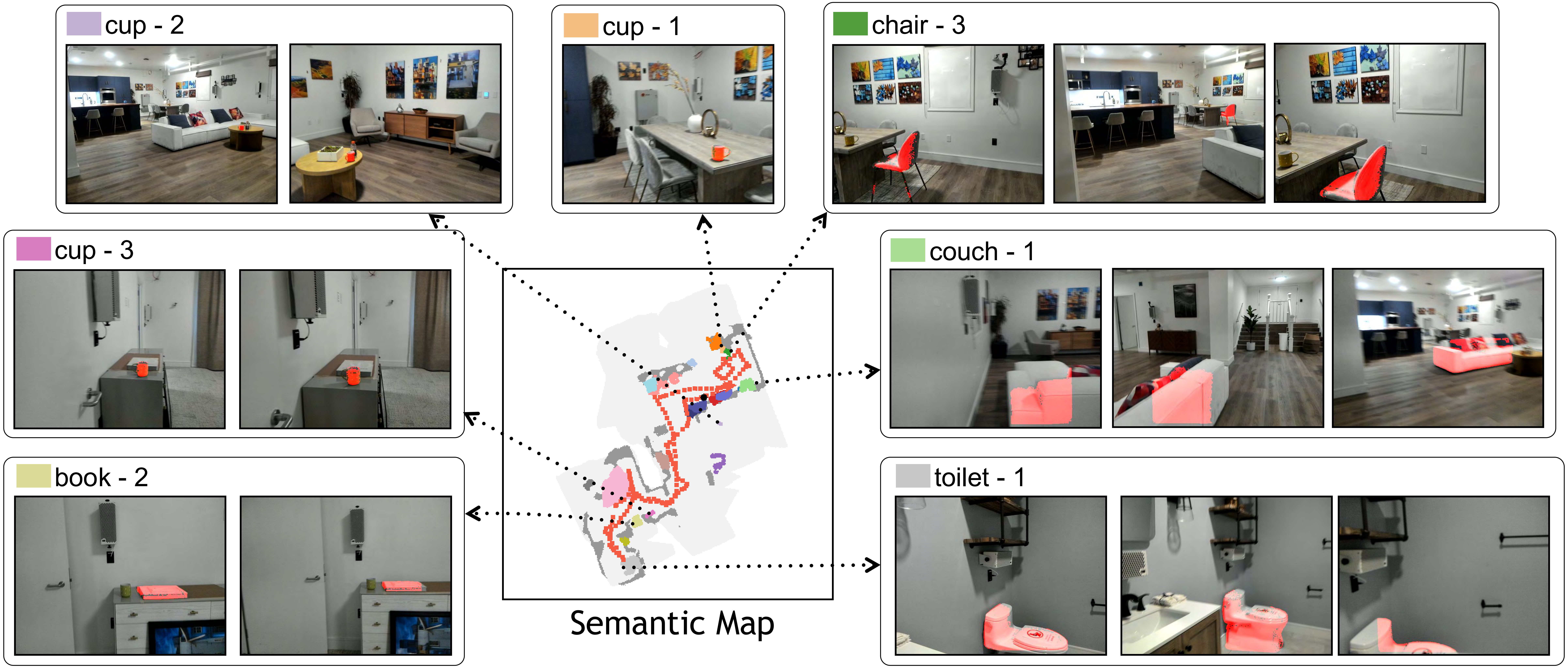
Perception System

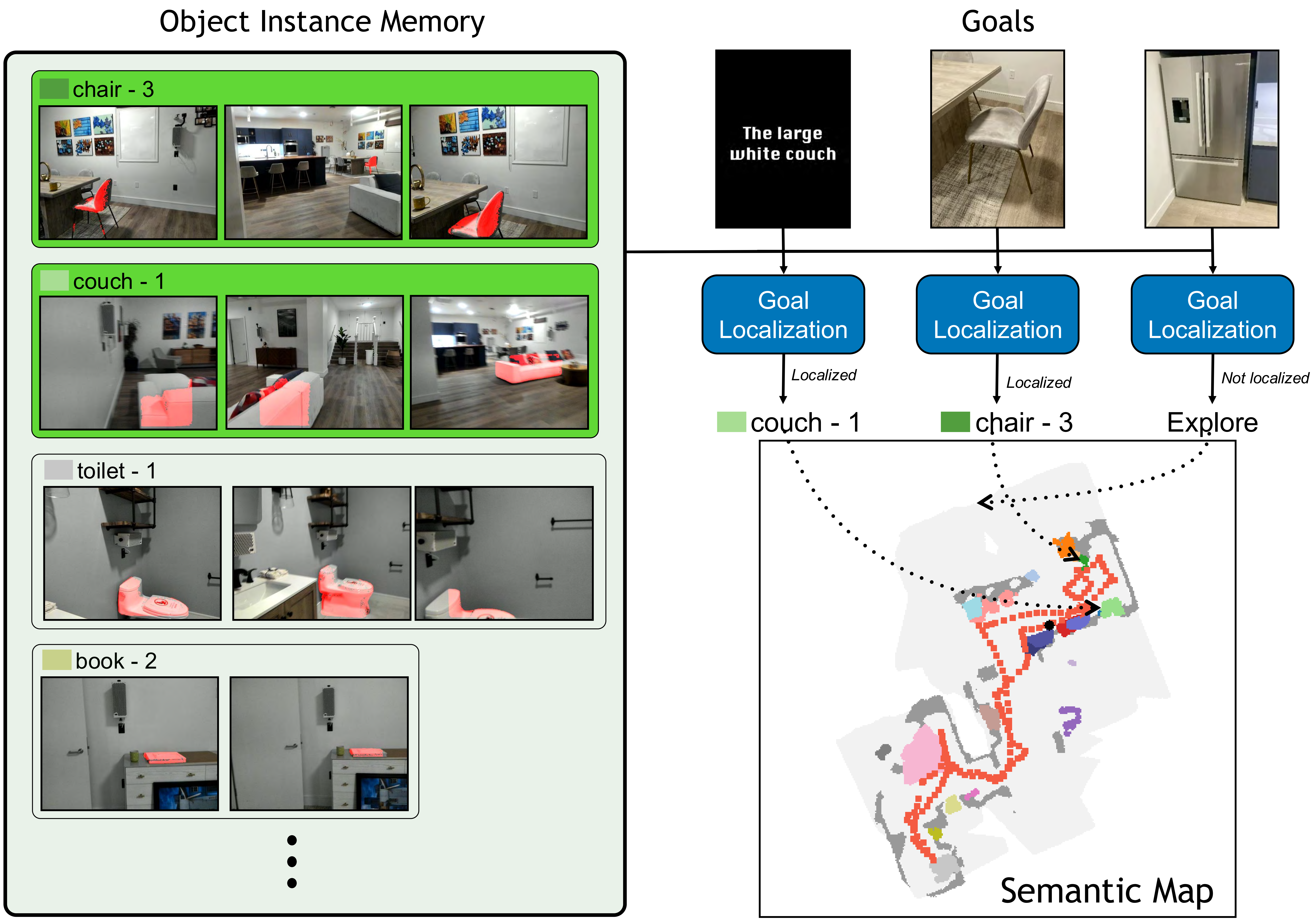


System Architecture
Language Goal
Brown office chair
Bed with white sheets
Cup on the kitchen counter
Image Goal
Category Goal
CHAIR
BOOK
CUP
Global Policy
π_G
Local Policy
π_L
Sensor Pose Reading (x_t)
Perception
Instance Segmentation
Depth Estimation
Geometric Projection
Dynamic Instance Mapping
Long-term goal (g_t)
Observation (s_t) (RGB-D)
chair - 3
couch - 1
Semantic Map (m_t)
Object Instance Memory (o_t)
Action (a_t)
GOAT: GO to Any Thing

GOAT Memory Representation

Semantic map** with associated **Object Instance memory





Object Instance Memory
chair - 3
couch - 1
toilet - 1
book - 2
Goals
The large white couch
Goal Localization
Goal Localization
Goal Localization
Localized
Localized
Not localized
couch - 1
chair - 3
Explore
Semantic Map

System Architecture

1. GOAT Problem
2. GOAT System Architecture
3. Results
4. Applications
   - Pick & Place
   - Social Navigation
   - Platform Agnostic

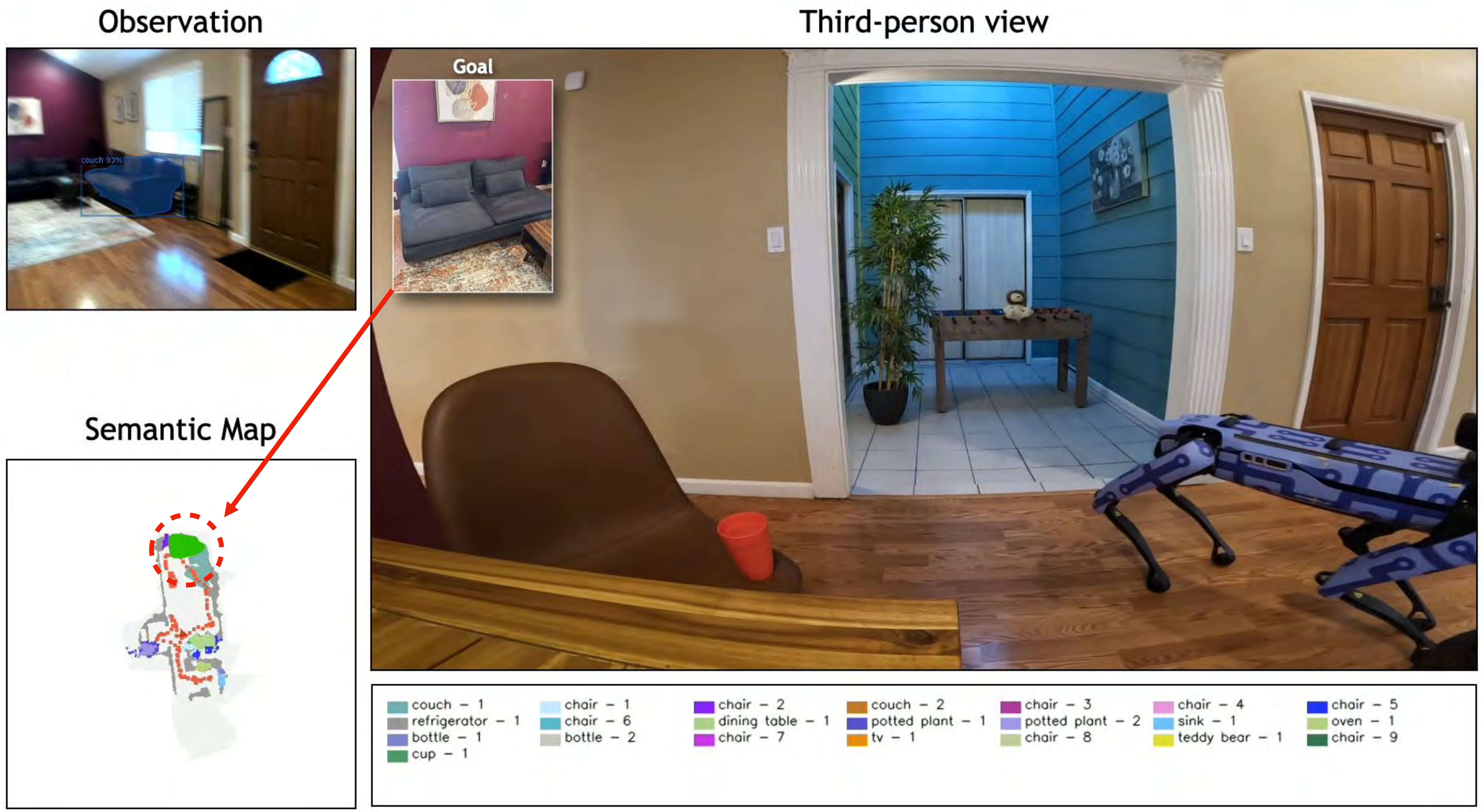
"In the Wild"
Empirical Evaluation

9 Unseen Homes
4 Methods
10 Trajectories per Home
5-10 Goals per Trajectory
~90h of Experiments

"In the Wild"
Empirical Evaluation
200+ Object Instances
The bed with the white blanket pulled back halfway and grey sheets.
The bunk bed with stars on the blanket.
The bed with blue and white sheets.
The with the blue and yellow painting above it and plant on the right.
The bed with the blue blanket and blue pillows.
The bed with grey sheets and light green trim.
The bed with the white blanket and red pillows
The large grey living room couch with many pillows.
The green couch.
The white rectangular couch with no pillows.
The large grey living room couch with many pillows.
The light brown couch with blue pillows.
The large grey couch in front of the yellow wall.
The couch covered in a white blanket.
The green cup on the kitchen counter
The red cup on the kitchen counter
The green mug on the plastic chair.
The grey cup on the nightstand.
The red cup on top of the glass table.
The red cup on the kitchen counter.
The light blue cup.
The plant in front of the window.
The potted plant next to the stairs.
The large potted plant in the bedroom.
The group of plants in front of the curtain.
The large potted plant next to the foosball table.
The large potted plant in front of the mirror.
The small potted plant on the hallway table
The grey dining table chair.
The green lawn chair.
The black office chair.
The small wooden chair at the desk.
The black leather chairs in the kitchen.
at the kitchen island.
a wooden seat.
wood desk.
the kitchen counter
The light blue bowl on the living room coffee table.
The lavender bowl on the kitchen counter.
The light blue bowl on the couch.
The green cover book on the coffee table.
The book on the desk. It has a car on the cover.
The beige teddy bear.
The stuffed lion toy.
The green dinosaur stuffed toy.
The stuffed lion.
The beige teddy bear.
The stuffed lion toy.
The green dinosaur stuffed toy.
The refrigerator.
The refrigerator.
The refrigerator.
The refrigerator.
The toilet next to the shower curtain with blue fish.
The toilet.
The toilet next to the glass shower door.
The kitchen sink.
The bathroom sink with marble top.
The kitchen sink.
The bathroom sink with green counter.
The bathroom sink.
The kitchen sink.
The bathroom sink with black and white counter.
The television.
The television mounted on a yellow wall.
The television mounted on a white wall.
The oven.
The oven.
The oven.
The oven.

Observation

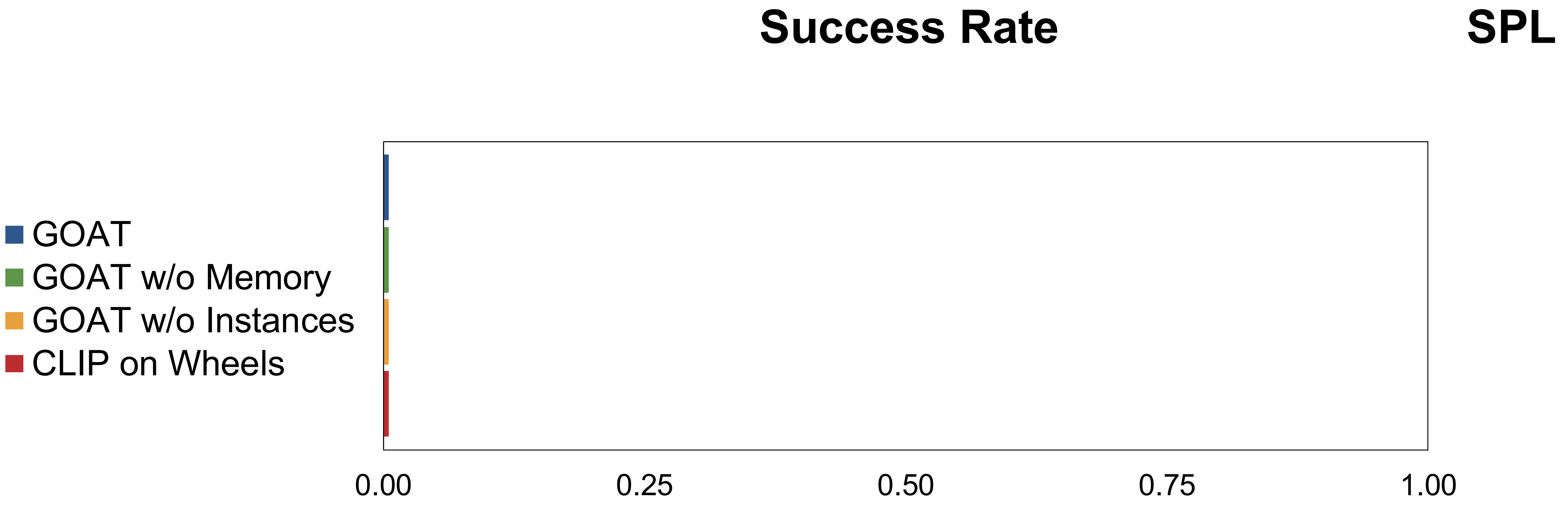
Third-person view



Observation
couch 93%
Third-person view
Goal
Semantic Map
couch – 1
refrigerator – 1
bottle – 1
cup – 1
chair – 1
chair – 6
bottle – 2
chair – 2
dining table – 1
chair – 7
couch – 2
potted plant – 1
tv – 1
chair – 3
potted plant – 2
chair – 8
chair – 4
sink – 1
teddy bear – 1
chair – 5
oven – 1
chair – 9

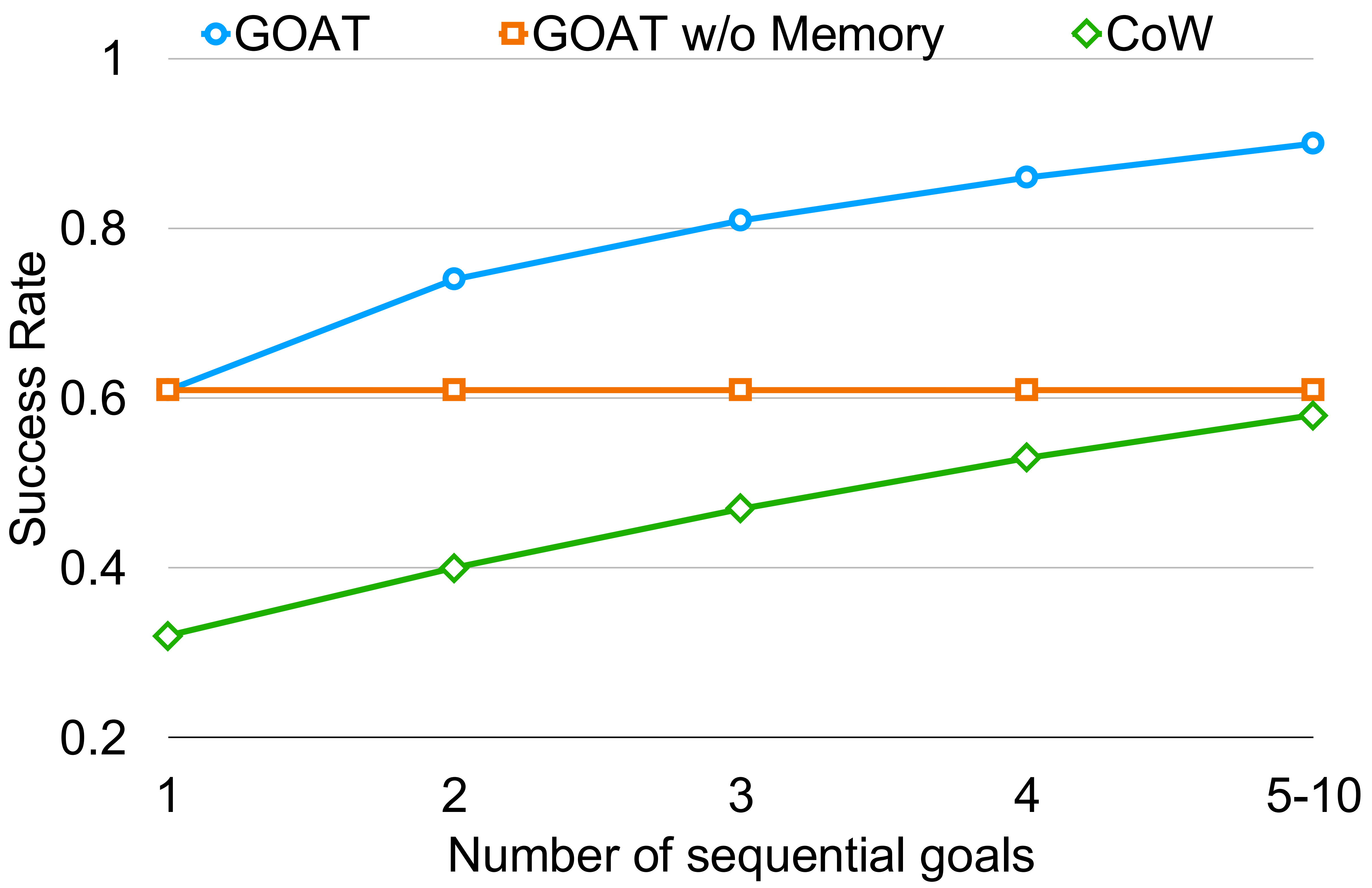
Third-person view

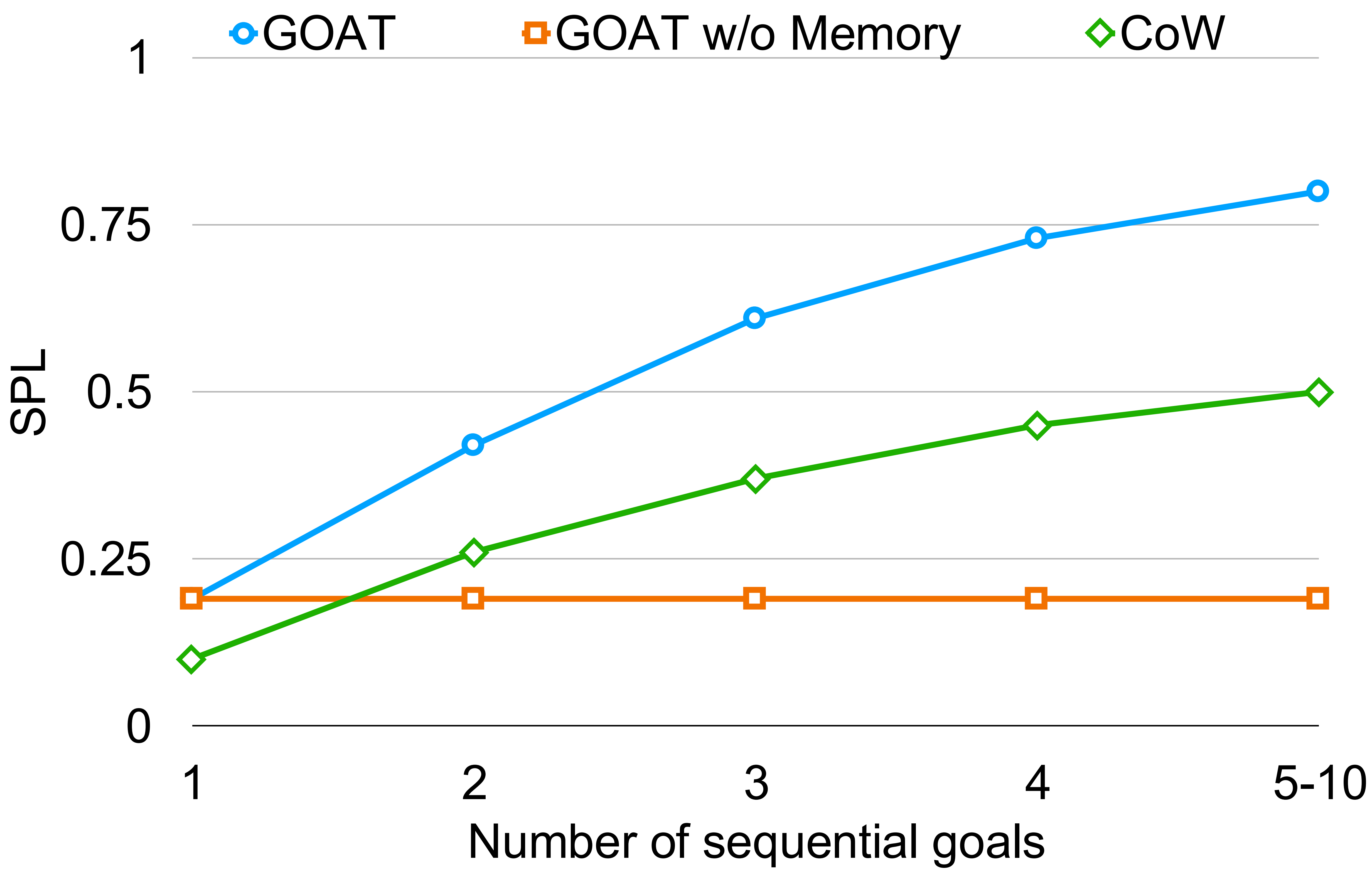
Results



Success Rate
SPL
GOAT
GOAT w/o Memory
GOAT w/o Instances
CLIP on Wheels
0.00
0.25
0.50
0.75
1.00

Performance Across Episode





Goal: 

Baselines

Ours

Observation Instance Map

Success: 6/6 SPL: 0.78

No Memory

Observation Instance Map

Success: 4/6 SPL: 0.40

CoW

Observation Instance Map

Success: 1/6 SPL: 0.16

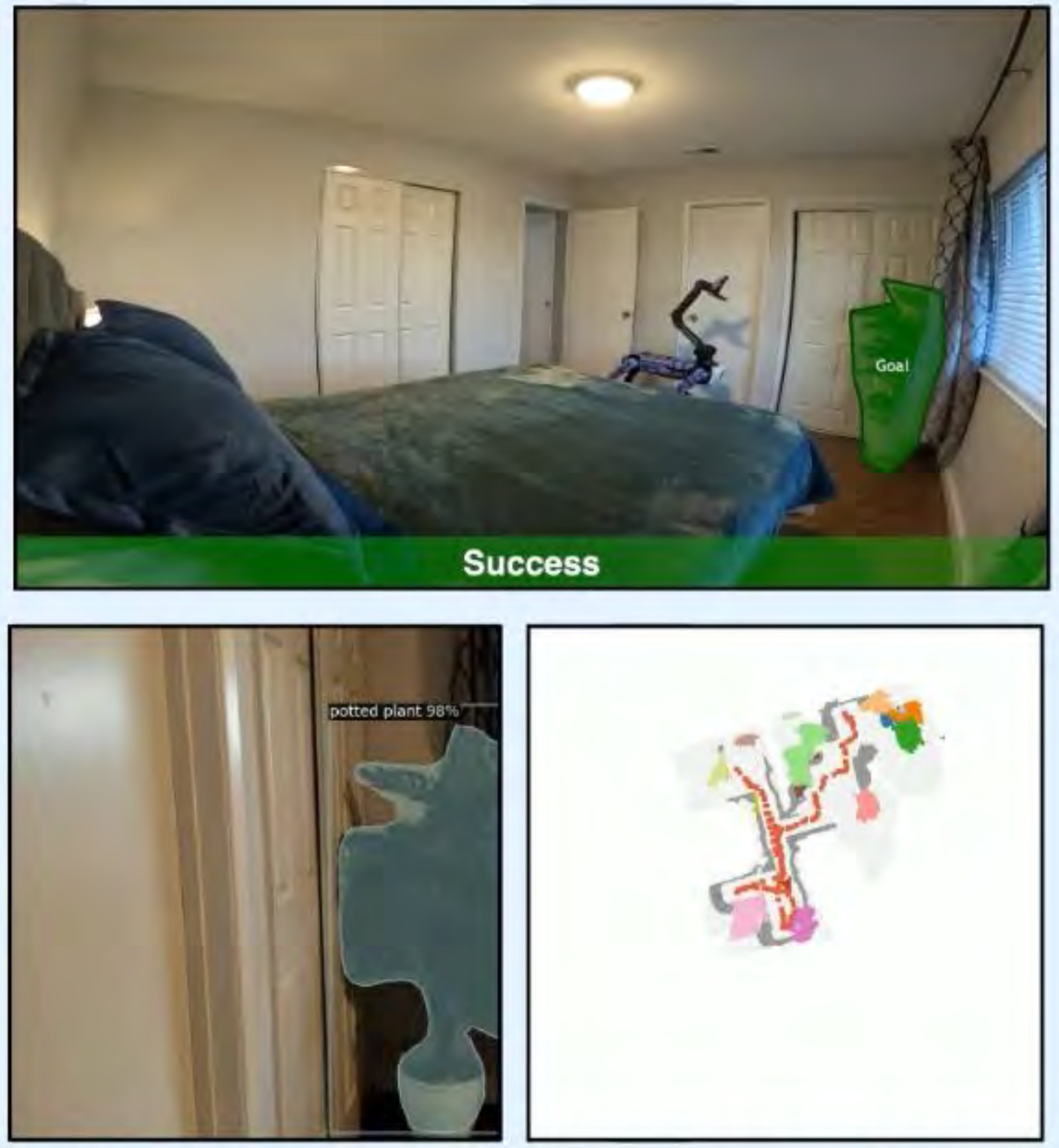
Goal:
Baselines
Ours
Success
Observation
Instance Map
No Memory
Success
Observation
Instance Map
CoW
Failure
Observation
Instance Map
Success: 6/6 SPL: 0.78
Success: 4/6 SPL: 0.40
Success: 1/6 SPL: 0.16

Goal:
Baselines
Ours
Success
Observation
Instance Map
No Memory
Success
Observation
Instance Map
CoW
Failure
Observation
Instance Map
Success: 6/6 SPL: 0.78
Success: 4/6 SPL: 0.40
Success: 1/6 SPL: 0.16

Goal: 

Baselines

Ours



Observation Instance Map

Success: 6/6 SPL: 0.78

No Memory

Observation Instance Map

Success: 4/6 SPL: 0.40

CoW

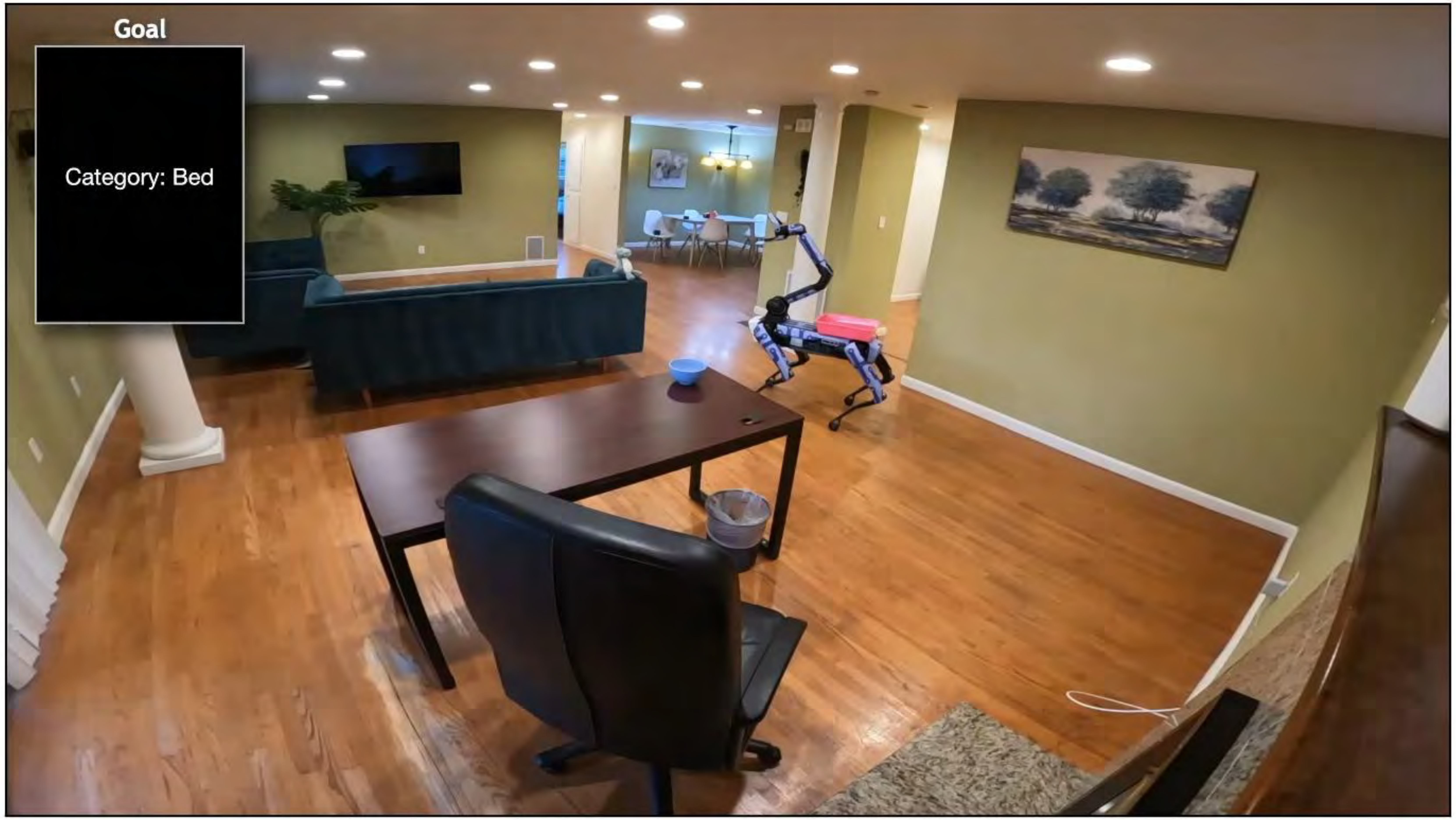
Observation Instance Map

Success: 1/6 SPL: 0.16

1. GOAT Problem
2. GOAT System Architecture
3. Results
4. Applications
   - Pick & Place
   - Social Navigation
   - Platform Agnostic

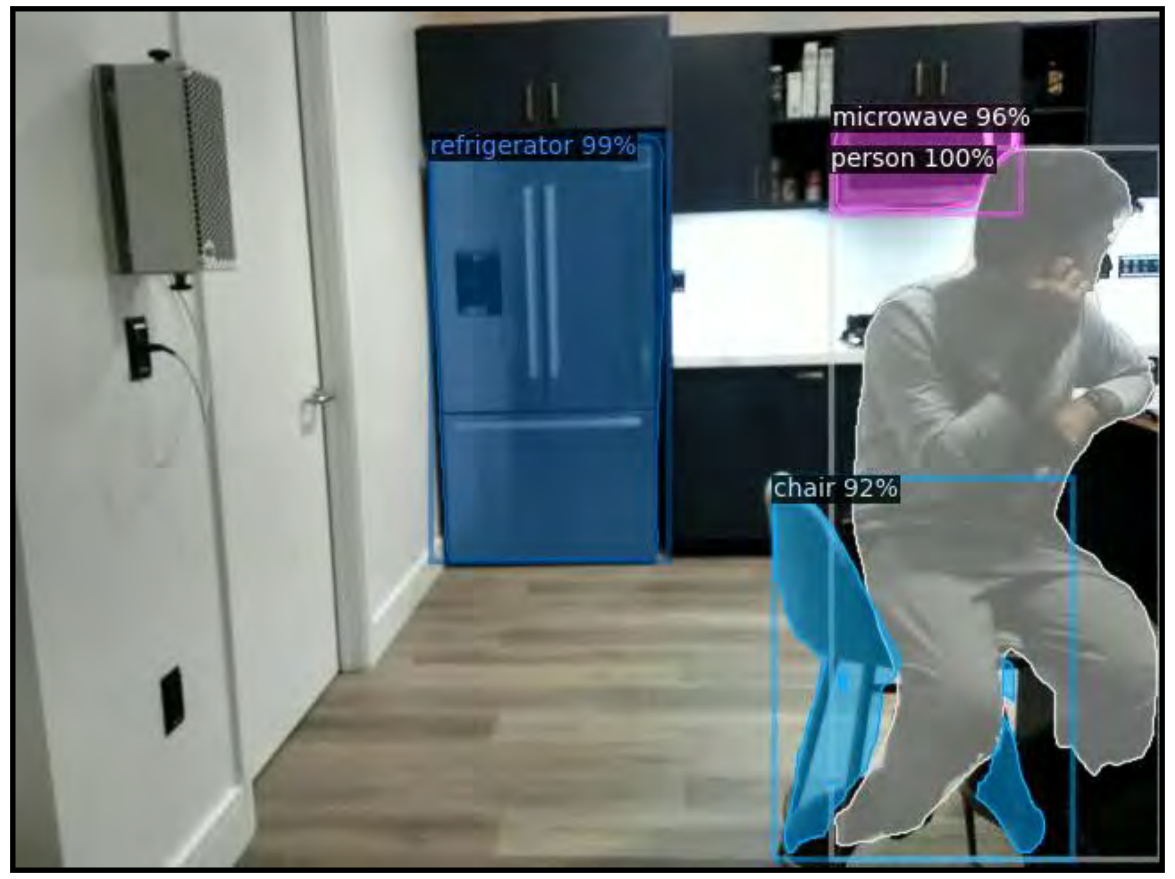
Pick & Place

Third-person view



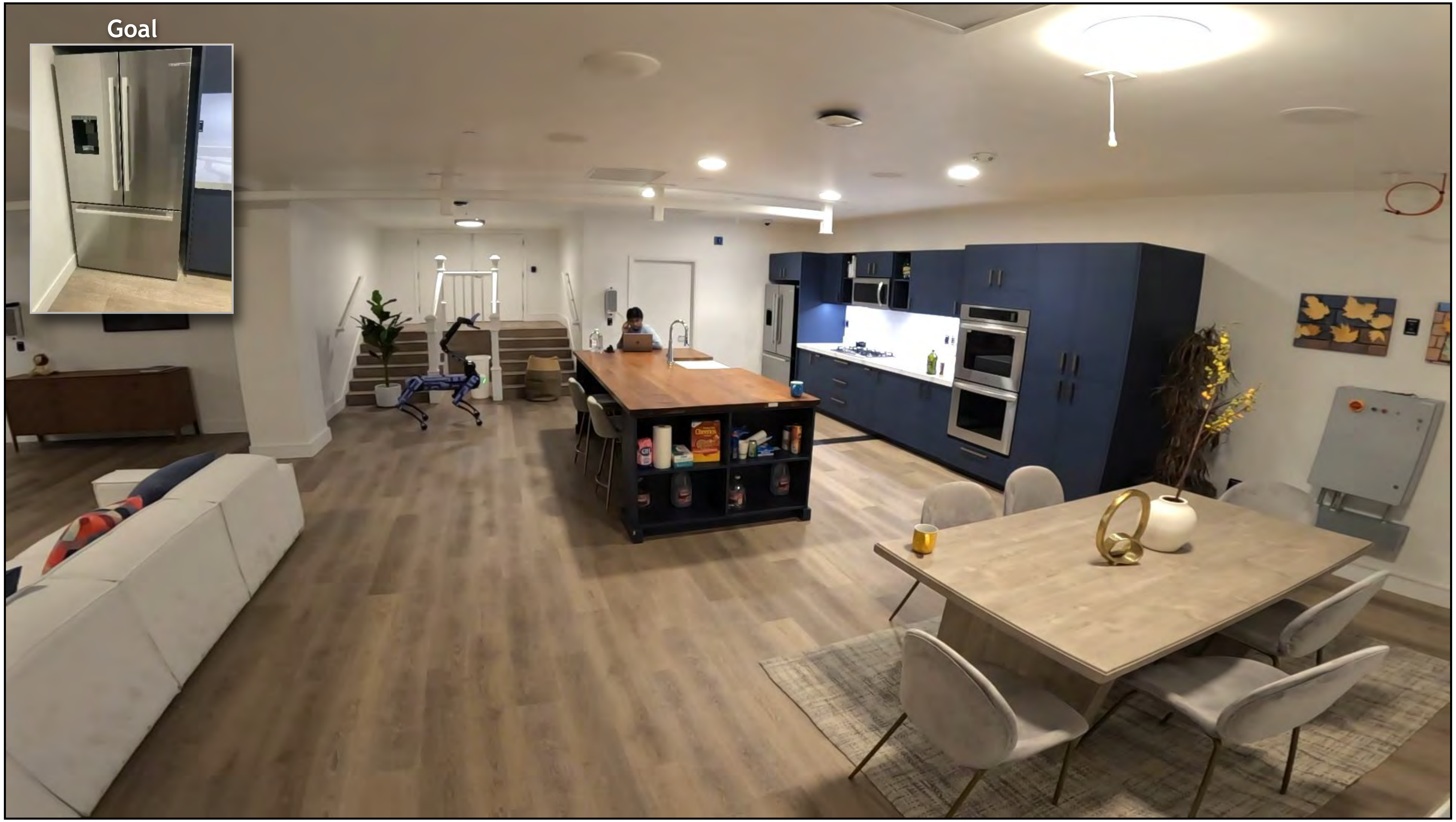
Social Navigation

Observation



Semantic Map

Third-person view

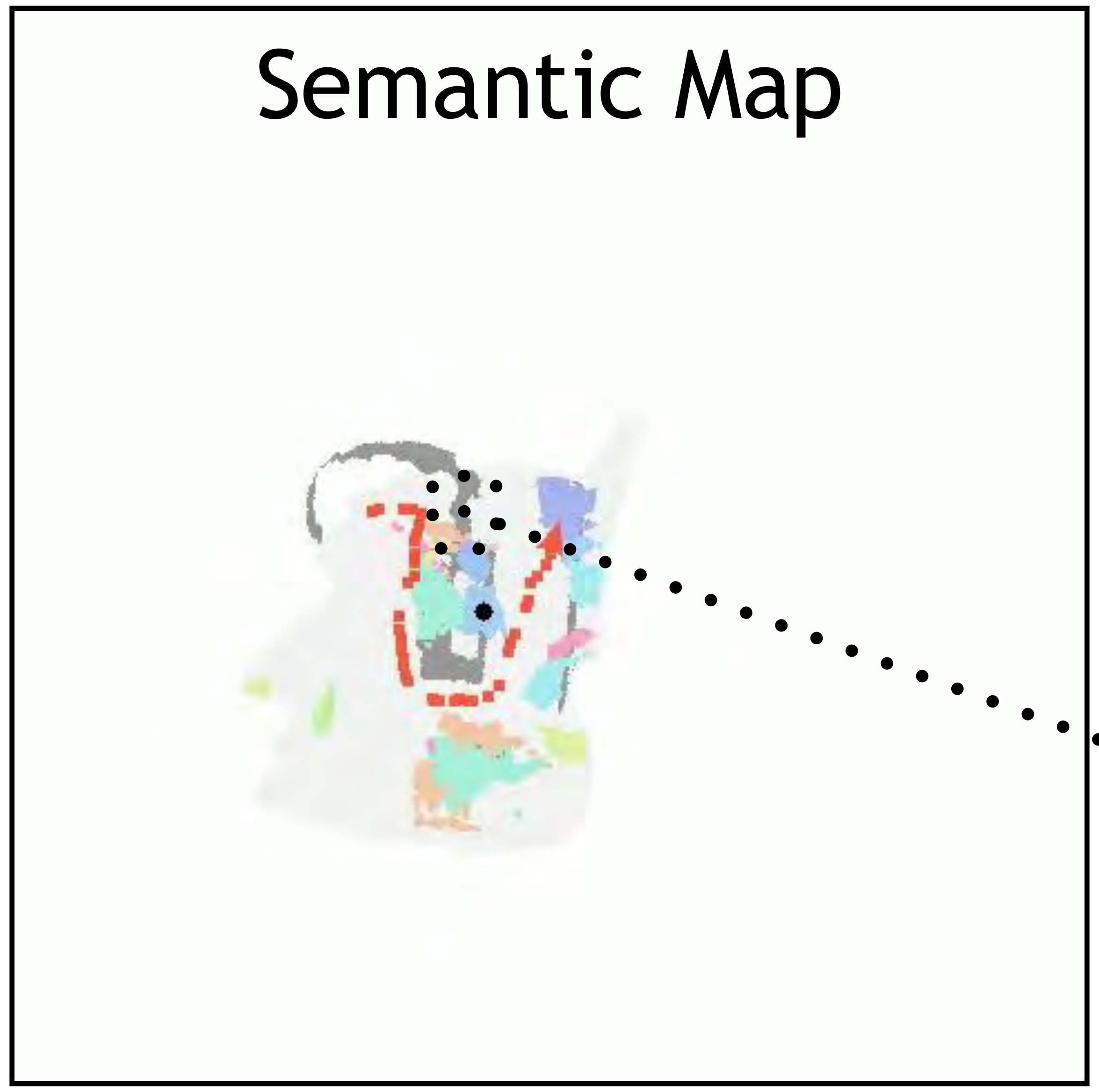


Robot plans around the dynamic obstacle (person) to go to the refrigerator

Social Navigation

Observation

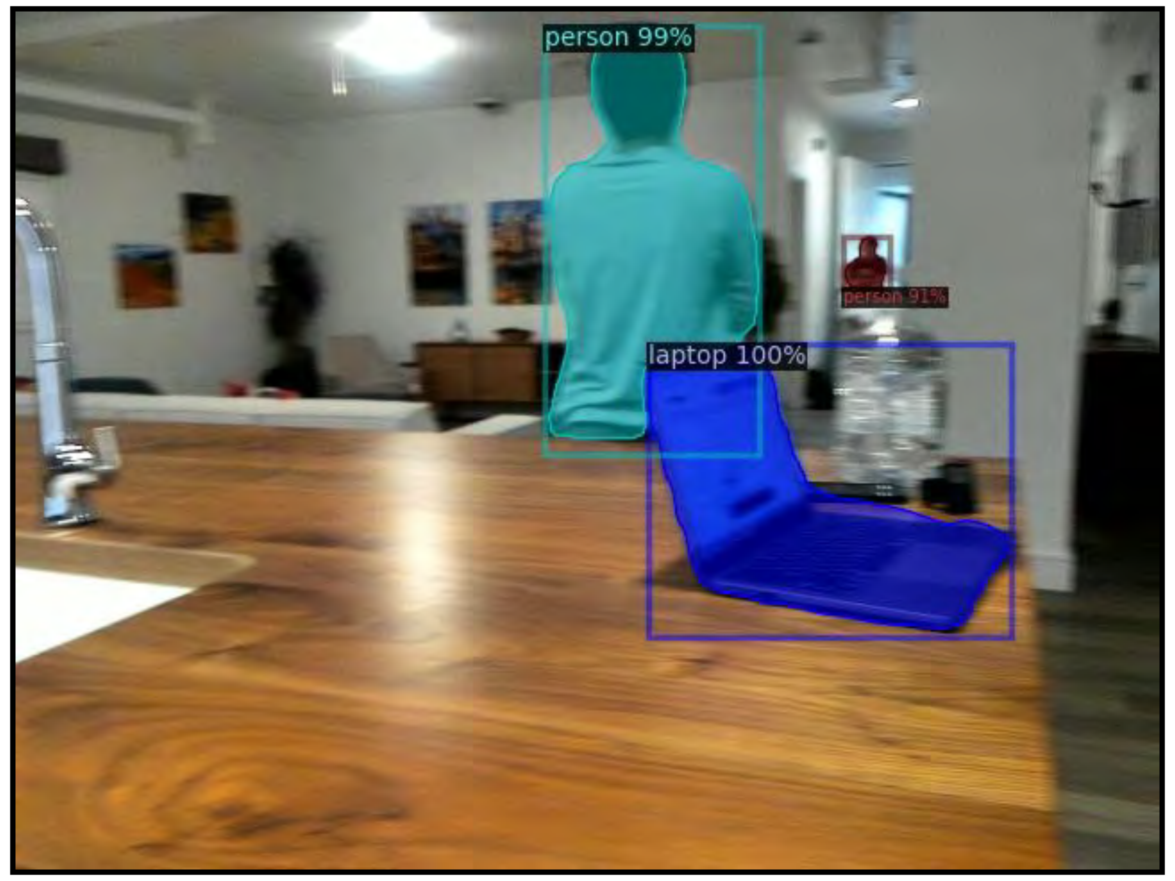
Third-person view



Robot removes previous location of person from the map

Social Navigation

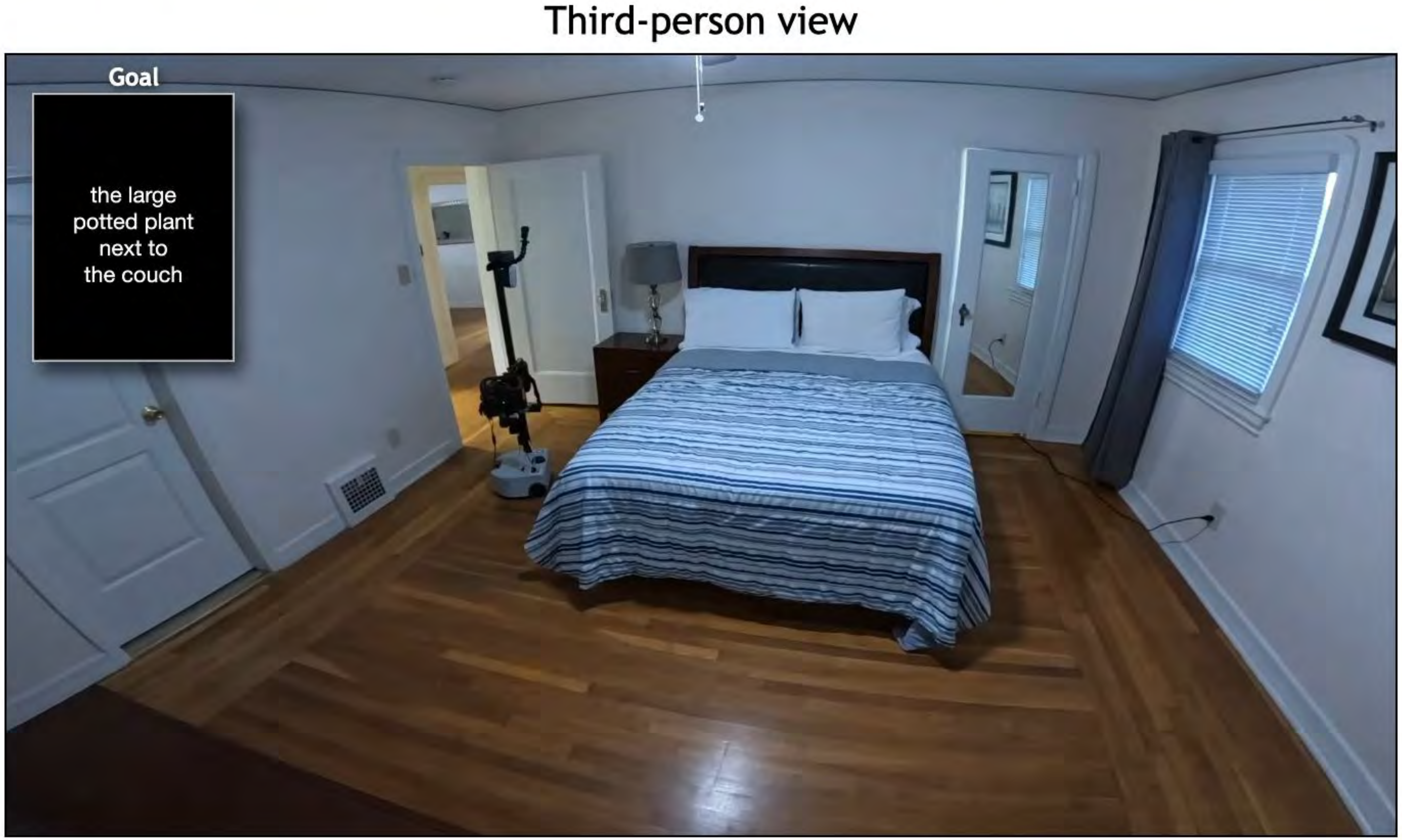
Observation



Third-person view

Robot follows the person while updating their location

Platform Agnostic



Summary

Universal navigation

- Multimodal

Image

Language

*Find **the fruit basket on the kitchen counter***

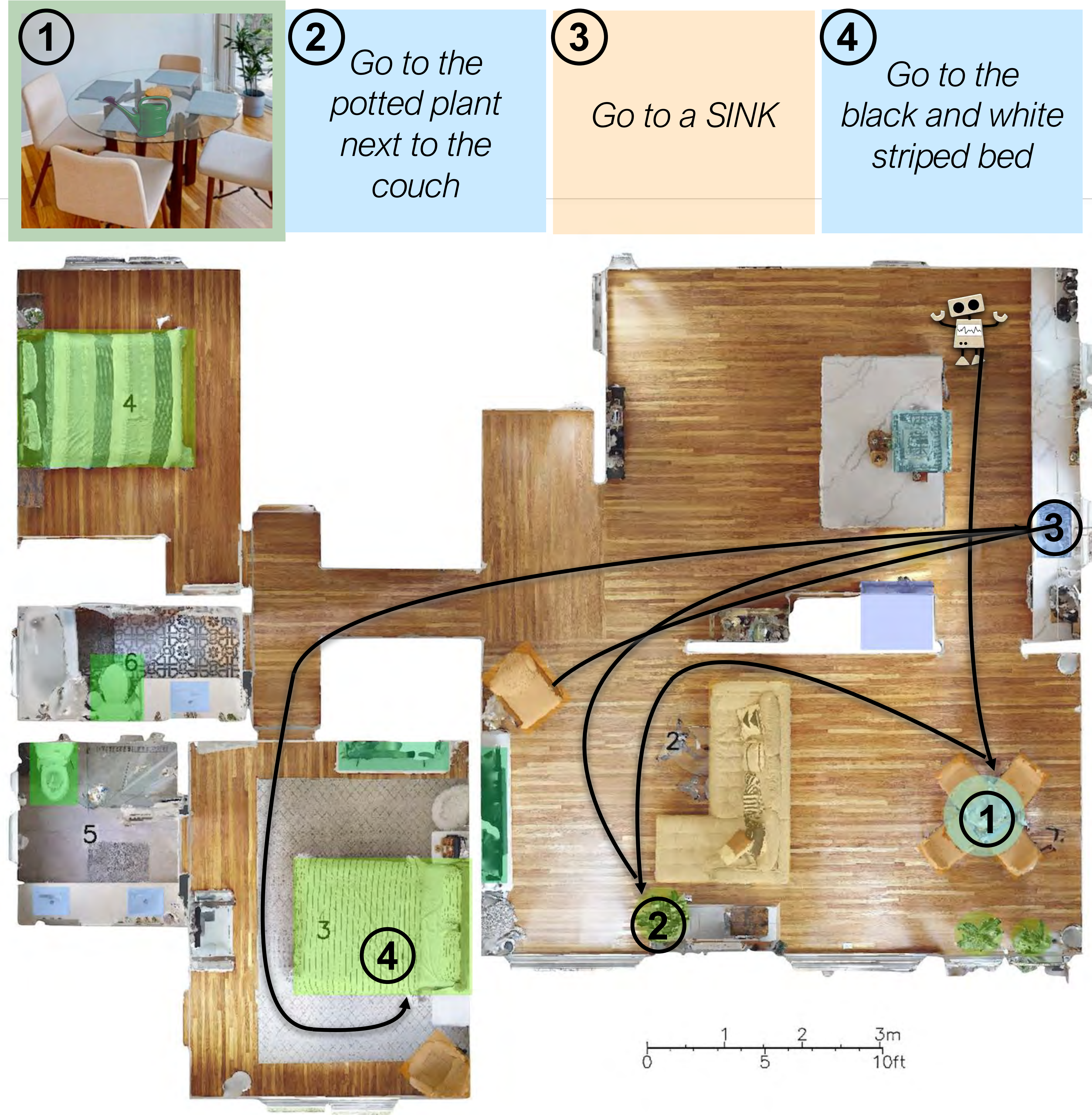
Category

*Bring me **a CUP***

Summary

Universal navigation

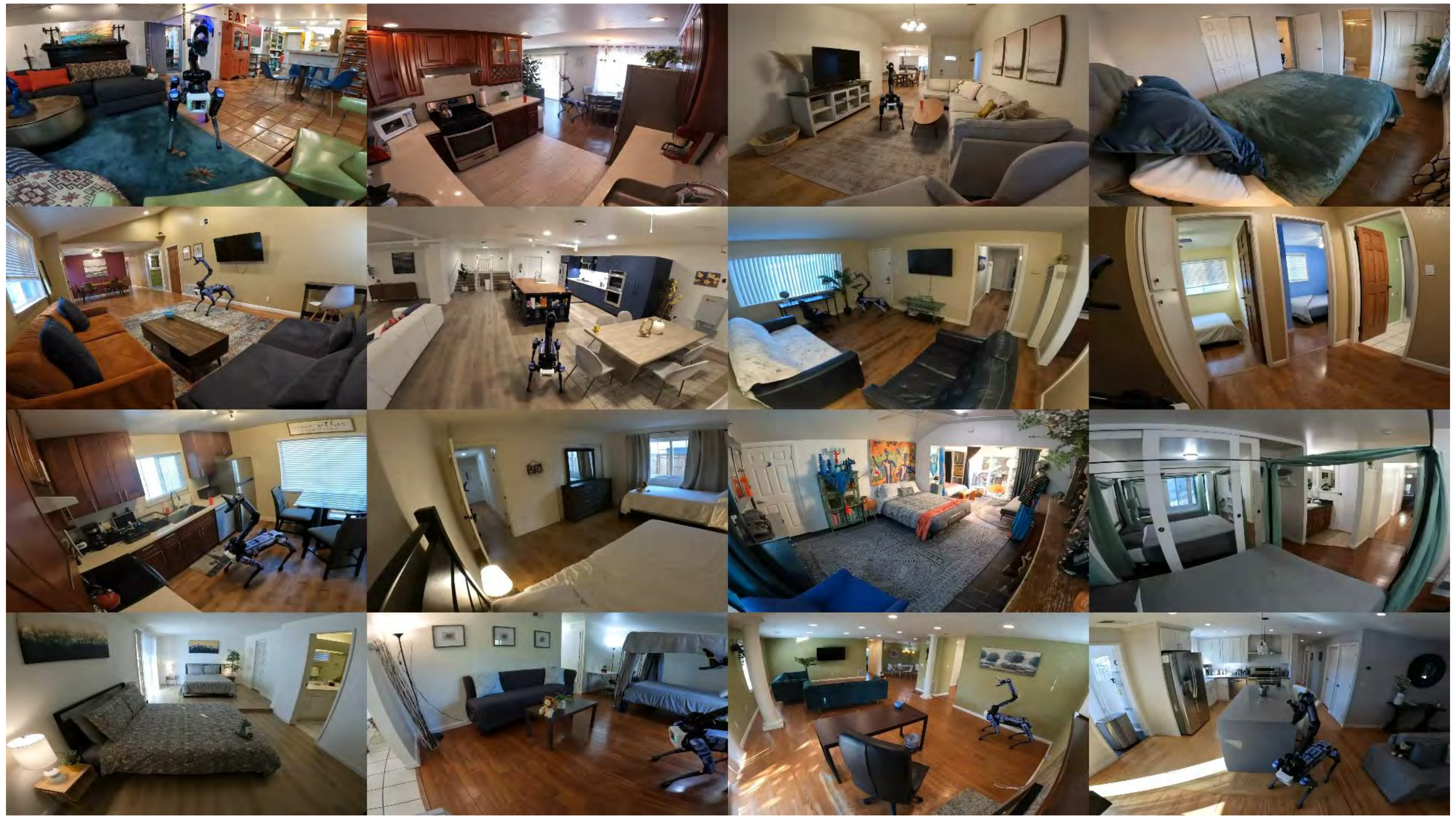
- Multimodal
- Lifelong



Summary

Universal navigation

- Multimodal
- Lifelong
- Unseen environments



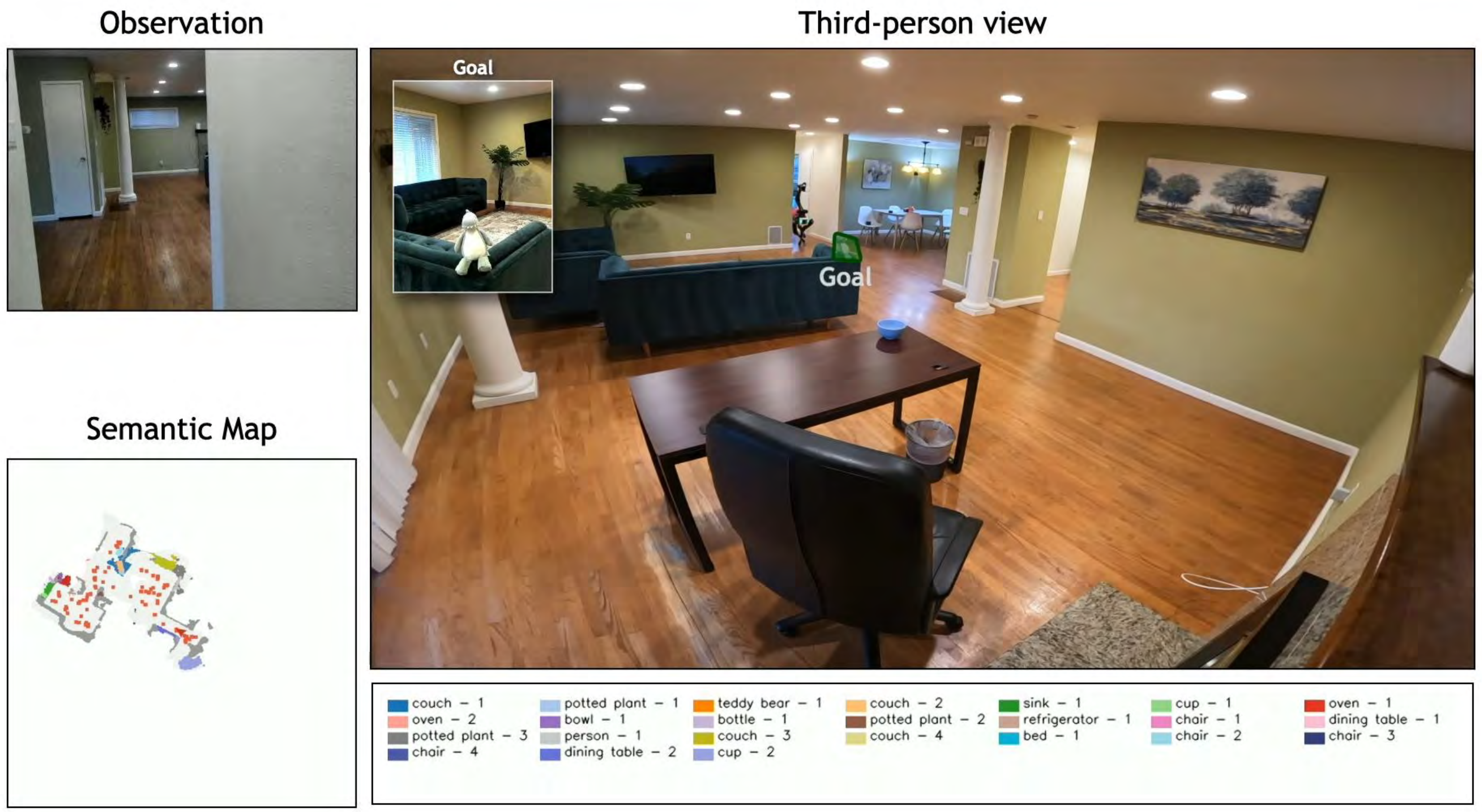
Summary

Universal navigation

- Multimodal
- Lifelong
- Unseen environments

Applications

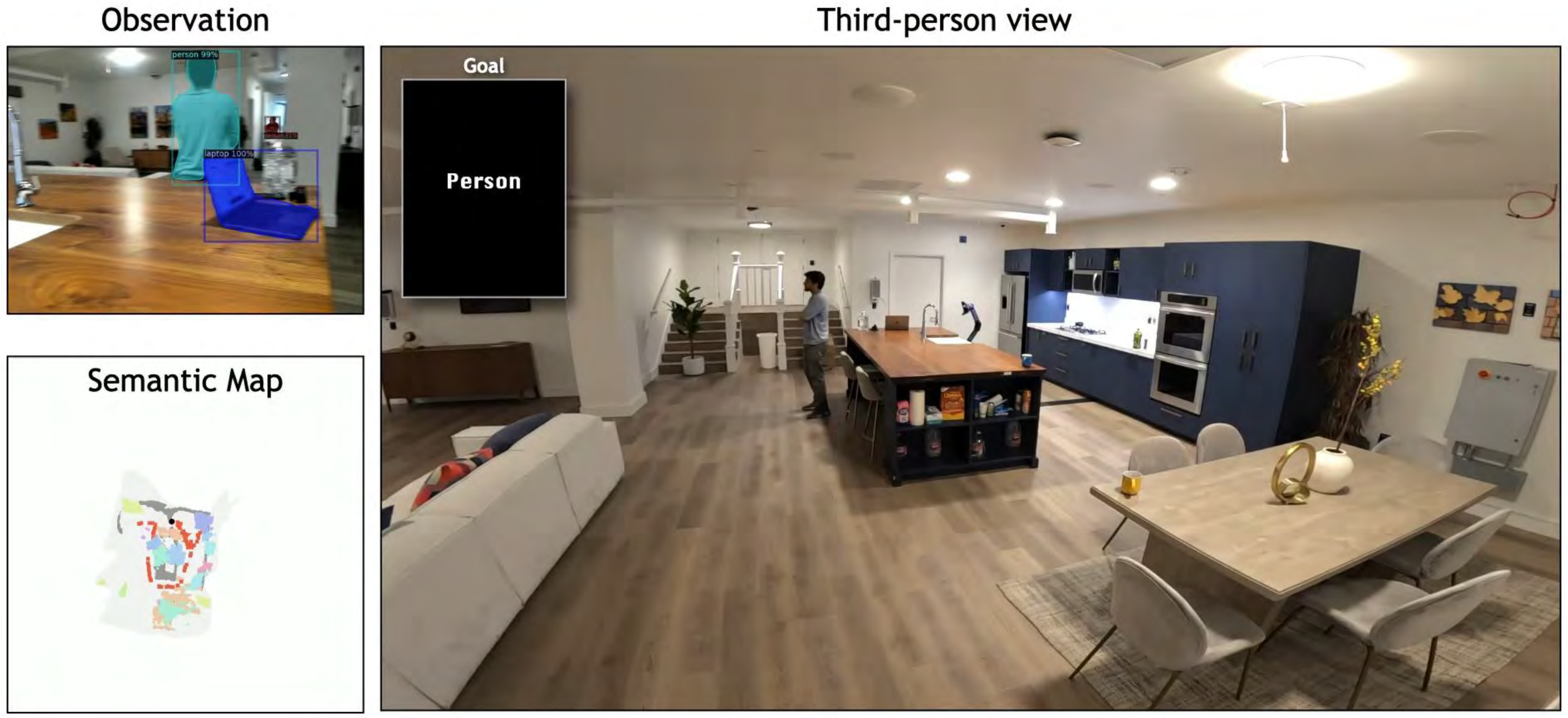
- Pick & Place



Summary

Universal navigation
- Multimodal
- Lifelong
- Unseen environments

Applications
- Pick & Place
- Social Navigation



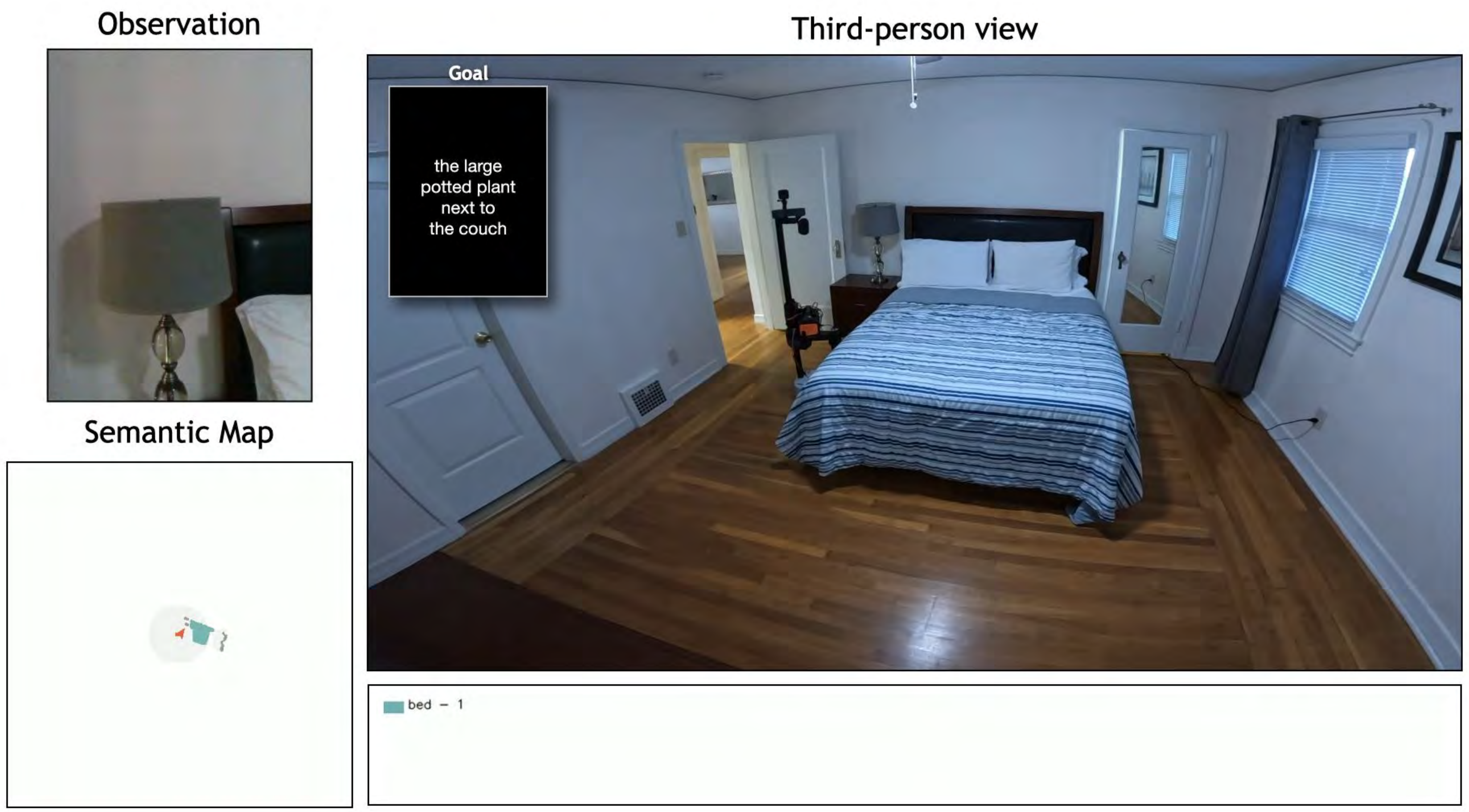
Summary

Universal navigation

- Multimodal
- Lifelong
- Unseen environments

Applications

- Pick & Place
- Social Navigation
- Platform Agnostic



Thank you!

Webpage: https://theophilegervet.github.io/projects/goat


Thank you!

Learning to Walk for Quadrupeds

Jitendra Malik
UC Berkeley


Recovery from leg obstruction
Rocky area next to river bed


Stairs on a hiking path


Unstable and constantly deforming ground
Loose Mud Pile at Construction Site

Vegetation on uneven surface


Bouncing (Gallop) Gait @ 1.5 m/s
bounce on carpet

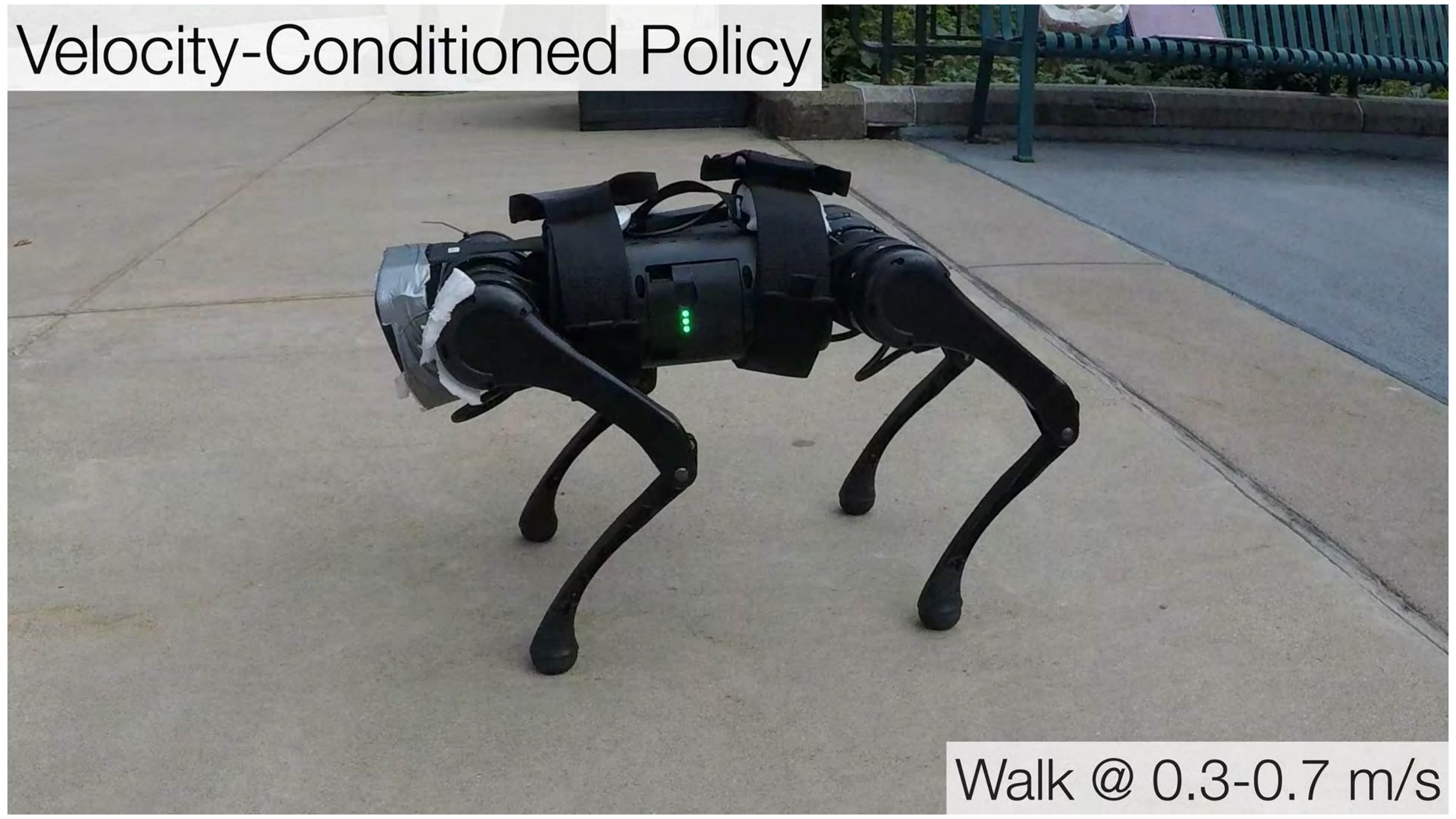
Navigation in Cluttered Indoor Setting

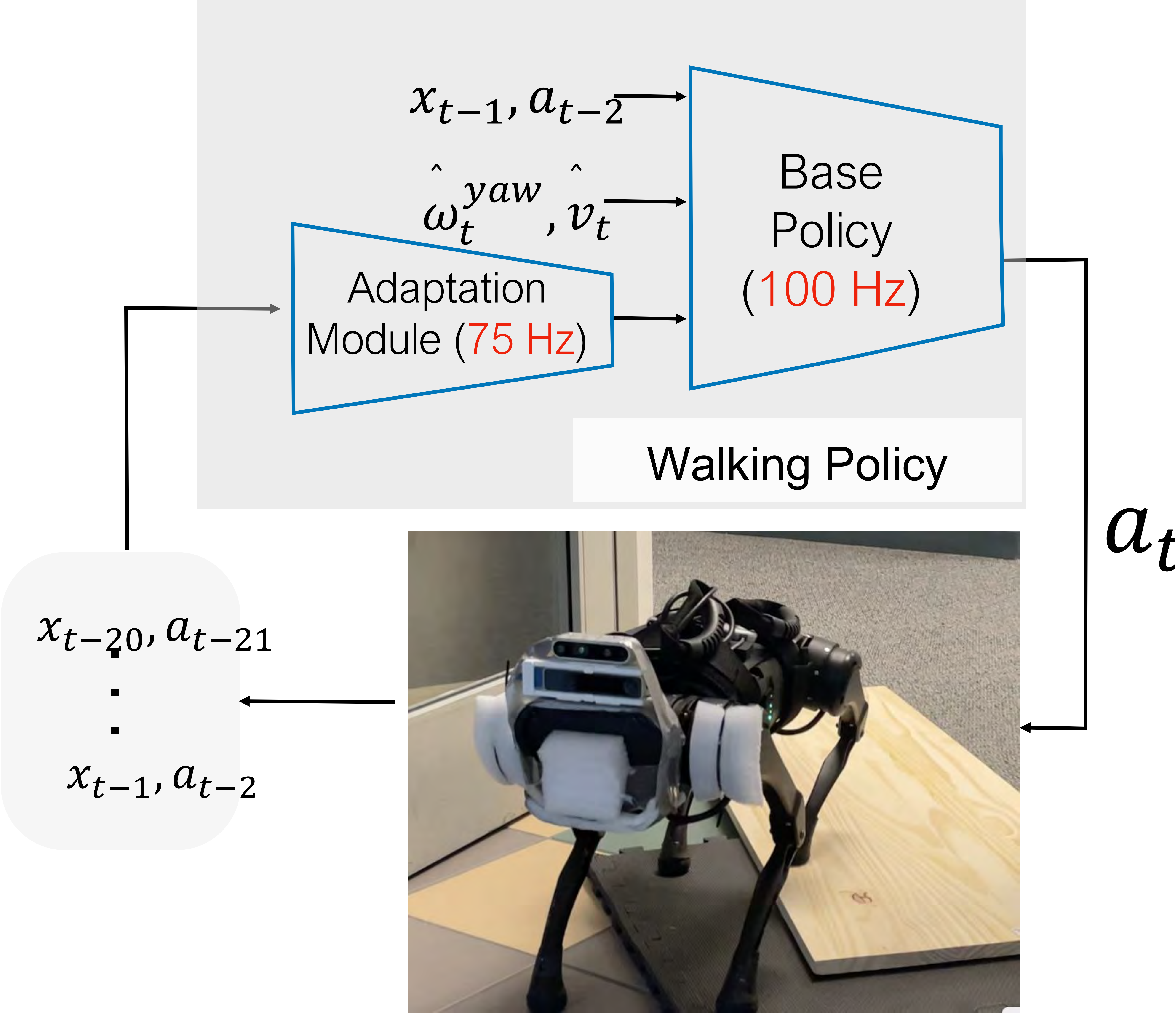
Coupling Vision and Proprioception for Navigation of Legged Robots

[1]Zipeng Fu*, [2]Ashish Kumar*, [1]Ananye Agarwal, [2]Haozhi Qi,
[2]Jitendra Malik, [1]Deepak Pathak

[1]CMU [2]UC Berkeley
CVPR 2022

Linear and Angular Velocity





Visual Path Planner

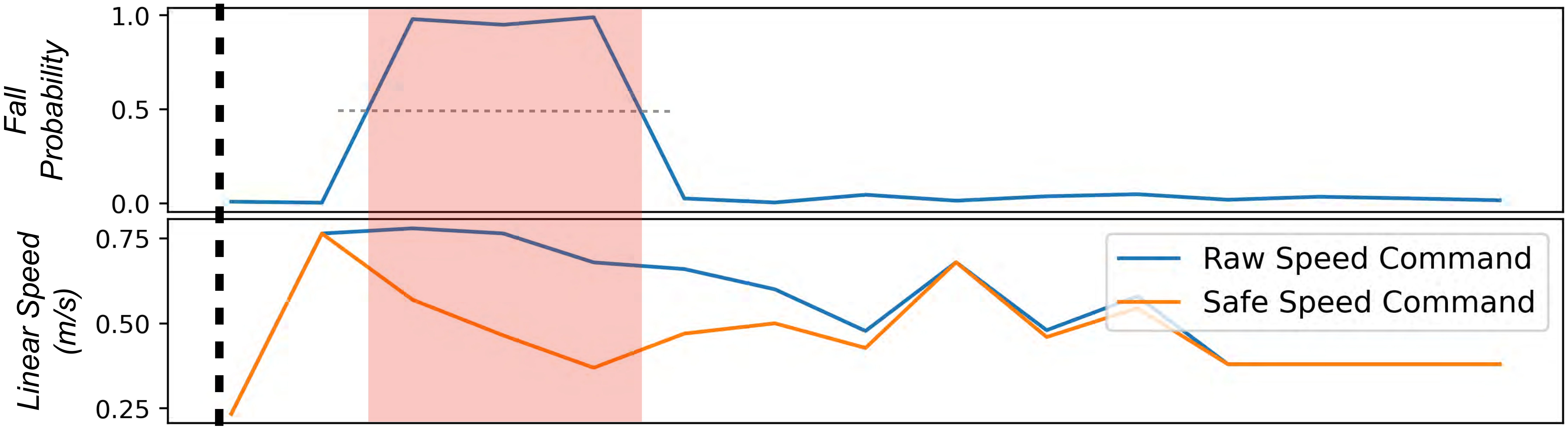
Navigation in Cluttered Indoor Setting

Whats Missing?

1. Navigation should be coordinated with locomotion

2. Vision should be coordinated with proprioception

Navigation coordinated with locomotion



Vision coordinated with proprioception

Coupled Navigation and Locomotion

Comparison of

Vision + Proprioception

Only Vision


Vision + Proprioception


Only Vision

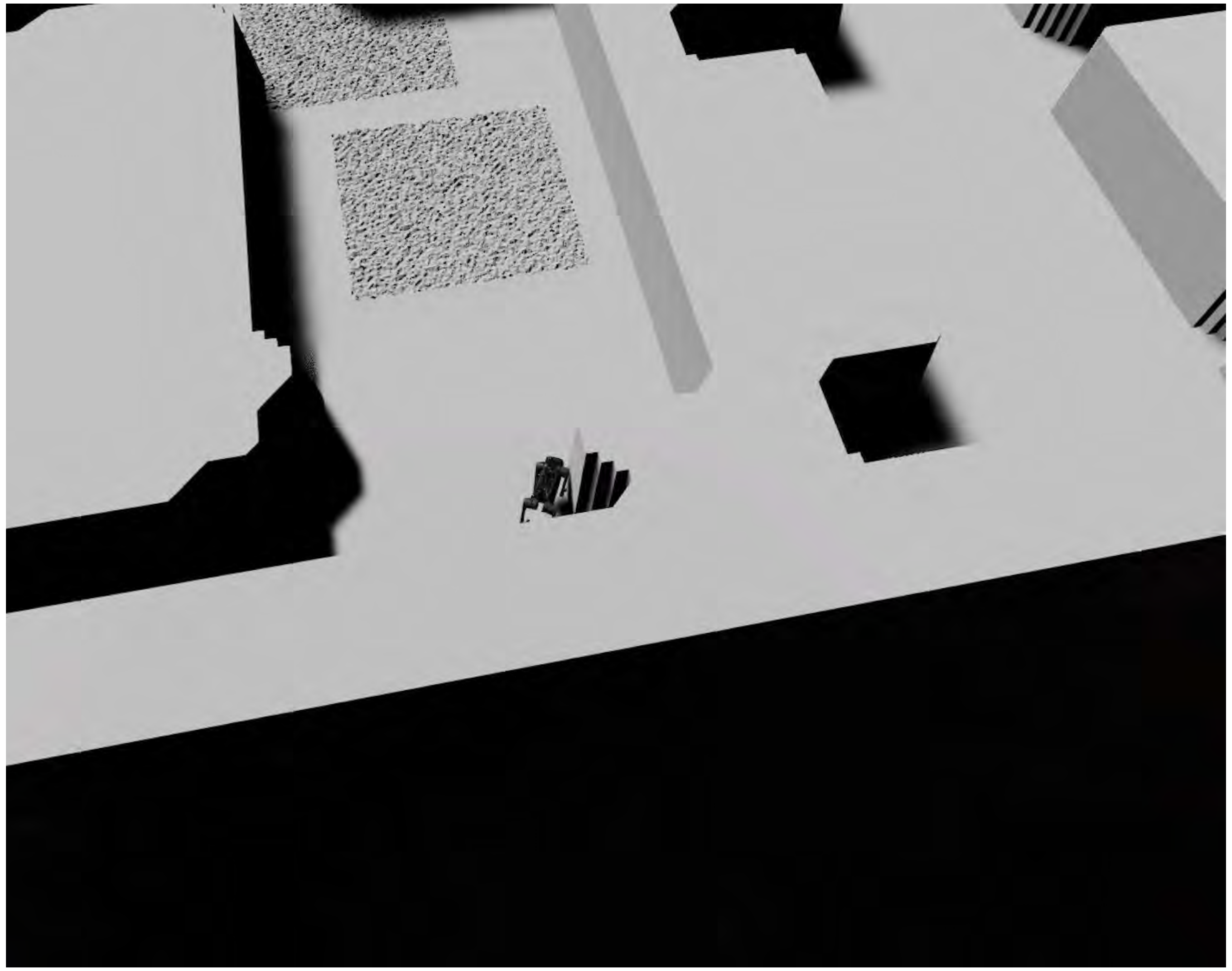

Vision + Proprioception
Only Vision

More Deployment Videos

Collision Detector
x0.5 slow
Collision Detector
1.0
0.5
0.0
Feet Contacts
RF
LF
RR
LR
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
time (s)


Navigation in Cluttered Indoor Setting
0.5
0.0
Command Speed using Only Vision
Actual Speed using Proprioception
Invisible Obstacle Detected

Navigation in the Wild



Simulation Comparisons

	Navigation System	Terrain Type	Success	SPL	Time (s)
(a)	w/o Proprio	Flat	95.20	0.79	80.28
(b)	w/o Proprio	2x Inv-Obstacle	68.45	0.57	119.80
(c)	Ours	2x Inv-Obstacle	74.15	0.61	111.93
(d)	w/o Proprio	4x Inv-Obstacle	45.85	0.38	152.39
(e)	Ours	4x Inv-Obstacle	59.20	0.49	134.70
(f)	w/o Proprio	8x Inv-Obstacle	24.35	0.20	184.07
(g)	Ours	8x Inv-Obstacle	39.25	0.32	164.95

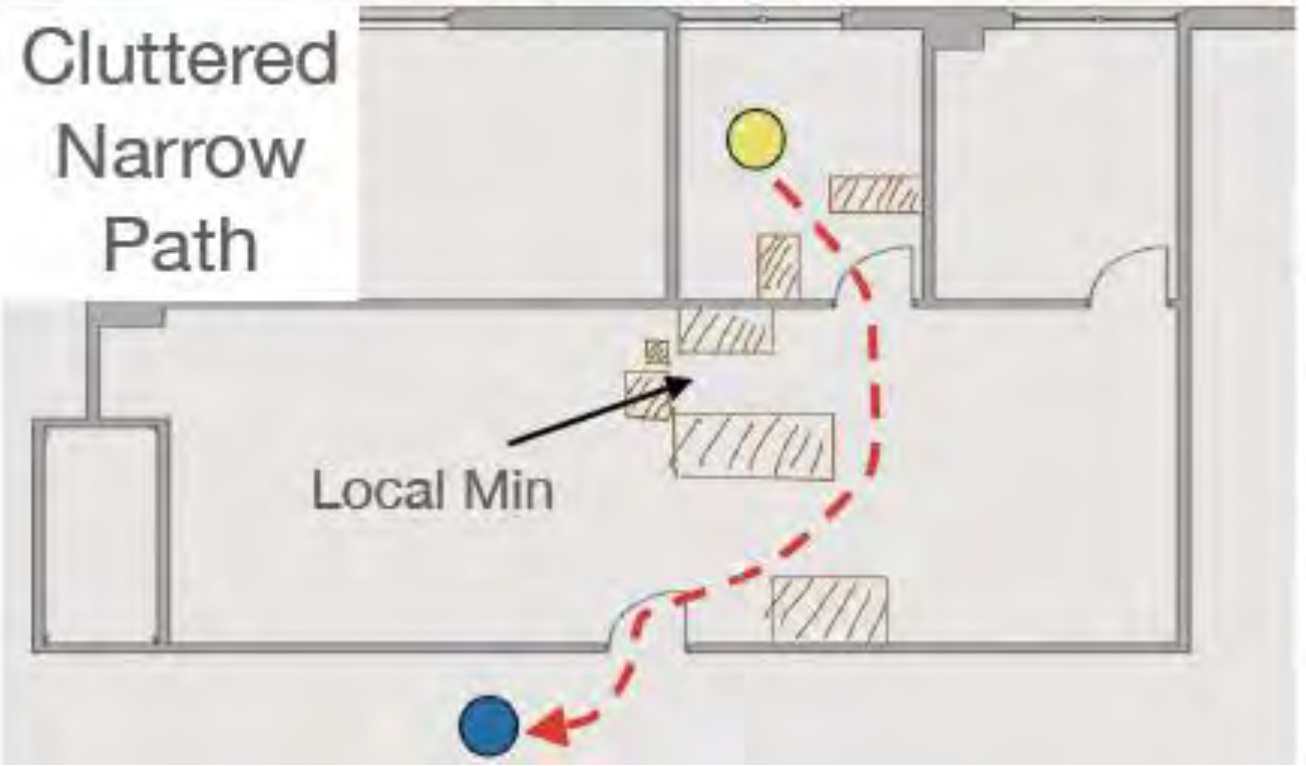
invisible obstacles

	Navigation System	Terrain Type	Success	SPL	Time (s)
(a)	w/o Proprio	Flat	95.20	0.79	80.28
(b)	w/o Proprio	Randomized	80.25	0.66	105.68
(c)	Ours	Randomized	87.40	0.73	117.65

rough, slip, payload randomization

Real world comparisons

	Success Rate	# Obstacles Hit
Vision + Proprio	100%	0.2
Vision	80%	0.2



	Success Rate	# Obstacles Hit
Vision + Proprio	80%	0.6
Vision	60%	0.6

	Success Rate	Time to Recover After Blocked (s)
Vision + Proprio	100%	0.9
Vision	0%	∞

	Success Rate	# Slips
Vision + Proprio	100%	3.2
Vision	80%	4.0

Legged Locomotion in Challenging Terrains using Egocentric Vision

Ananye Agarwal*

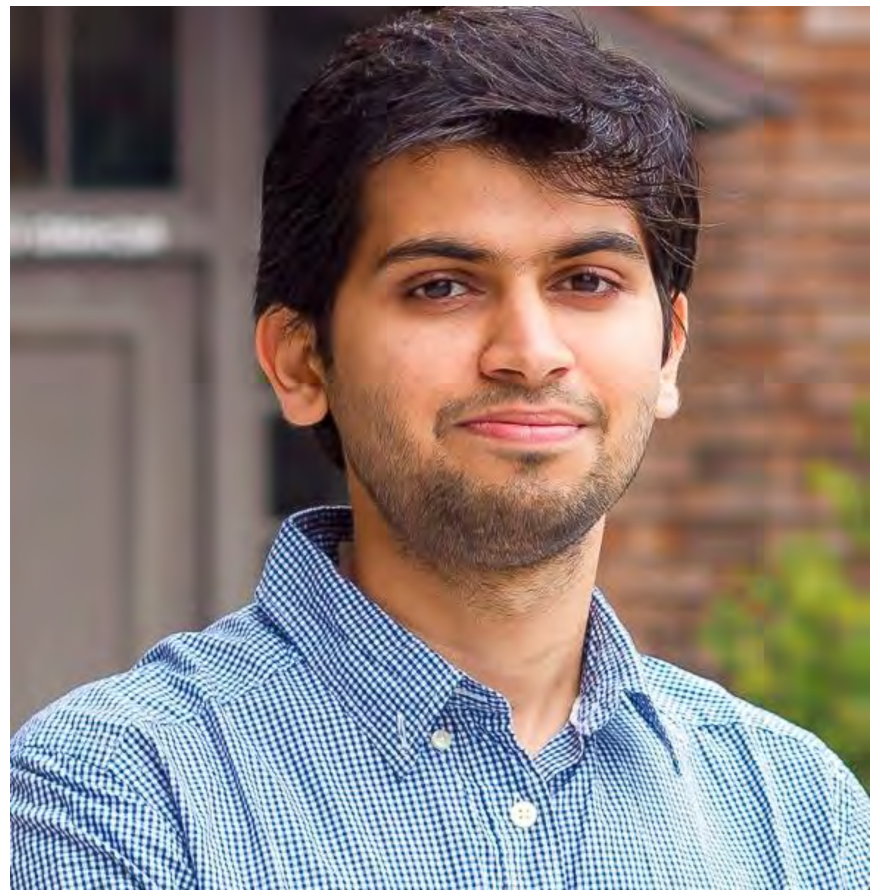
CMU

Ashish Kumar*

UC Berkeley

Jitendra Malik†

UC Berkeley

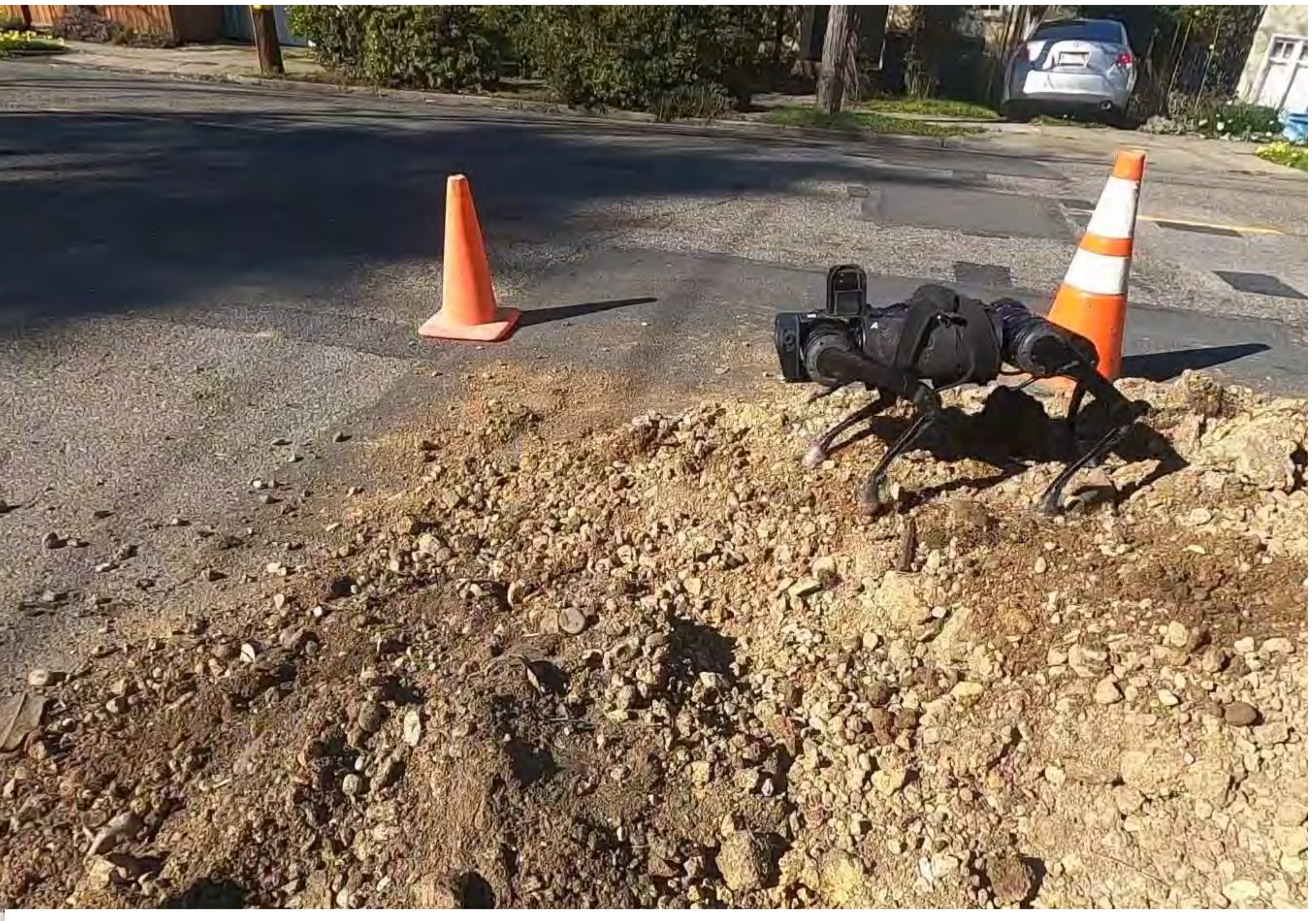


Deepak Pathak†

CMU

Robots can walk on challenging terrain

These robots are blind!



Kumar, Ashish, et al. "Rma: Rapid motor adaptation for legged robots." RSS (2021).

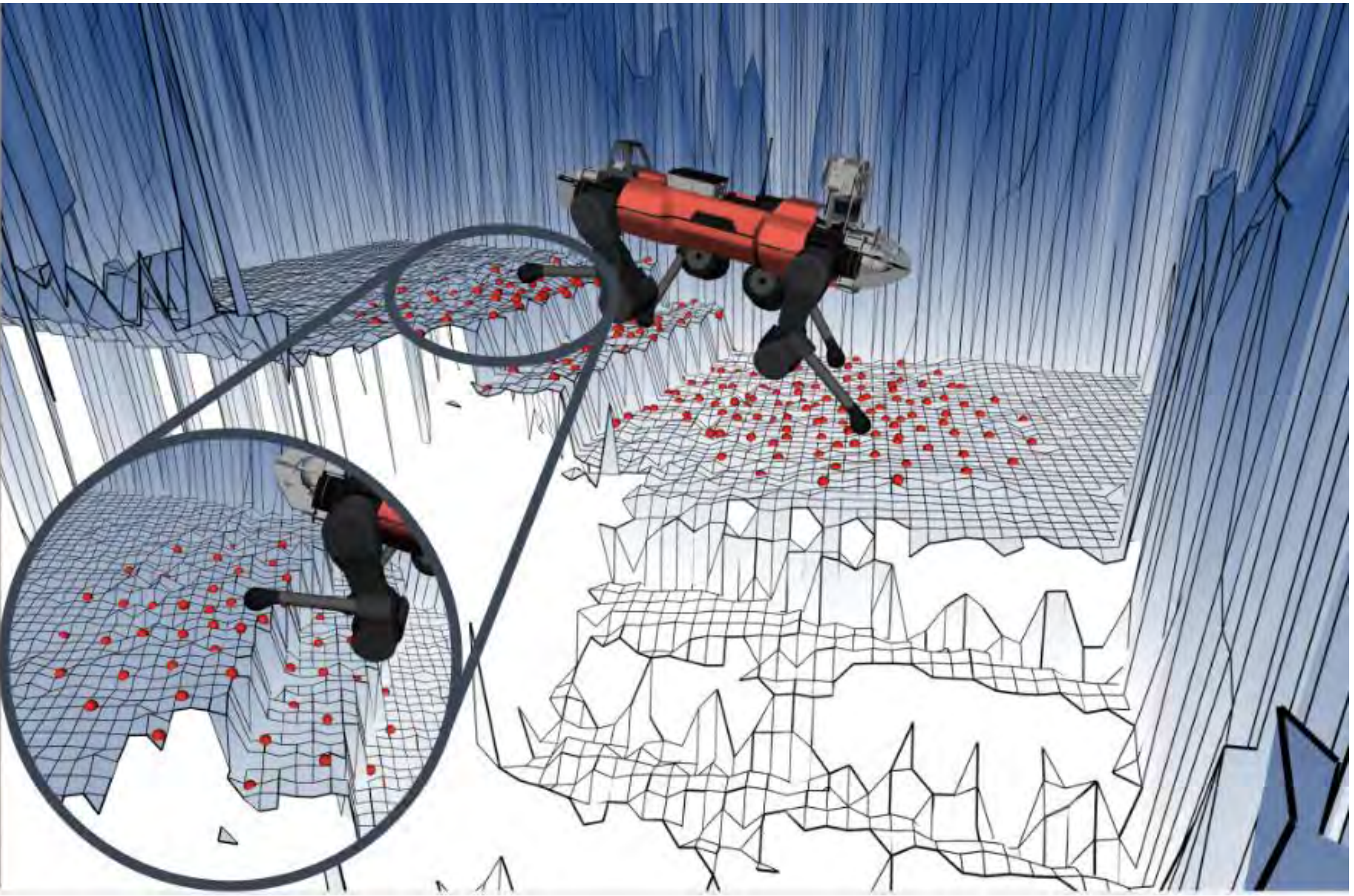
Siekmann, Jonah, et al. "Blind bipedal stair traversal via sim-to-real reinforcement learning." RSS (2021)

Lee, Joonho, et al. "Learning quadrupedal locomotion over challenging terrain." Science robotics 5.47 (2020)

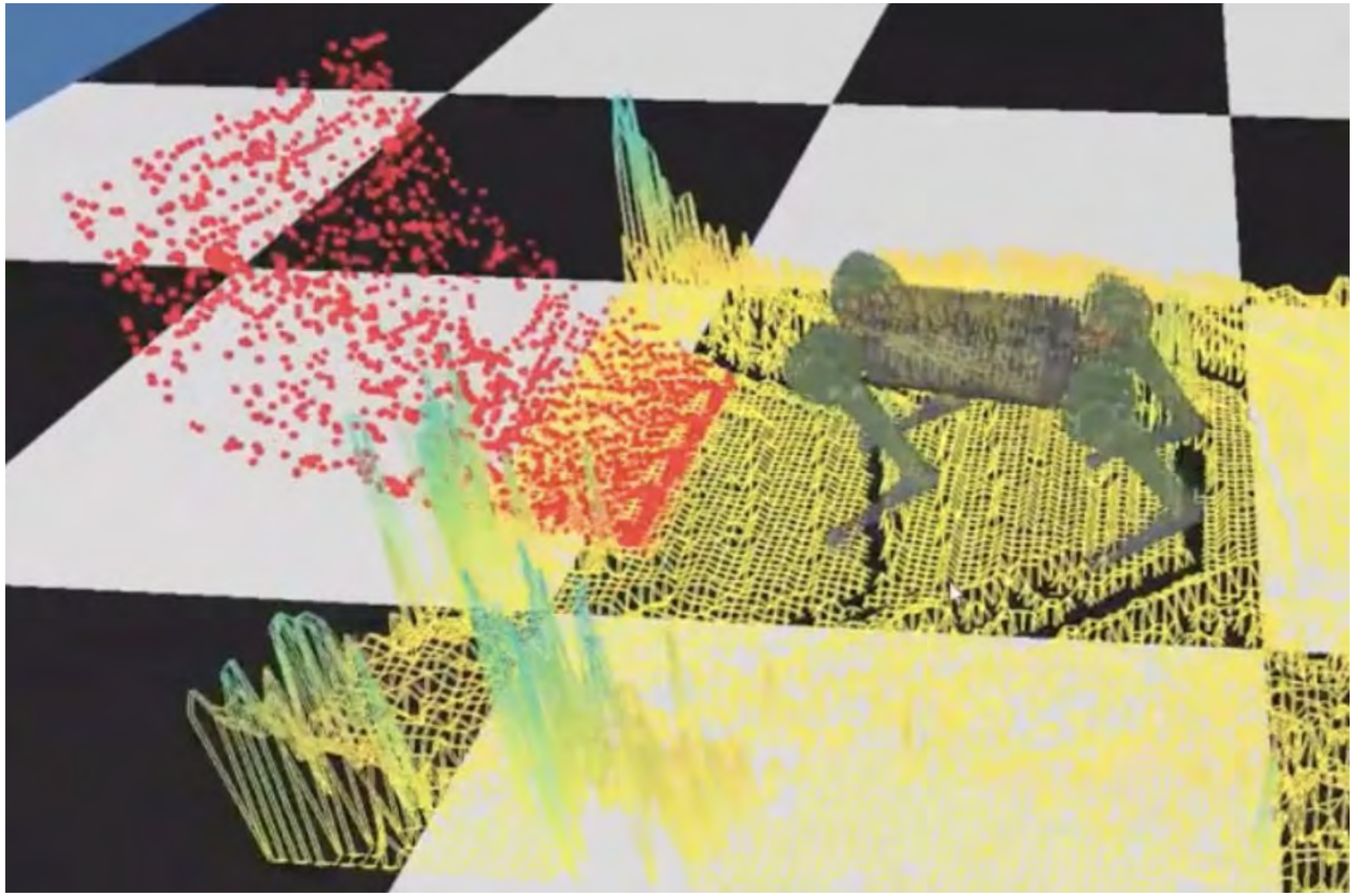
Why do we need vision?

Perception enables *precise* locomotion

Typical approach: build terrain maps from vision



Miki, T., Lee, J., Hwangbo, J., Wellhausen, L., Koltun, V., & Hutter, M. (2022). Learning robust perceptive locomotion for quadrupedal robots in the wild. Science Robotics



Kim, Donghyun, et al. "Vision aided dynamic exploration of unstructured terrain with a small-scale quadruped robot." ICRA 2020.

Real world maps accumulate noise

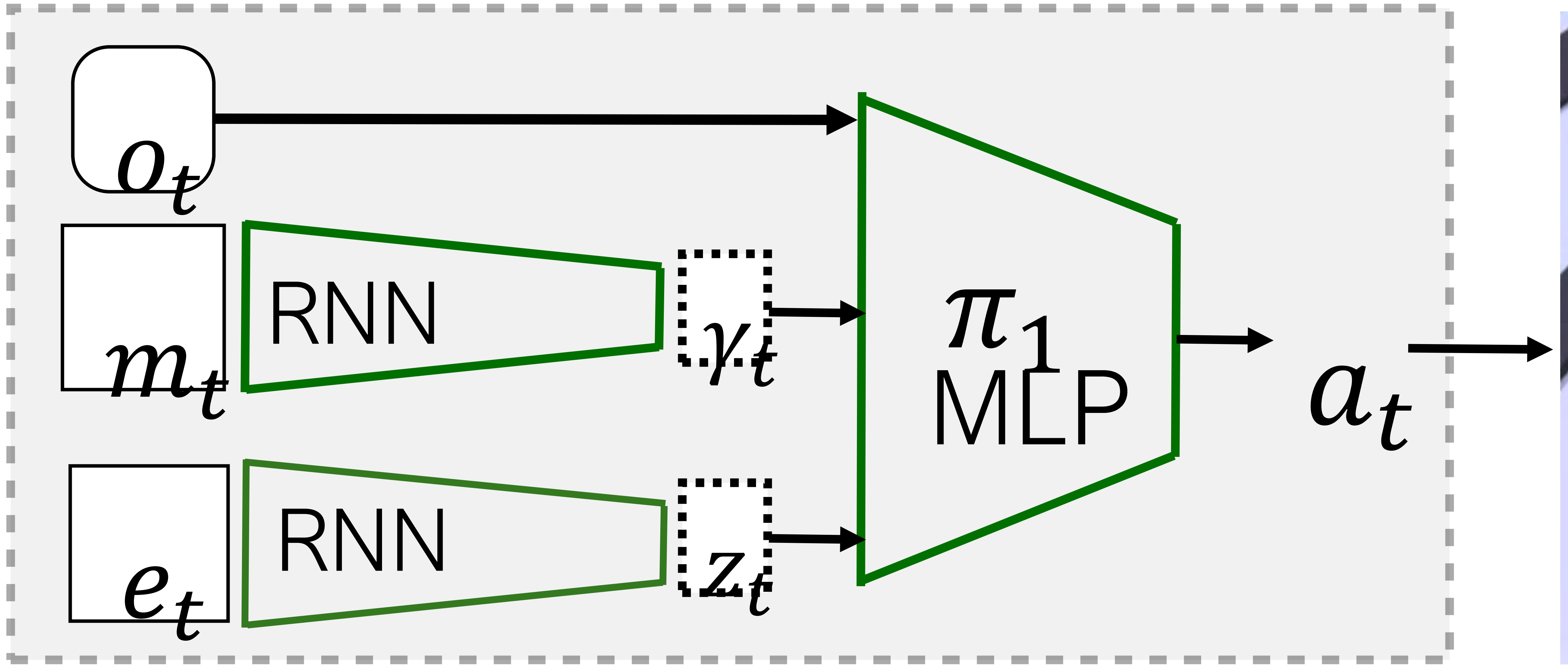
Miki, Takahiro, et al. "Learning robust perceptive locomotion for quadrupedal robots in the wild." *Science Robotics* (2022)

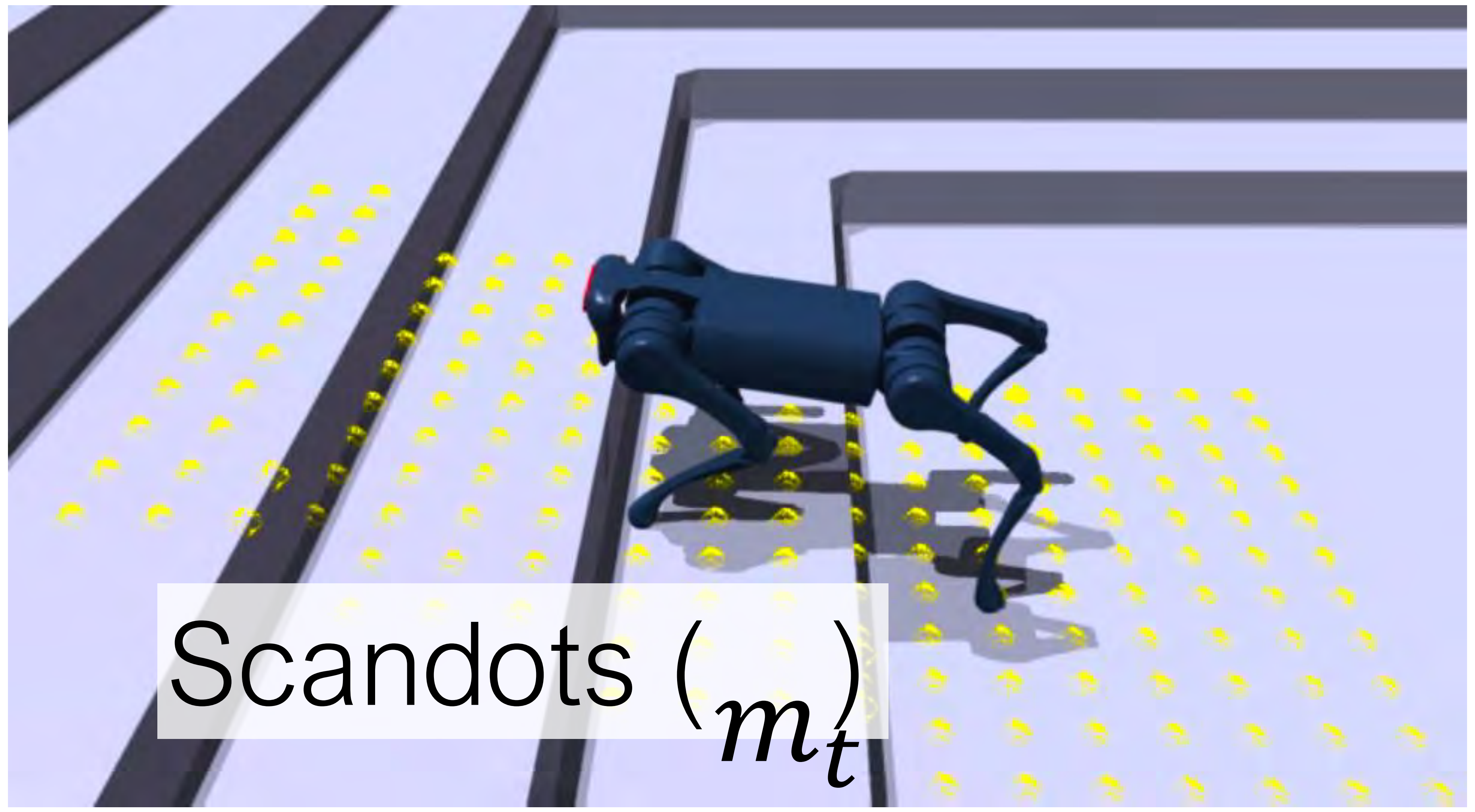
Do we really need terrain maps?

We directly go from vision to control


Stepping Stones (~15 cm apart)

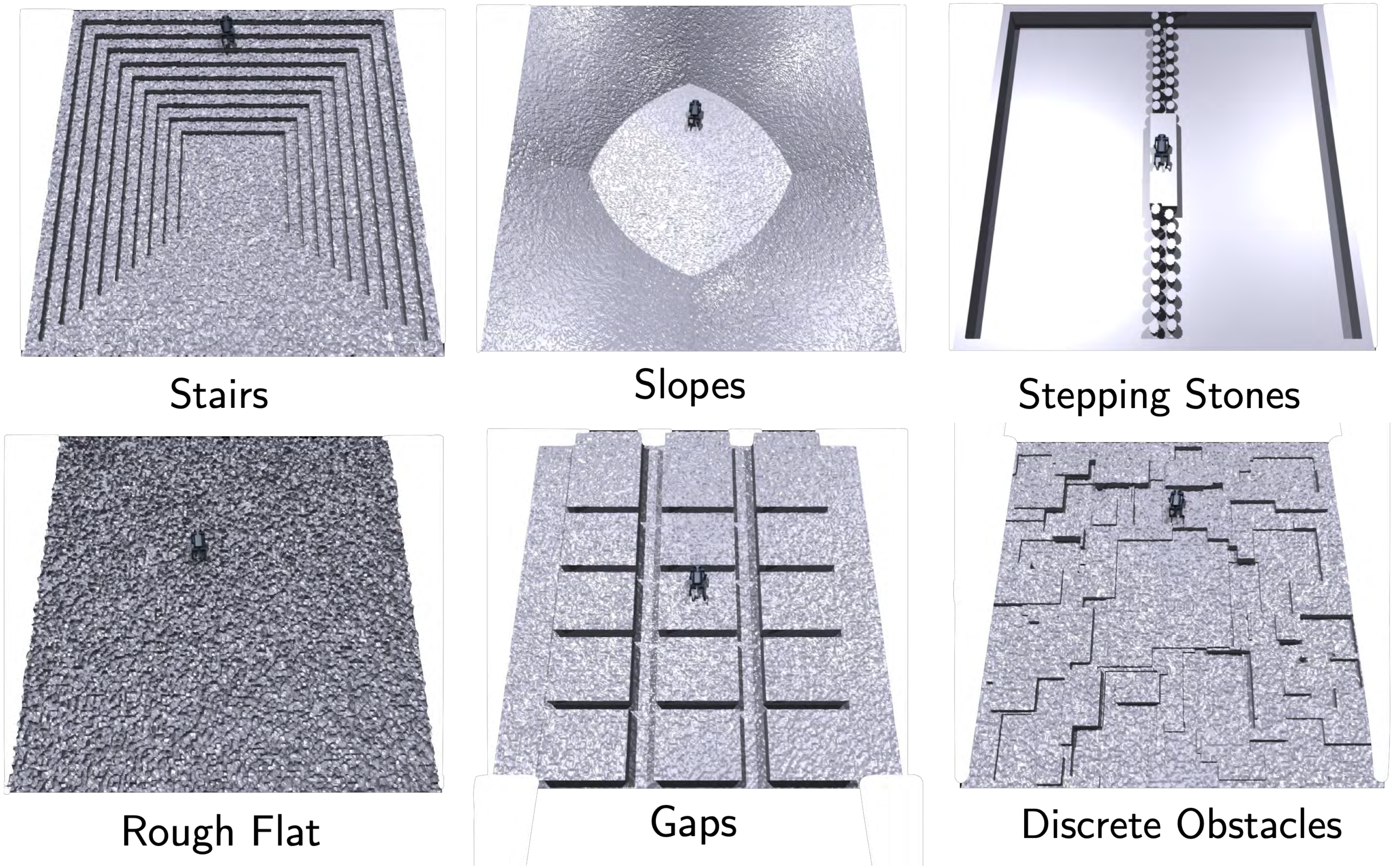
Phase 1: Learning to Walk with Privileged Terrain Information





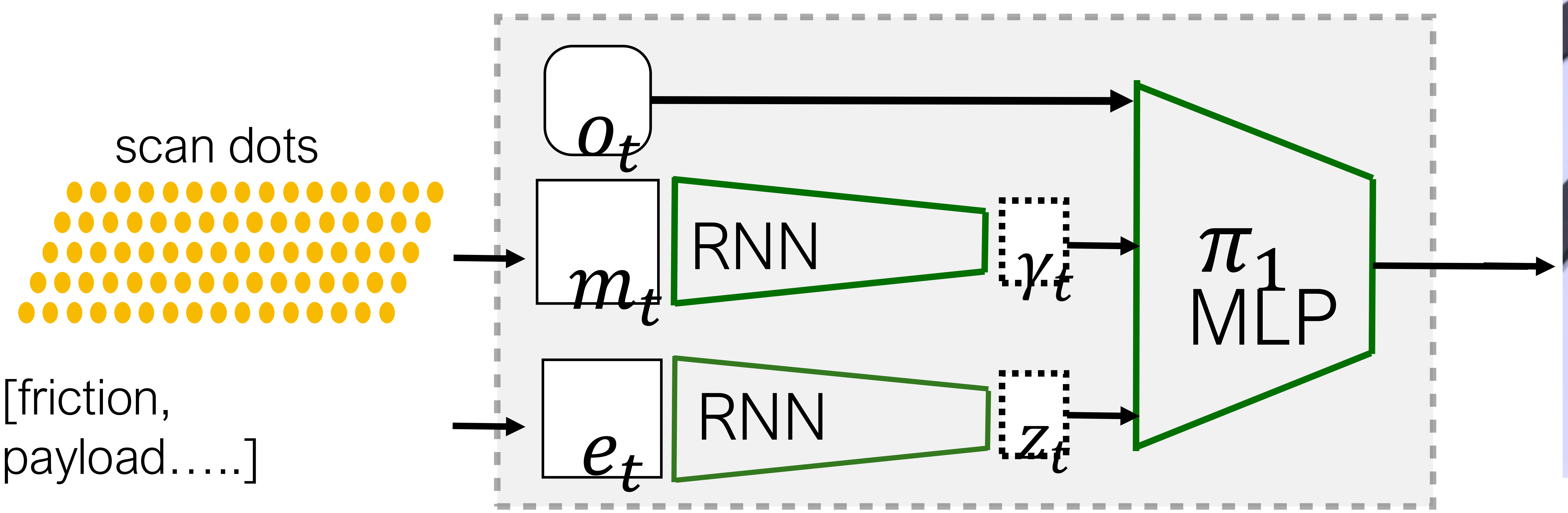
- Simulation: IsaacGym
- Trained with PPO with smoothness+task rewards
- ~10 Billion Samples simulated in 3 days on a single GPU

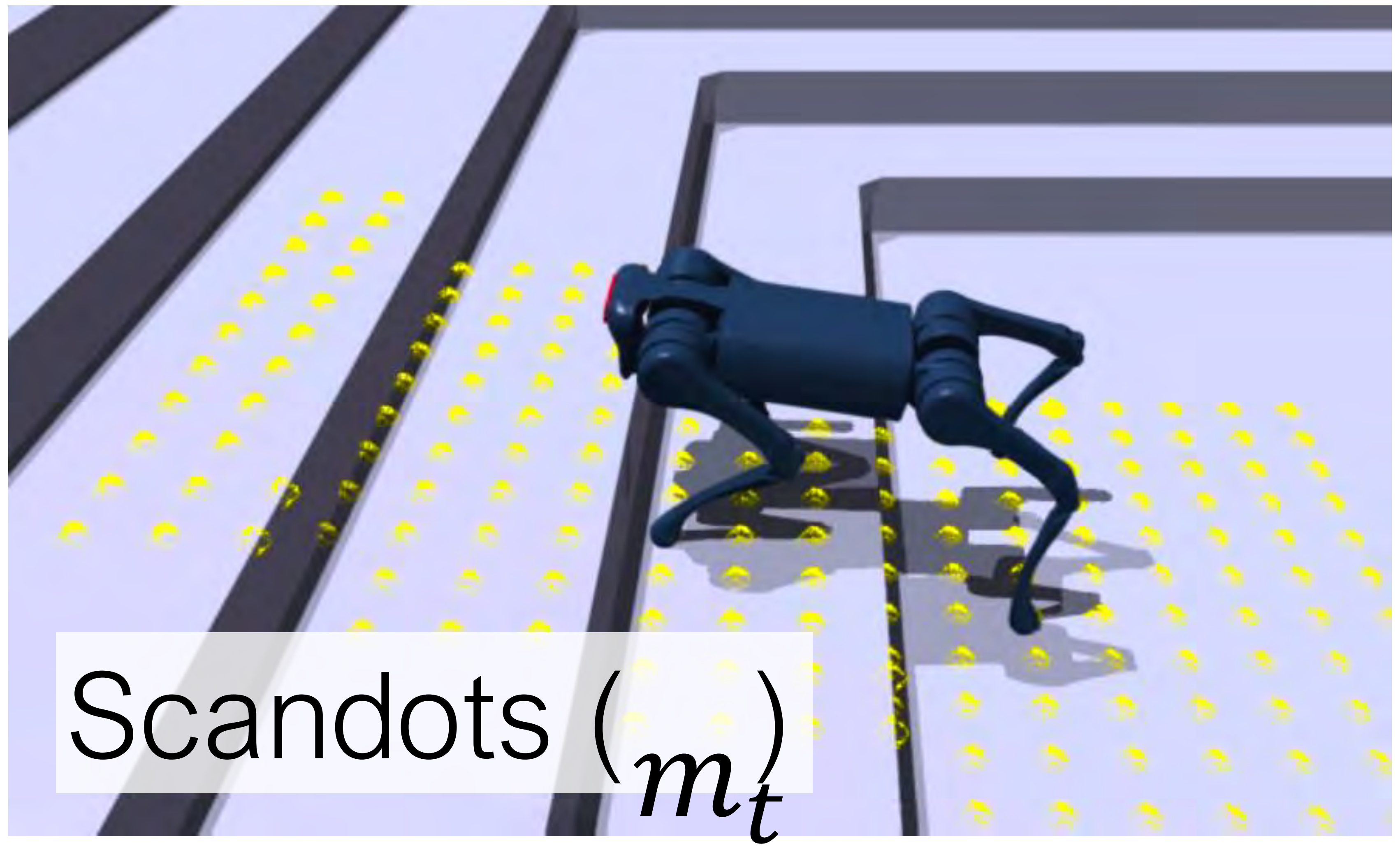
Training Terrains



Phase 1 Policy

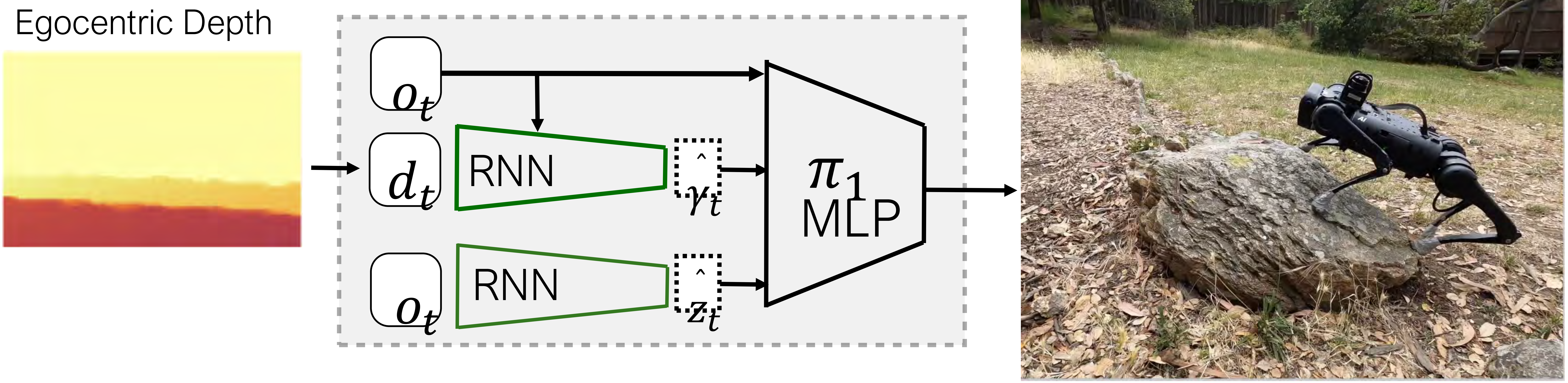
How do we deploy it?





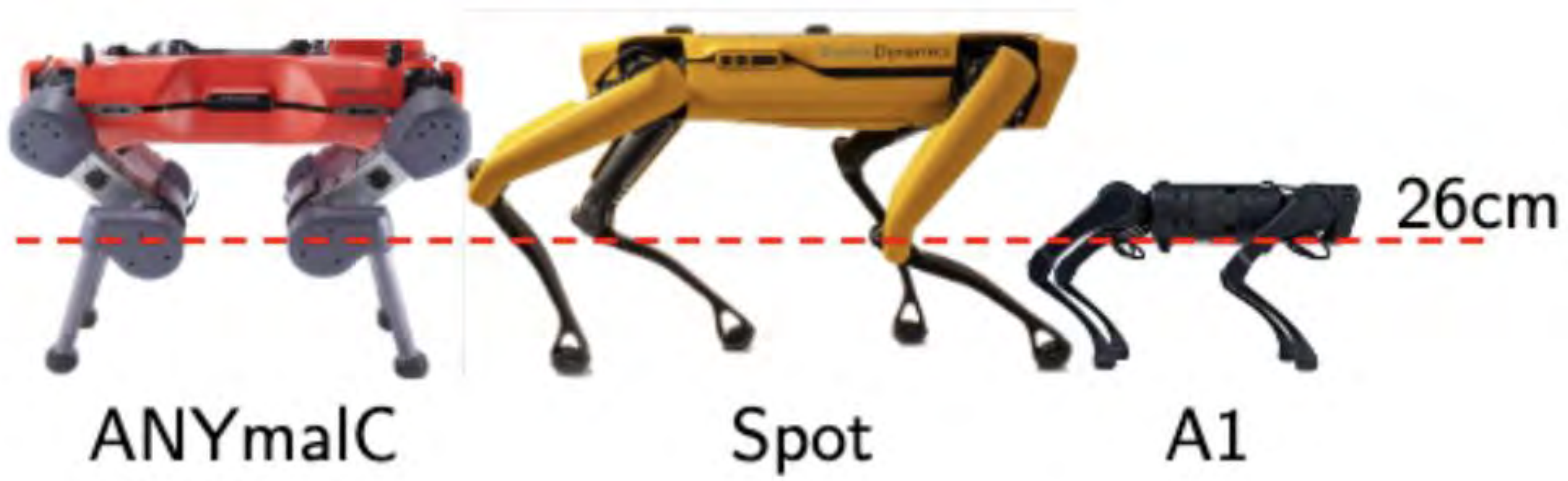
Phase 2: Learning to Walk with Egocentric Depth

Deployment Policy

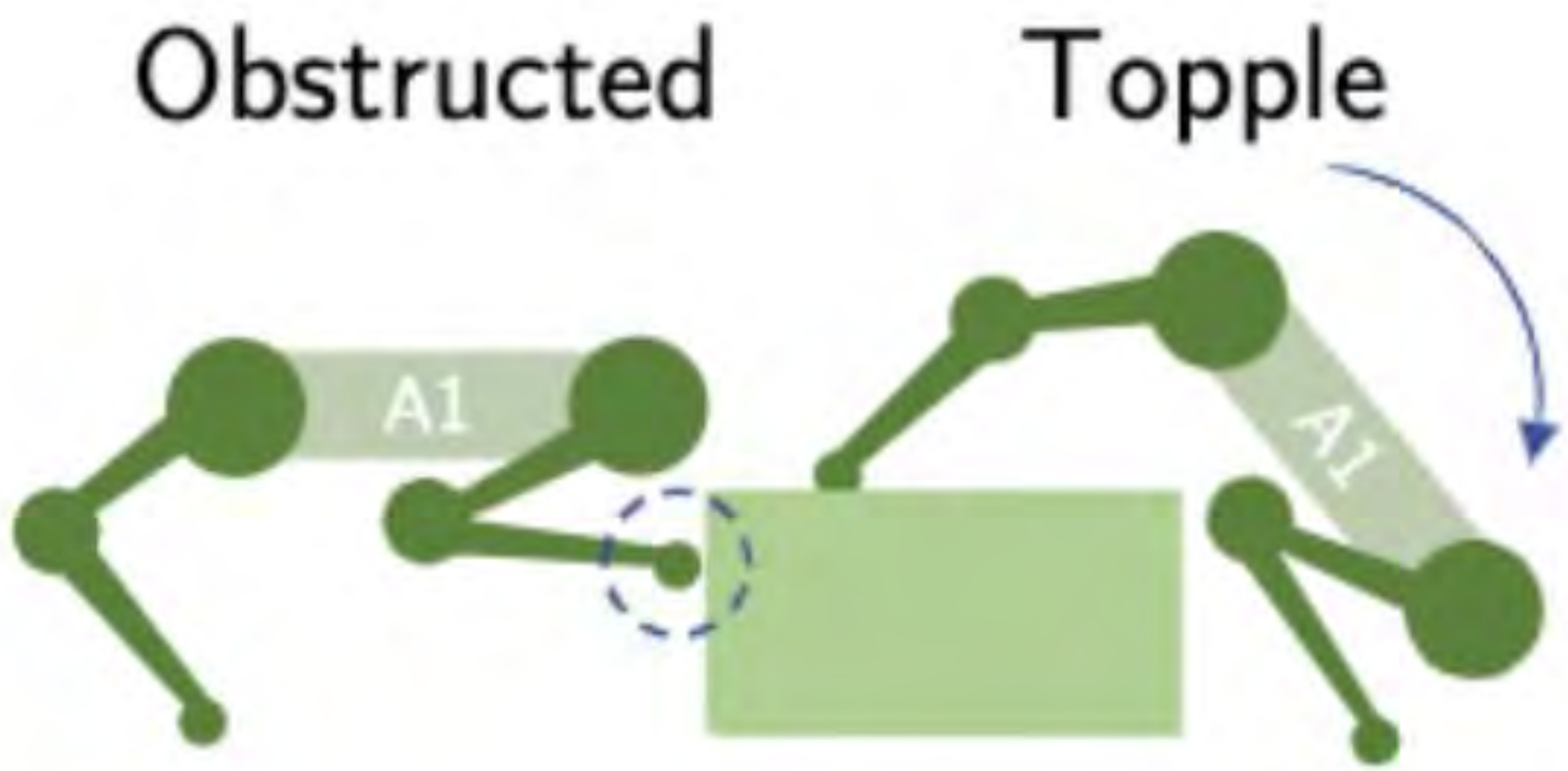



Stepping Stones (~20 cm apart)
Emergent footstep planning

Gait heuristics fail on a small robot



Size Comparison



Challenges Due to Size

Emergent Hip Abduction

Map Free, Gait Free

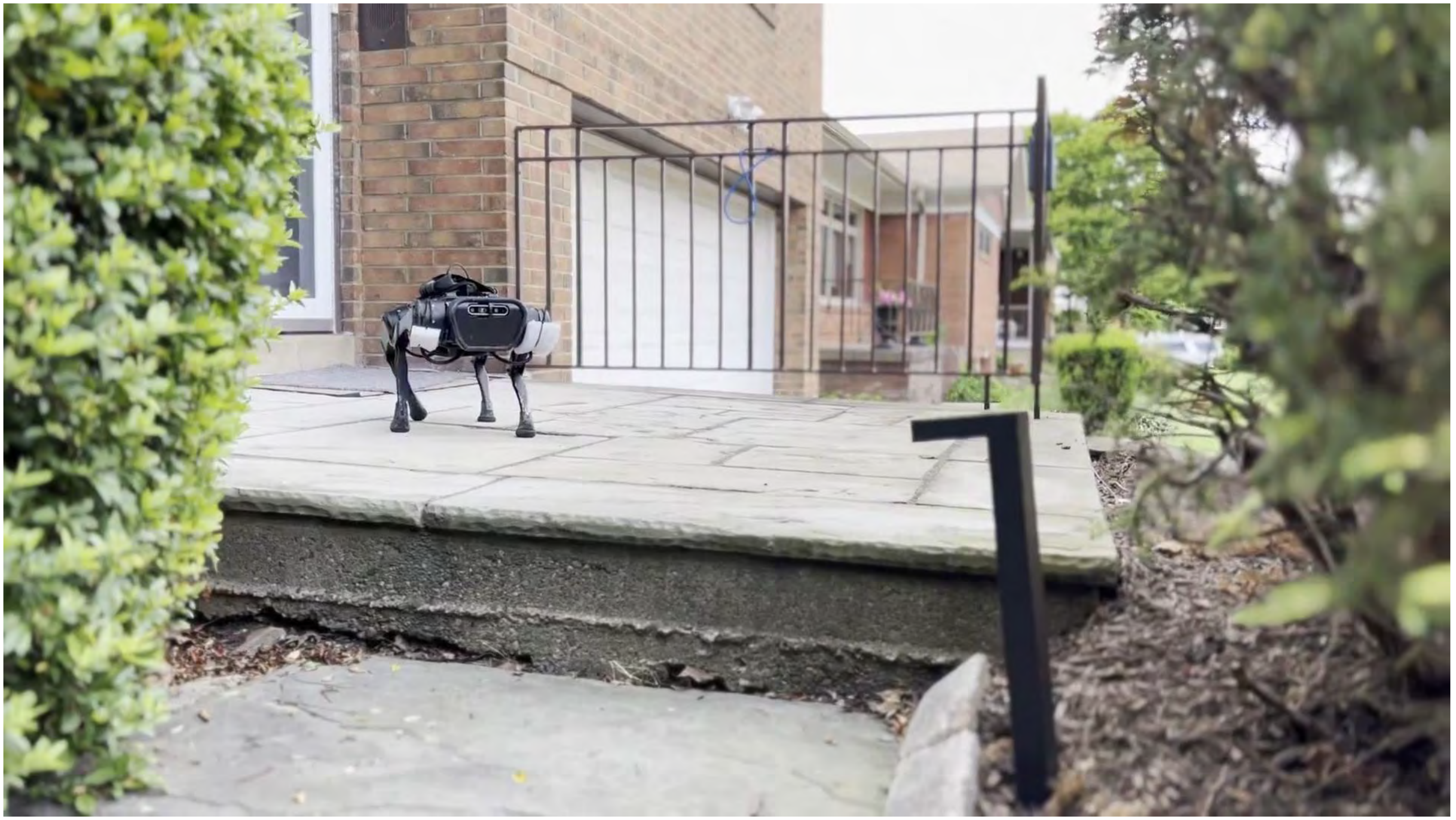
Performance is better without maps

Average Distance

	Blind	Noisy	Ours
Slopes	34.72	36.14	43.98
Stepping Stones	1.02	1.09	18.83
Stairs	16.64	6.74	31.24
Discrete Obstacles	32.41	29.08	40.13

Limitations

Front Camera Failure (unable to see a dip)



Implicit planning failure


Thank You