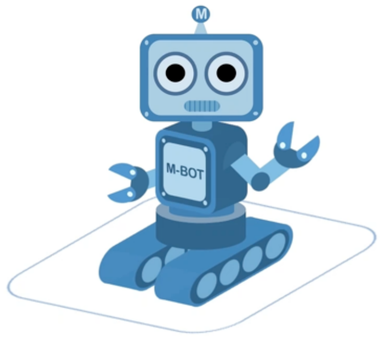
Habitat



Habitat: A Platform for Embodied AI Research

Dhruv Batra

Georgia Tech

facebook Artificial Intelligence Research


BRACE YOURSELF
CHANGE IS COMING!

Long-term Goal

Physical agent capable of taking actions in the world and talking to humans in natural language

Is there smoke in any room around you?

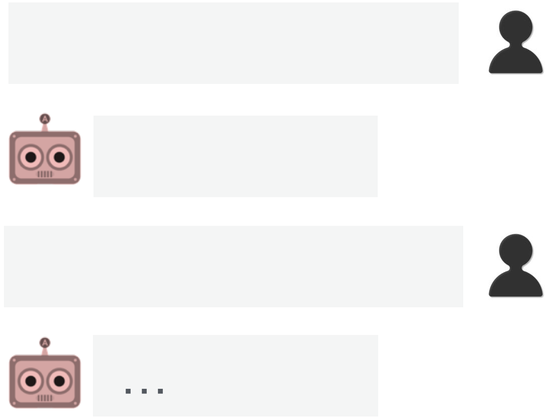
Yes, in one room

Go there and look for people

…

Slide credit: Abhishek Das

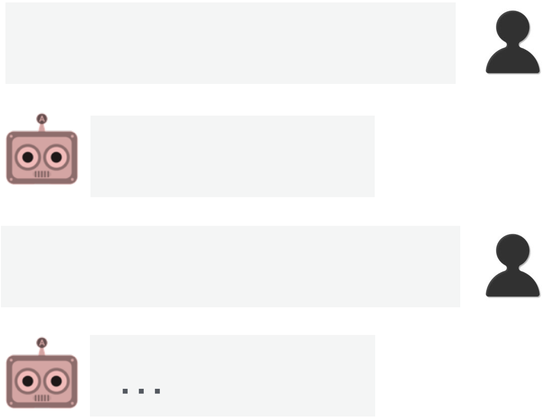
Image Credit: Lockheed Martin; DARPA Robotics Challenge



Slide credit: Abhishek Das

Image Credit: Lockheed Martin; DARPA Robotics Challenge

Internet AI → Embodied AI

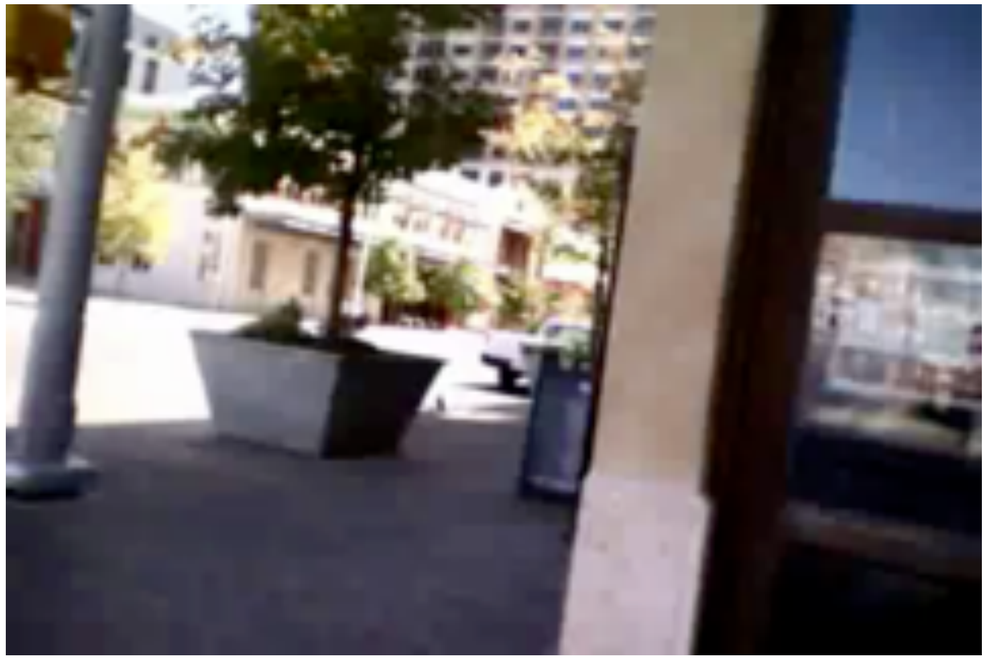


Slide credit: Abhishek Das

Image Credit: Image-Net
Image Credit: Lockheed Martin; DARPA Robotics Challenge

Internet AI → Embodied AI

Egocentric vision



No access to well-composed, curated images

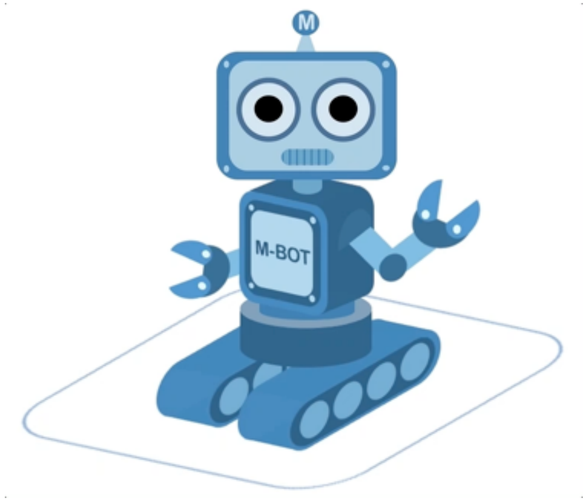
Slide credit: Abhishek Das

Image Credit: Image-Net, Video Credit: Lee et al., 2012

Internet AI → Embodied AI

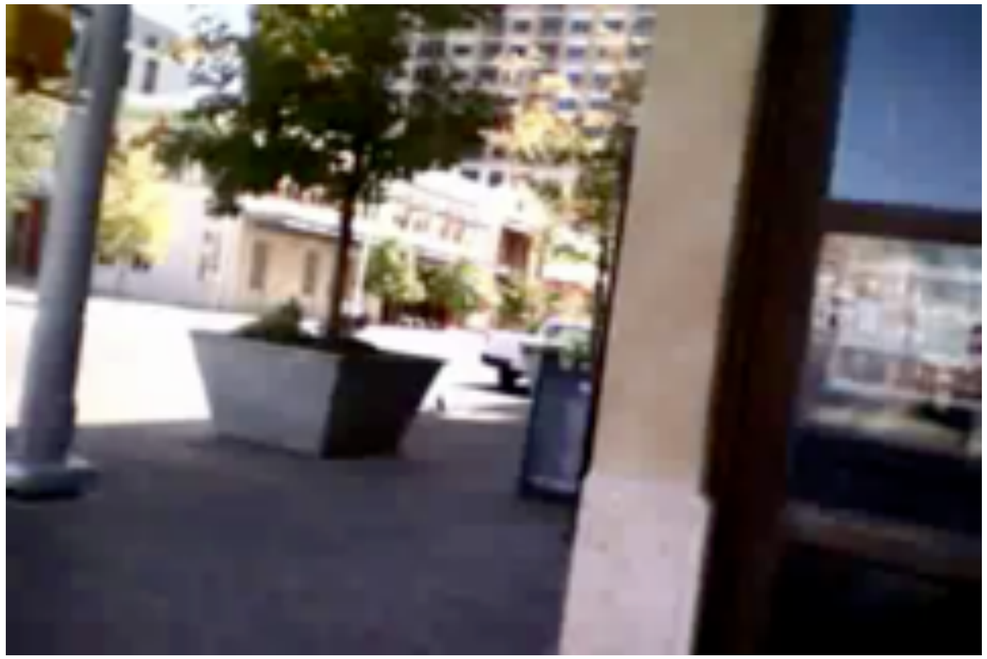
Egocentric vision

Active perception



Action

Observation



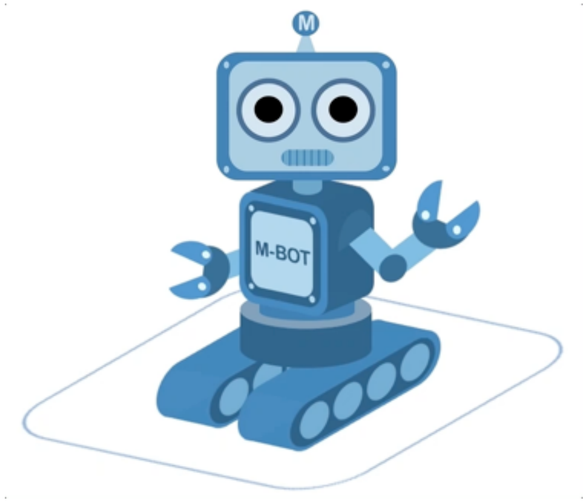
Agent controls incoming data distribution

Slide credit: Abhishek Das

Video Credit: Lee et al., 2012

Internet AI → Embodied AI

Egocentric vision
Active perception
Sparse rewards

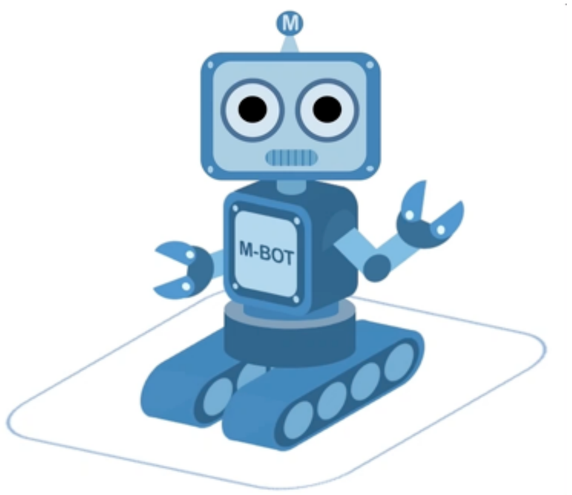


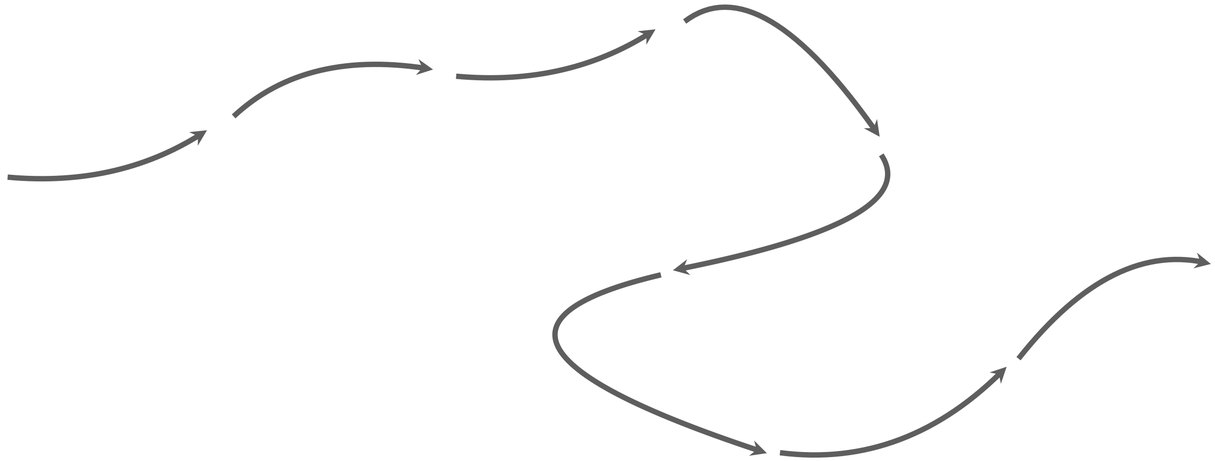
Slide credit: Abhishek Das

Image Credit: Image-Net

Internet AI → Embodied AI

Egocentric vision
Active perception
Sparse rewards



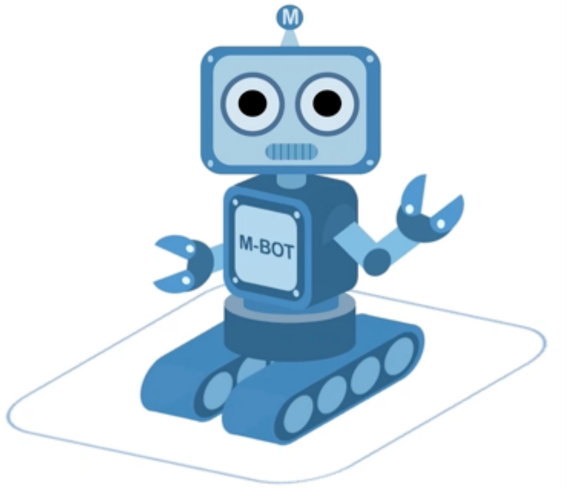


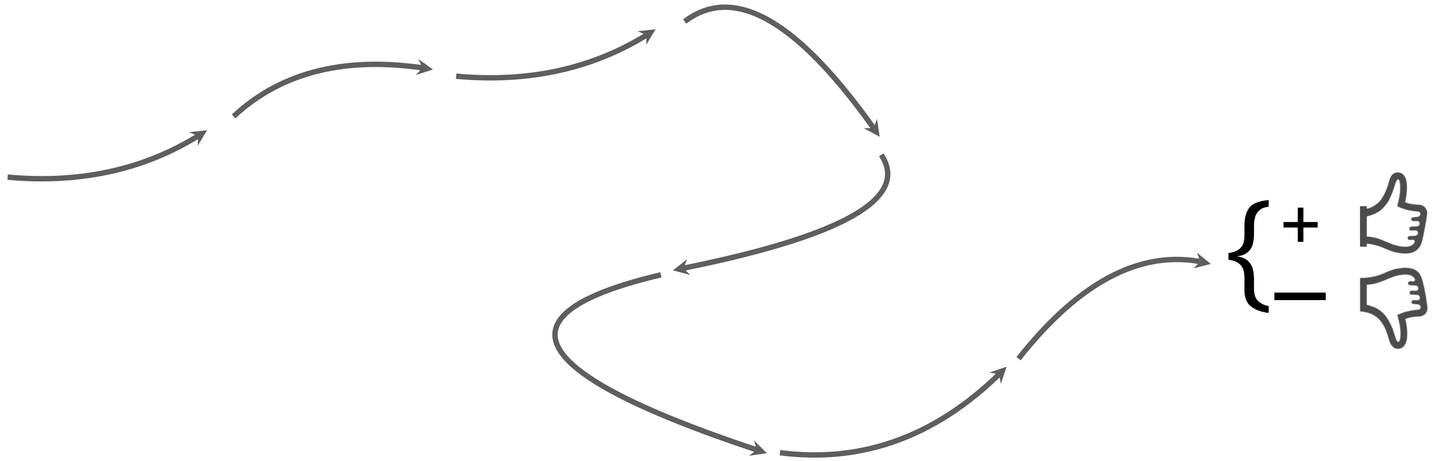
Slide credit: Abhishek Das

Image Credit: Image-Net

Internet AI → Embodied AI

Egocentric vision
Active perception
Sparse rewards





Slide credit: Abhishek Das

Image Credit: Image-Net

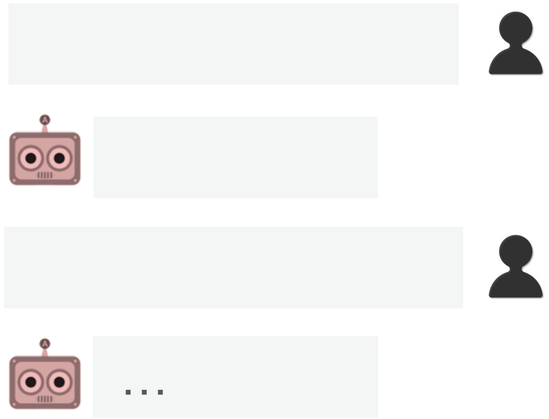
Internet AI → Embodied AI

Egocentric vision
Active perception
Sparse rewards
Language understanding

Slide credit: Abhishek Das

Problems with reality

- Slow
- Dangerous
- Expensive
- Difficult to control
- Not easy reproducible



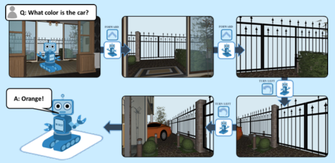
Our Approach: Sim2Real

- Slow
- Dangerous
- Expensive
- Difficult to control
- Not easy reproducible

Resurrection of Embodied AI

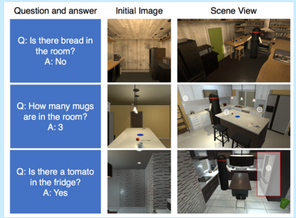
>= 2017 (!)

Tasks



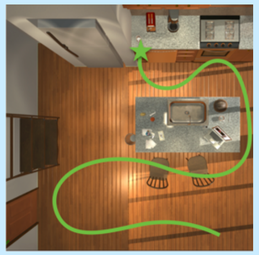
EmbodiedQA

Language grounding
(Chaplot et al., 2017,
Hermann & Hill et al., 2017)



Interactive QA
(Gordon et al., 2018)

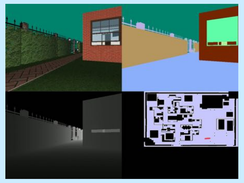
Vision-Language Navigation
(Anderson et al., 2018)



Visual Navigation
(Zhu & Gordon et al., 2017,
Savva et al., 2017,
Wu et al., 2017)

Simulators

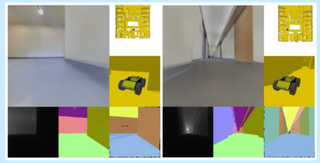
HoME (Brodeur et al., 2018)



House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)



Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

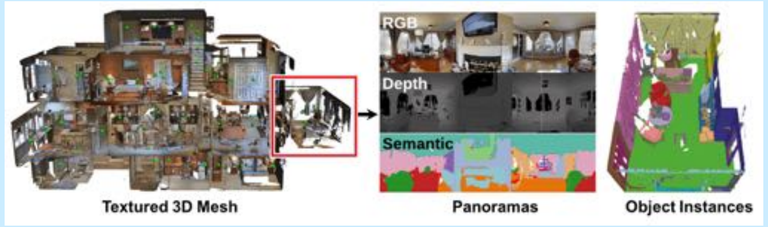
VirtualHome
(Puig et al., 2018)

AdobeIndoorNav
(Mo et al., 2018)

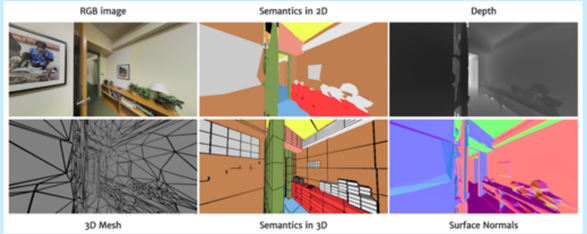
Matterport3DSim
(Anderson et al., 2018)

Datasets

SUNCG (Song et al., 2017)



Matterport3D (Chang et al., 2017)



Stanford 2D-3D-S (Armeni et al., 2017)

Slide credit: Abhishek Das

Embodied Question Answering

[CVPR ’18 Oral]

Abhishek Das
(Georgia Tech)

Samyak Datta
(Georgia Tech)

Georgia Gkioxari
(FAIR)

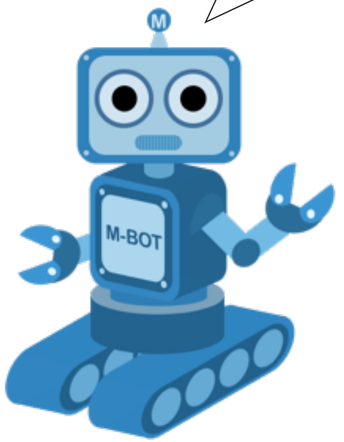
Stefan Lee
(Georgia Tech)

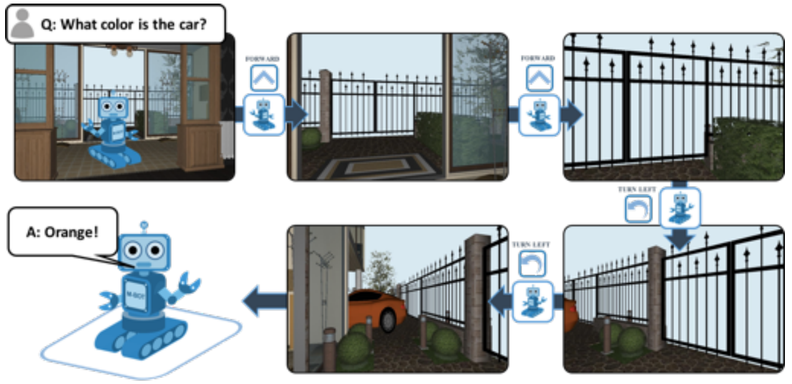
Devi Parikh
(FAIR/Georgia Tech)

Dhruv Batra
(FAIR/Georgia Tech)

What is to the left of the shower?

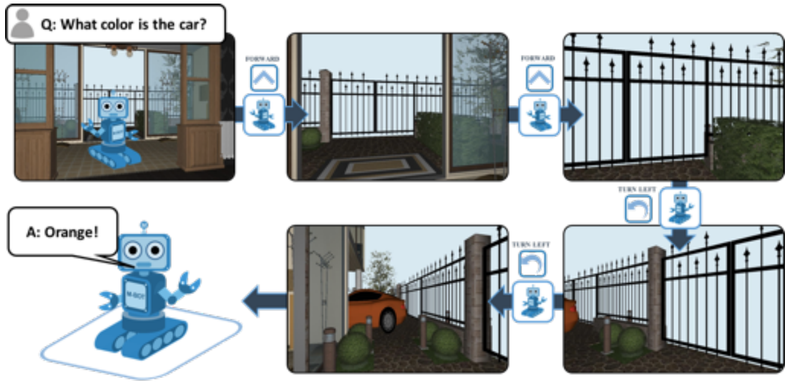
Cabinet



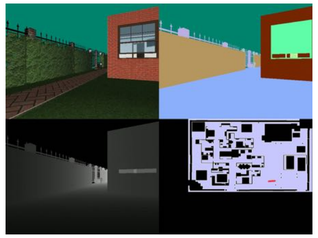


EmbodiedQA

Slide credit: Abhishek Das

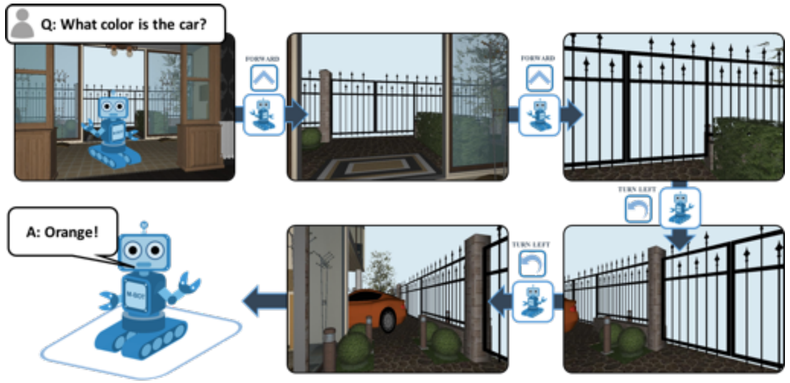


EmbodiedQA

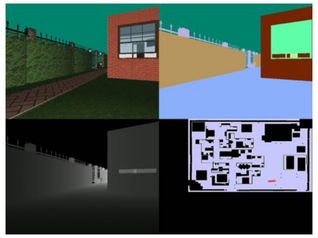


House3D (Wu et al., 2017)

Slide credit: Abhishek Das



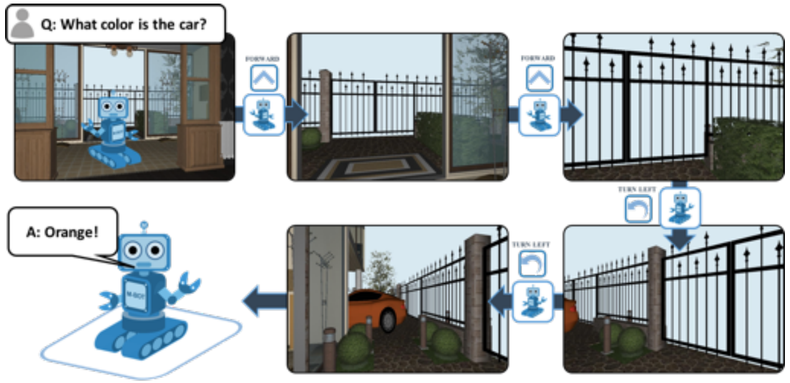
EmbodiedQA



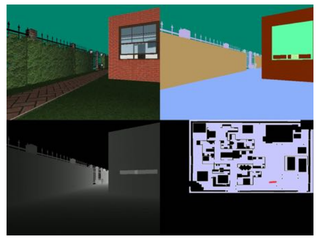
House3D (Wu et al., 2017)

SUNCG (Song et al., 2017)

Slide credit: Abhishek Das



EmbodiedQA

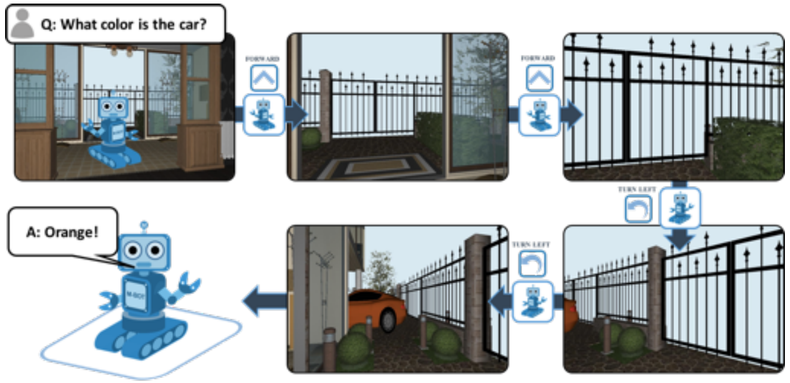


House3D (Wu et al., 2017)

Dataset

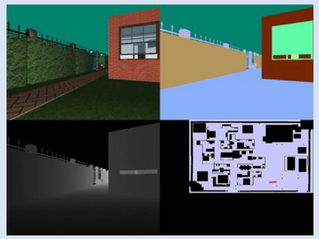
SUNCG (Song et al., 2017)

Slide credit: Abhishek Das



EmbodiedQA

Simulator



House3D (Wu et al., 2017)

Dataset

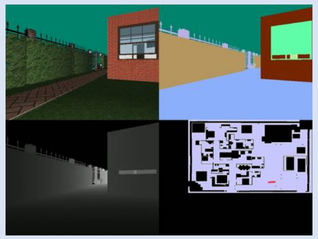
SUNCG (Song et al., 2017)

Slide credit: Abhishek Das

Task

EmbodiedQA

Simulator



House3D (Wu et al., 2017)

Dataset

SUNCG (Song et al., 2017)

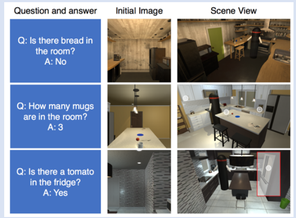
Slide credit: Abhishek Das

>= 2017 (!)

Tasks

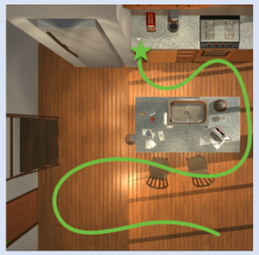
EmbodiedQA

Language grounding
(Chaplot et al., 2017,
Hermann & Hill et al., 2017)



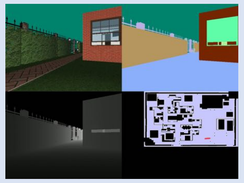
Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)



Visual Navigation
(Zhu & Gordon et al., 2017,
Savva et al., 2017,
Wu et al., 2017)

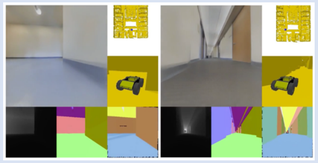
Simulators



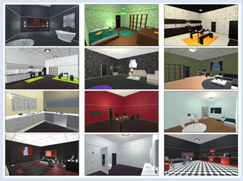
House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)



Gibson
(Zamir et al., 2018)



CHALET
(Yan et al., 2018)

HoME (Brodeur et al., 2018)

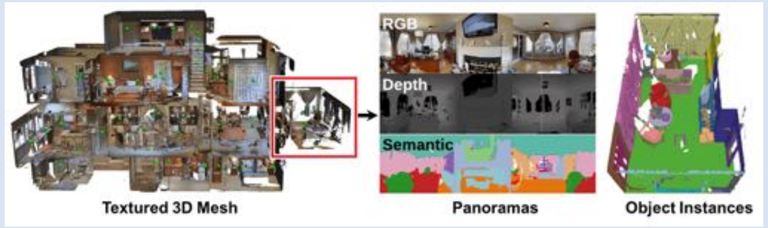
VirtualHome
(Puig et al., 2018)

AdobeIndoorNav
(Mo et al., 2018)

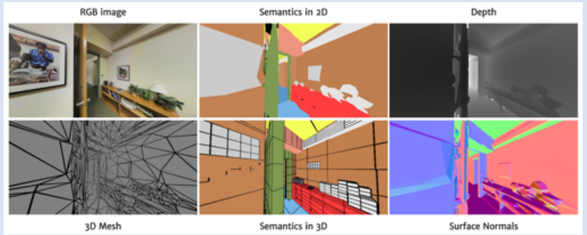
Matterport3DSim
(Anderson et al., 2018)

Datasets

SUNCG (Song et al., 2017)



Matterport3D (Chang et al., 2017)



Stanford 2D-3D-S (Armeni et al., 2017)

Slide credit: Abhishek Das

Tasks

EmbodiedQA

Language grounding
(Chaplot et al., 2017,
Hermann & Hill et al., 2017)

Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu & Gordon et al., 2017,
Savva et al., 2017,
Wu et al., 2017)

Simulators

House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)

Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

HoME (Brodeur et al., 2018)

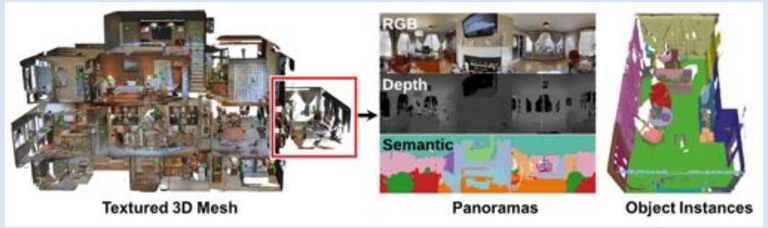
VirtualHome
(Puig et al., 2018)

AdobeIndoorNav
(Mo et al., 2018)

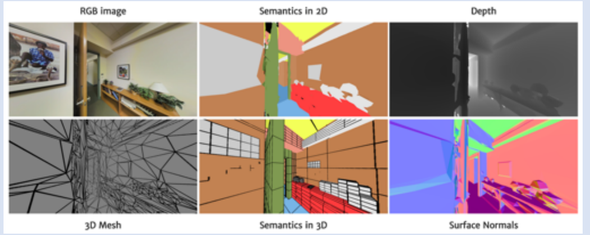
Matterport3DSim
(Anderson et al., 2018)

Datasets

SUNCG (Song et al., 2017)



Matterport3D (Chang et al., 2017)



Stanford 2D-3D-S (Armeni et al., 2017)

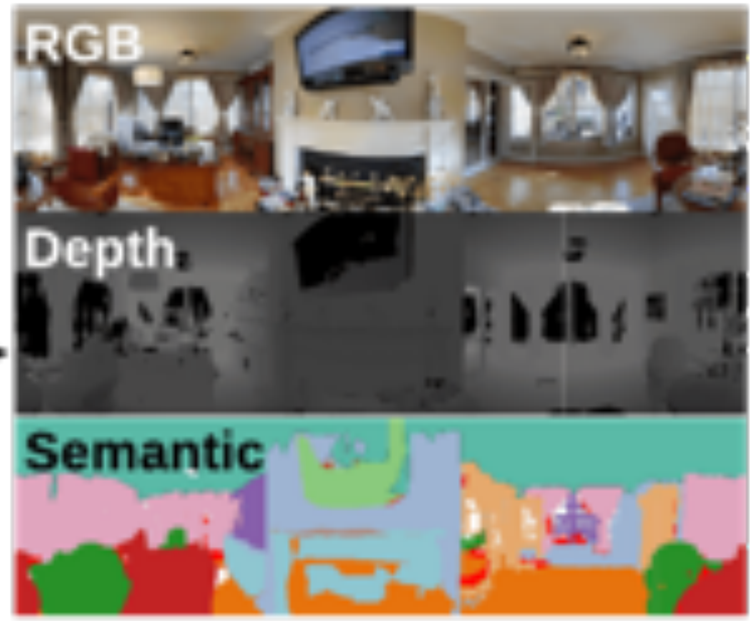
Slide credit: Abhishek Das

Datasets: Matterport3D

Datasets: Matterport3D

Datasets: Matterport3D

Textured 3D Mesh



Panoramas

Object Instances

10,800 panoramic views
194,400 RGB-D images of
90 building-scale scenes

Matterport3d dataset

[Chang 3DV 2017]

Tasks

EmbodiedQA

Language grounding
(Chaplot et al., 2017,
Hermann & Hill et al., 2017)

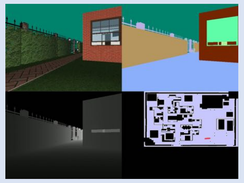
Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu & Gordon et al., 2017,
Savva et al., 2017,
Wu et al., 2017)

Simulators

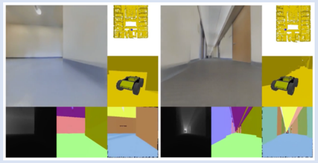
HoME (Brodeur et al., 2018)



House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)



Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

VirtualHome
(Puig et al., 2018)

AdobeIndoorNav
(Mo et al., 2018)

Matterport3DSim
(Anderson et al., 2018)

Datasets

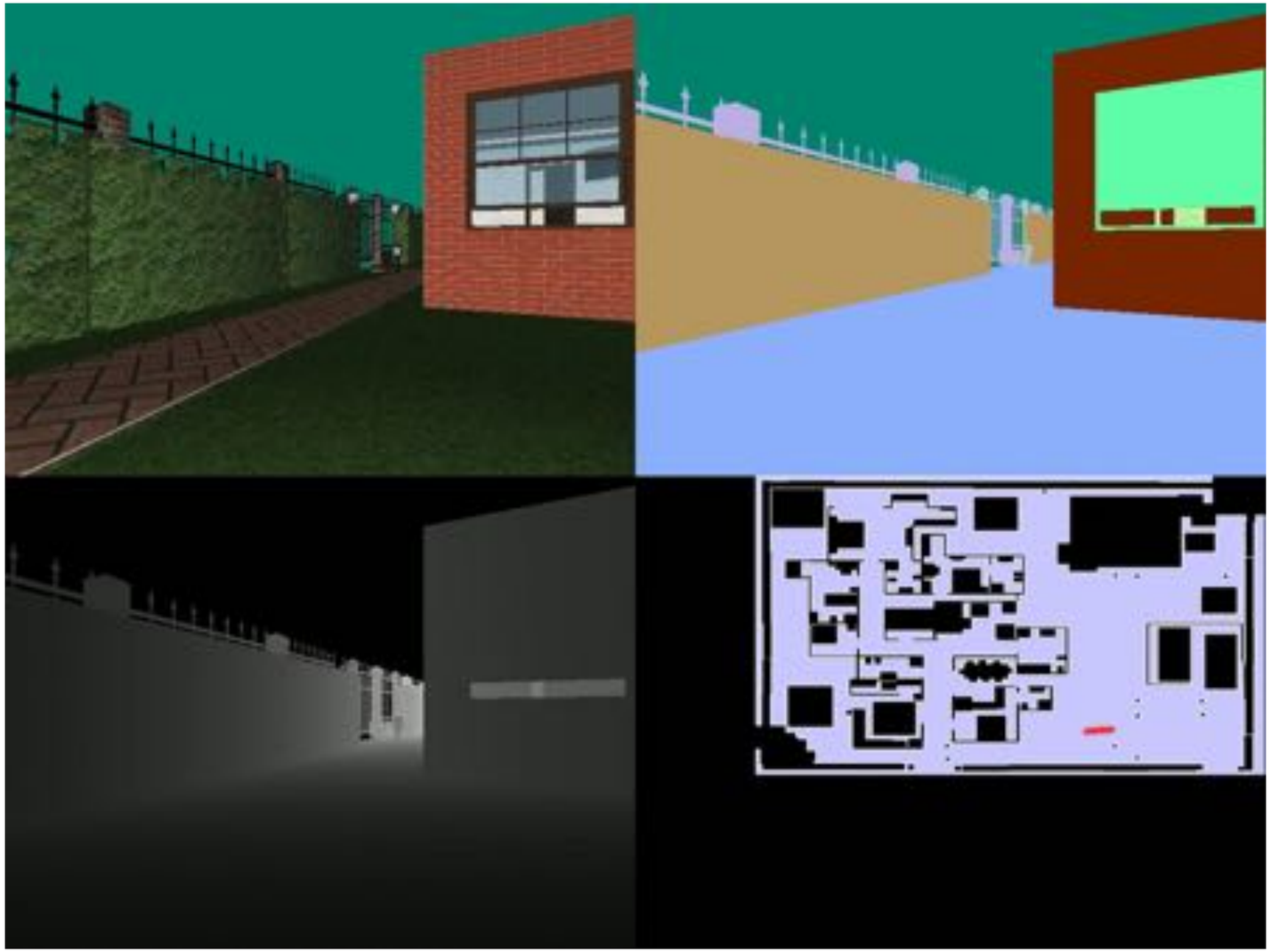
SUNCG (Song et al., 2017)

Matterport3D (Chang et al., 2017)

Stanford 2D-3D-S (Armeni et al., 2017)

Slide credit: Abhishek Das

Example: House3D



[Wu et al. 2017]

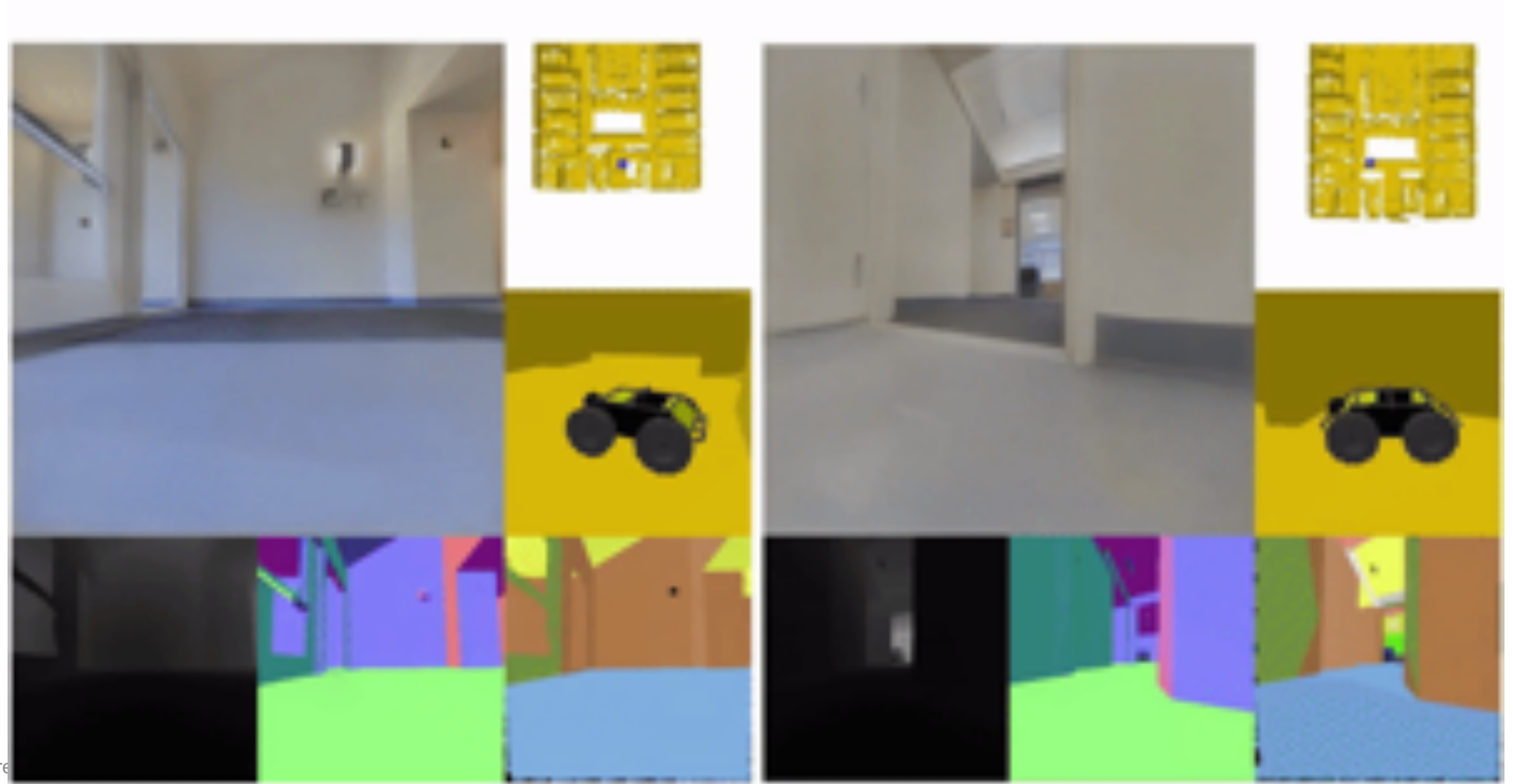
Slide credit: Manolis Savva

Example: MINOS

[Savva et al. 2017]

Slide credit: Manolis Savva

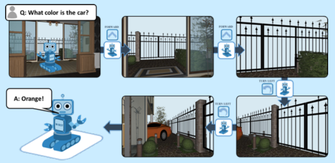
Gibson [Xia et al. 2018]



AI2 THOR [Kolve et al. 2017]

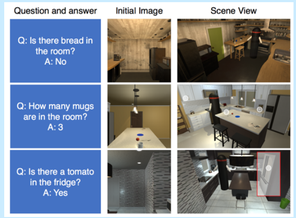
DeepMind Lab [Beattie et al. 2016]

Tasks



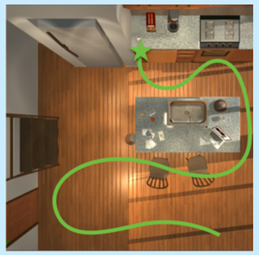
EmbodiedQA

Language grounding
(Chaplot et al., 2017,
Hermann & Hill et al., 2017)



Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)



Visual Navigation
(Zhu & Gordon et al., 2017,
Savva et al., 2017,
Wu et al., 2017)

Simulators

House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)

Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

HoME (Brodeur et al., 2018)

VirtualHome
(Puig et al., 2018)

AdobeIndoorNav
(Mo et al., 2018)

Matterport3DSim
(Anderson et al., 2018)

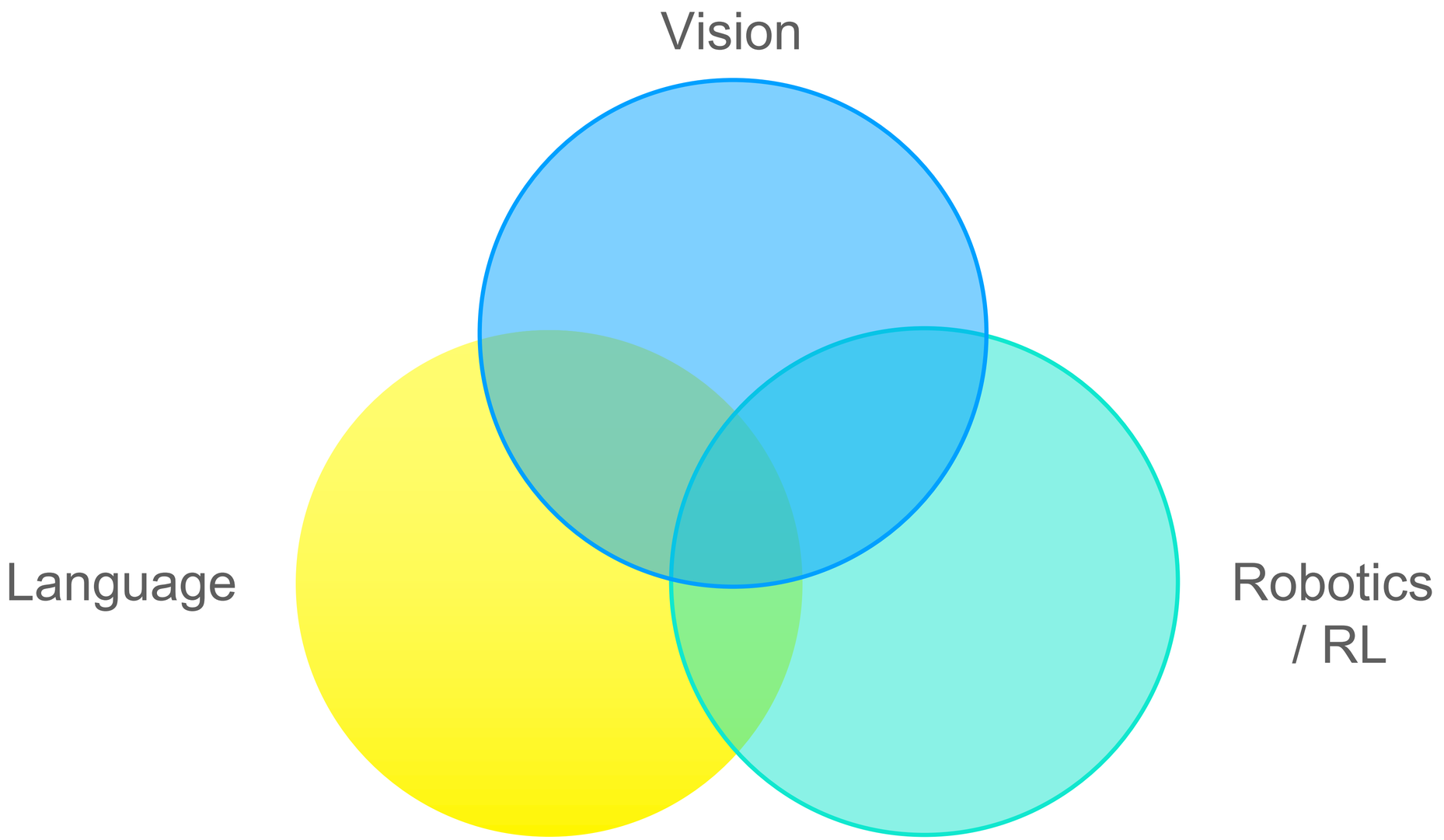
Datasets

SUNCG (Song et al., 2017)

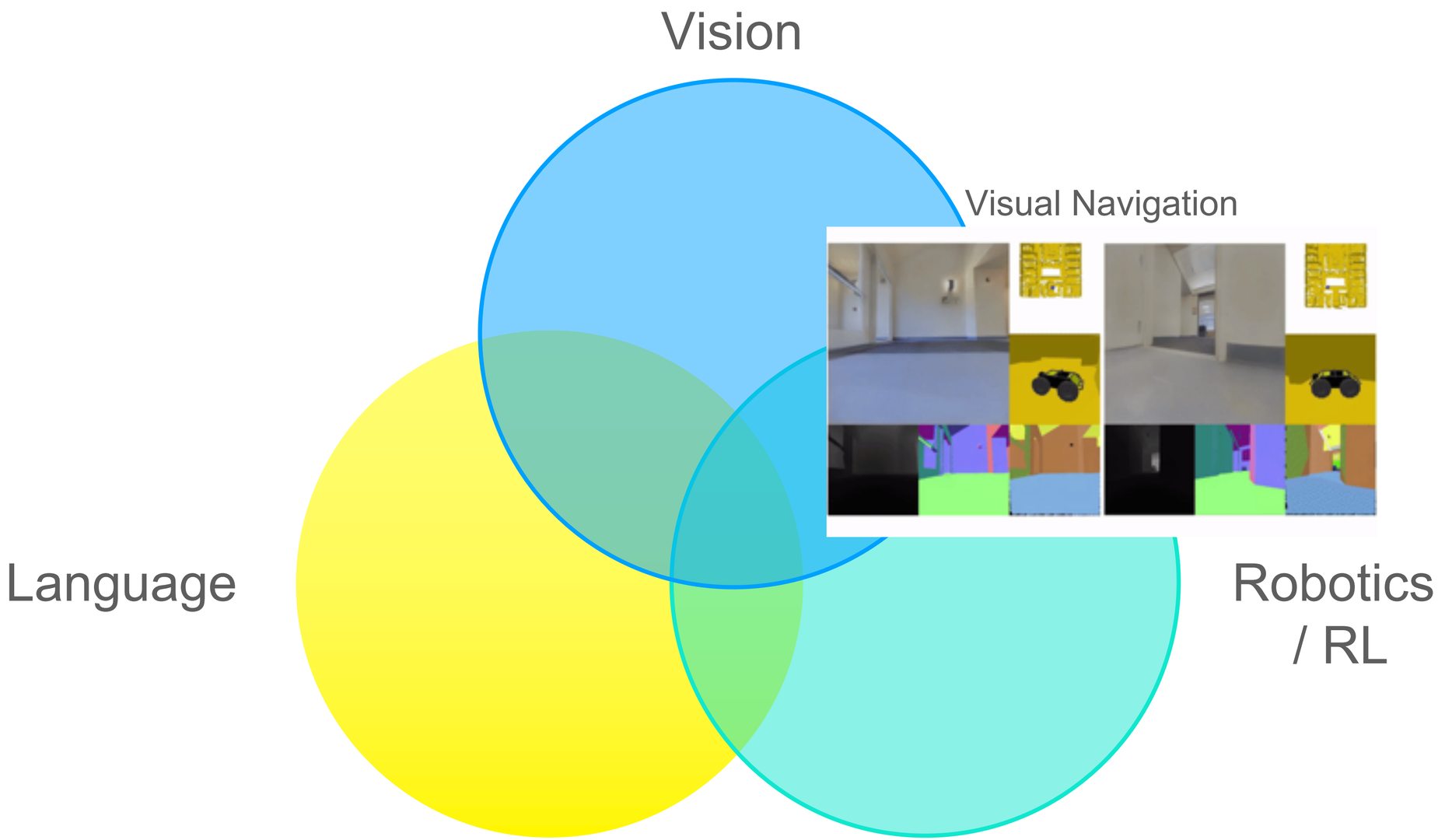
Matterport3D (Chang et al., 2017)

Stanford 2D-3D-S (Armeni et al., 2017)

Slide credit: Abhishek Das



Slide credit: Dhruv Batra



Slide credit: Dhruv Batra

Vision

V&L Navigation

Instruction: Head upstairs and walk past the piano through an archway directly in front. Turn right when the hallway ends at pictures and table. Wait by the moose antlers hanging on the wall.

Embodied QA

Language

Robotics
/ RL

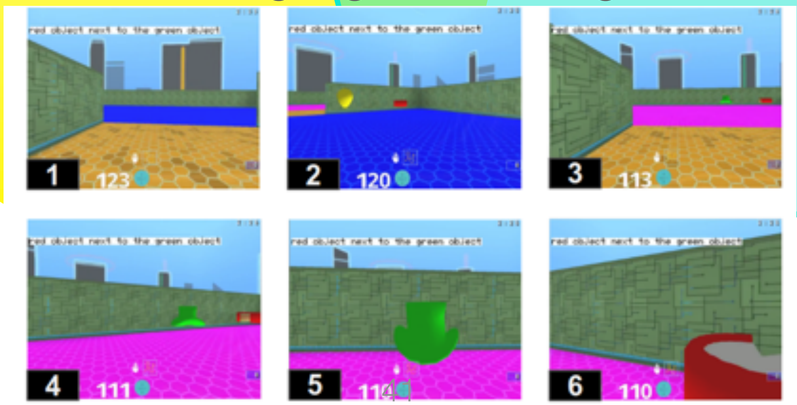
Slide credit: Dhruv Batra

Vision

Language

Robotics
/ RL

Language Grounding



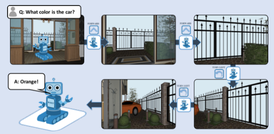
Slide credit: Dhruv Batra

Our Vision

- Create the ImageNet/COCO/VQA of Embodied AI
  - Dataset → Simulator → Task → Benchmark Challenge

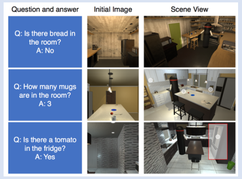
Standardizing the Embodied Agent Stack

Tasks



EmbodiedQA
(Das et al., 2018)

Language grounding
(Hill et al., 2017)

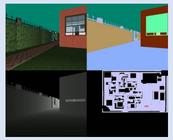


Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu et al., 2017, Gupta et al., 2017)

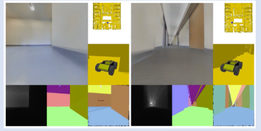
Simulators



House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)

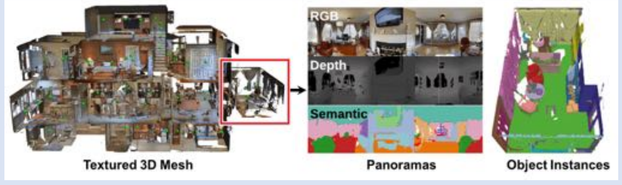


Gibson
(Zamir et al., 2018)

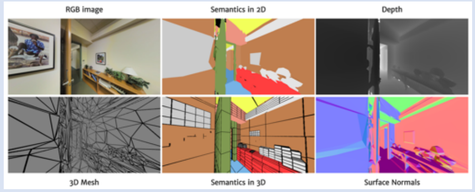
CHALET
(Yan et al., 2018)

Datasets

SUNCG (Song et al., 2017)



Matterport3D (Chang et al., 2017)



2D-3D-S (Armeni et al., 2017)

Habitat Platform

Habitat API

Habitat Sim

Generic Dataset Support

Standardizing the Embodied Agent Stack

Tasks

EmbodiedQA
(Das et al., 2018)

Language grounding
(Hill et al., 2017)

Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu et al., 2017, Gupta et al., 2017)

Simulators

House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

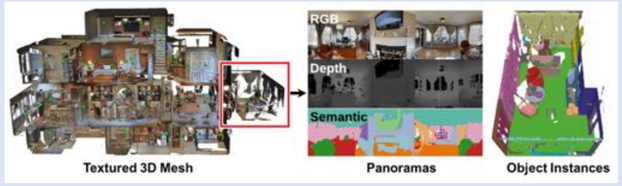
MINOS
(Savva et al., 2017)

Gibson
(Zamir et al., 2018)

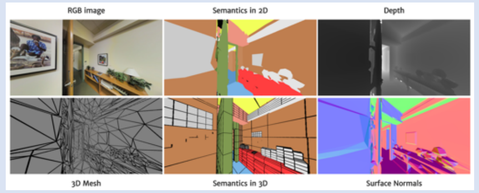
CHALET
(Yan et al., 2018)

Datasets

SUNCG (Song et al., 2017)



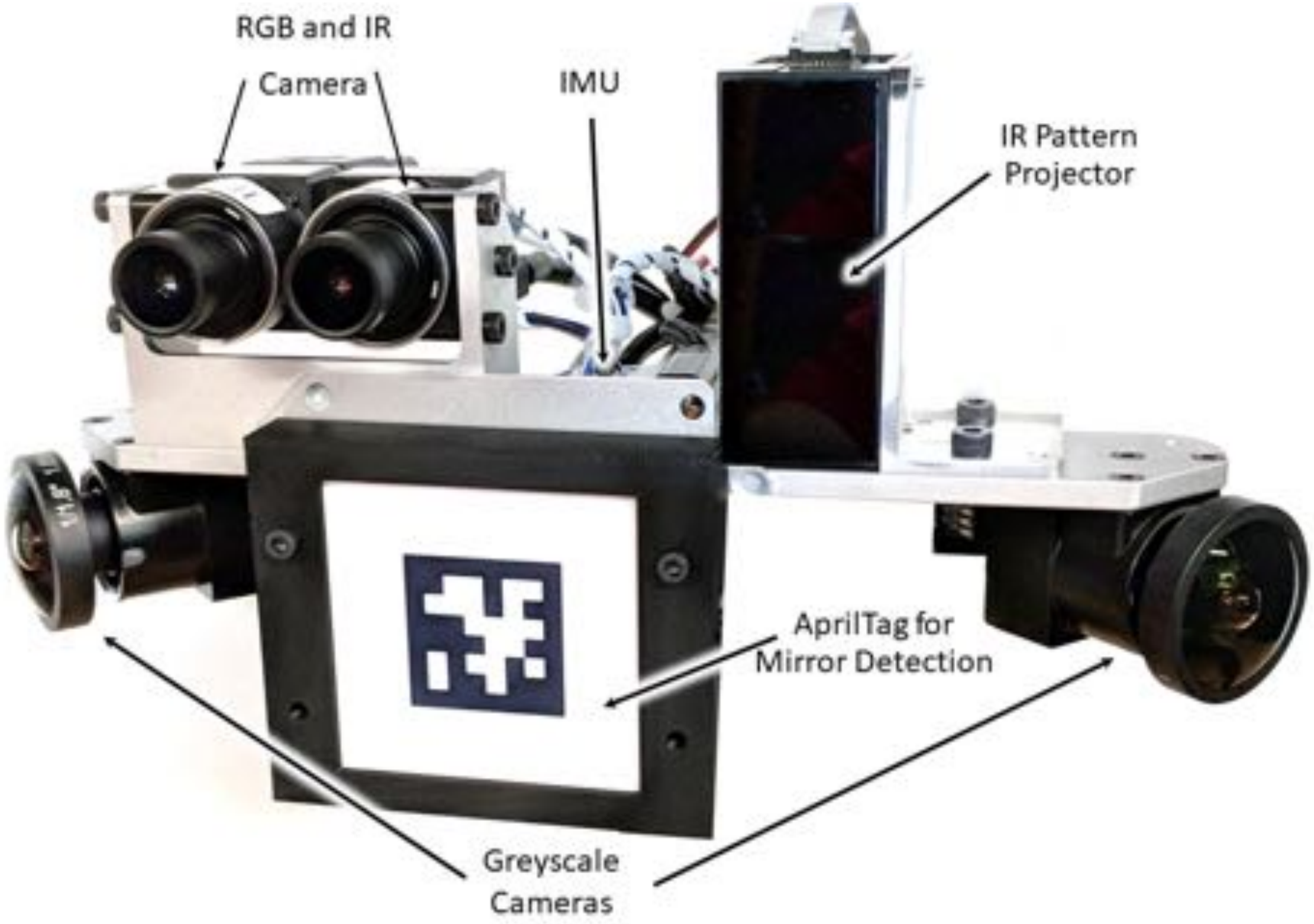
Matterport3D (Chang et al., 2017)



2D-3D-S (Armeni et al., 2017)

Julian Straub
(FRL)

Richard Newcombe
(FRL)



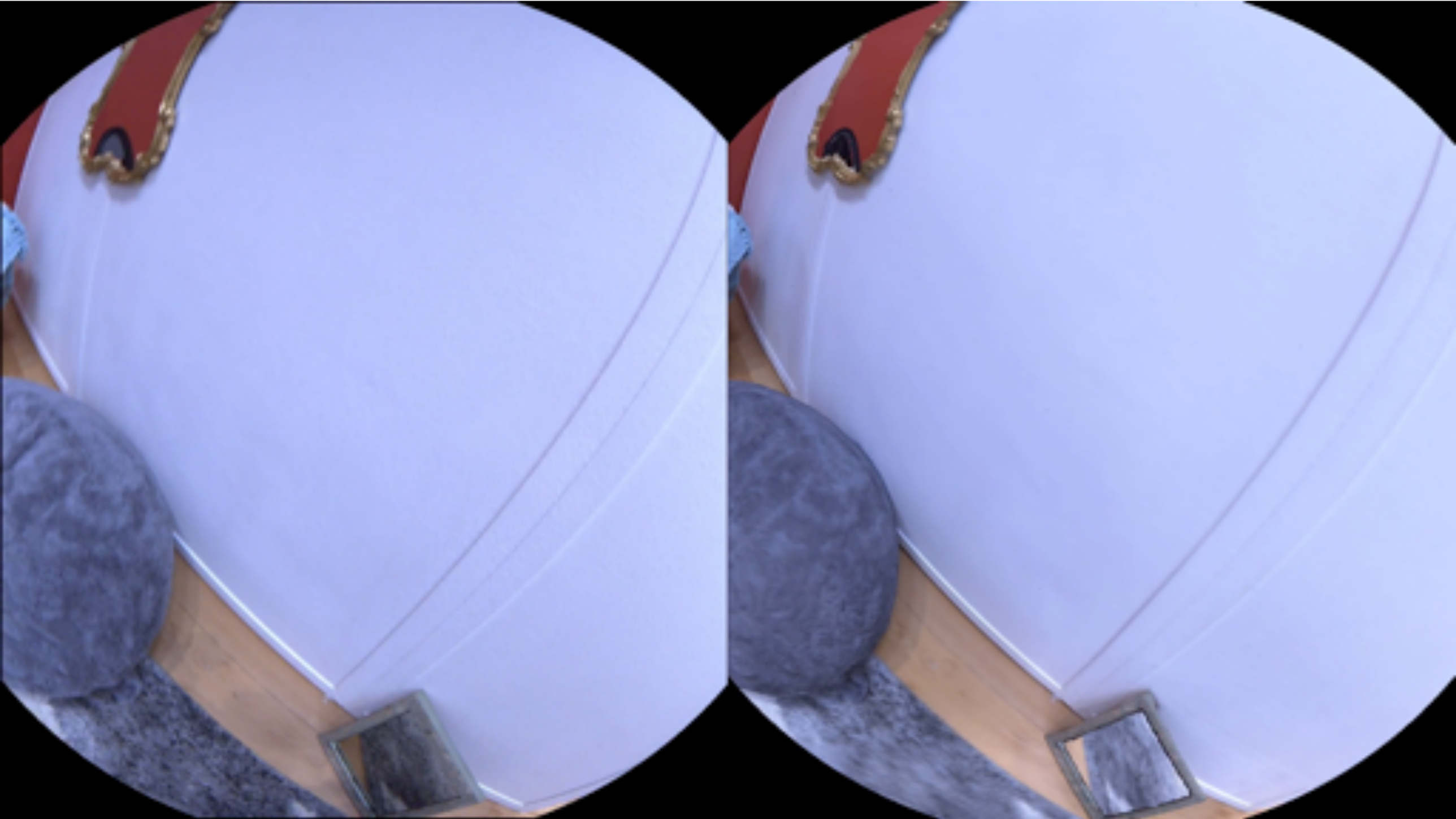
Julian Straub (FRL)

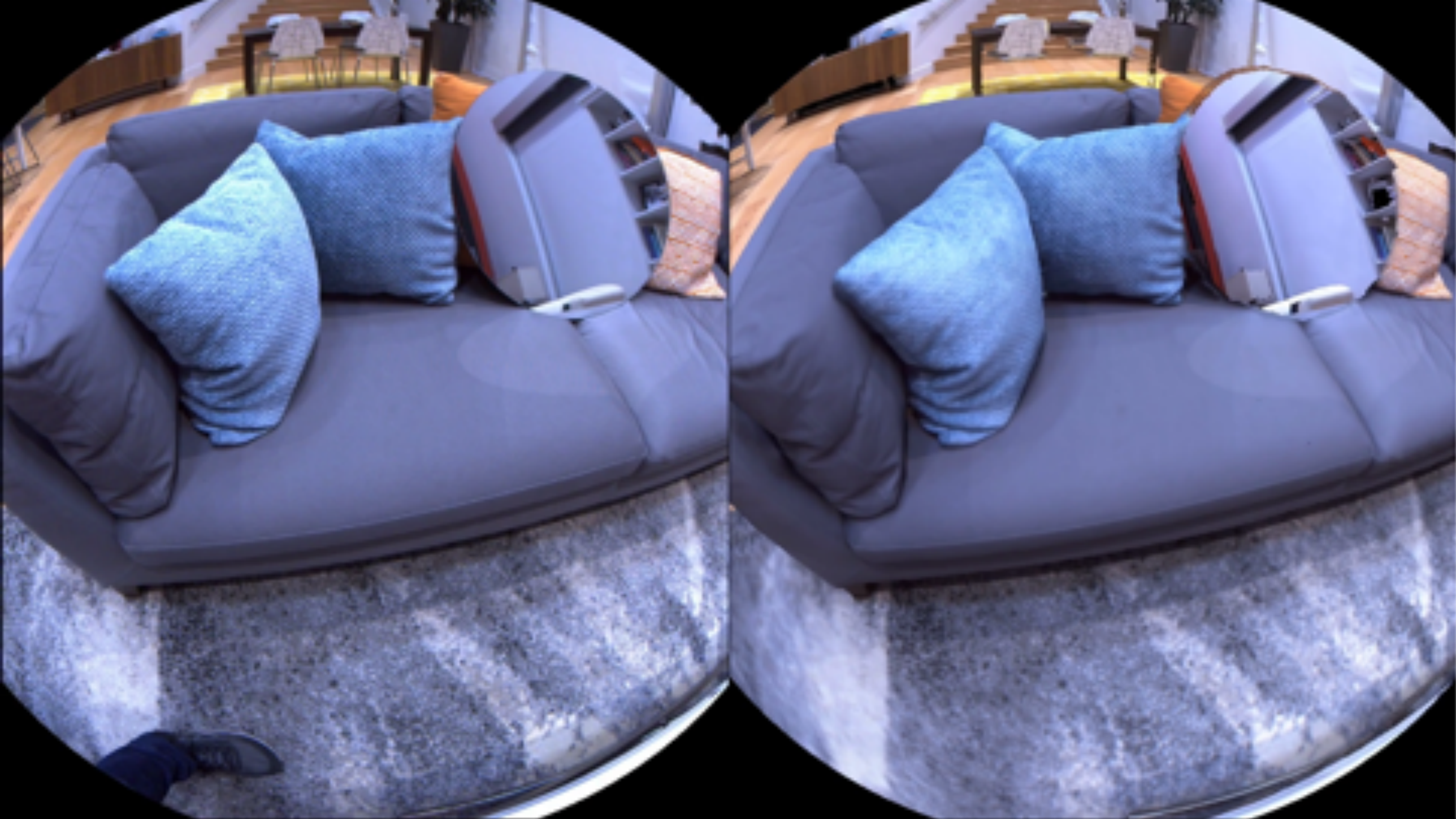
Richard Newcombe (FRL)

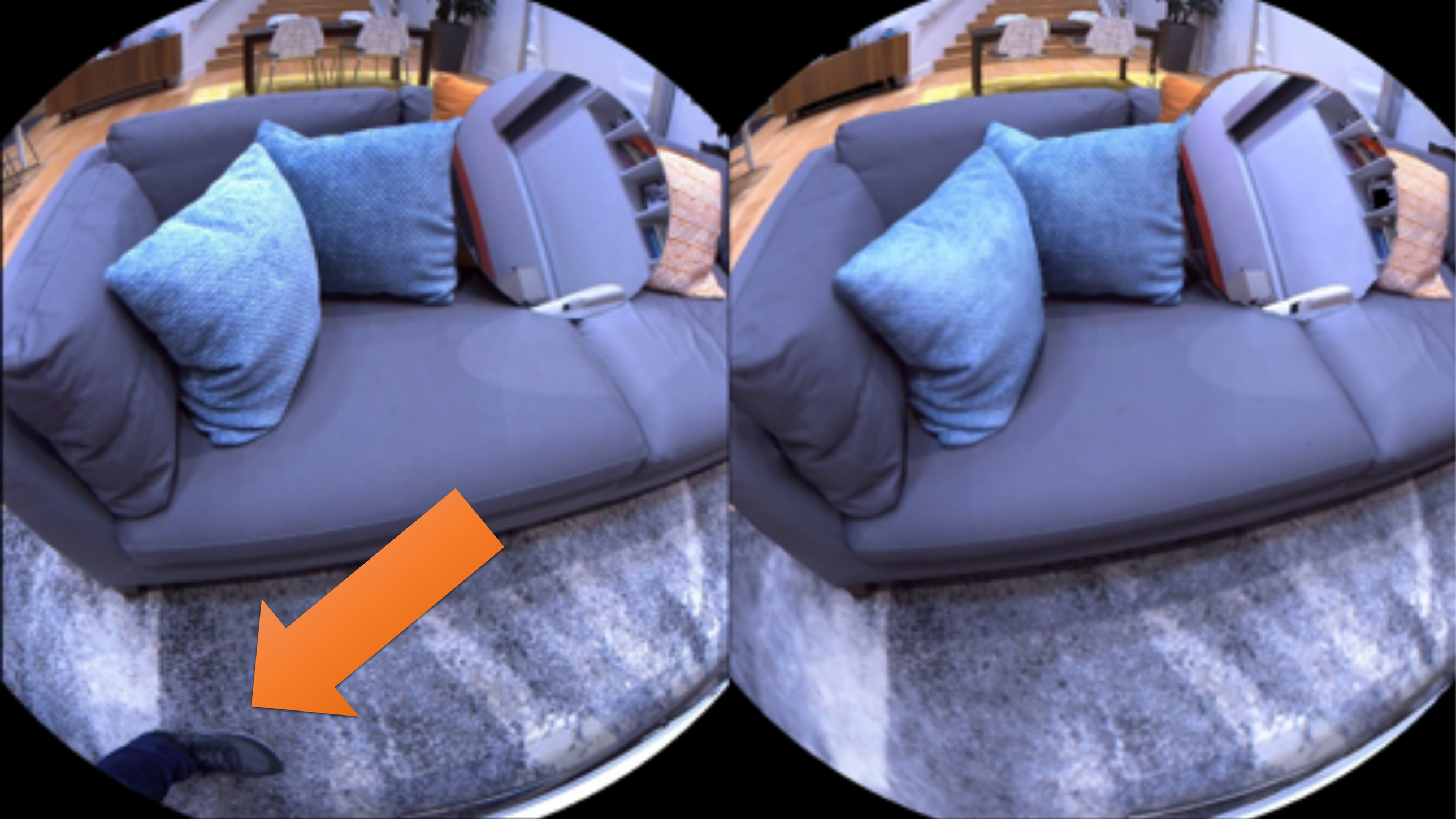
facebook
Reality Labs

FRL Surreal Team: high quality 3D reconstructions

The Replica Dataset: A Digital Replica of Indoor Spaces [Straub et al. 2019]







Standardizing the Embodied Agent Stack

Tasks

EmbodiedQA
(Das et al., 2018)

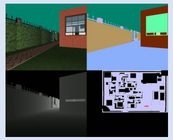
Language grounding
(Hill et al., 2017)

Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu et al., 2017, Gupta et al., 2017)

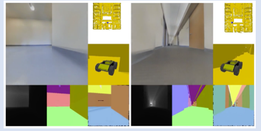
Simulators



House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)



Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

Manolis Savva (FAIR)

Yili Zhao (FAIR)

Datasets

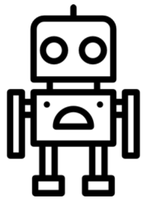
SUNCG (Song et al., 2017)

Matterport3D (Chang et al., 2017)

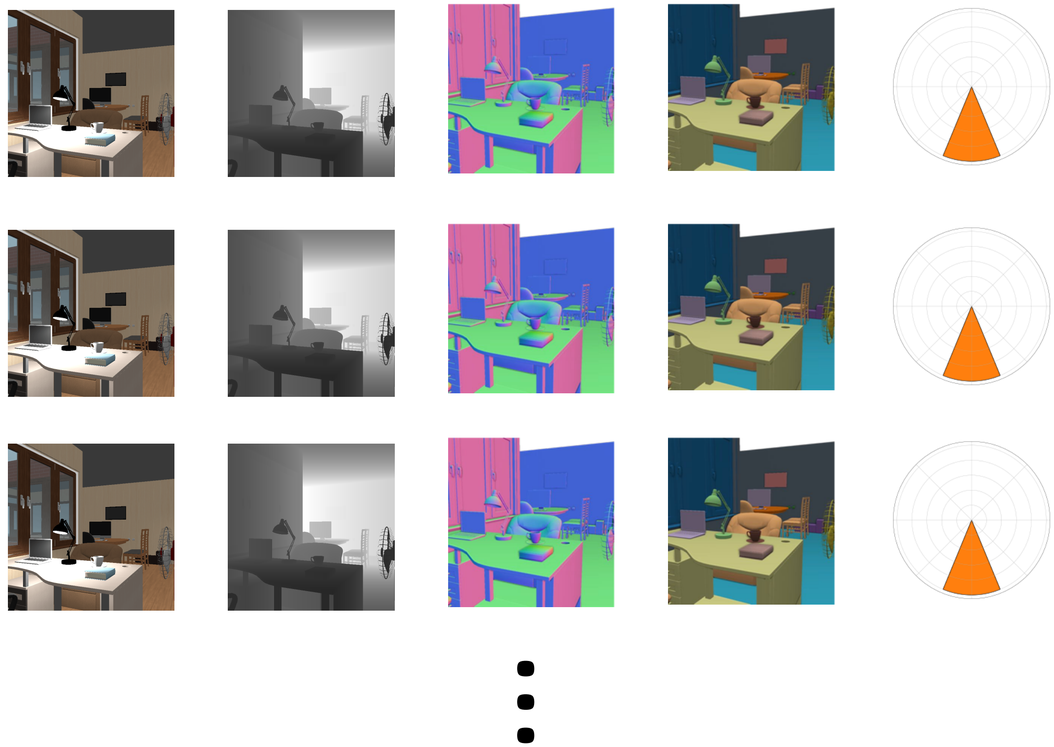
2D-3D-S (Armeni et al., 2017)

Challenge: human vs machine needs

1080p @ 60Hz



256x256 @ 1000+ Hz



Slide credit: Manolis Savva


HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

Habitat-Sim

- Photorealistic 3D simulator

  (C++ with pybind11)

- Generic 3D dataset support

  (Replica, Gibson, MP3D, +more)

- Fast: over 1,000 FPS single-threaded

  10,000 FPS multi-process (single GPU)

Habitat-Sim: Datasets agnostic!

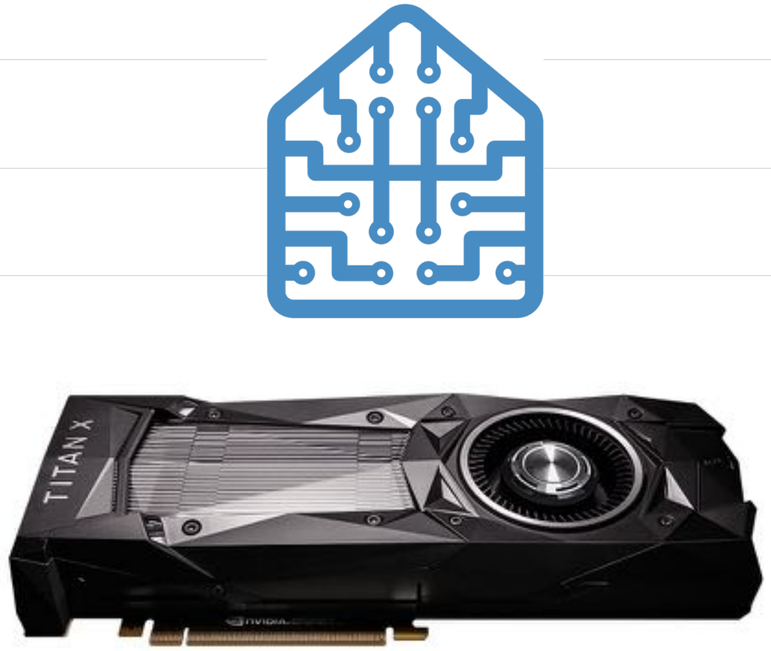
Frames Per Second

1100
1000
900
800
700
600
500
400
300
200
100
0

Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim


1100
1000
900
800
700
600
500
400
300
200
100
0
Gibson
BUT WAIT
THERE'S MORE
Habitat-Sim

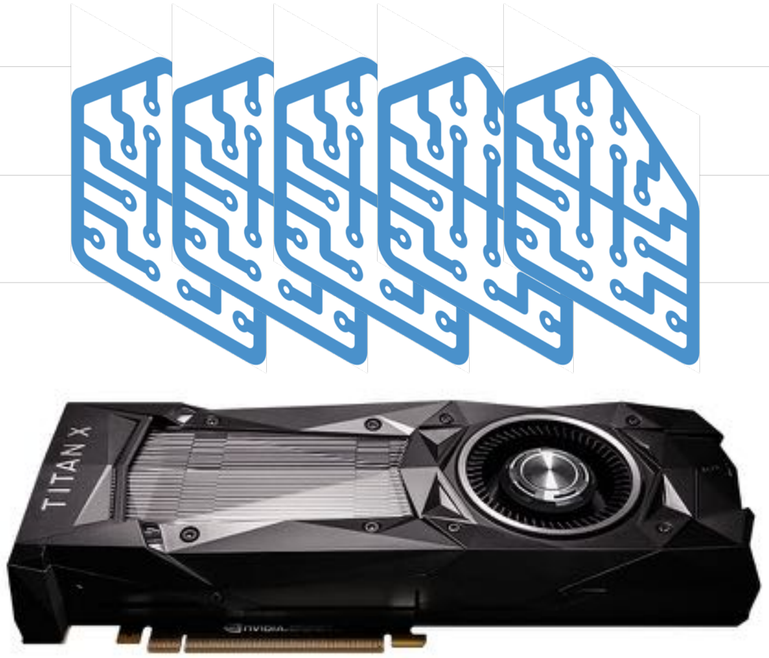
Frames Per Second



1100
1000
900
800
700
600
500
400
300
200
100
0

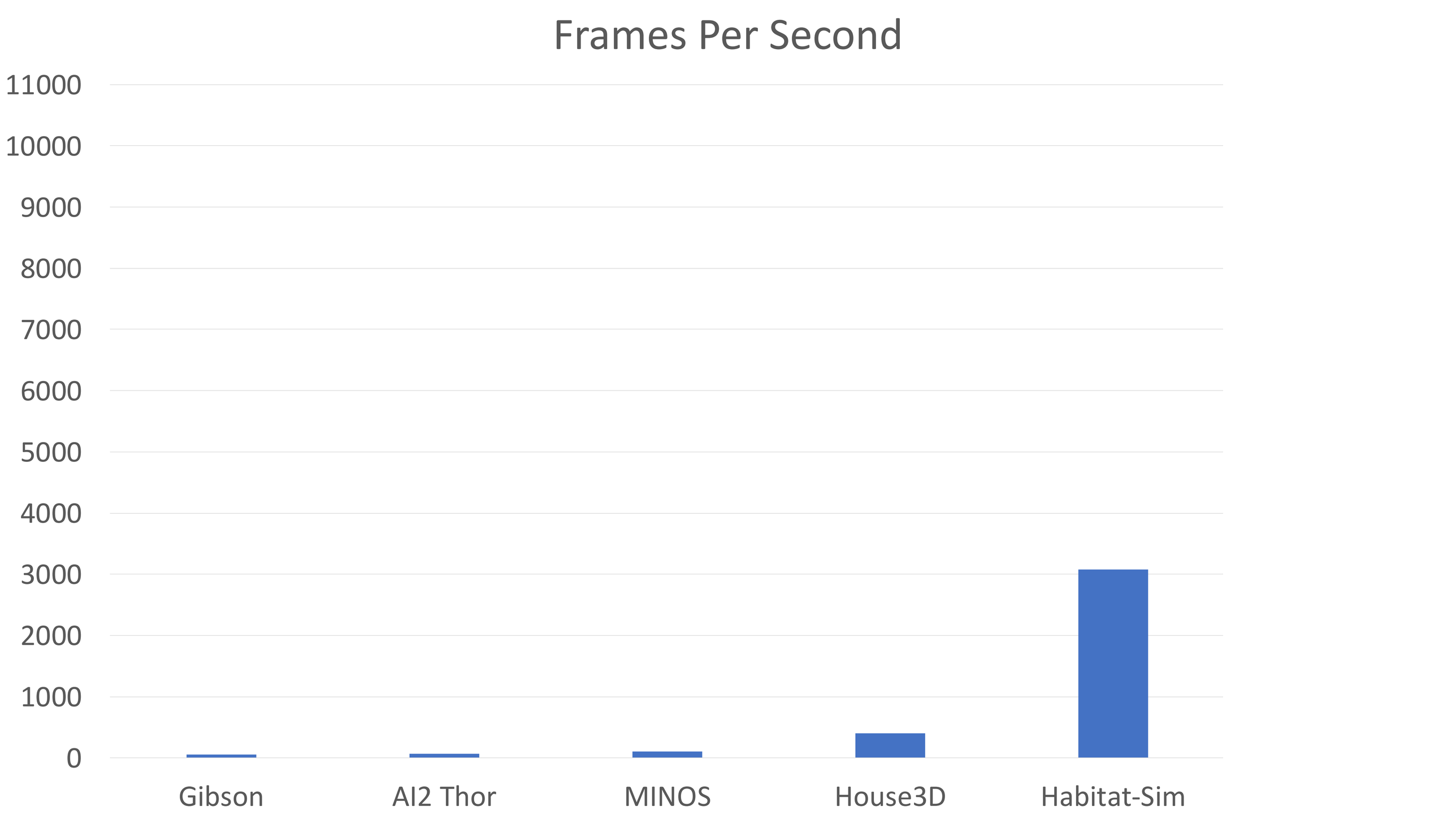
Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim

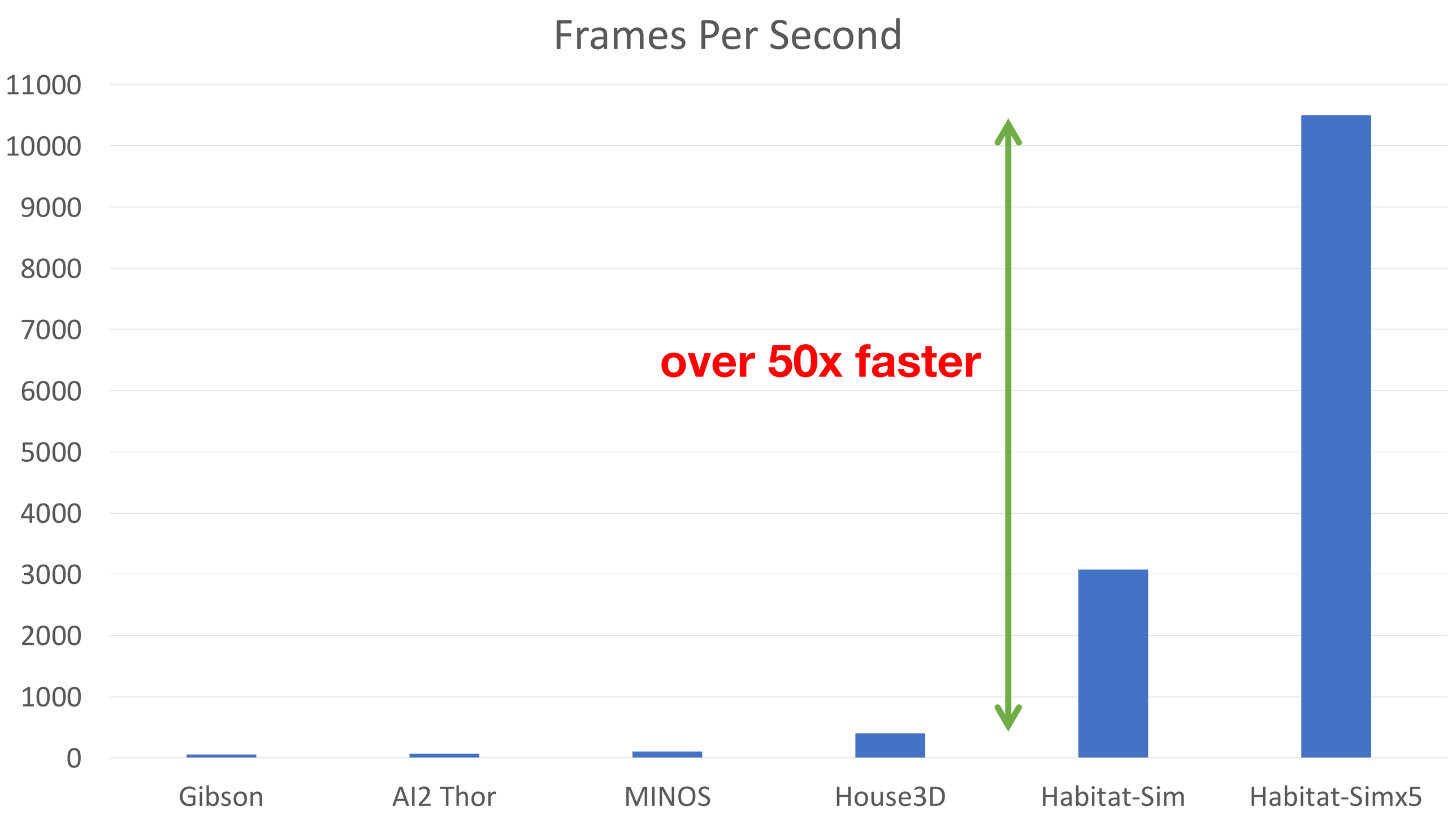
Frames Per Second

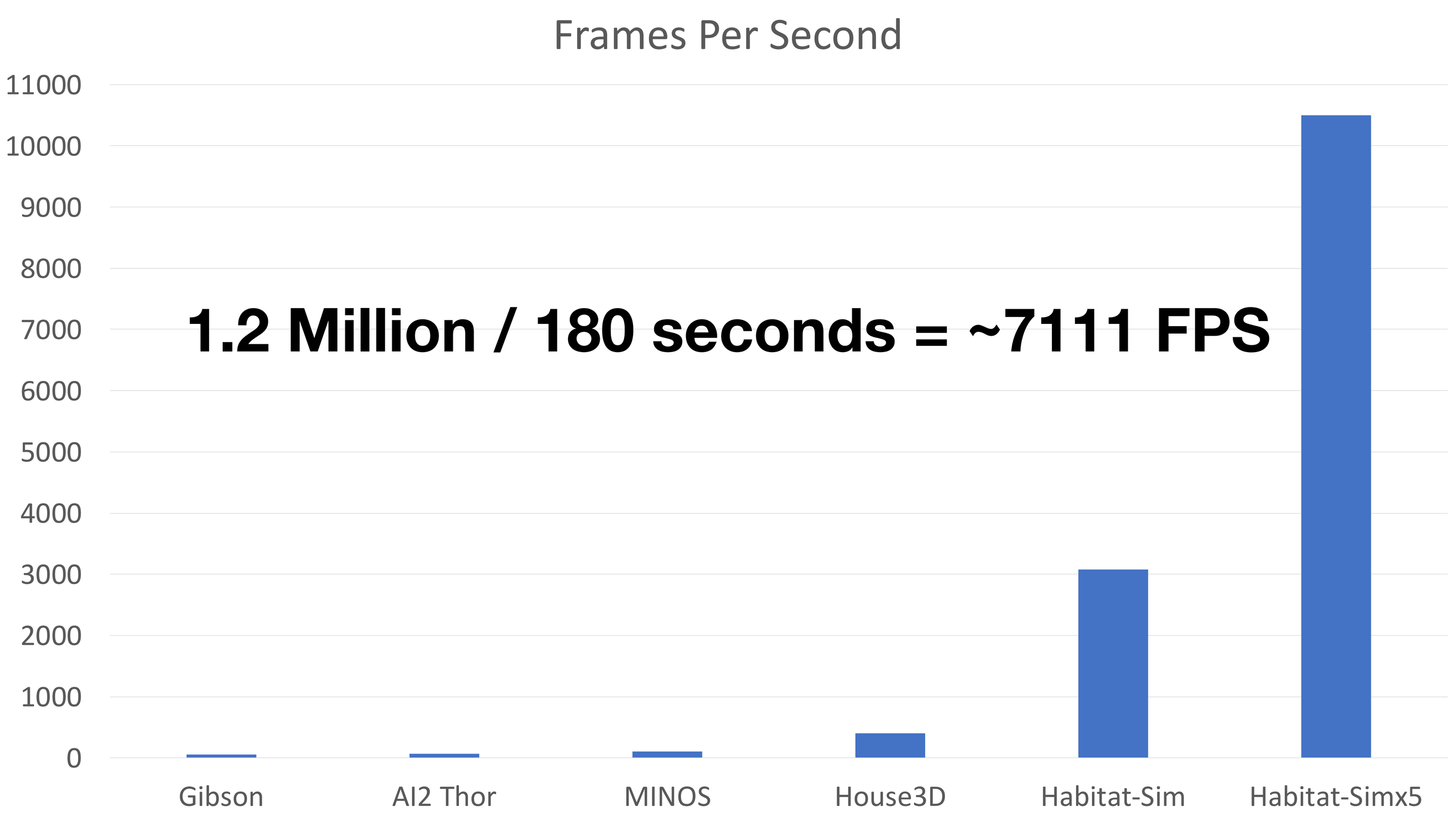


1100
1000
900
800
700
600
500
400
300
200
100
0

Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim

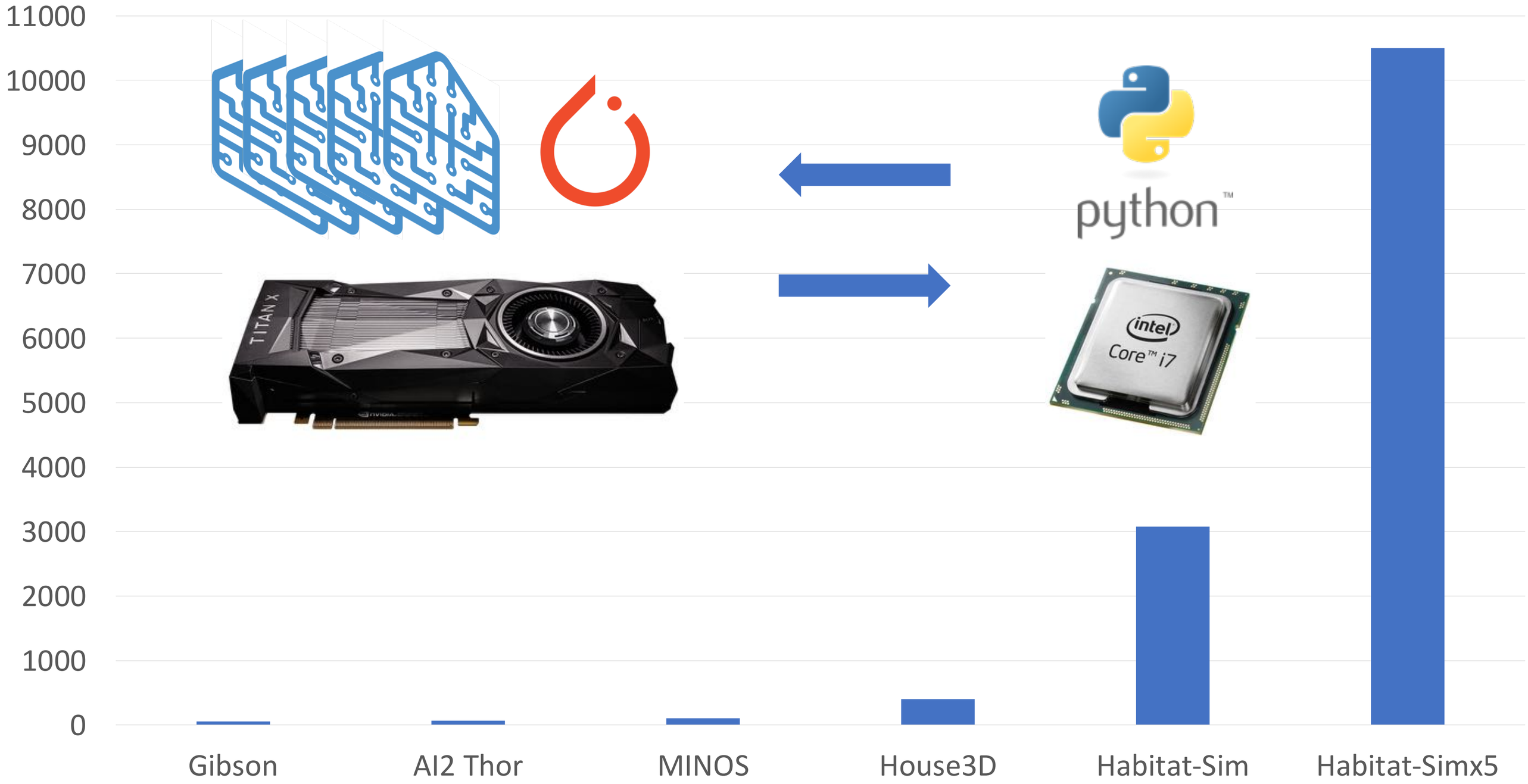

Frames Per Second
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim


Frames Per Second
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
over 50x faster
Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim
Habitat-Simx5

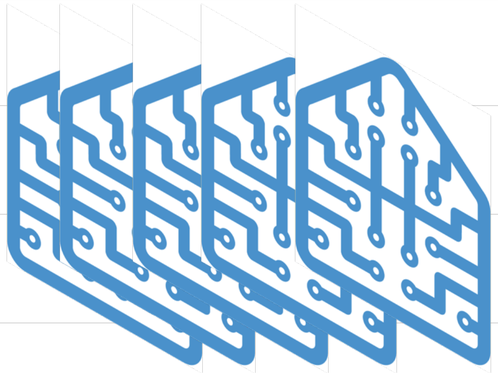

Frames Per Second
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1.2 Million / 180 seconds = ~7111 FPS
Gibson
AI2 Thor
MINOS
House3D
Habitat-Sim
Habitat-Simx5


11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Gibson
Habitat-Sim
Habitat-Simx5
BUT WAIT
THERE'S MORE

Frames Per Second

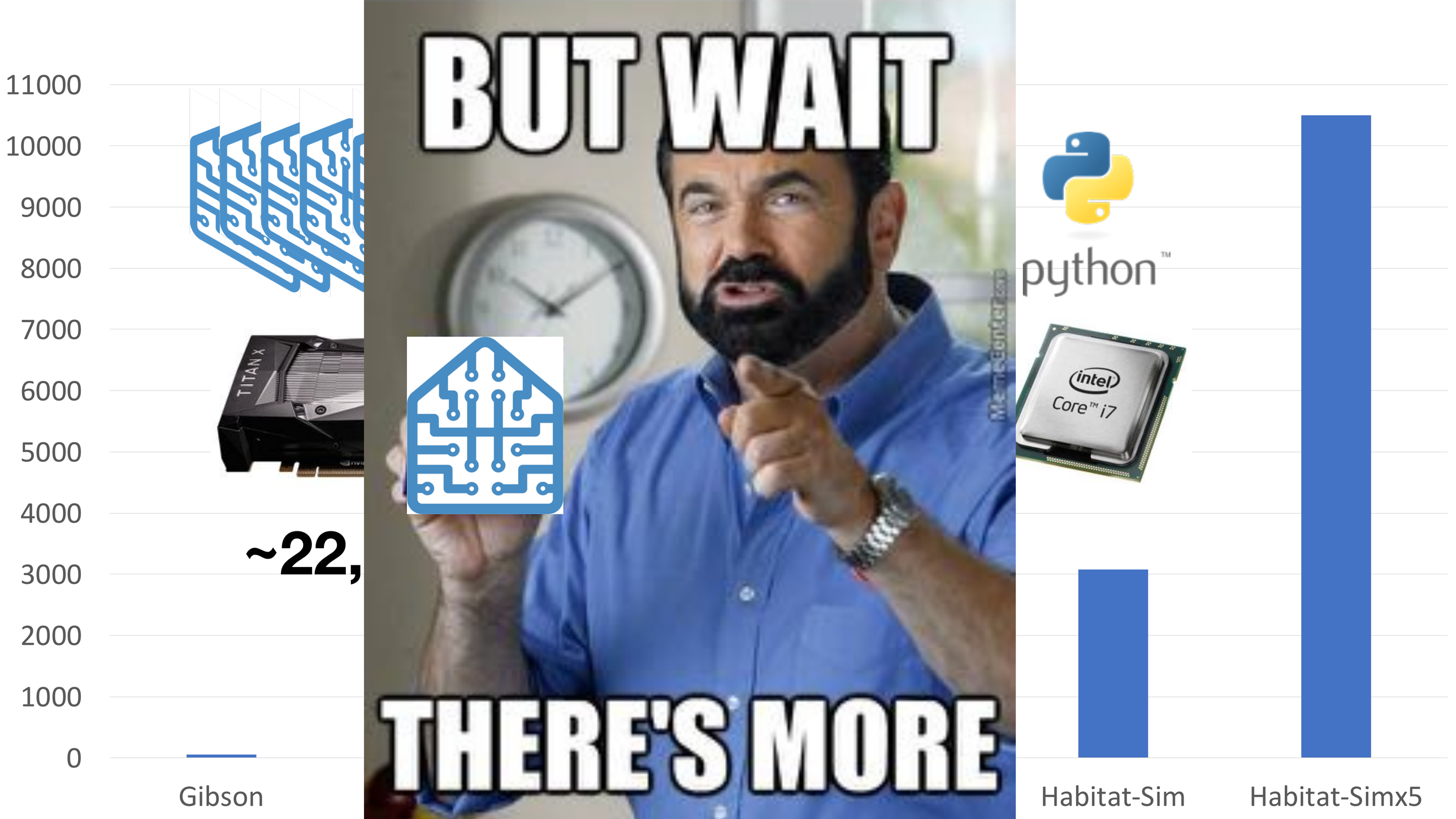


Frames Per Second



~22,000 FPS

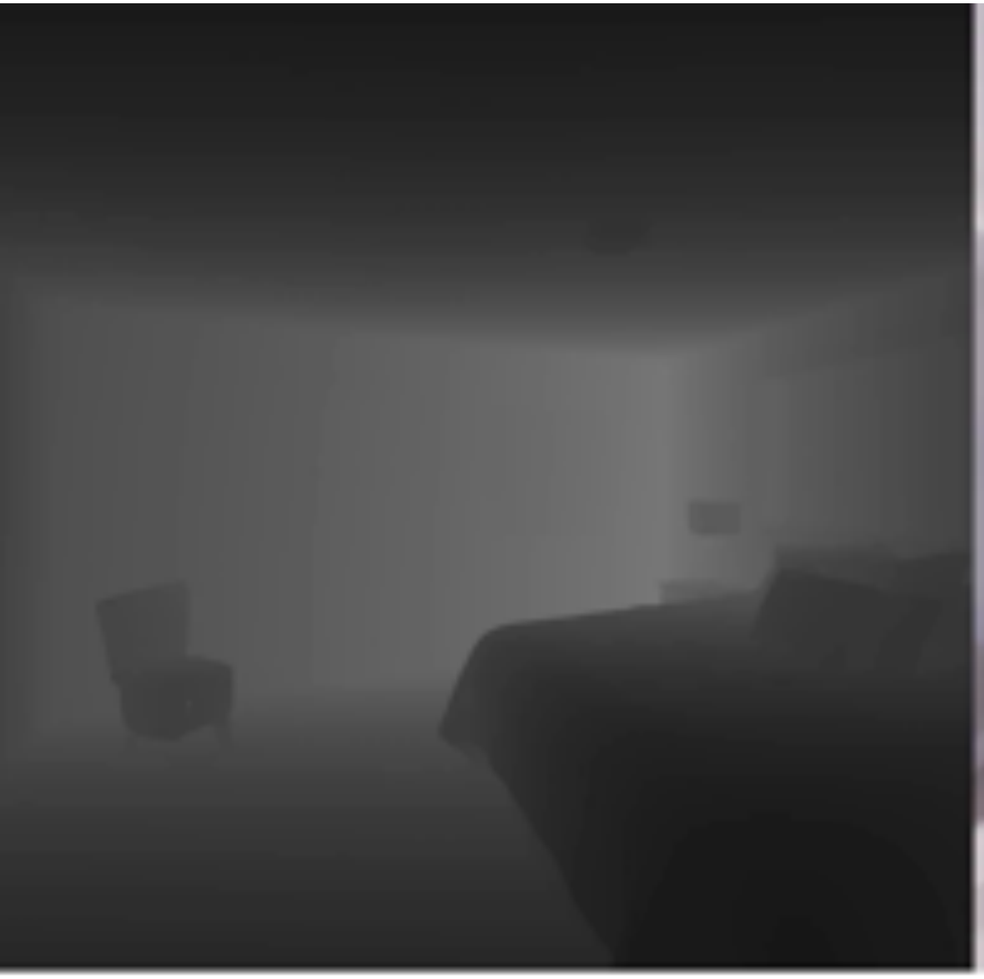
python™


11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
~22,
BUT WAIT
THERE'S MORE
python
intel
Core™ i7
Gibson
Habitat-Sim
Habitat-Simx5

Why does speed matter?

Because you can now run experiments you couldn’t before.

PointGoal Navigation



Depth

RGB and GPS+Compass

Top Down Map

Depth

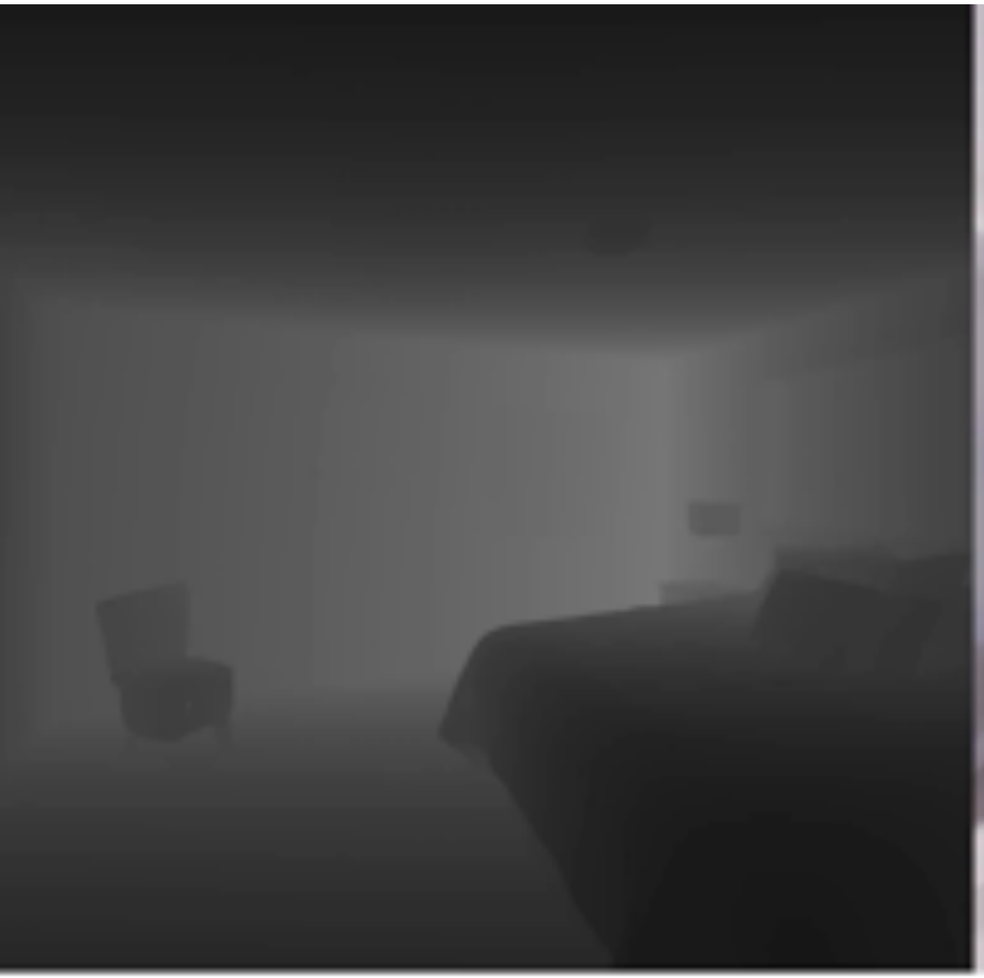
RGB and GPS+Compass

Top Down Map

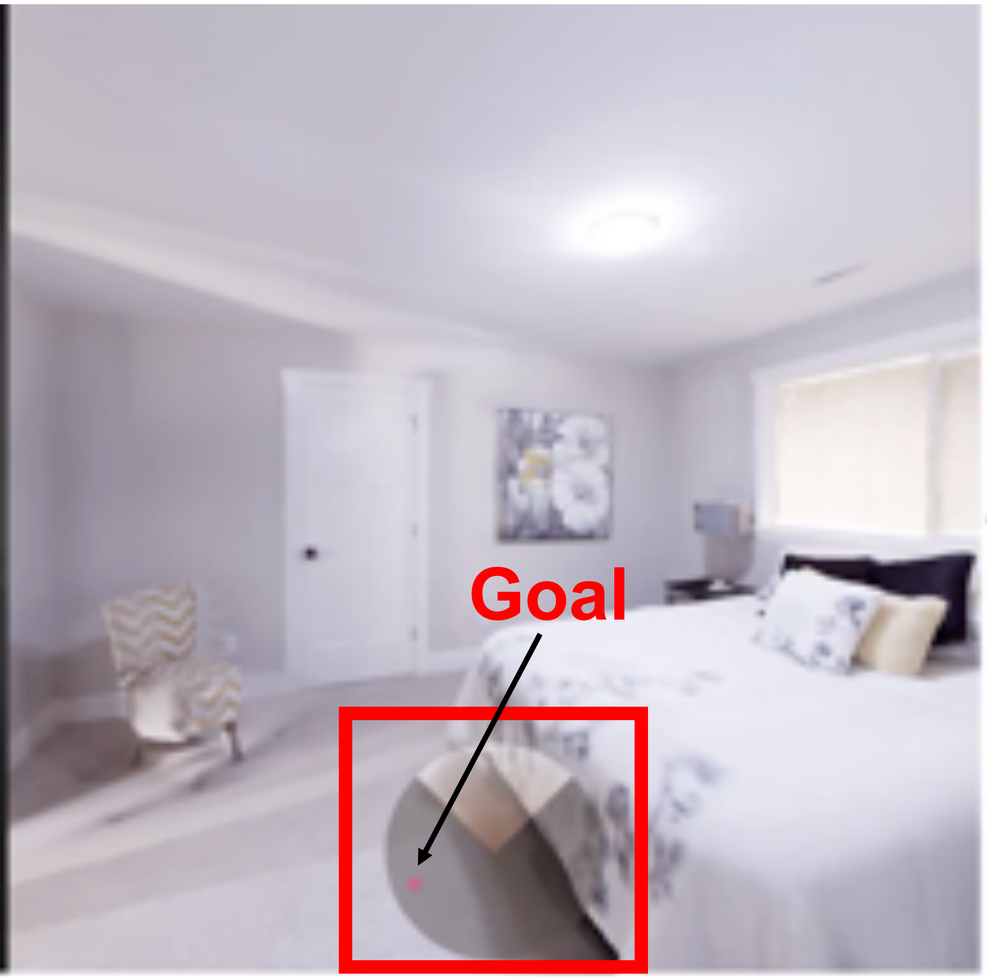
Depth

RGB and GPS+Compass

Top Down Map

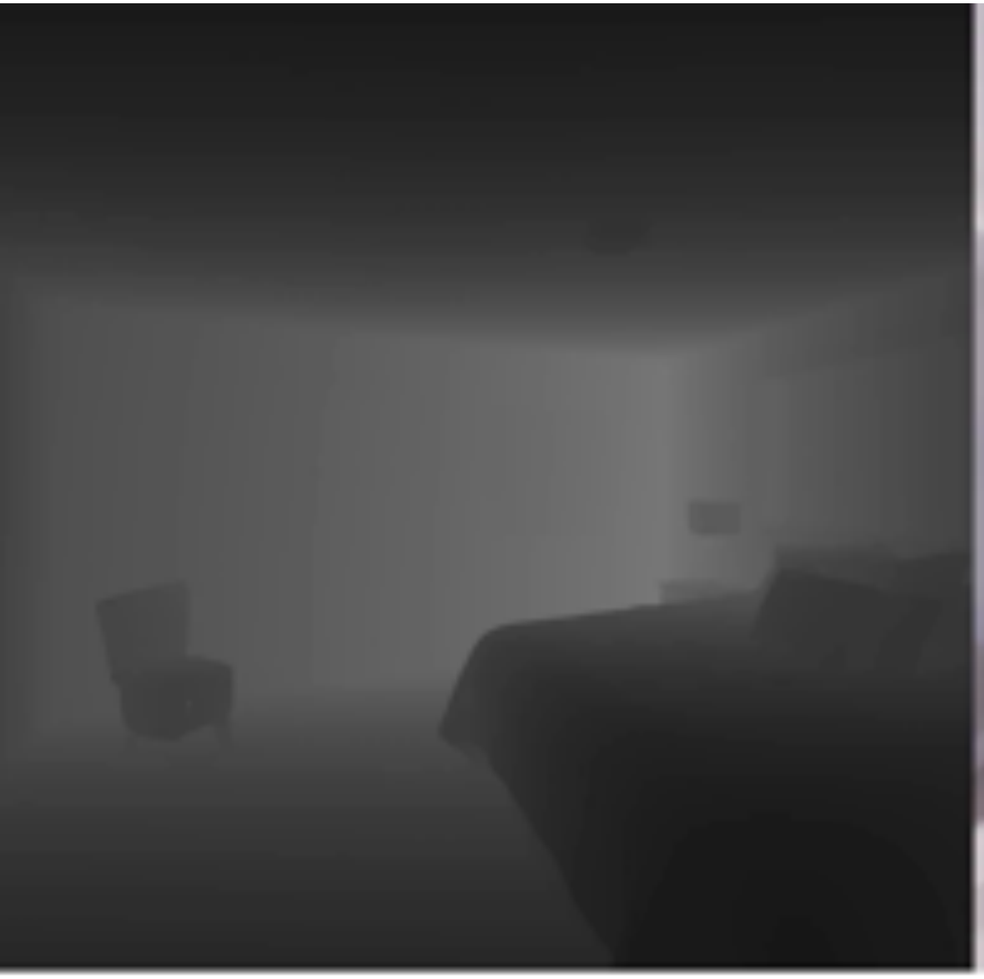


Depth



RGB and GPS+Compass

Top Down Map



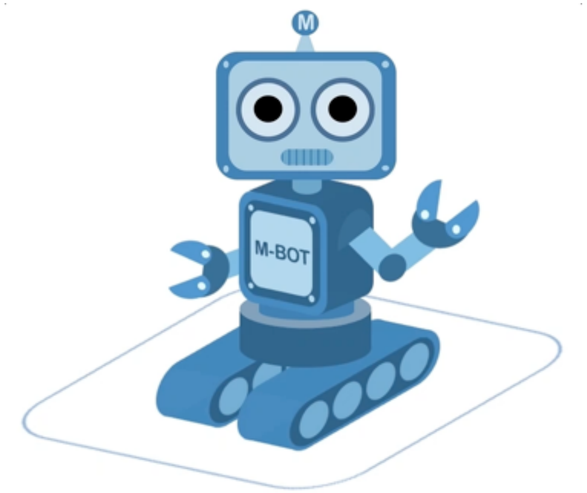
Depth

RGB and GPS+Compass

Top Down Map

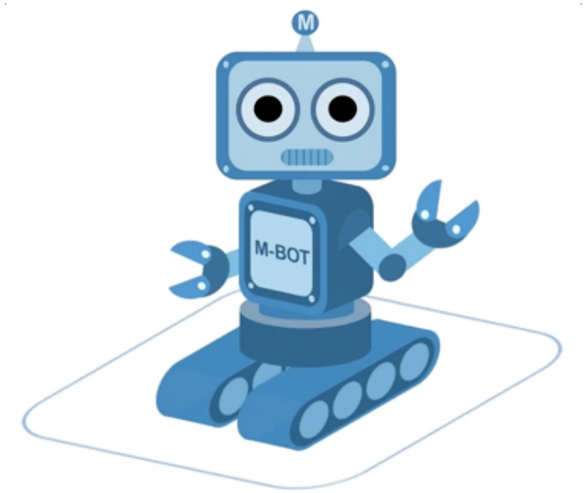
Agent and Model Design

Agent and Model Design

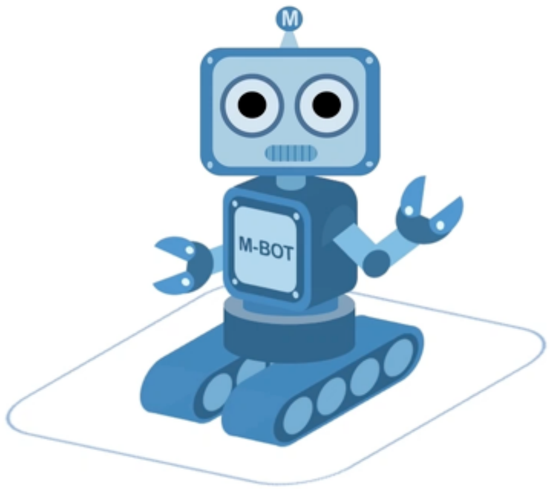


Agent and Model Design

- 1.25m tall cylinder with 0.1m radius

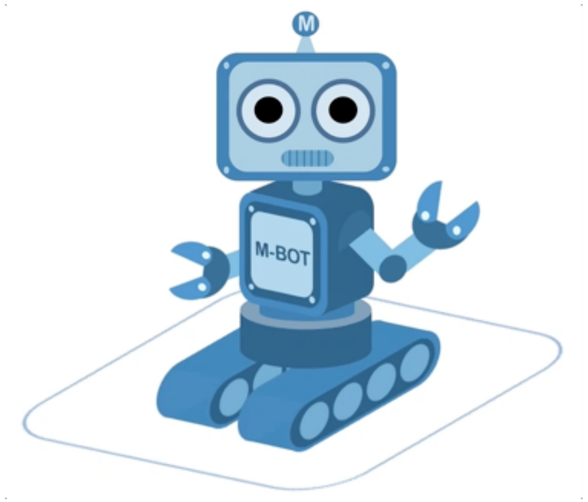


Agent and Model Design

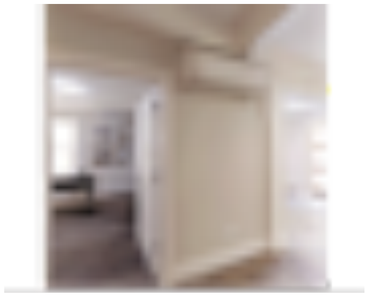


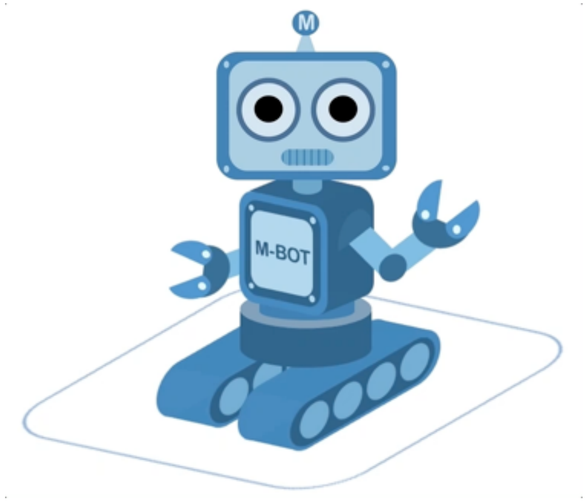
- 1.25m tall cylinder with 0.1m radius
- Actions:
  - <stop>: Indicates the agent believes it has completed the task
  - <forward>: Moves 0.25m forward
  - <left>, <right>: Turn 10 degrees

Agent and Model Design

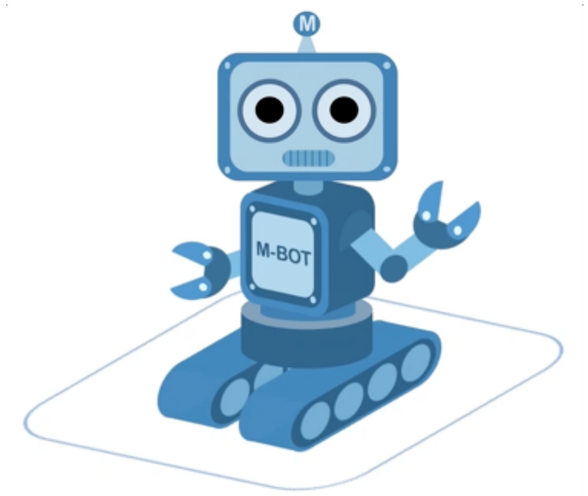


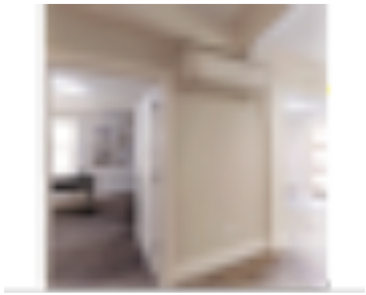
Agent and Model Design



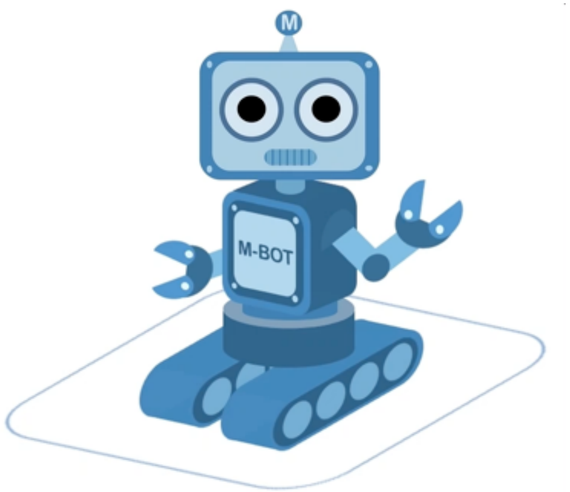


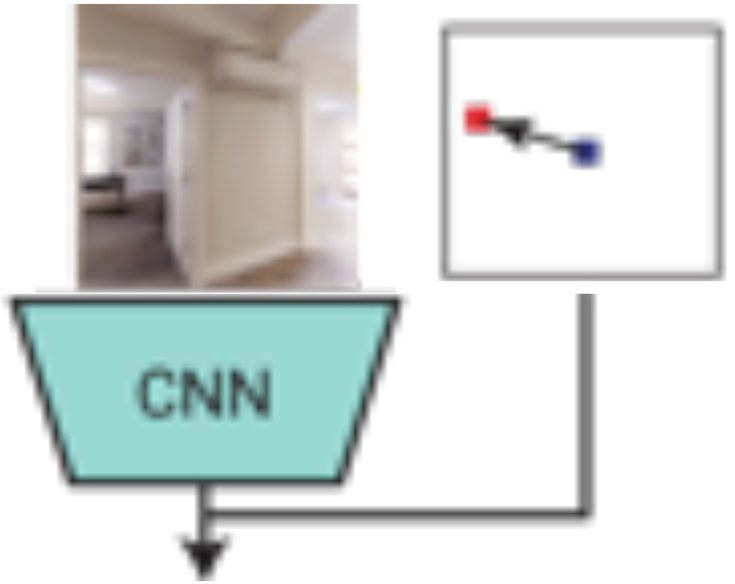
Agent and Model Design



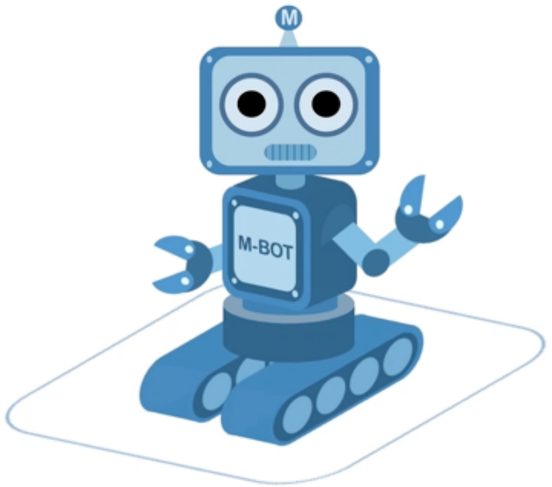


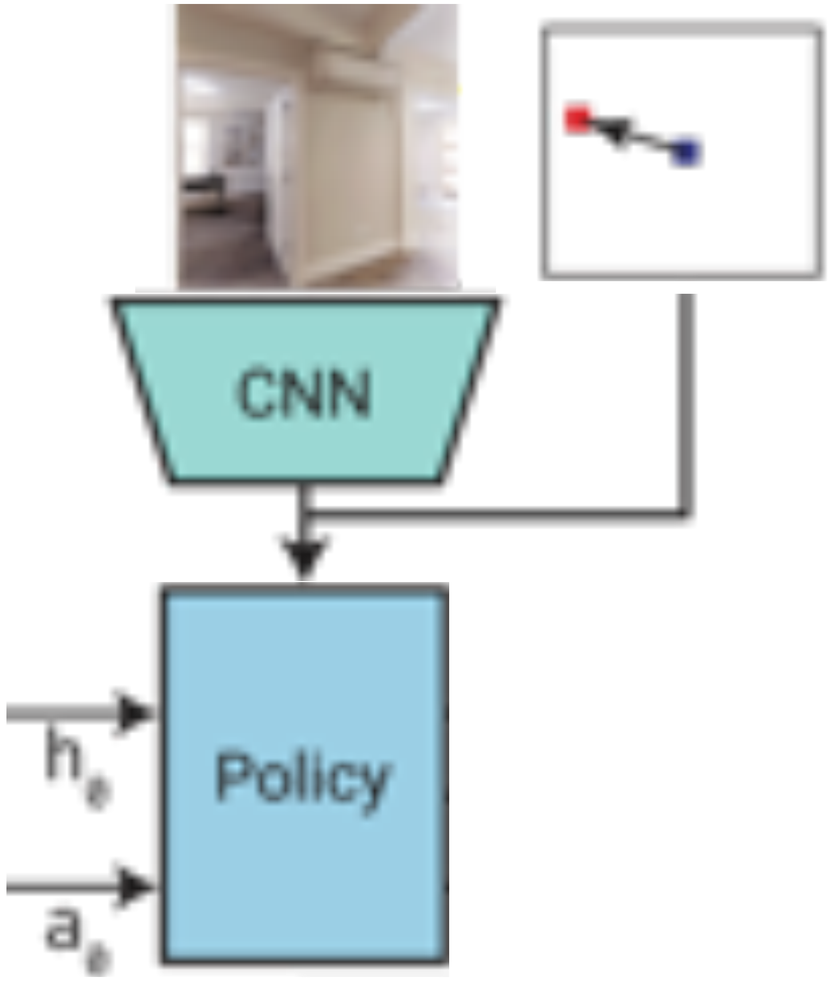
Agent and Model Design



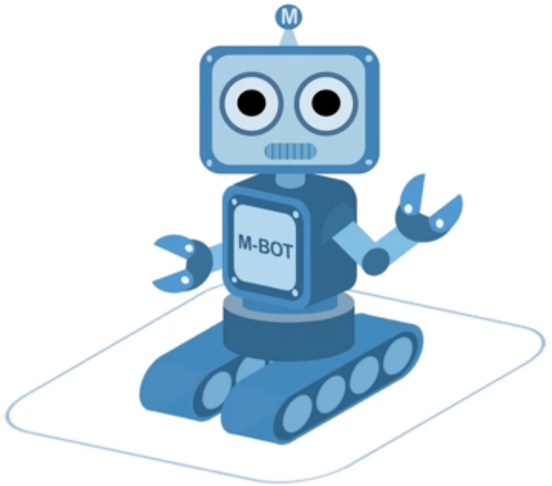


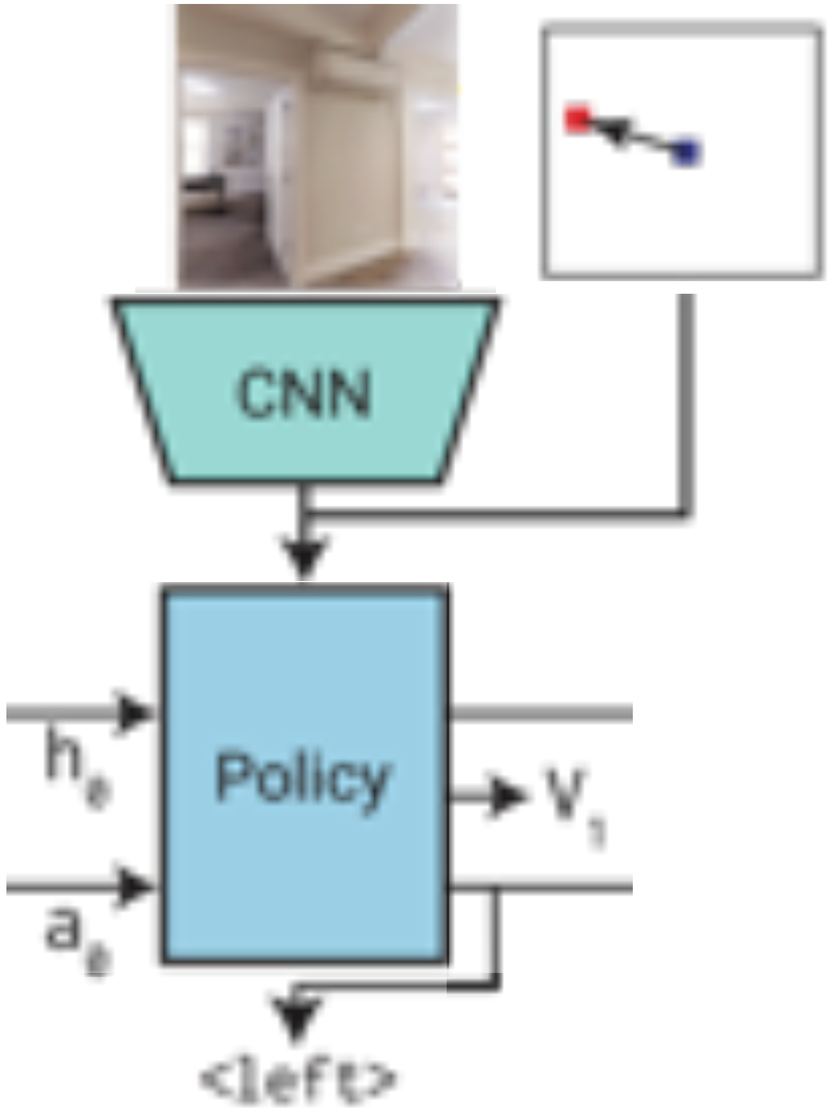
Agent and Model Design



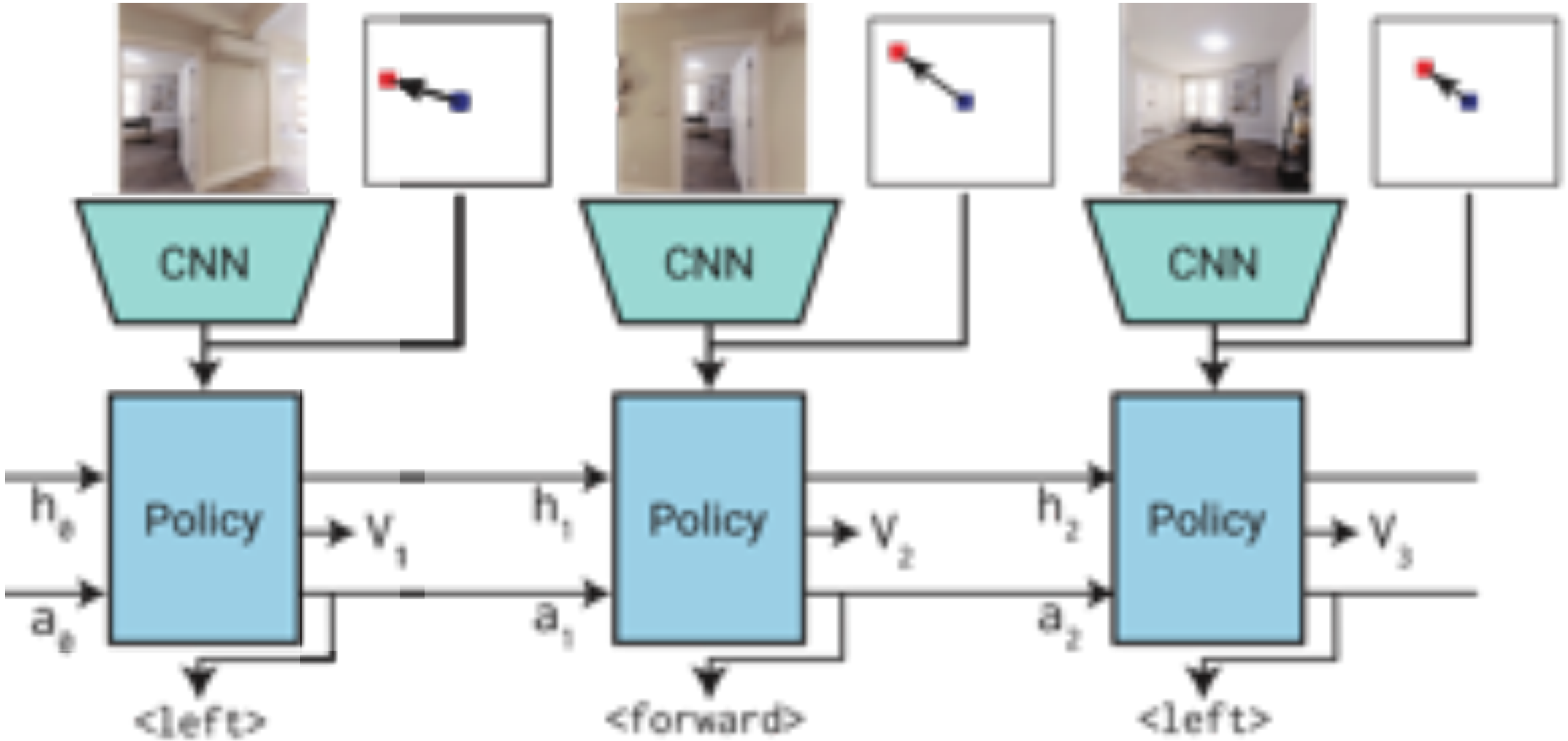


Agent and Model Design

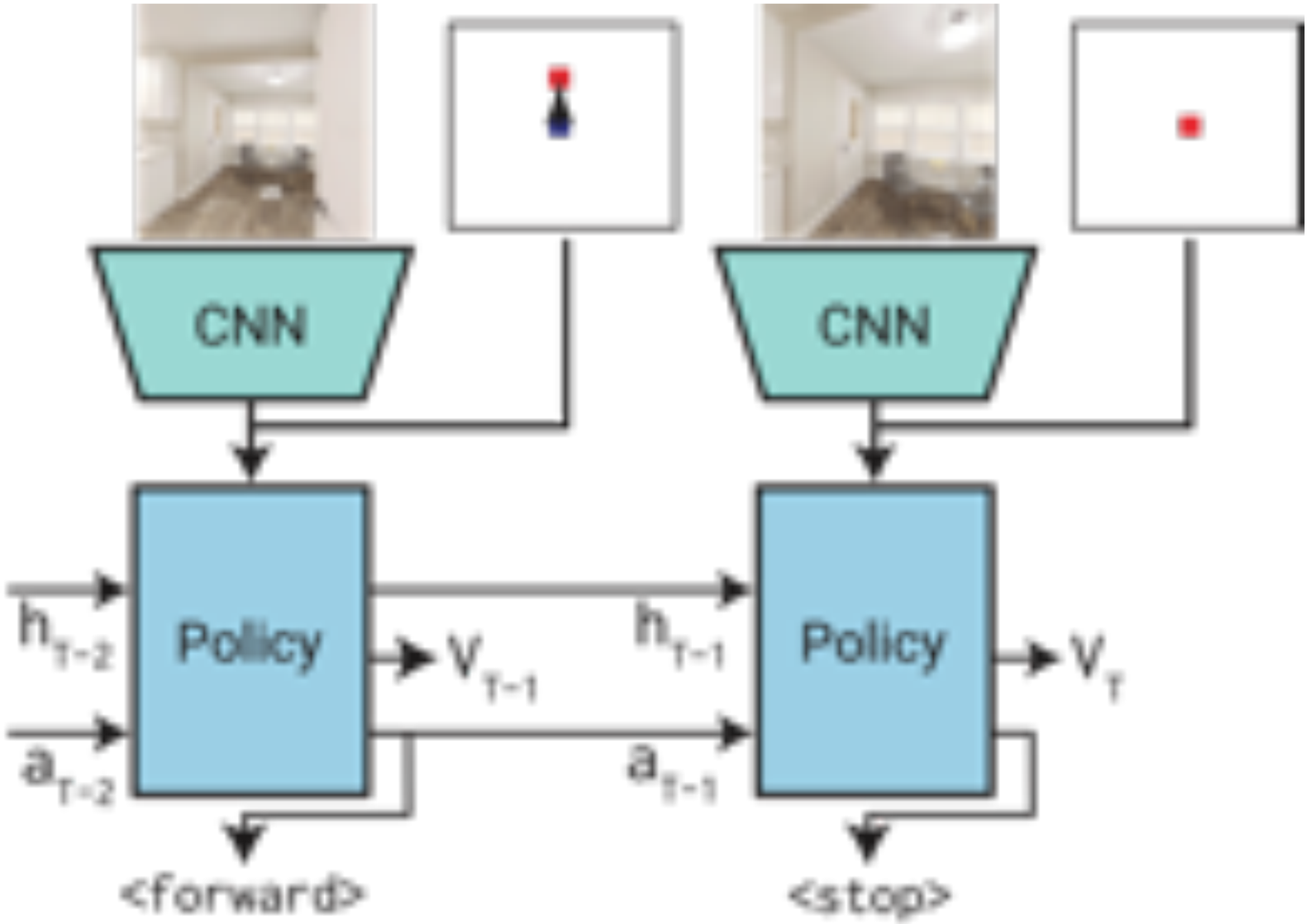




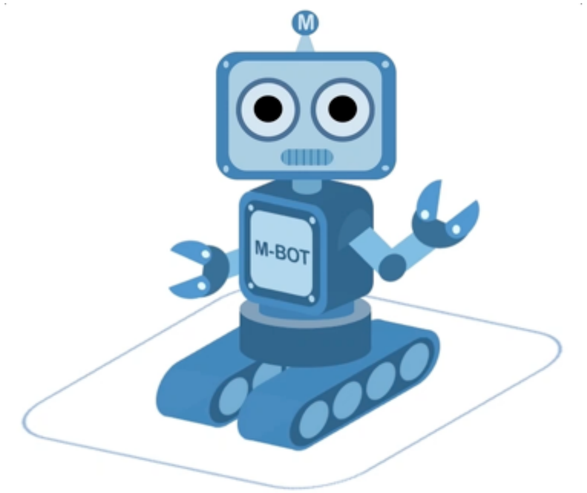
Agent and Model Design



Agent and Model Design

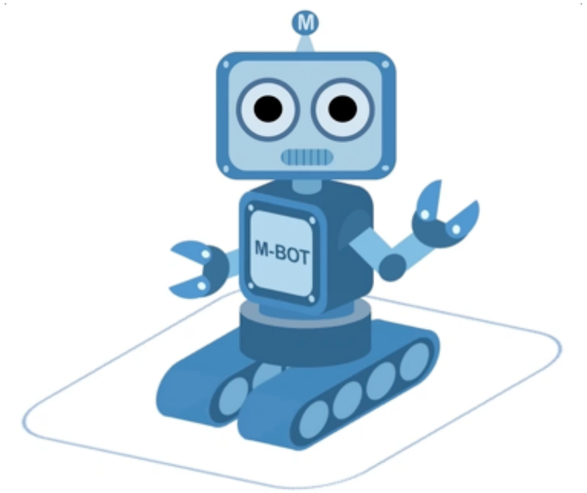


Agent and Model Design



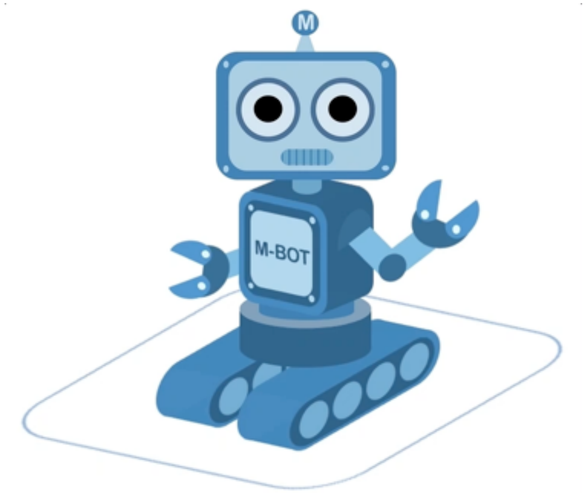
- How do we train this agent?

Agent and Model Design



- How do we train this agent?
- Both actions (they are discrete) and the simulation are non-differential-able

Agent and Model Design



- How do we train this agent?
- Both actions (they are discrete) and the simulation are non-differential-able
- Use reinforcement learning!

Learning vs SLAM

To Learn or Not to Learn: Analyzing the Role of Learning for Navigation in Virtual Environments

Noriyuki Kojima
University of Michigan
2260 Hayward St, Ann Arbor, MI 48109
kojimano@umich.edu

Jia Deng
Princeton University
35 Olden St 423, Princeton, NJ 08540
jiadeng@cs.princeton.edu

Abstract

In this paper we focus on the task of geometric navigation, i.e. navigation when ground-truth 3D information is available. Specifically, we explore the dichotomy between "learning" and "coding" for this task. We construct a hand-coded navigating agent, and demonstrate that it outperforms state-of-the-art learning based agents on two popular benchmarks, MINOS[37] and Stanford large-scale 3D Indoor Spaces (S3DIS)[2]. We also observe that as the environment becomes more challenging, the performance gap between learning-based and hand coded-agent increases.

ods. Therefore, in the context of geometric navigation, the strengths and weaknesses of "learning" over "coding" are not clear. In this paper, we attempt to clarify this so that intelligent choices can be made while developing real-world systems.

We construct a hand-coded agent for the task of geometric navigation and compare its performance with state-of-the-art learning based methods on two challenging benchmarks: S3DIS [2] and MINOS [37]. On MINOS, the UNREAL agent [37] (which is based on deep reinforcement learning) and on S3DIS, the CMP agent [14] (which uses imitation learning to jointly train a mapper and plan-

Benchmarking Classic and Learned Navigation in Complex 3D Environments

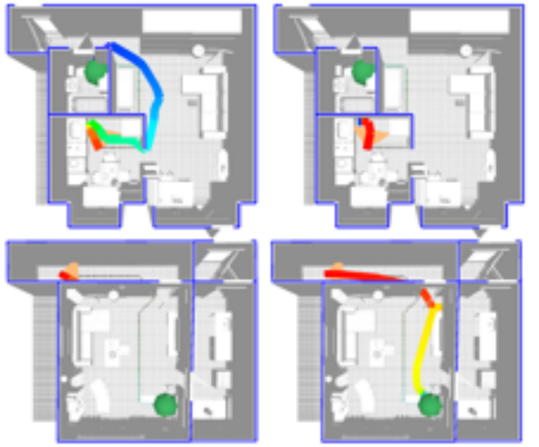
Dmytro Mishkin*
Czech Technical University

Alexey Dosovitskiy
Intel Labs

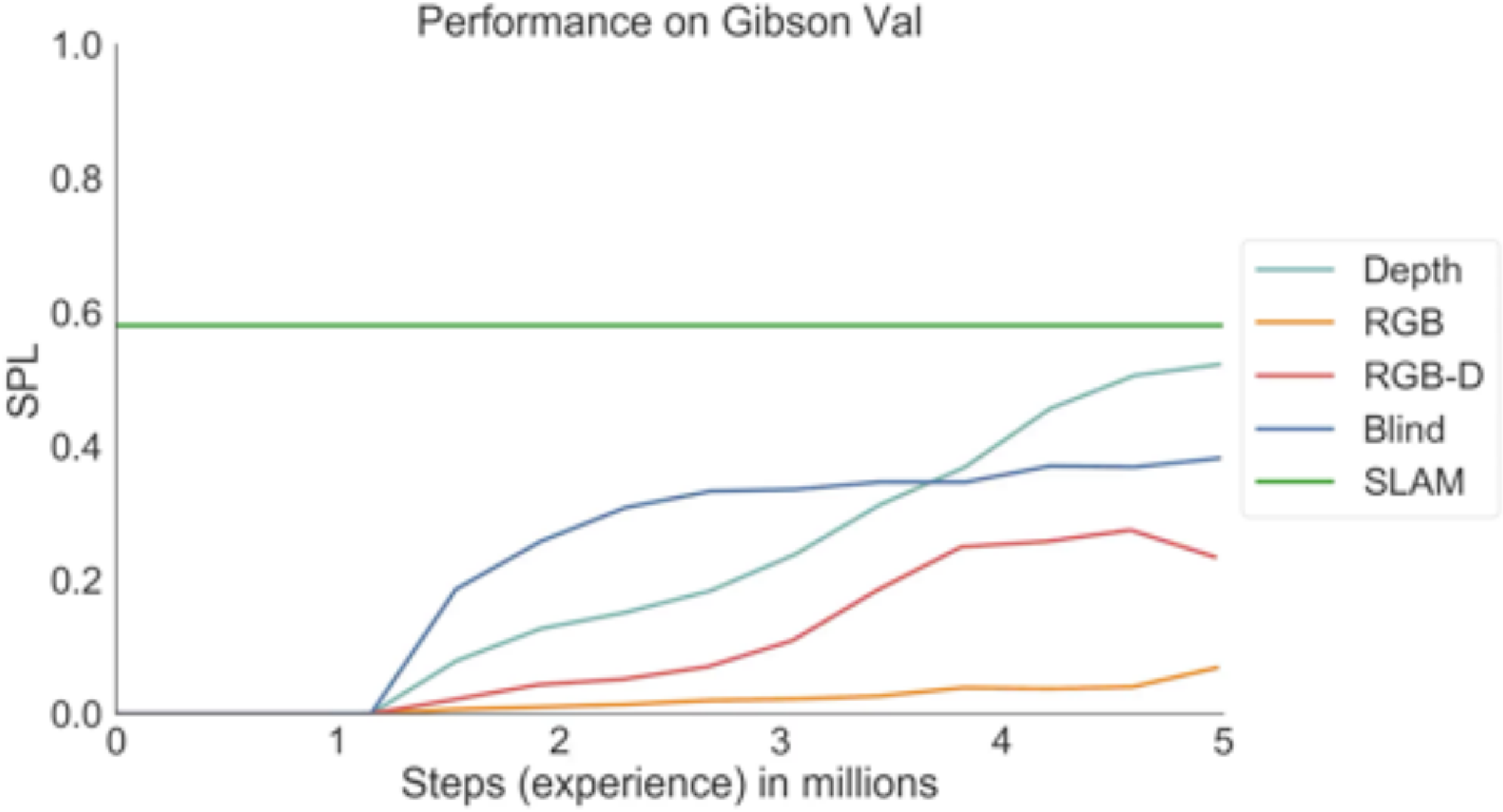
Vladlen Koltun
Intel Labs

Abstract

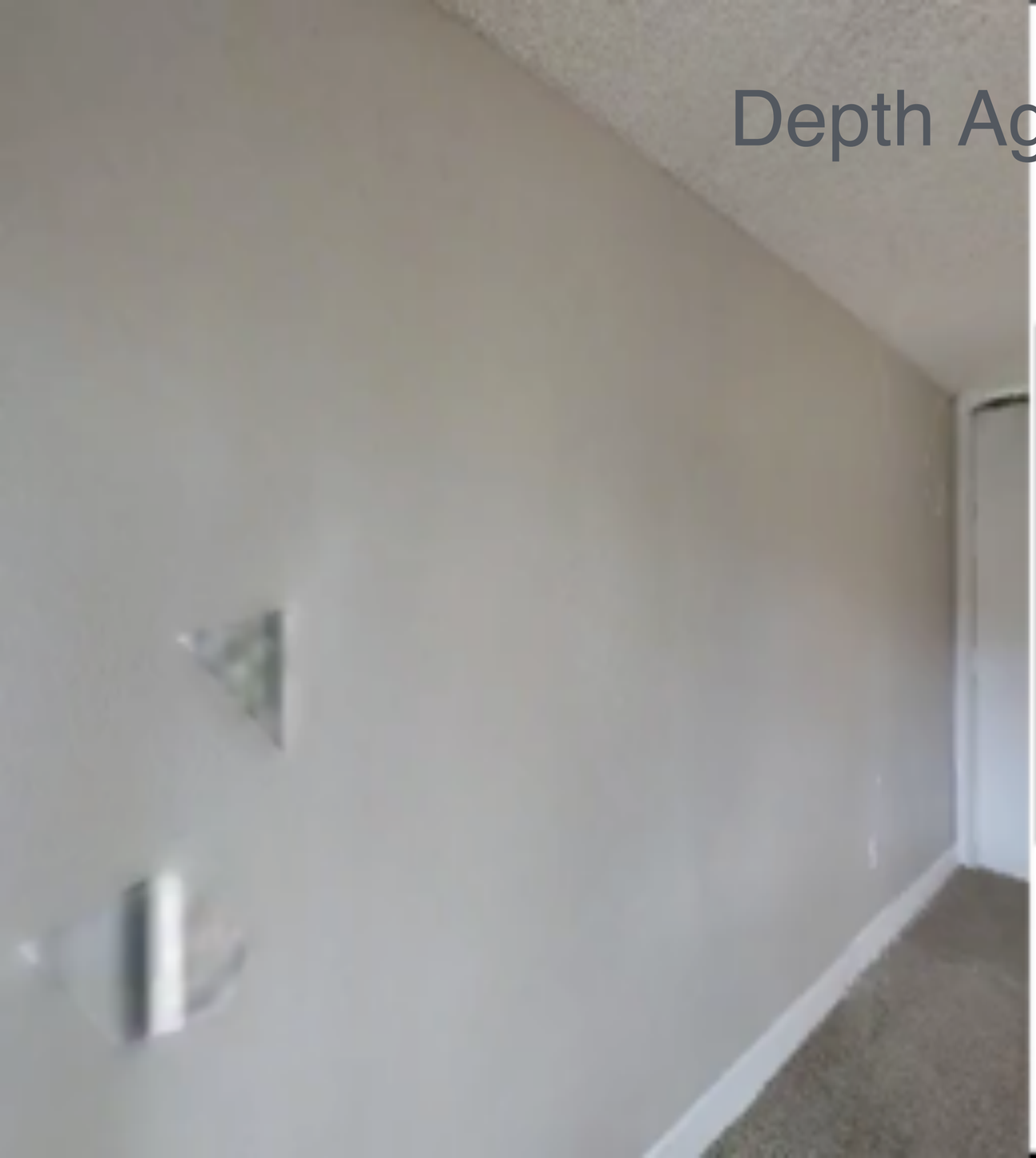
Navigation research is attracting renewed interest with the advent of learning-based methods. However, this new line of work is largely disconnected from well-established classic navigation approaches. In this paper, we take a step towards coordinating these two directions of research. We set up classic and learning-based navigation systems in common simulated environments and thoroughly evaluate them in indoor spaces of varying complexity, with access to different sensory modalities. Additionally, we measure human performance in the same environments. We find that a classic pipeline, when properly tuned, can perform very well in complex cluttered environments. On the other hand, learned systems can operate more robustly with a limited sensor suite. Both approaches are still far from human-level performance.



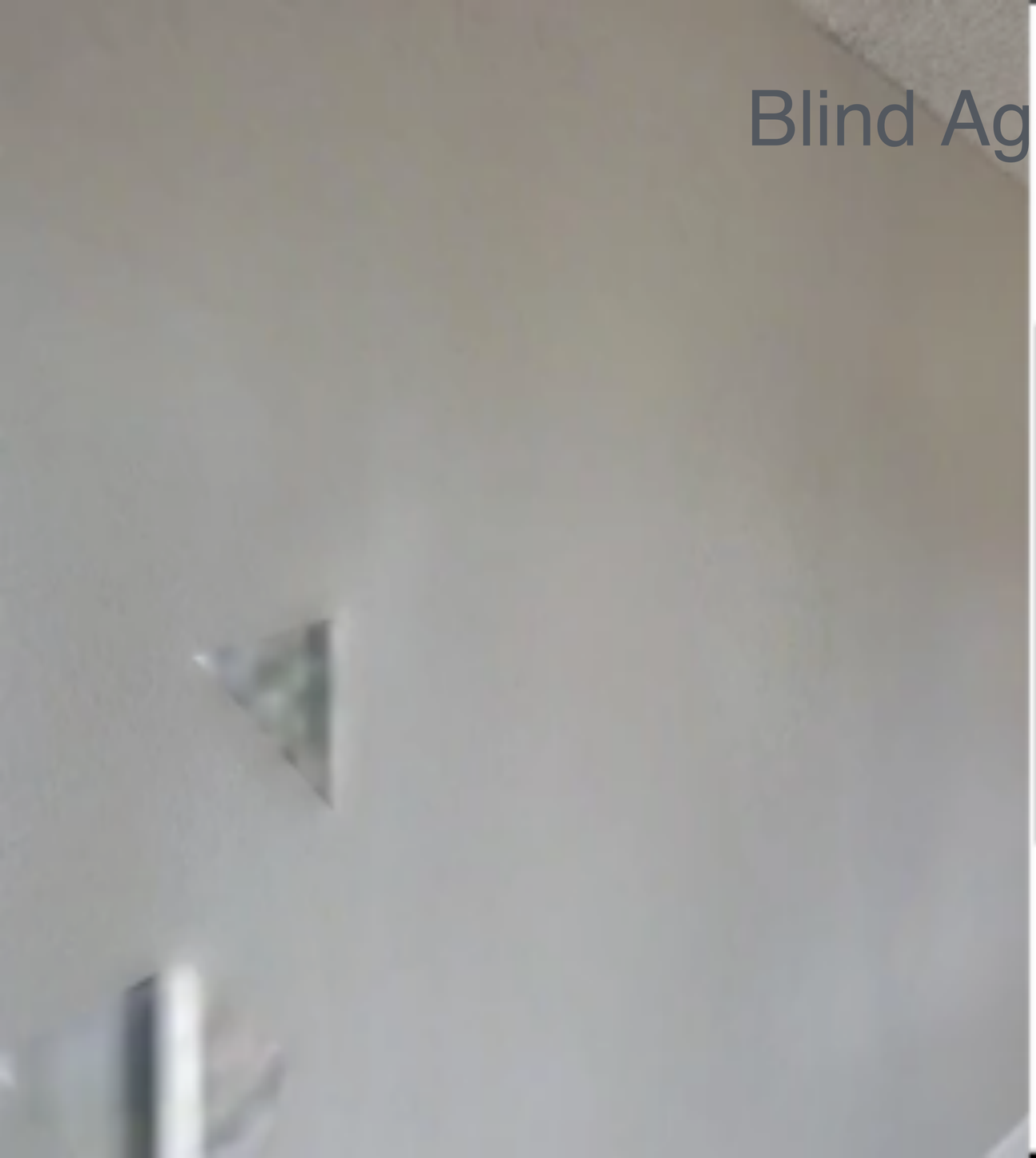
Classic pipeline Learned agent


Performance on Gibson Val
1.0
0.8
0.6
0.4
0.2
0.0
SPL
0
1
2
3
4
5
Steps (experience) in millions
Depth
RGB
RGB-D
Blind
SLAM

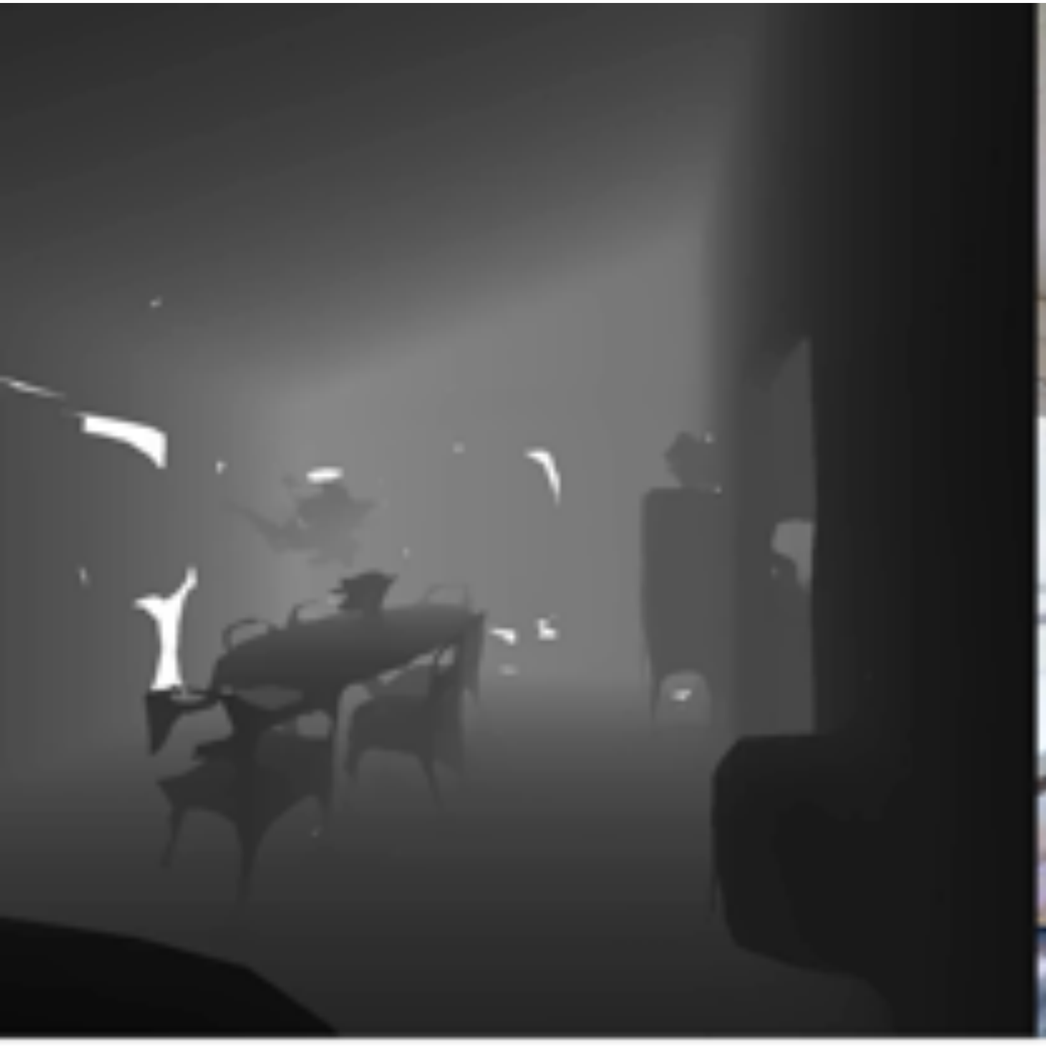
Depth Agent (RL)



Blind Agent (RL)



Depth Agent (RL)



Depth

RGB and GPS+Compass

Top Down Map

Depth RGB and GPS+Compass Top Down Map

The agent must decide between left, right, and straight at the end of the kitchen

Depth RGB and GPS+Compass Top Down Map

Goal Sensor (GPS+Compass) indicates straight

Depth

RGB and GPS+Compass

Top Down Map

However, it can see there is wall straight

and a wall on the left

Depth RGB and GPS+Compass Top Down Map

It correctly predicts that right is the direction to pursue

Depth

RGB and GPS+Compass

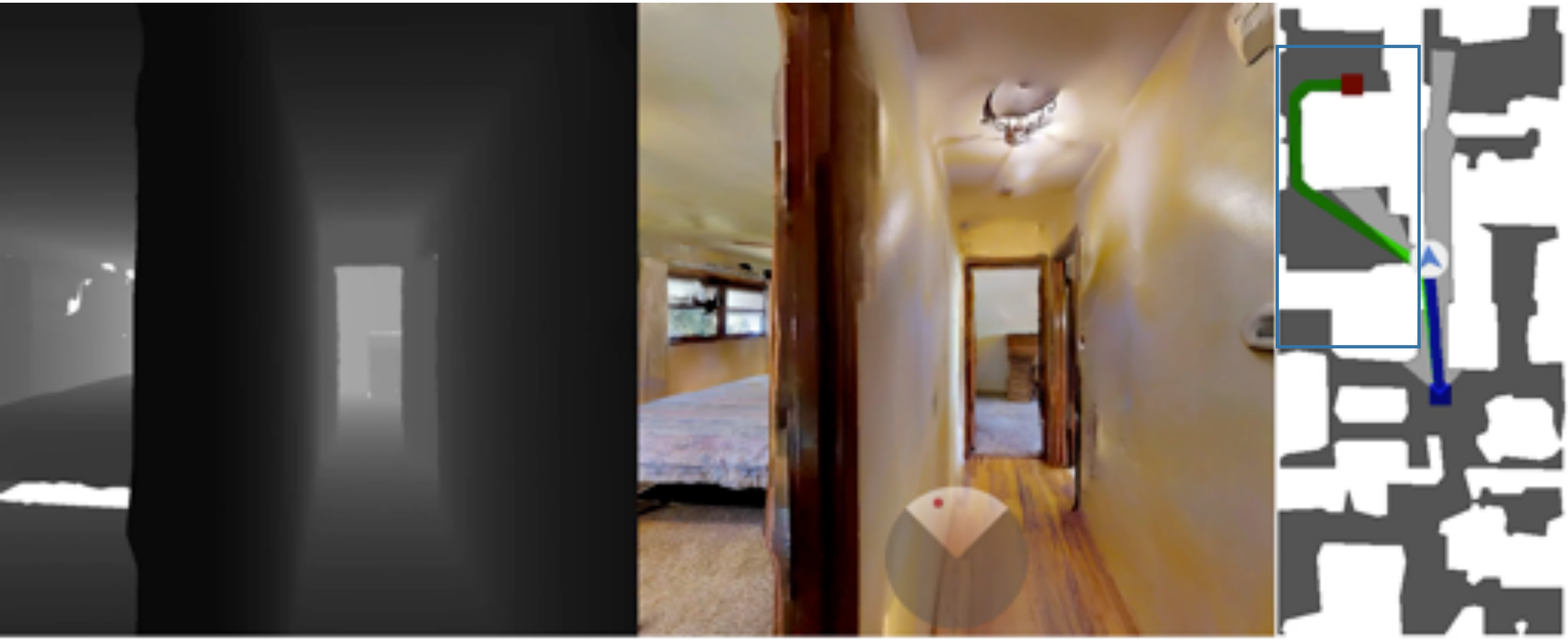
Top Down Map

Backtracking

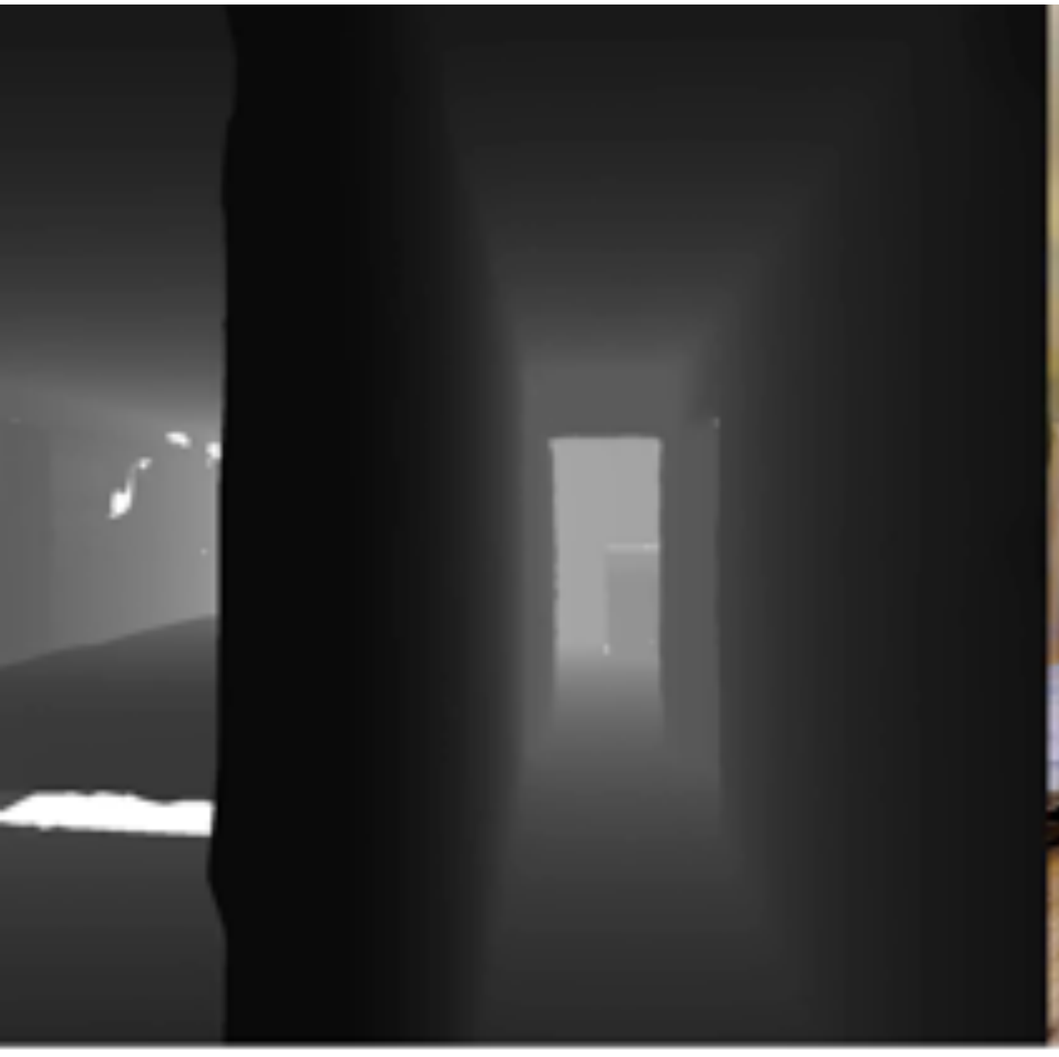
Depth

RGB and GPS+Compass

Top Down Map

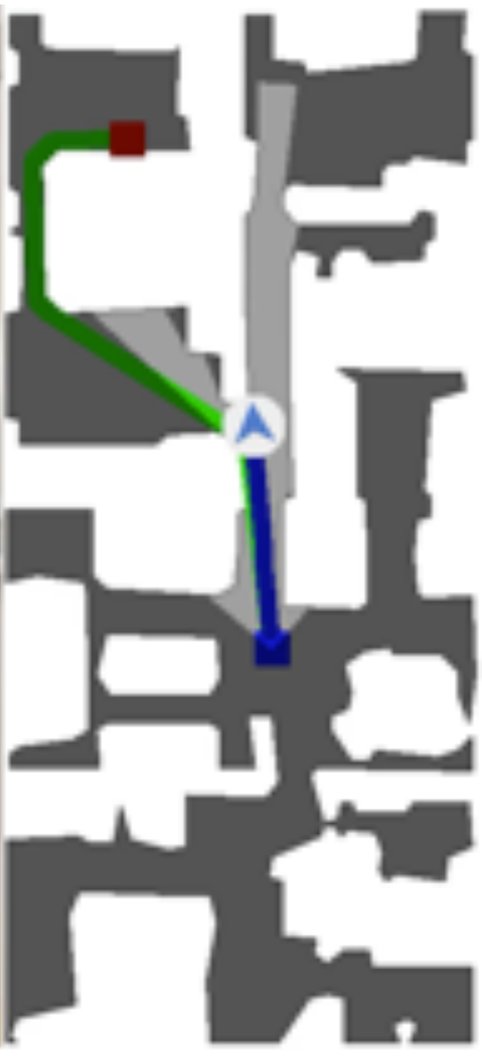


Depth RGB and GPS+Compass Top Down Map

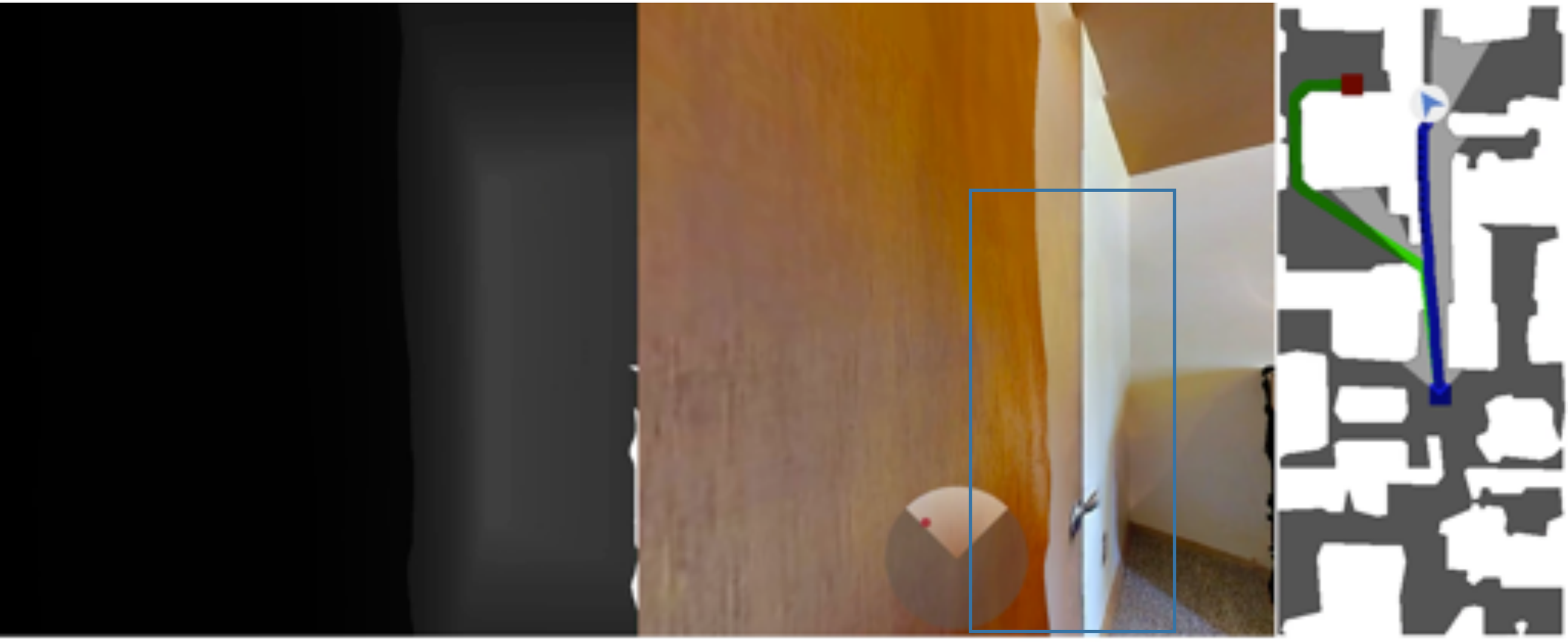


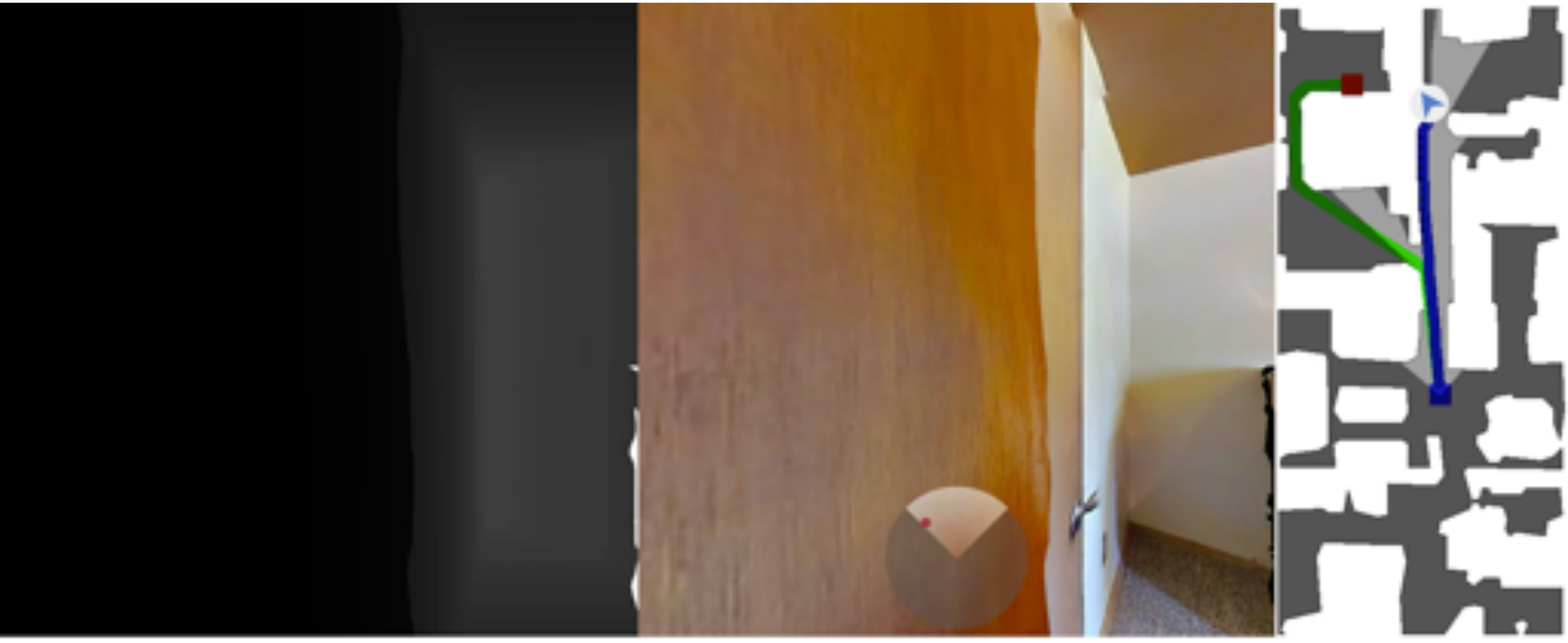
Depth

RGB and GPS+Compass



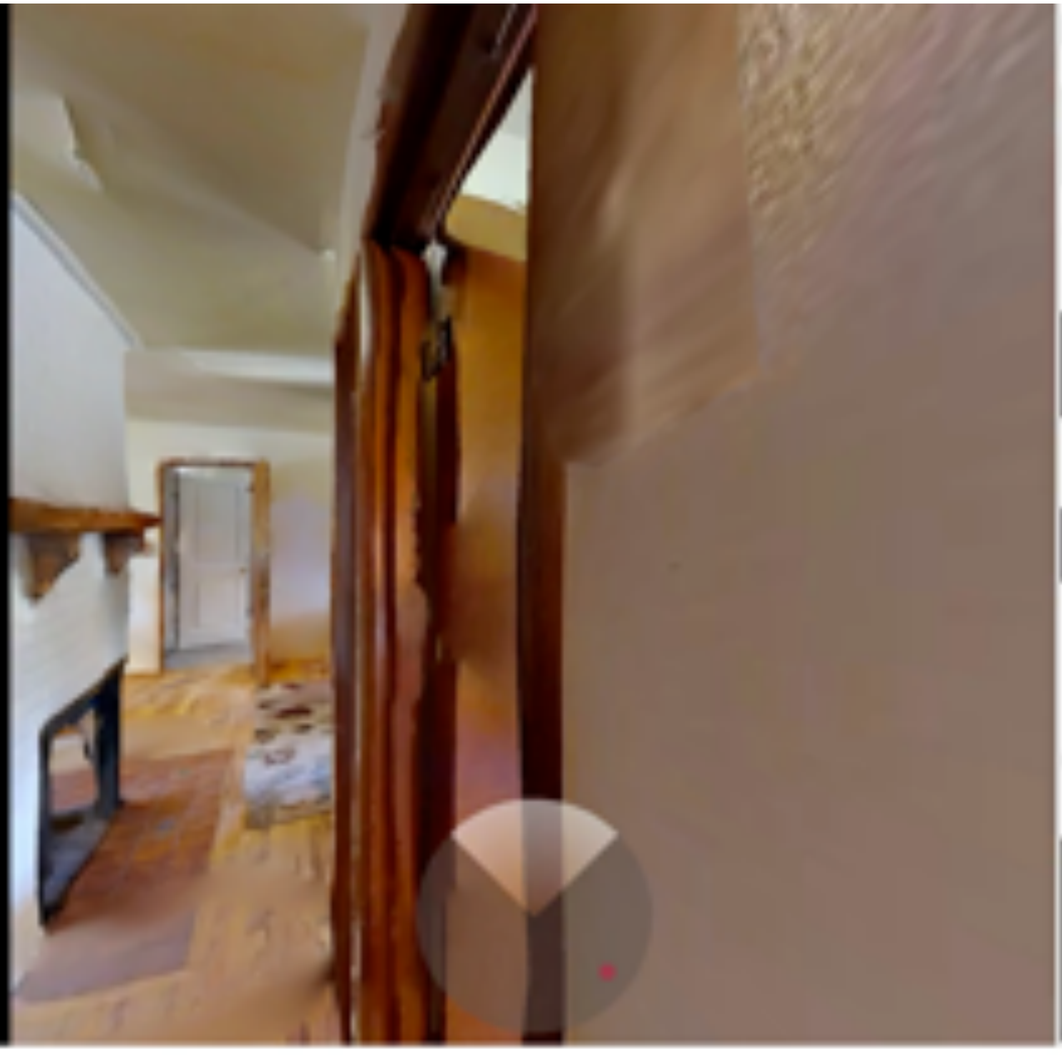
Top Down Map


Depth
RGB and GPS+Compass
Top Down Map

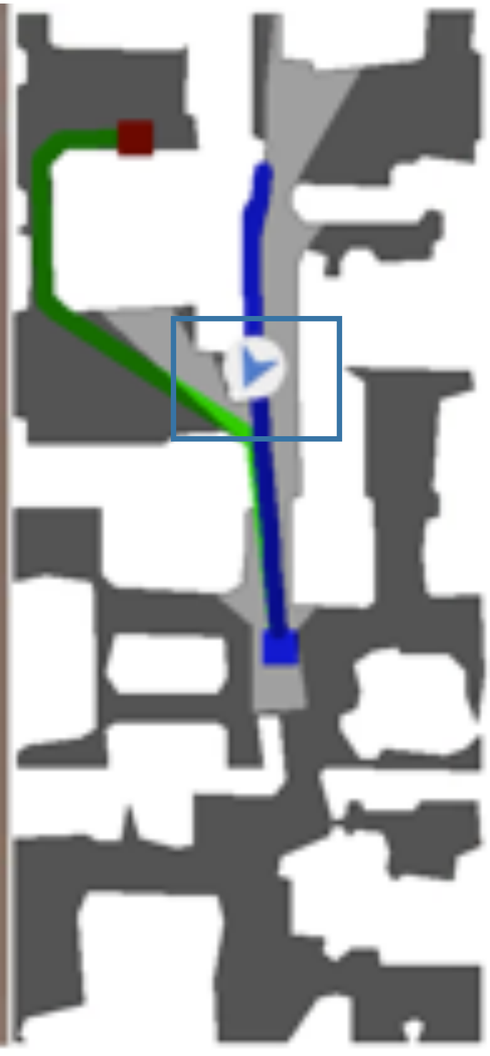


Depth RGB and GPS+Compass Top Down Map

Depth

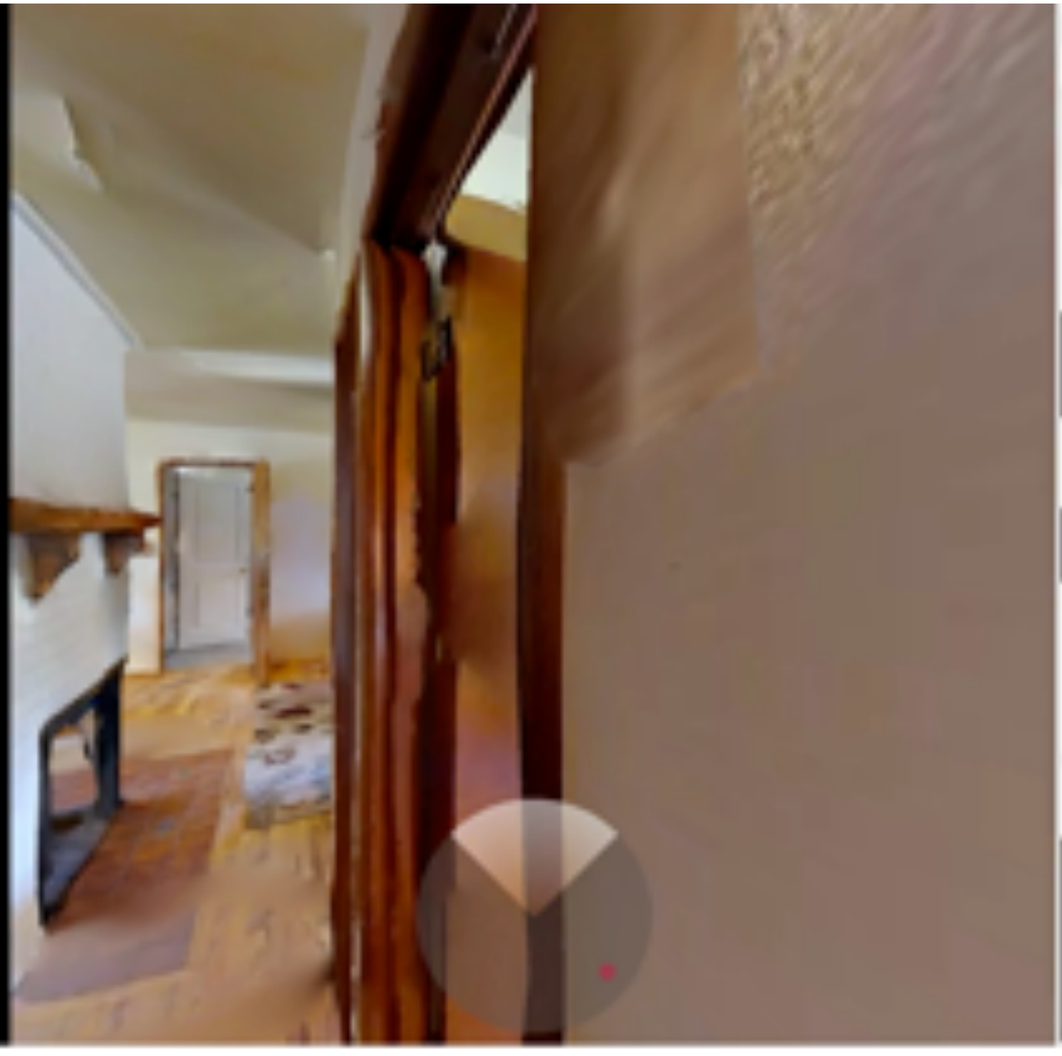


RGB and GPS+Compass

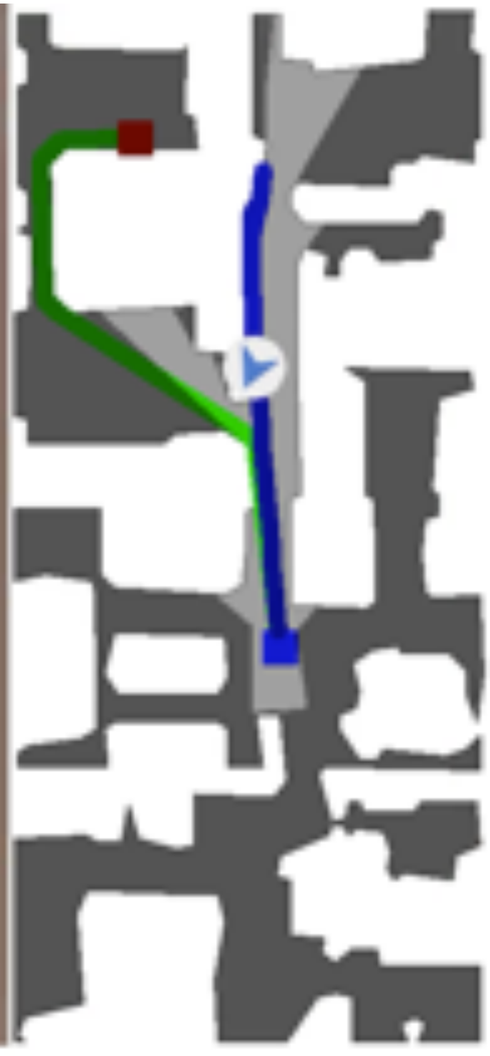


Top Down Map

Depth



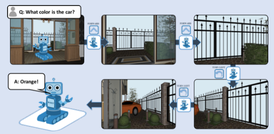
RGB and GPS+Compass



Top Down Map

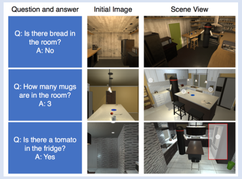
Standardizing the Embodied Agent Stack

Tasks



EmbodiedQA
(Das et al., 2018)

Language grounding
(Hill et al., 2017)



Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu et al., 2017, Gupta et al., 2017)

Abhishek Kadian
(FAIR)

Oleksandr Maksymets
(FAIR)

Simulators

House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

MINOS
(Savva et al., 2017)

Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

Datasets

SUNCG (Song et al., 2017)

Matterport3D (Chang et al., 2017)

2D-3D-S (Armeni et al., 2017)

Habitat-API

- Modular high-level Python library
- Easy to define virtual robot configurations
- Multiple Embodied AI tasks
  - PointGoal, ObjectGoal, VLN, EmbodiedQA
- Baselines: Classical Robotics (SLAM),
  Imitation and Reinforcement Learning


Object Navigation
"Go to toilet"

```python
import habitat

# Load embodied AI task (PointNav) and a pre-s
env = habitat.Env(
    config=habitat.get_config(config_file="tas
)

# Step through environment with random actions
while not env.episode_over:
    observations = env.step(env.action_space.s
```

```yaml
ENVIRONMENT:
  MAX_EPISODE_STEPS: 500
SIMULATOR:
  AGENT_0:
    SENSORS: ['RGB_SENSOR']
  HABITAT_SIM_V0:
    GPU_DEVICE_ID: 0
  RGB_SENSOR:
    WIDTH: 256
    HEIGHT: 256
  DEPTH_SENSOR:
    WIDTH: 256
    HEIGHT: 256
TASK:
  TYPE: Nav-v0
  SUCCESS_DISTANCE: 0.2
  SENSORS: ['POINTGOAL_SENSOR']
  POINTGOAL_SENSOR:
    TYPE: PointGoalSensor
    GOAL_FORMAT: POLAR
  MEASUREMENTS: ['SPL']
  SPL:
    TYPE: SPL
    SUCCESS_DISTANCE: 0.2
```

Habitat Challenge and Workshop @ CVPR ‘19

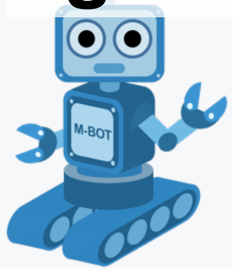
PointGoal Navigation: Go to (x,y)

Slide credit: Abhishek Kadian

Habitat Challenge and Workshop @ CVPR ‘19

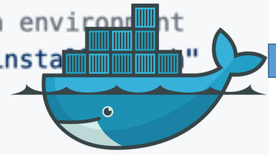
```python
import habitat

class ForwardOnlyAgent(habitat.Agent):
    def reset(self):
        pass

    def act(self, observations):
        action = SIM_NAME_TO_ACTION[SimulatorActions.FO
        return action

def main():
    agent = ForwardOnlyAgent()
    challenge = habitat.Challenge()
    challenge.submit(agent)

if __name__ == "__main__":
    main()
```

Agent



Rank	Participant team		SPL
1	Map and Plan Baseline		0.84
2	habitat (Depth-RL-PPO)	B	0.58
3	hela-ppo-baseline		0.52
4	CHROMA		0.47
5	Pansy		0.42
6	habitat (Blind-RL-PPO)	B	0.41

```
FROM fairembodied/habitat-challenge:latest

# install dependencies in the habitat conda environment
RUN /bin/bash -c ". activate habitat; pip insta
"

ADD myagent /myagent
ADD submission.sh /submission.sh
```



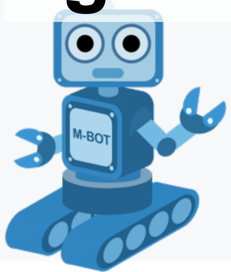

```
# Installing EvalAI Command Line Interface
pip install "evalai>=1.2.3"

# Set EvalAI account token
evalai set_token <your EvalAI participant token>

# Push docker image to EvalAI docker registry
evalai push my_submission:latest --phase <phase-name>
```

EvalAI

Habitat Challenge and Workshop @ CVPR ‘19

```python
import habitat

class ForwardOnlyAgent(habitat.Agent):
    def reset(self):
        pass

    def act(self, observations):
        action = SIM_NAME_TO_ACTION[SimulatorActions.FORWARD.value]
        return action

def main():
    agent = ForwardOnlyAgent()
    challenge = habitat.Challenge()
    challenge.submit(agent)

if __name__ == "__main__":
    main()
```

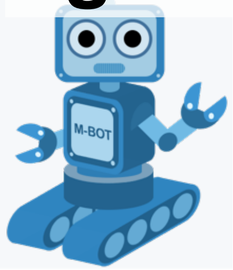
Agent



Habitat Challenge and Workshop @ CVPR ‘19

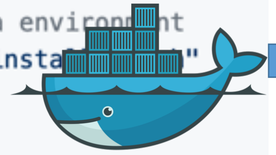
```python
import habitat

class ForwardOnlyAgent(habitat.Agent):
    def reset(self):
        pass

    def act(self, observations):
        action = SIM_NAME_TO_ACTION[SimulatorActions.FO
        return action

def main():
    agent = ForwardOnlyAgent()
    challenge = habitat.Challenge()
    challenge.submit(agent)

if __name__ == "__main__":
    main()
```

Agent



Rank	Participant team		SPL
1	Map and Plan Baseline		0.84
2	habitat (Depth-RL-PPO)	B	0.58
3	hela-ppo-baseline		0.52
4	CHROMA		0.47
5	Pansy		0.42
6	habitat (Blind-RL-PPO)	B	0.41

```
FROM fairembodied/habitat-challenge:latest

# install dependencies in the habitat conda environment
RUN /bin/bash -c ". activate habitat; pip insta
"

ADD myagent /myagent
ADD submission.sh /submission.sh
```

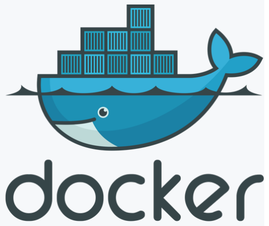



```
# Installing EvalAI Command Line Interface
pip install "evalai>=1.2.3"

# Set EvalAI account token
evalai set_token <your EvalAI participant token>

# Push docker image to EvalAI docker registry
evalai push my_submission:latest --phase <phase-name>
```

EvalAI

Habitat Challenge and Workshop @ CVPR ‘19

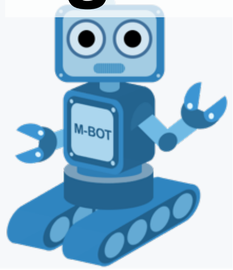
```
FROM fairembodied/habitat-challenge:latest

# install dependencies in the habitat conda environment
RUN /bin/bash -c ". activate habitat; pip install torch"

ADD myagent /myagent
ADD submission.sh /submission.sh
```



Habitat Challenge and Workshop @ CVPR ‘19

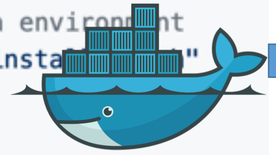
```python
import habitat

class ForwardOnlyAgent(habitat.Agent):
    def reset(self):
        pass

    def act(self, observations):
        action = SIM_NAME_TO_ACTION[SimulatorActions.FO
        return action

def main():
    agent = ForwardOnlyAgent()
    challenge = habitat.Challenge()
    challenge.submit(agent)

if __name__ == "__main__":
    main()
```

Agent



Rank	Participant team		SPL
1	Map and Plan Baseline		0.84
2	habitat (Depth-RL-PPO)	B	0.58
3	hela-ppo-baseline		0.52
4	CHROMA		0.47
5	Pansy		0.42
6	habitat (Blind-RL-PPO)	B	0.41

```
FROM fairembodied/habitat-challenge:latest

# install dependencies in the habitat conda environment
RUN /bin/bash -c ". activate habitat; pip insta
"

ADD myagent /myagent
ADD submission.sh /submission.sh
```




```
# Installing EvalAI Command Line Interface
pip install "evalai>=1.2.3"

# Set EvalAI account token
evalai set_token <your EvalAI participant token>

# Push docker image to EvalAI docker registry
evalai push my_submission:latest --phase <phase-name>
```

EvalAI

Habitat Challenge and Workshop @ CVPR ‘19

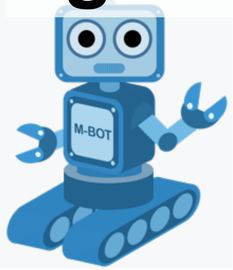
```
# Installing EvalAI Command Line Interface
pip install "evalai>=1.2.3"

# Set EvalAI account token
evalai set_token <your EvalAI participant token>

# Push docker image to EvalAI docker registry
evalai push my_submission:latest --phase <phase-name>
```

EvalAI

Habitat Challenge and Workshop @ CVPR ‘19

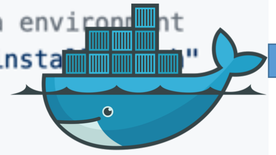
```python
import habitat

class ForwardOnlyAgent(habitat.Agent):
    def reset(self):
        pass

    def act(self, observations):
        action = SIM_NAME_TO_ACTION[SimulatorActions.FO
        return action

def main():
    agent = ForwardOnlyAgent()
    challenge = habitat.Challenge()
    challenge.submit(agent)

if __name__ == "__main__":
    main()
```

Agent



Rank	Participant team		SPL
1	Map and Plan Baseline		0.84
2	habitat (Depth-RL-PPO)	B	0.58
3	hela-ppo-baseline		0.52
4	CHROMA		0.47
5	Pansy		0.42
6	habitat (Blind-RL-PPO)	B	0.41

```
FROM fairembodied/habitat-challenge:latest

# install dependencies in the habitat conda environment
RUN /bin/bash -c ". activate habitat; pip insta
"

ADD myagent /myagent
ADD submission.sh /submission.sh
```




```
# Installing EvalAI Command Line Interface
pip install "evalai>=1.2.3"

# Set EvalAI account token
evalai set_token <your EvalAI participant token>

# Push docker image to EvalAI docker registry
evalai push my_submission:latest --phase <phase-name>
```

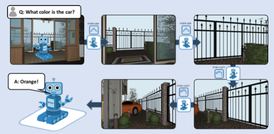
EvalAI

Habitat Challenge and Workshop @ CVPR ‘19

Rank	Team	SPL
1	Arnold	0.948
2	Pansy	0.927
3	titardrew	0.868
4	Hiccup	0.846
5	CHROMA	0.843
6	Mid-level-Features	0.815
7	ARF-RL	0.788
8	mkk1	0.738
9	cnmooc	0.693
10	hela-ppo-baseline	0.569

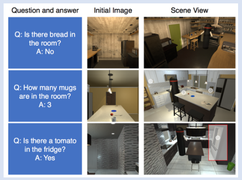
Standardizing the Embodied AI Stack

Tasks



EmbodiedQA
(Das et al., 2018)

Language grounding
(Hill et al., 2017)

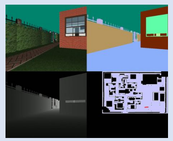


Interactive QA
(Gordon et al., 2018)

Vision-Language Navigation
(Anderson et al., 2018)

Visual Navigation
(Zhu et al., 2017, Gupta et al., 2017)

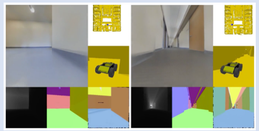
Simulators



House3D
(Wu et al., 2017)

AI2-THOR
(Kolve et al., 2017)

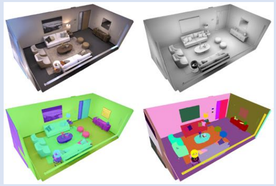
MINOS
(Savva et al., 2017)



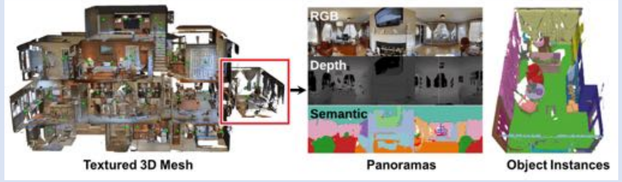
Gibson
(Zamir et al., 2018)

CHALET
(Yan et al., 2018)

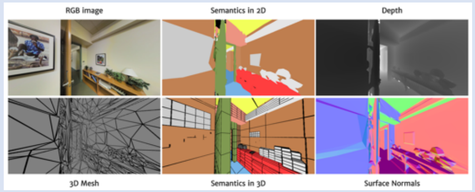
Datasets



Replica (Straub et al., 2019)



Matterport3D (Chang et al., 2017)



2D-3D-S (Armeni et al., 2017)

Habitat Platform

Habitat API

Habitat Sim

Generic Dataset Support

Our Vision

- Create the ImageNet/COCO/VQA of Embodied AI
  - Dataset → Simulator → Task → Benchmark Challenge

ICCV ‘19
[Best Paper Award Nominee]

Habitat: A Platform for Embodied AI Research

Manolis Savva[1,4]*, Abhishek Kadian[1]*, Oleksandr Maksymets[1]*, Yili Zhao[1], Erik Wijmans[1,2,3], Bhavana Jain[1], Julian Straub[2], Jia Liu[1], Vladlen Koltun[5], Jitendra Malik[1,6], Devi Parikh[1,3], Dhruv Batra[1,3]

[1]Facebook AI Research, [2]Facebook Reality Labs, [3]Georgia Institute of Technology,
[4]Simon Fraser University, [5]Intel Labs, [6]UC Berkeley
https://aihabitat.org

Abstract

We present Habitat, a new platform for research in embodied artificial intelligence (AI). Habitat enables training embodied agents (virtual robots) in highly efficient photorealistic 3D simulation, before transferring the learned skills to reality. Specifically, Habitat consists of the following –

1. Habitat-Sim: a flexible, high-performance 3D

1. Introduction

Imagine walking up to a home robot and asking *‘Hey – can you go check if my laptop is on my desk? And if so, bring it to me.’* In order to be successful, such a robot would need a range of skills – visual perception (to recognize scenes and objects), language understanding (to translate questions and instructions into actions), and navigation in complex environ-

Decentralized Distributed PPO: Mastering PointGoal Navigation

Erik Wijmans

Abhishek Kadian

Ari Morcos

Stefan Lee

Irfan Essa

Devi Parikh

Manolis Savva

Dhruv Batra

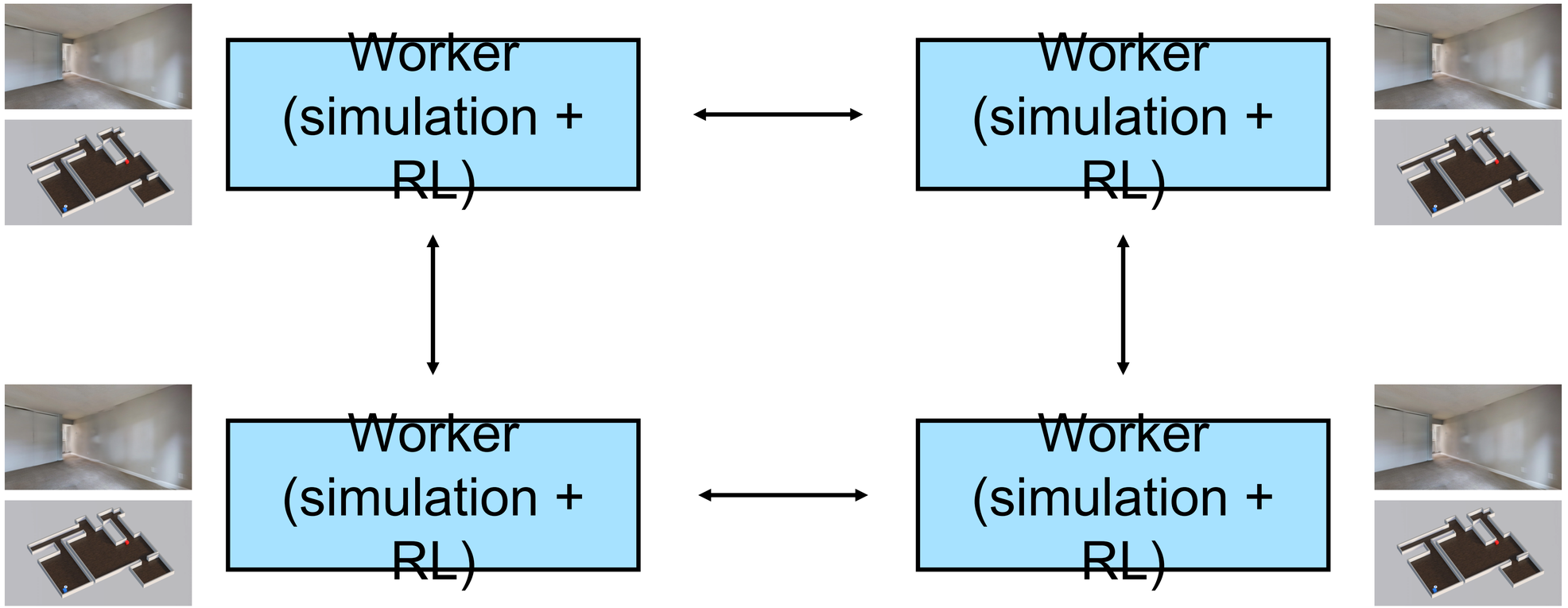
facebook
Artificial Intelligence Research

Georgia Tech

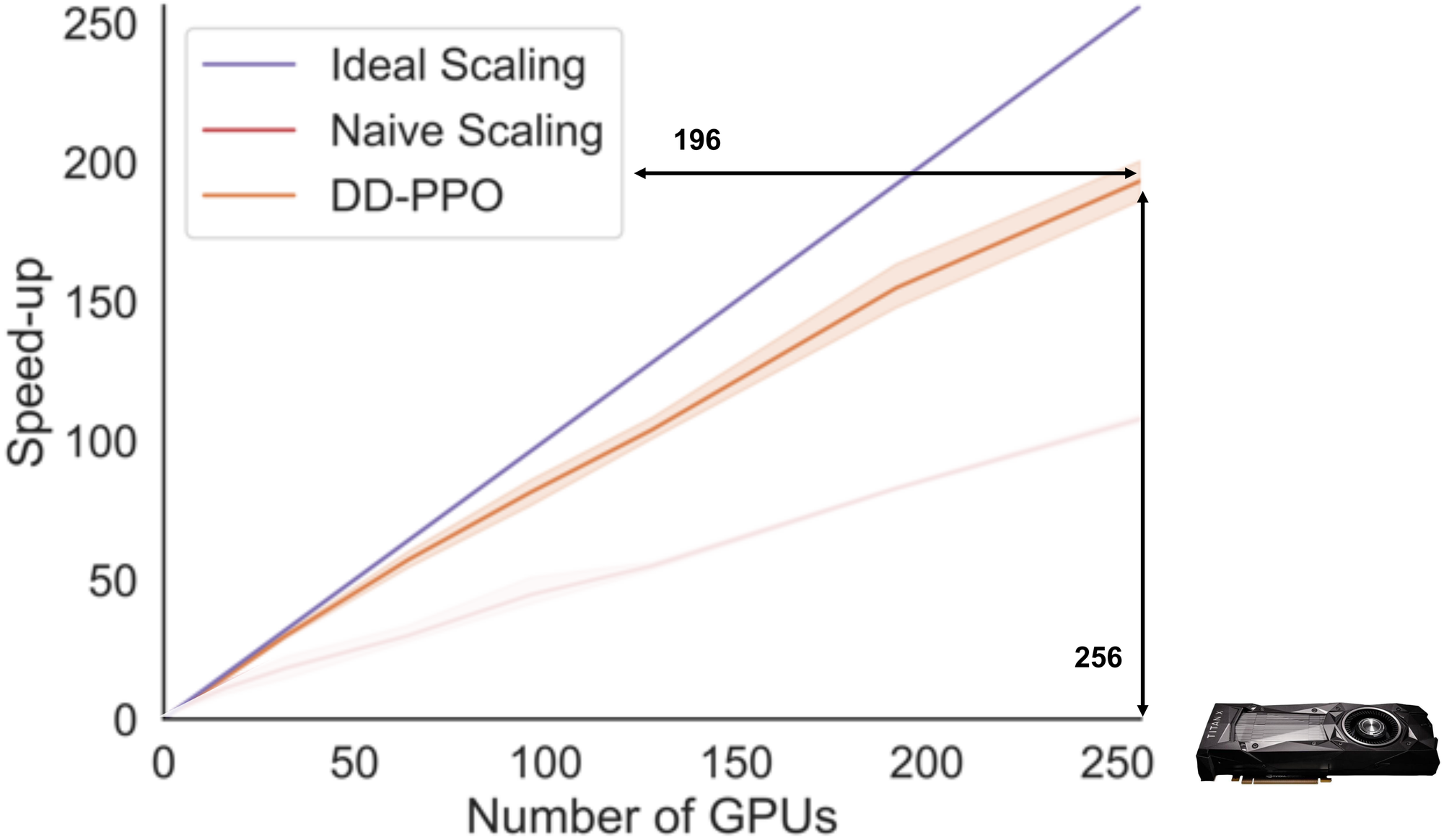
SFU

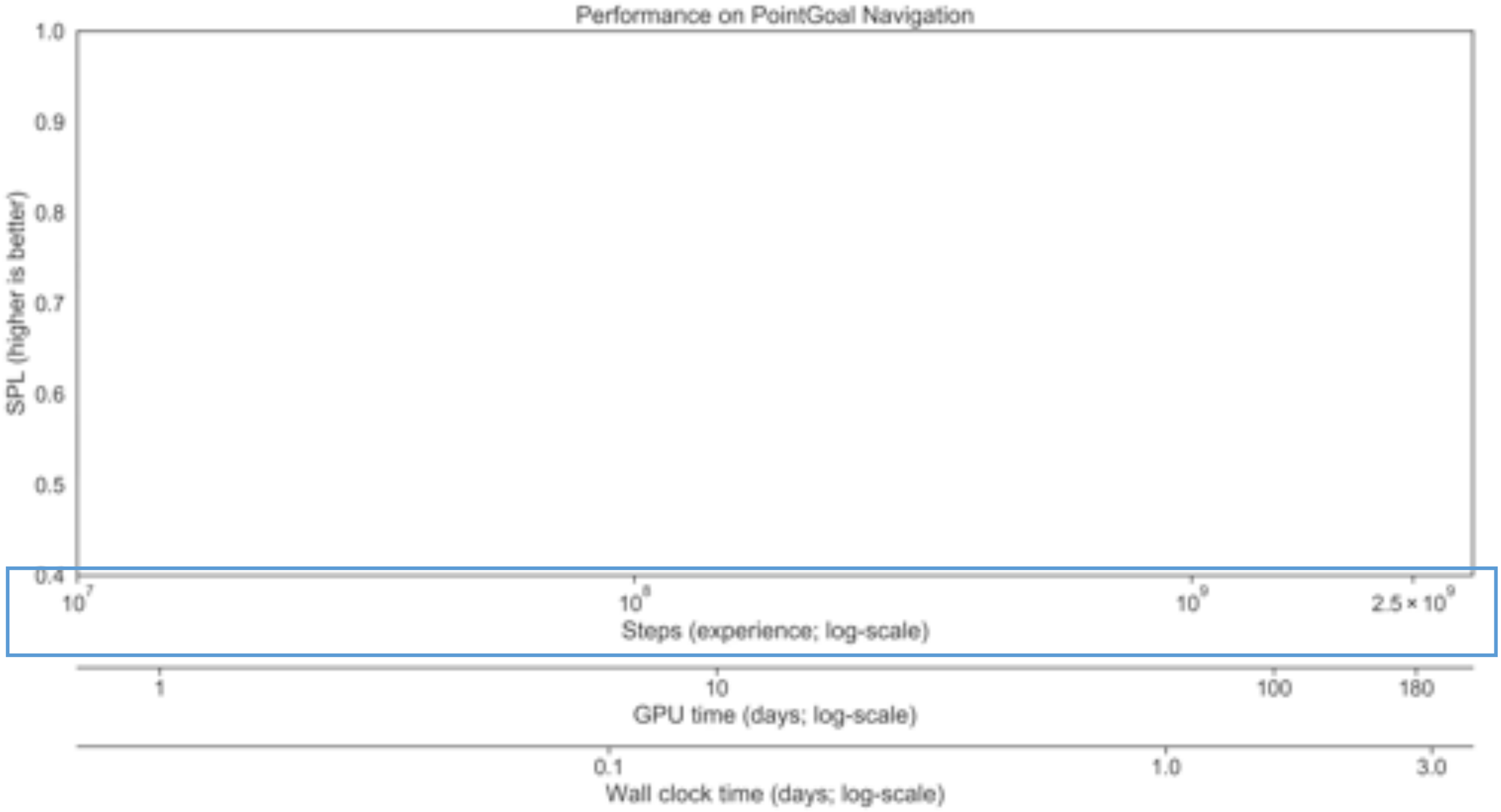
Oregon State University

Decentralized Distributed PPO

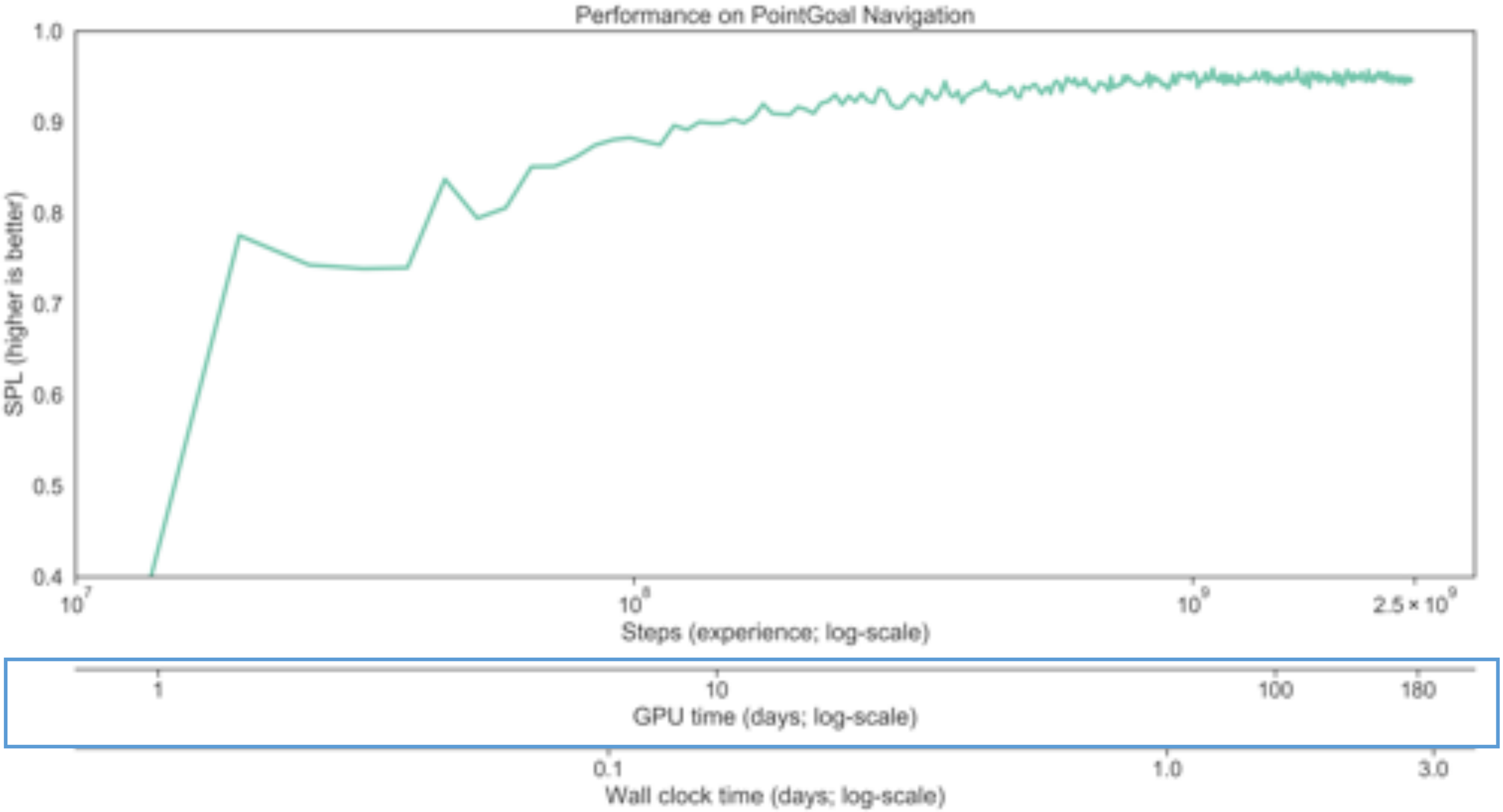


Decentralized Distributed PPO

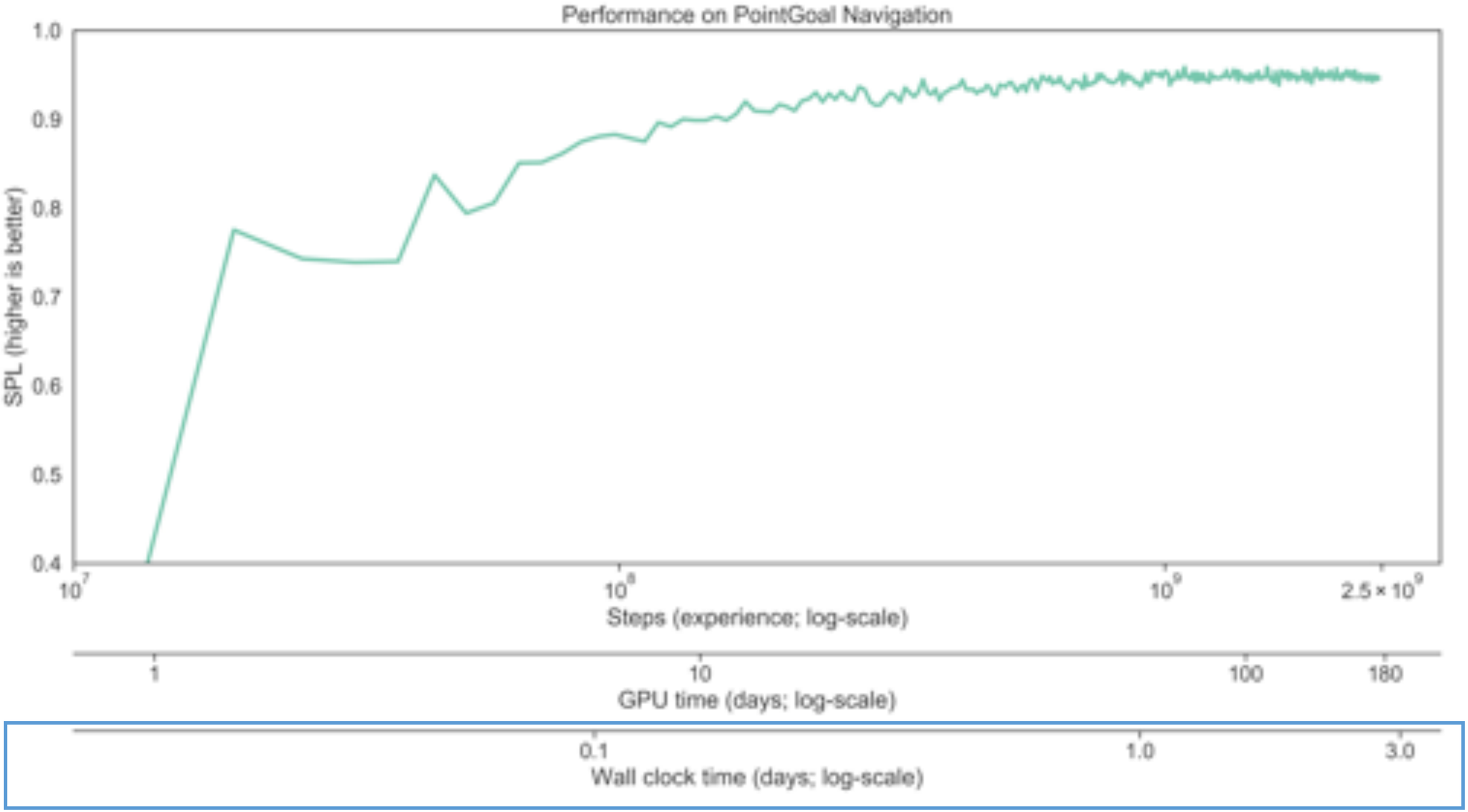




We utilize DD-PPO to train an agent for 2.5 Billion steps of experience



over 180 days of GPU-time training

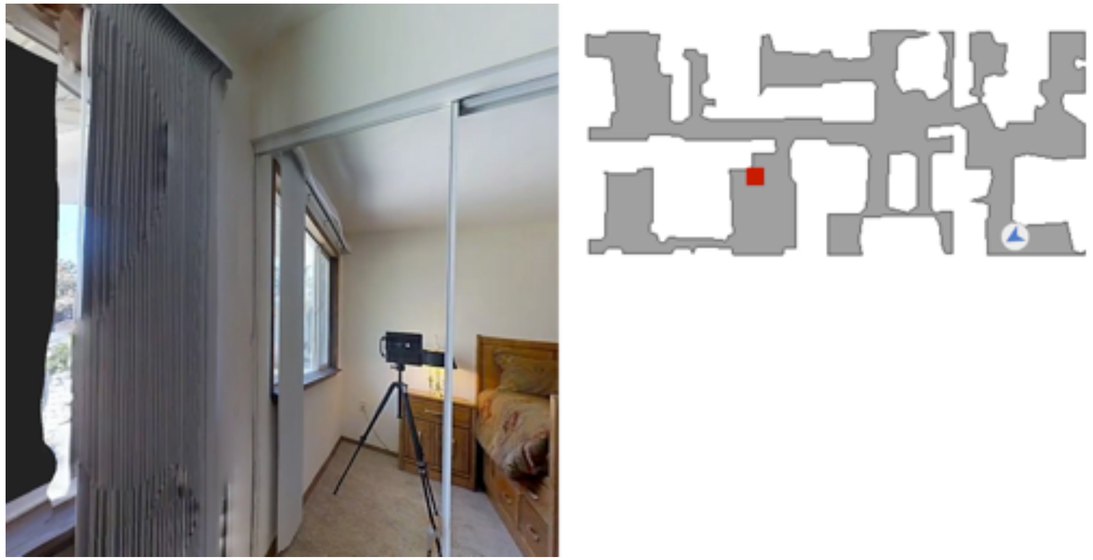


in under 3 days of wall-clock time

Visual Turing Test

Option 1

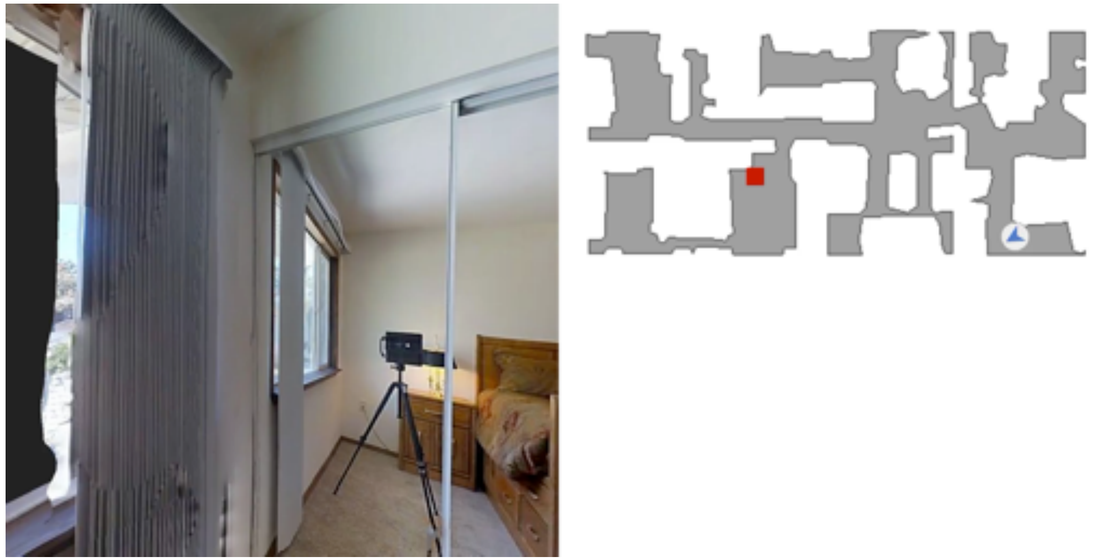
Option 2



Option 1

Option 1

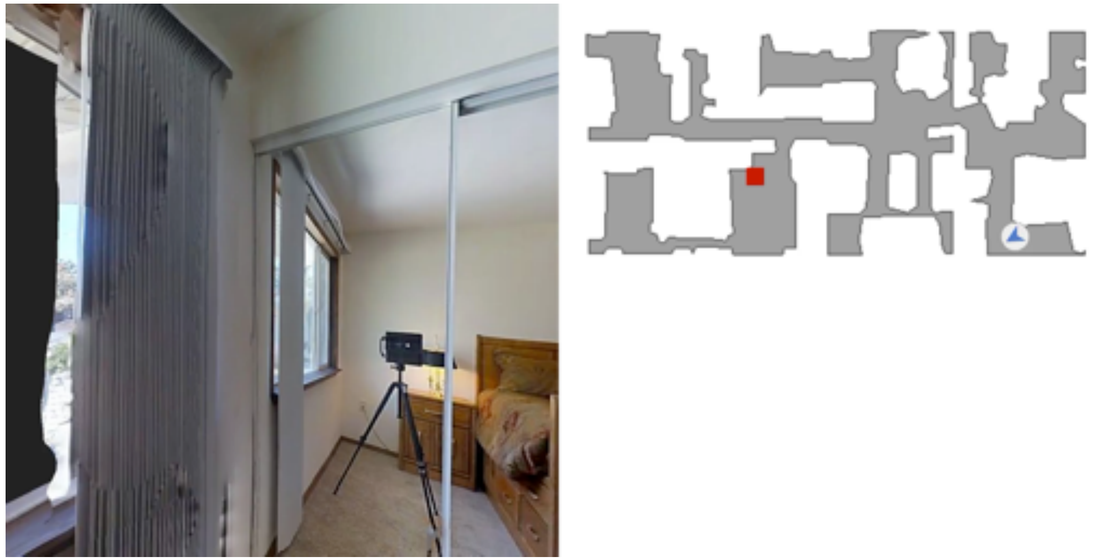
Option 2



Option 2

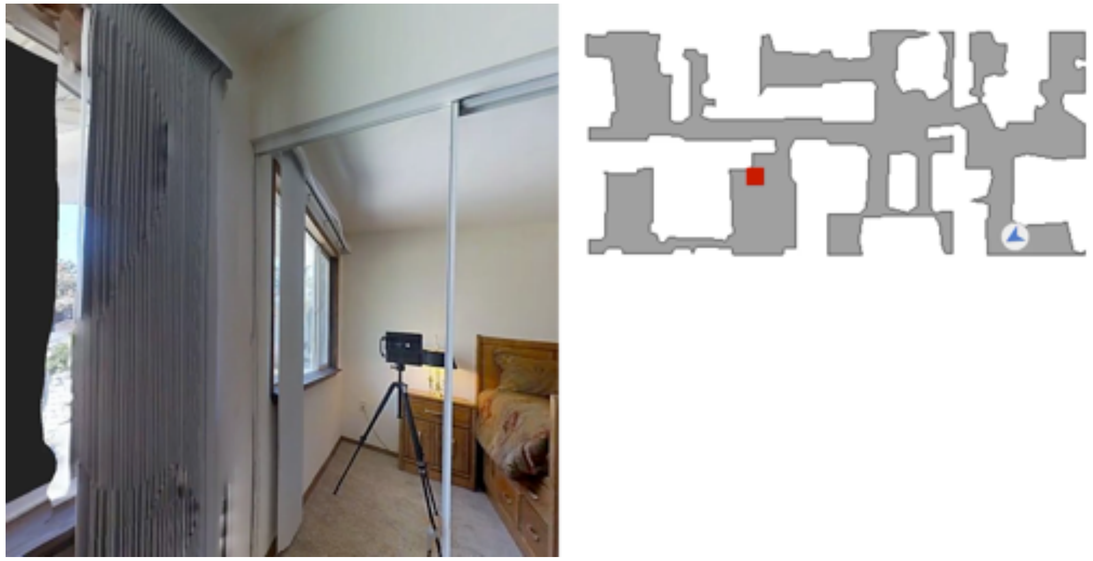
Option 1

Option 2



Learned Agent

Shortest Path Oracle



Are We Making Real Progress in Simulated Environments? Measuring the Sim2Real Gap in Embodied Visual Navigation

Abhishek Kadian*

Joanne Truong*

Aaron Gokasalan

Alex Clegg

Erik Wijmans

Manolis Savva

Stefan Lee

Sonia Chernova

Dhruv Batra

Does progress in simulation translate to progress on real robots?

(In the context of embodied navigation)


16x

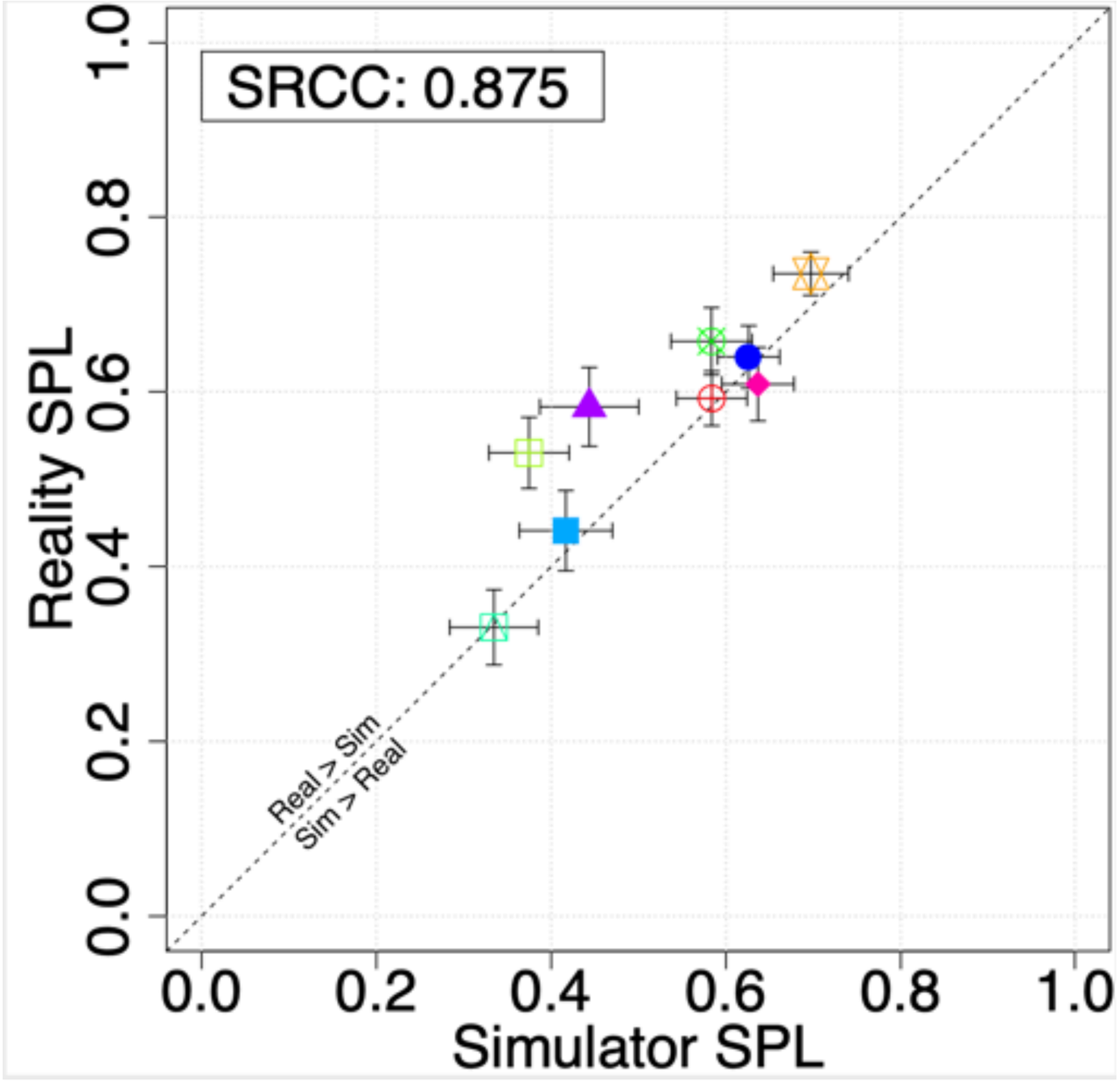
Georgia Tech CODA Building Scans

- https://my.matterport.com/show/?m=yZVvKaJZghh

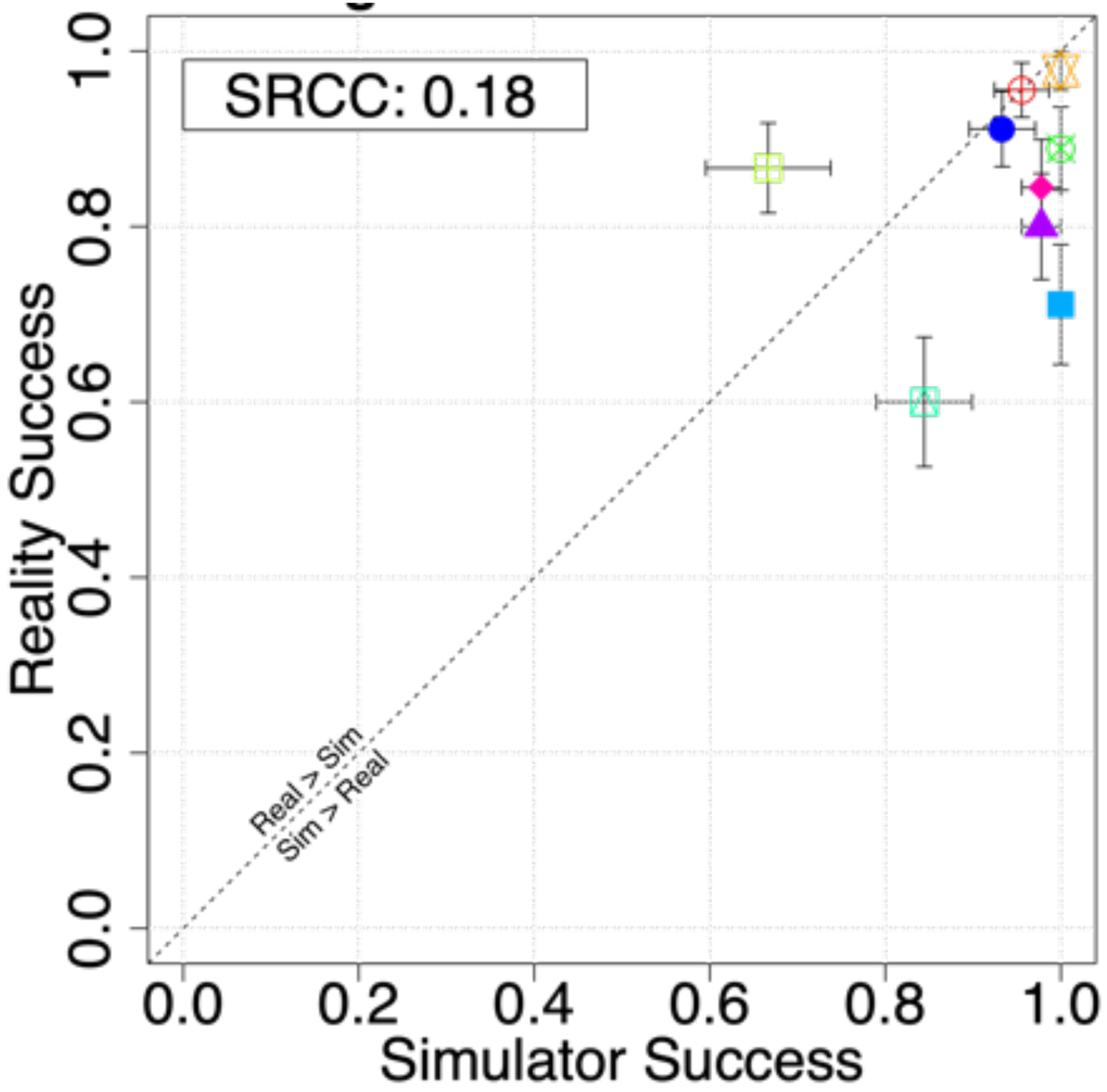

Goal
Spawn

```python
import habitat

# What is the Embodied AI task?
config = habitat.get_config("pointnav.yaml")

# Are we in sim or reality?
if args.use_simulation: # Use Habitat-Sim
    config.SIMULATOR.TYPE = "Habitat-Sim-v0"
else:                   # Use LoCoBot via PyRobot
    config.SIMULATOR.TYPE = "PyRobot-Locobot-v0"

# Create environment (sim or real doesn't matter)
env = habitat.Env(config)
observations = env.reset()

# Which model are we testing?
model = torch.load("my_model.pth")

# Let's act!
while not env.episode_over:
```

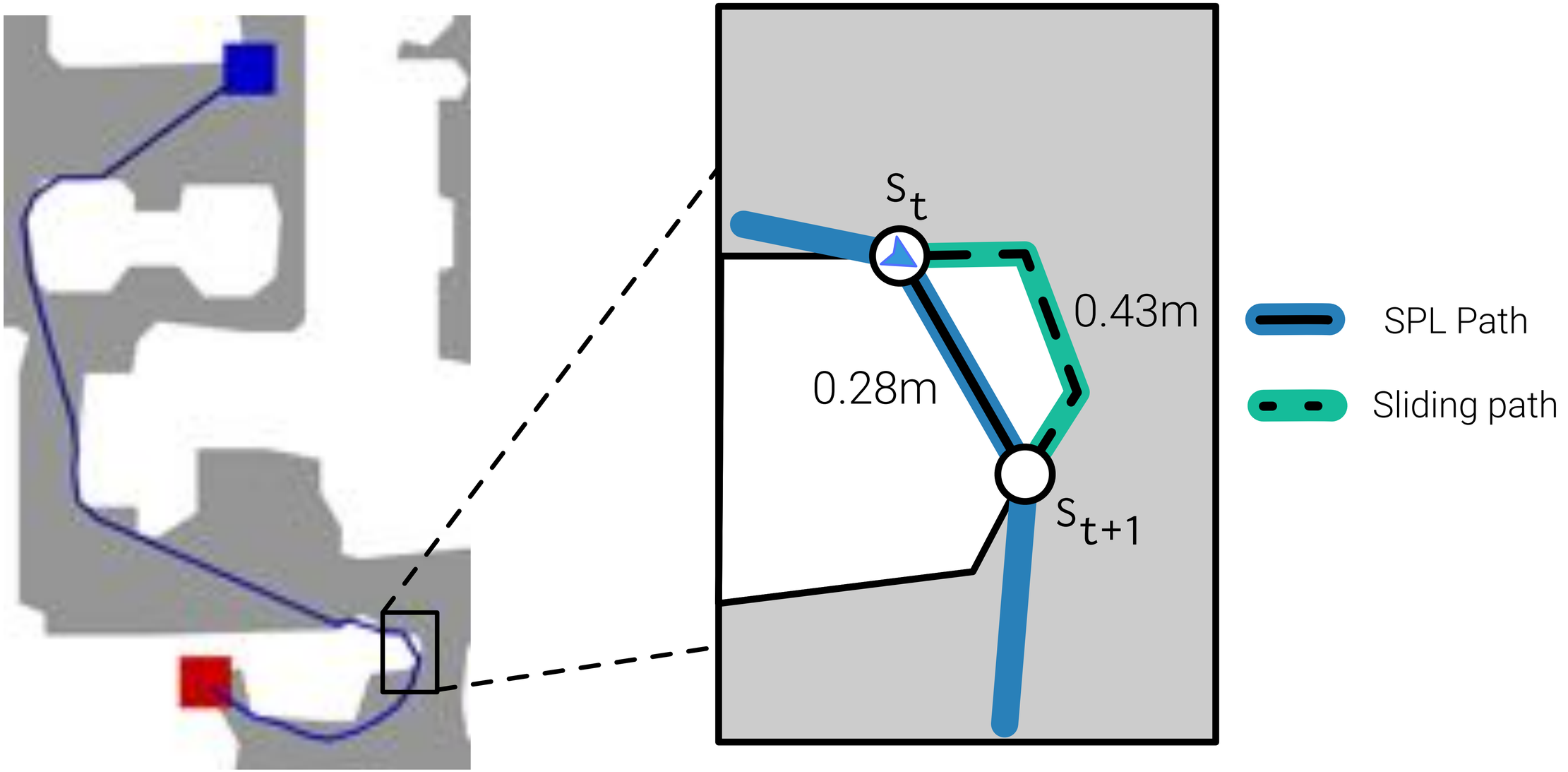
Sim-vs-Real Correlation Coefficient (SRCC)



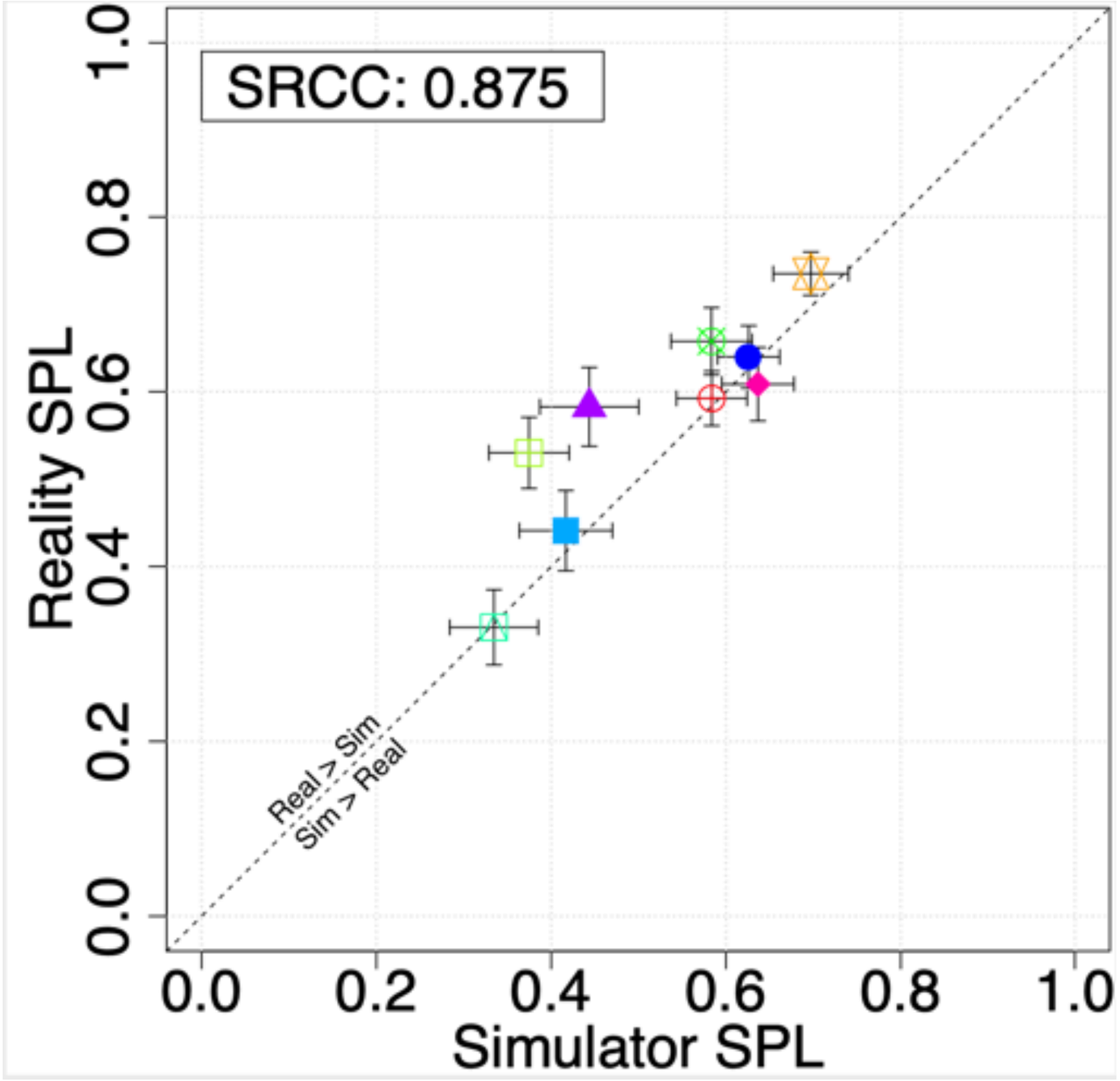
Sim-vs-Real Correlation Coefficient (SRCC)



Cheating by Sliding



Sim-vs-Real Correlation Coefficient (SRCC)



Import Objects

- Why?
- Egocentric CV
- Domain randomization

Physics

- Why?
- Intuitive physics
- Robotics, sim2real
- Egocentric CV

Habitat in Browser

- Why?
- Grounded Dialog via

  2-player data collection

- **Demo:**
  - **https://aihabitat.org/iccv2019-demo/**

Plans

- Full support for object interaction + physics
  - Physics is slow! Need to spend time optimizing.
  - Articulated robot integration (URDF)
- Humans-as-agents (Web + VR)
- CVPR20 Challenge
  - PointGoal Navigation w/ GPS+Compass
  - ObjectGoal Navigation

Habitat Core Team

Manolis Savva[1,4*]

Abhishek Kadian[1*]

Oleksandr Maksymets[1*]

Yili Zhao[1]

Erik Wijmans[1,2,3]

Bhavana Jain[1]

Julian Straub[2]

Jia Liu[1]

Vladlen Koltun[5]

Jitendra Malik[1,6]

Devi Parikh[1,3]

Dhruv Batra [1,3]

* denotes equal contribution

facebook Artificial Intelligence Research

1

facebook Reality Labs

2

Georgia Tech

3

SFU

intel

4

Berkeley UNIVERSITY OF CALIFORNIA

5

facebook
Artificial Intelligence Research

Georgia Gkioxari

Marcus Rohrbach

Saurabh Gupta

Amanpreeet Singh

Xinlei Chen

Leo Guibas

Or Litany

facebook
Reality Labs

Richard Newcombe

Michael Shvartsman

James Hillis

Naga Venkata Medathati

Steven Lovegrove

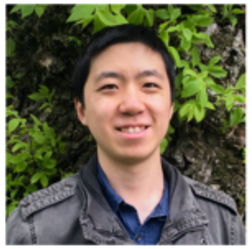
In collaboration with Devi Parikh's and my groups at Georgia Tech

Research Scientists

Stefan Lee

Peter Anderson

Postdoc



Zhile Ren

Georgia Tech

PhD Students

Aishwarya Agrawal
(2014 — Present)

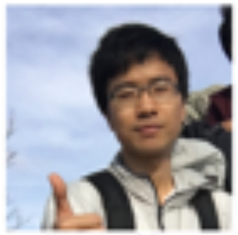
Yash Goyal
(2014 — Present)

Michael Cogswell
(2015 — Present)

Ramakrishna Vedantam
Ph.D. Student

Arjun Chandrasekaran
Ph.D. Student

Jiasen Lu
Ph.D. Student



Jianwei Yang
Ph.D. Student

Abhishek Das
(2016 — Present)

Ashwin Kalyan
(2016 — Present)

Nirbhay Modhe
(2017 — Present)

Samyak Datta
Ph.D. Student

Prithvijit Chattopadhyay
M.S. Student

Viraj Prabhu
M.S. Student

Ramprasaath Selvaraju
Ph.D. Student

Erik Wijmans
(2017 — Present)

Harsh Agrawal
(2018 — Present)

Deshraj Yadav
(2017 — Present)

MS


Thank you
Internet AI → Embodied AI