Reinforcement Learning for Visual Semantic Navigation

Universidad de Alcalá

PhD. Program in Information and Communication Technologies

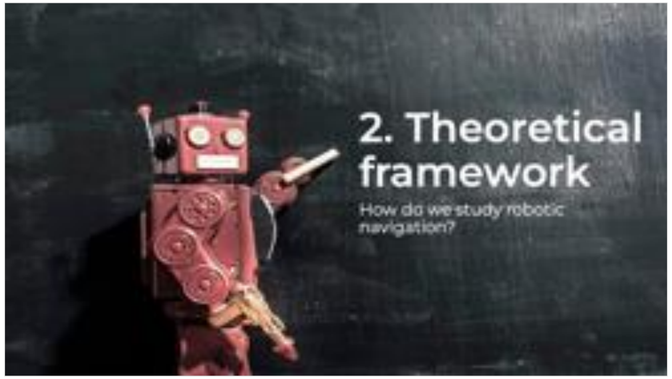
Thesis presentation by Carlos Gutiérrez Álvarez

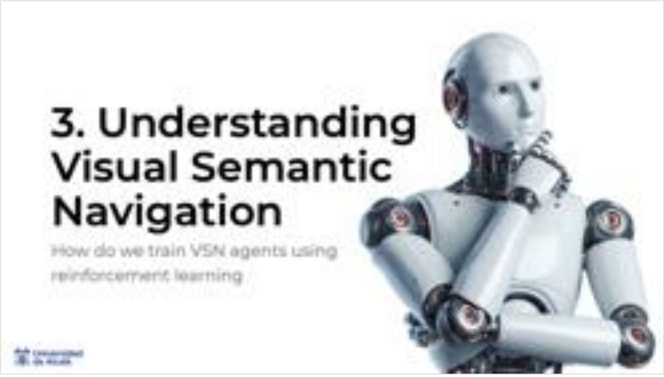
Directed by Roberto Javier López Sastre

Alcalá de Henares, 22 of January of 2026

Summary


1. Motivation
Why do we study robotic navigation?


2. Theoretical framework
How do we study robotic navigation?


3. Understanding Visual Semantic Navigation
How do we train VSN agents using reinforcement learning


4. Real World VSN


5. Bridging the gap


6. Final closure
Scientific trajectory, impact and final conclusions

1. Motivation

Why do we study robotic navigation?

Why Navigation Matters

Interact

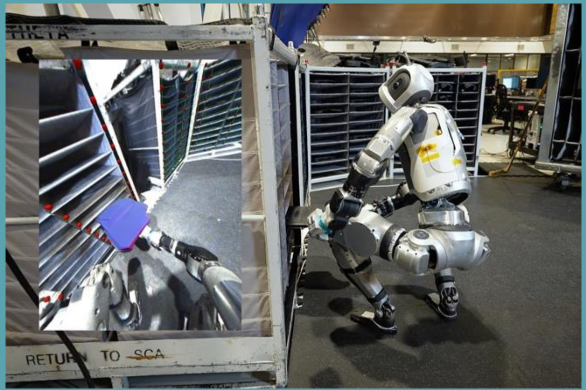
Explore

Move

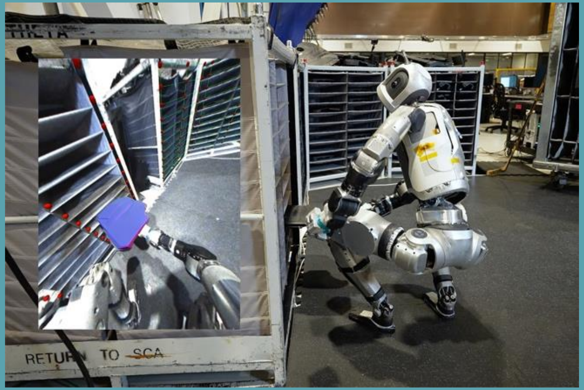
Why Navigation Matters

Embodied intelligent entities

⟹

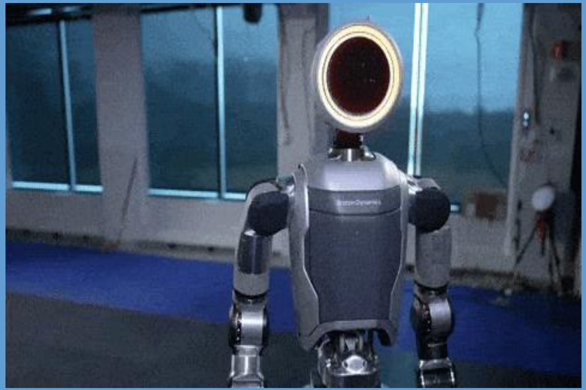


Interaction with the real world



Interaction with the real world

⟹



Movement

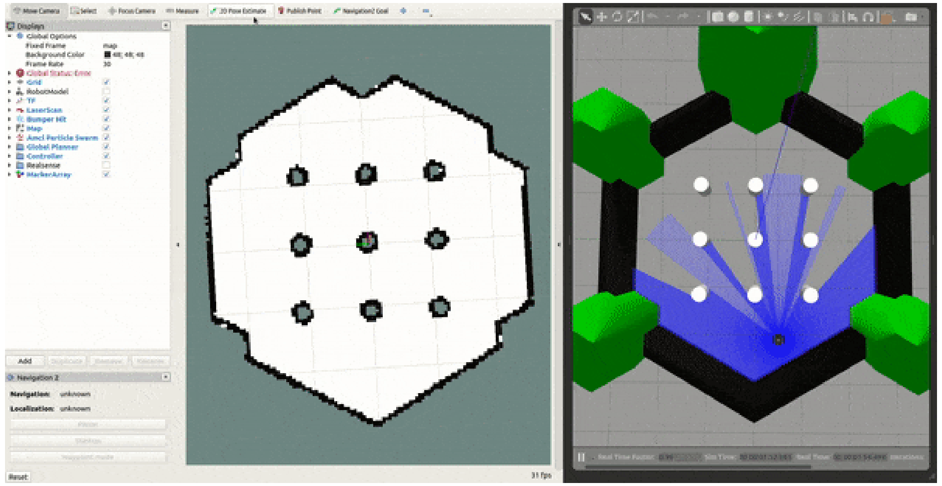
Why Navigation Matters

Without navigation there is no embodied intelligence

Different types of robotic navigation

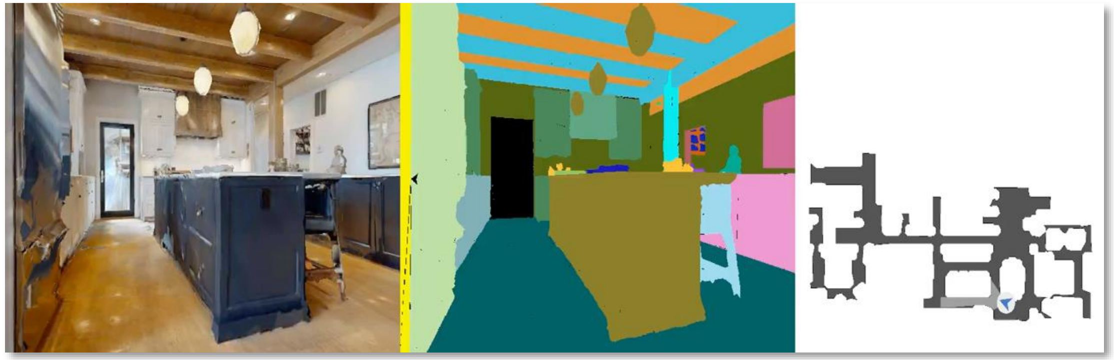
Classical Navigation

- Navigation based on the use of geometrical information to calculate most optimal routes.
- It needs a previously existing map of the environment or the creation of it on the fly.

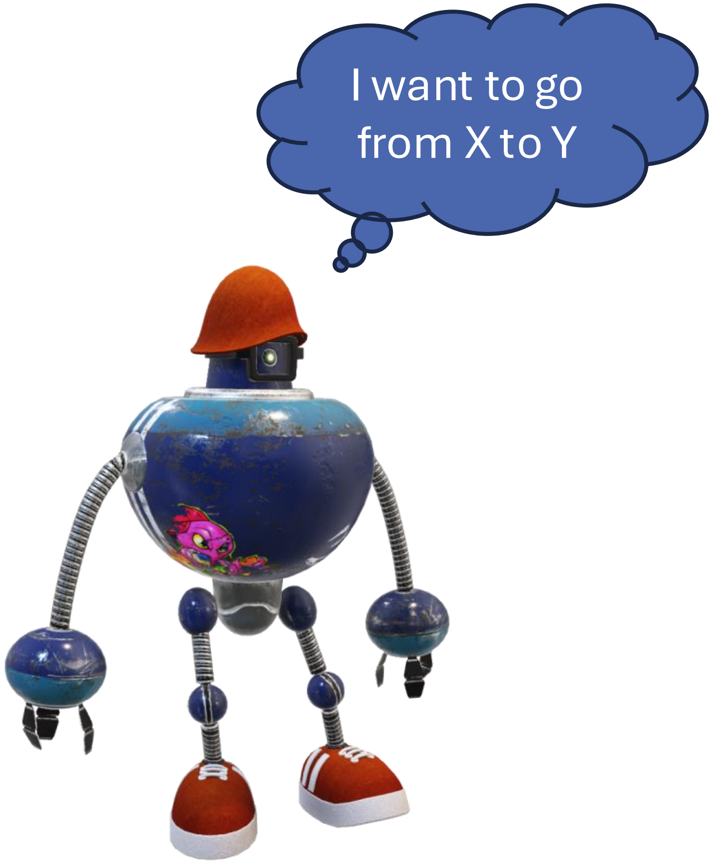


Visual Semantic Navigation

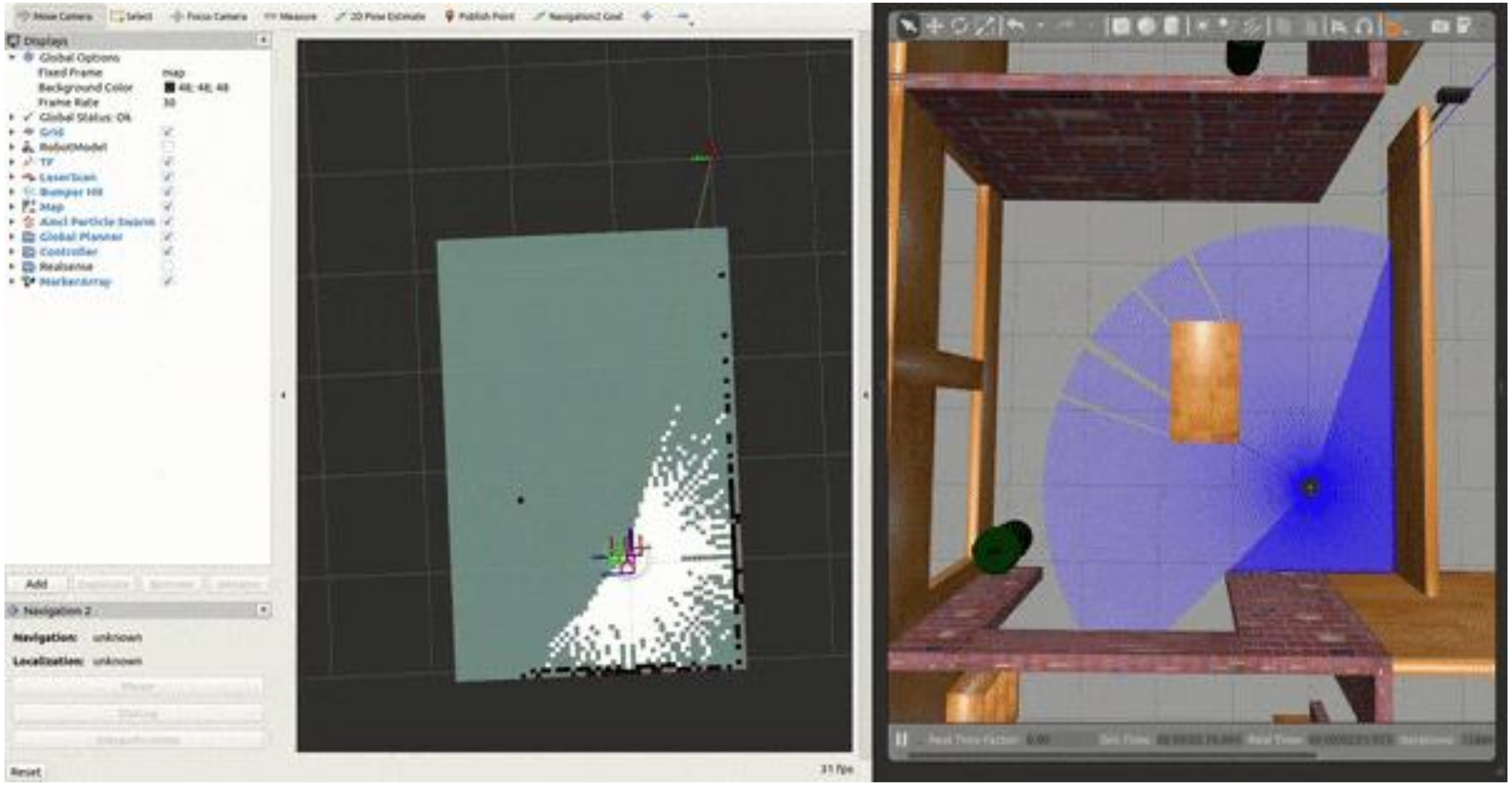
- Based on the use of egocentric images of the agent to decide where to navigate.
- This approach does not necessarily need any map of the environment, but some approaches create it on the fly.



Classical Navigation



SLAM – Simultaneous Localization and Mapping

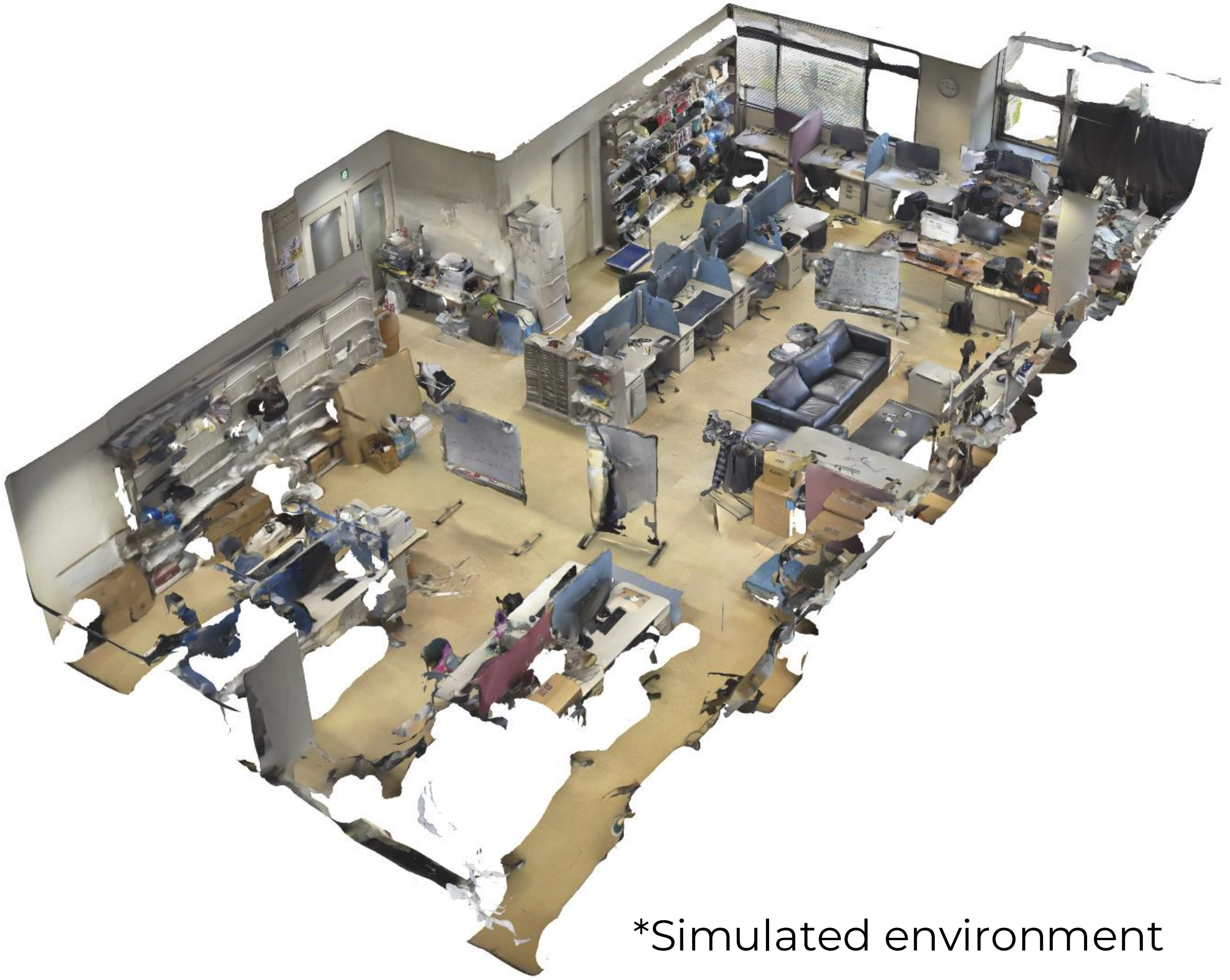


Universidad de Alcalá

* video from nav2 documentation.

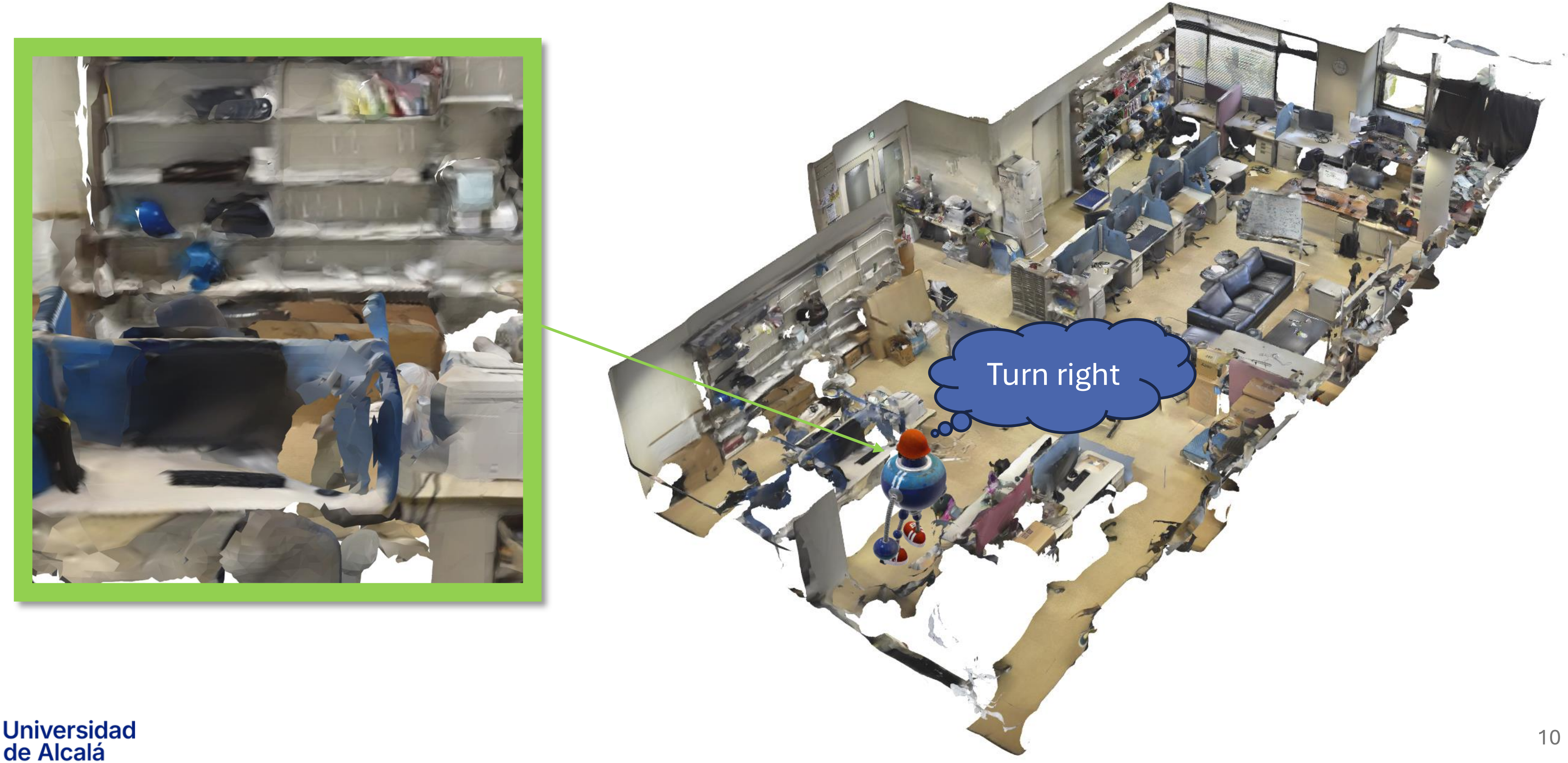
Diferenciar claramente entre classical y visual semantic y mencionar slam

Visual Semantic Navigation



*Simulated environment

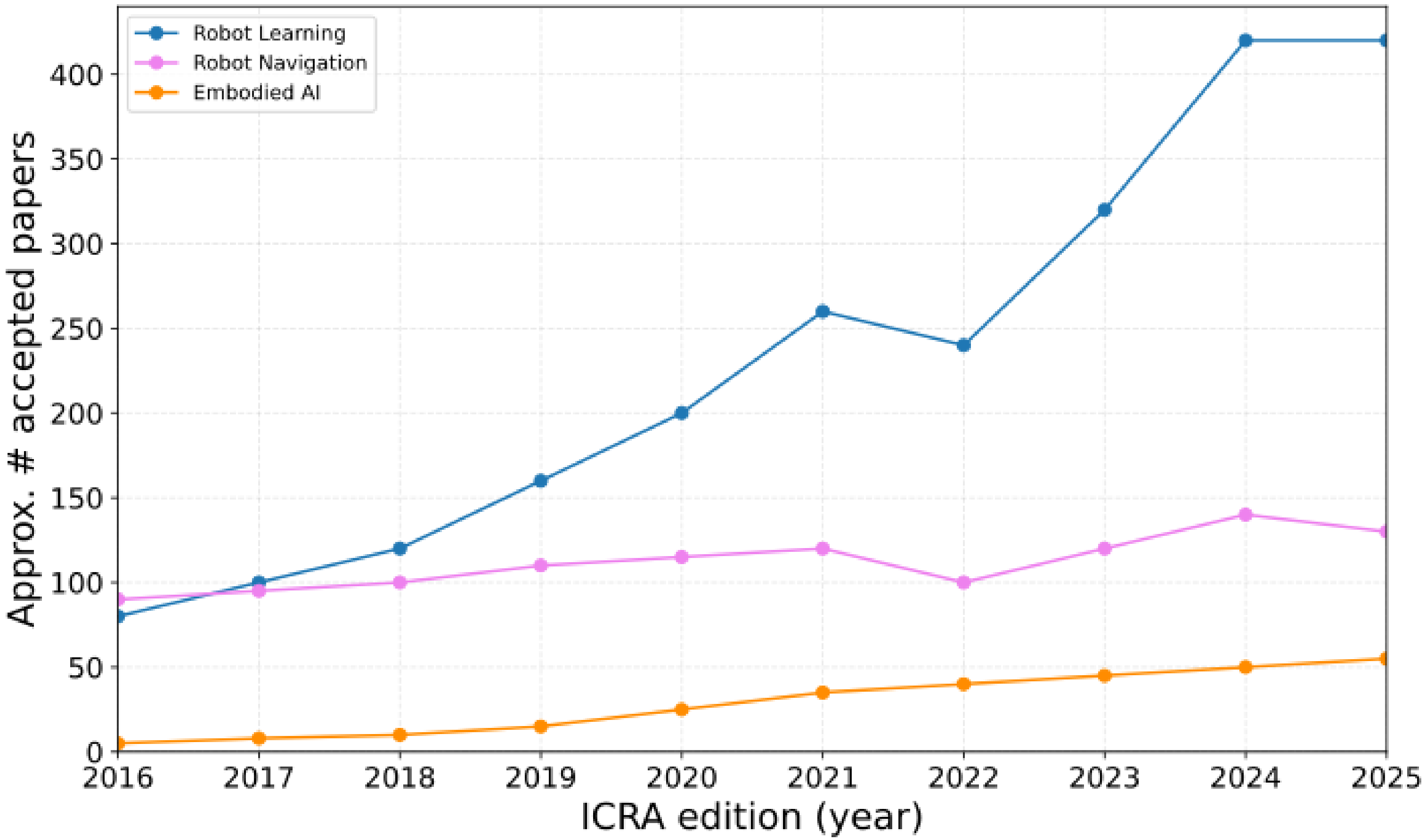
Visual Semantic Navigation



Visual Semantic Navigation

Visual Semantic Navigation

This is a Hot Research Topic



Universidad de Alcalá

This is a Hot Research Topic

The Scientific Challenges

1. Exploration vs exploitation

How to decide when to stop exploring and exploiting the knowledge of the scene.

2. Generalization

How to transfer the knowledge from one environment to another.

3. Sim-to-real

How to transfer the knowledge from simulated environments to real ones.

The Scientific Challenges

1. Exploration vs exploitation

The Scientific Challenges

1. Exploration vs exploitation

The Scientific Challenges

1. Exploration vs exploitation

- Exploration trajectory.
- Not optimal but probably will get to the target.

The Scientific Challenges

1. Exploration vs exploitation

Universidad de Alcalá

The Scientific Challenges

1. Exploration vs exploitation

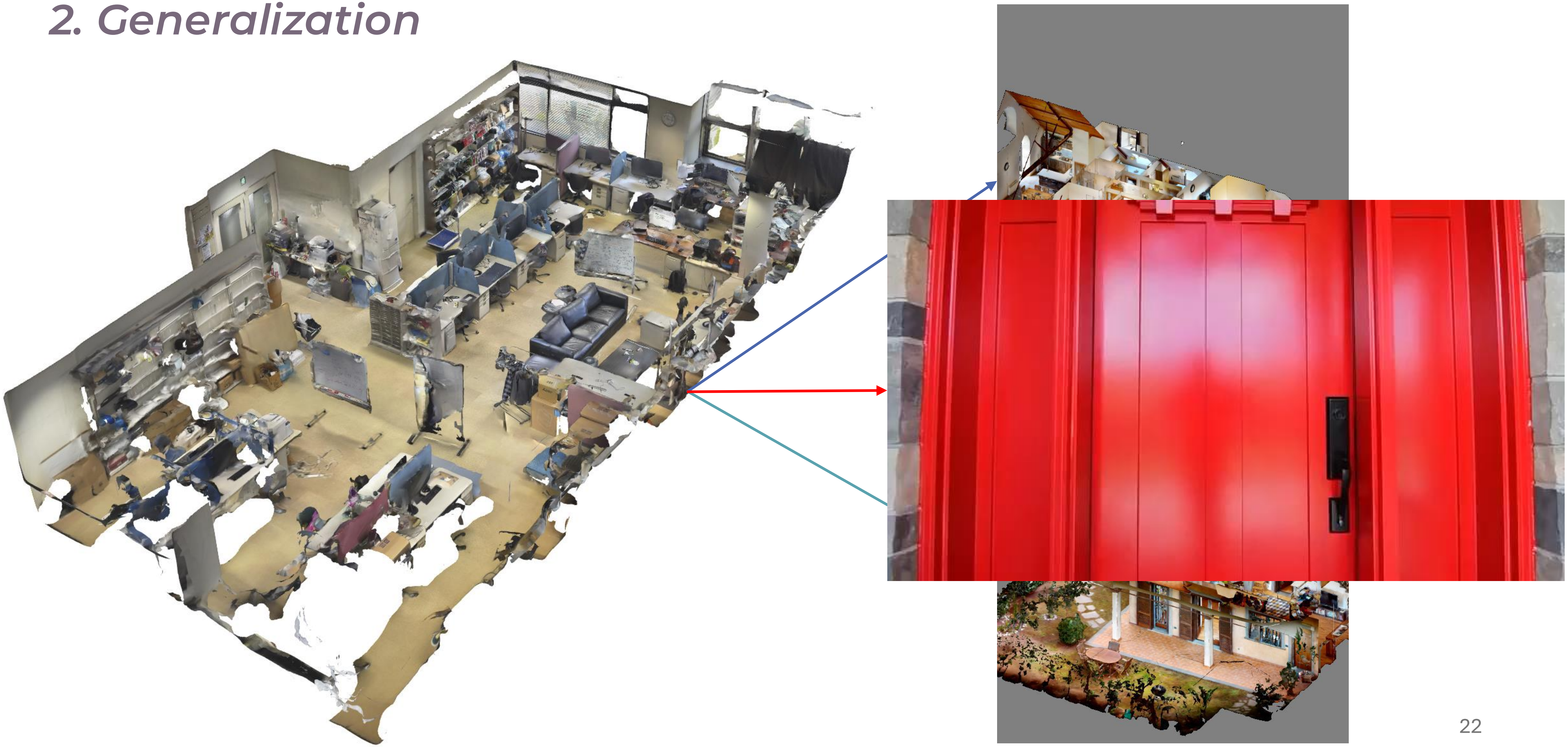
- Exploitation trajectory.
- Close to optimal path length.
- However, it needs previous knowledge of the environment.

The Scientific Challenges

2. Generalization

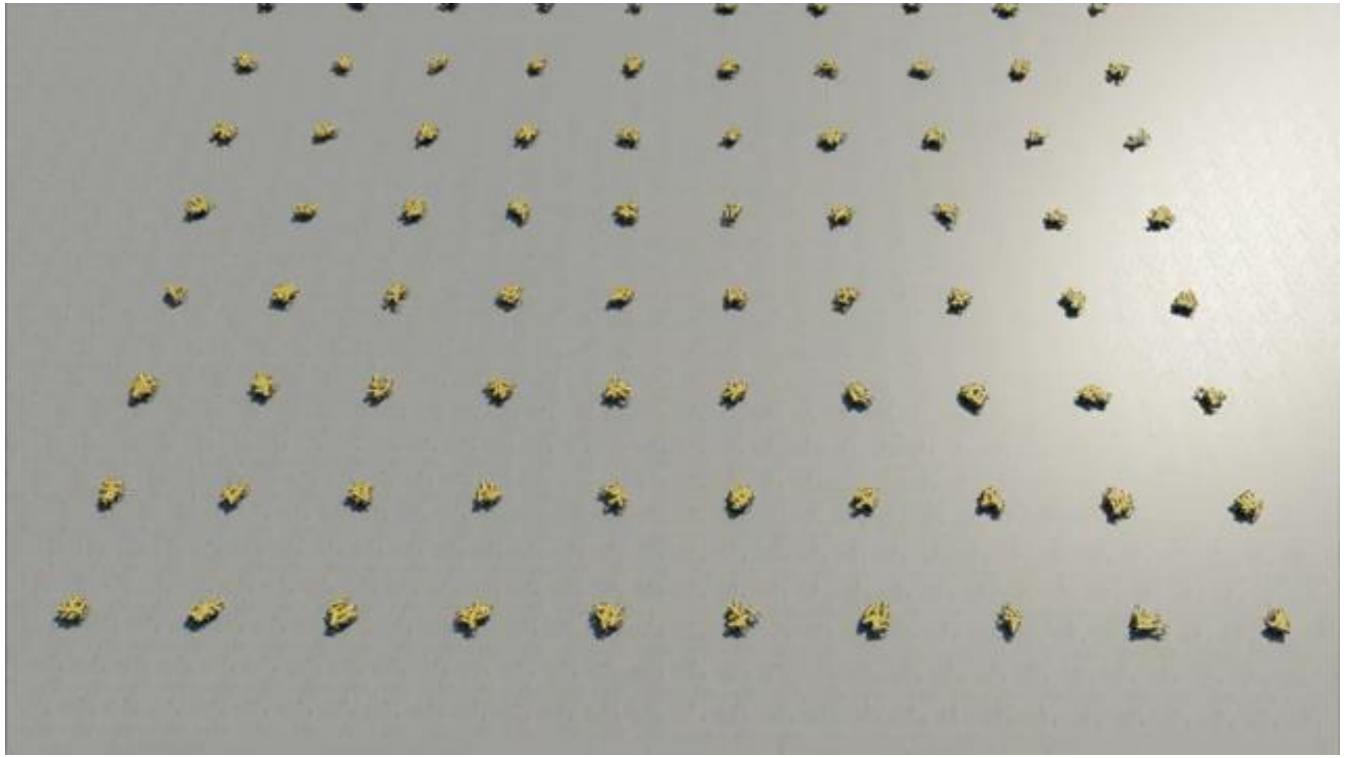
The Scientific Challenges

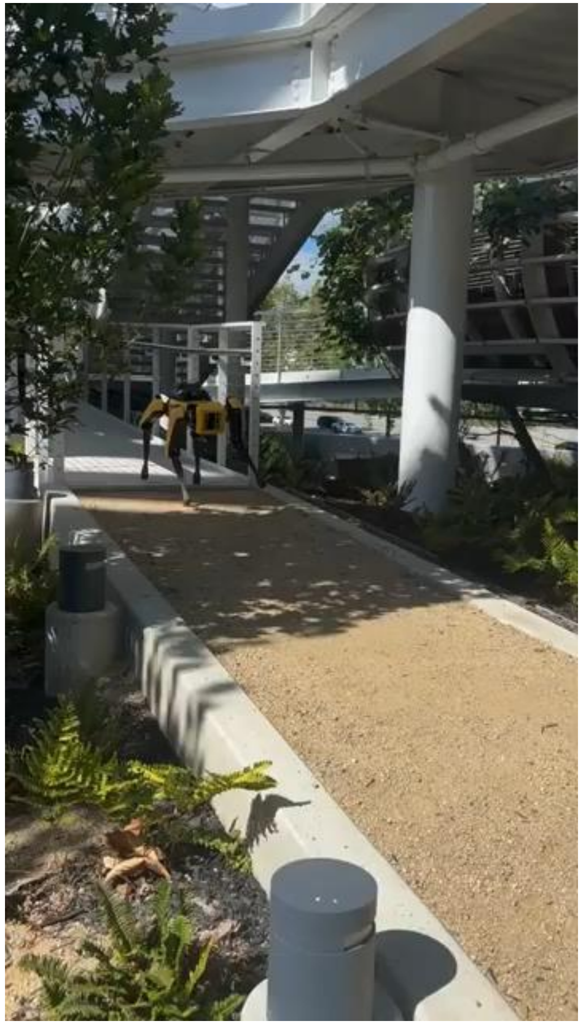
2. Generalization



The Scientific Challenges

3. Sim-to-real

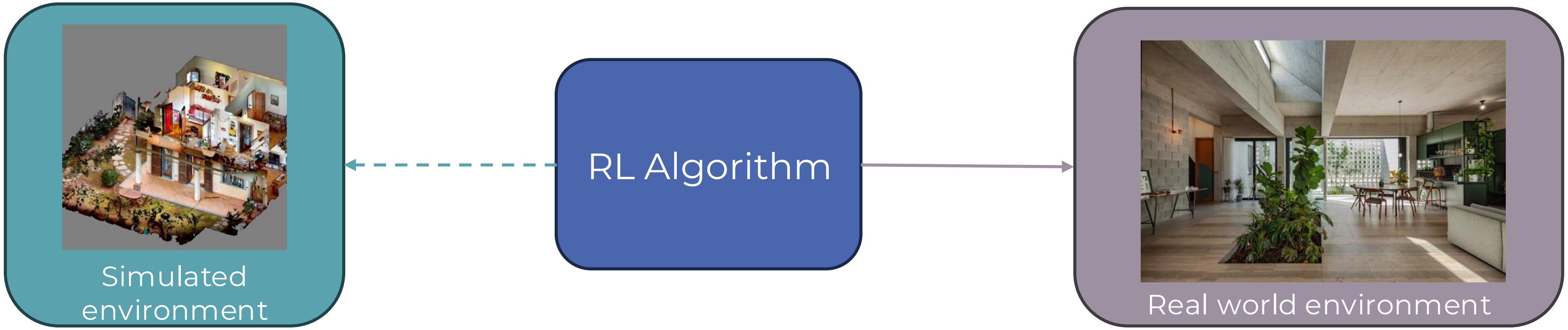


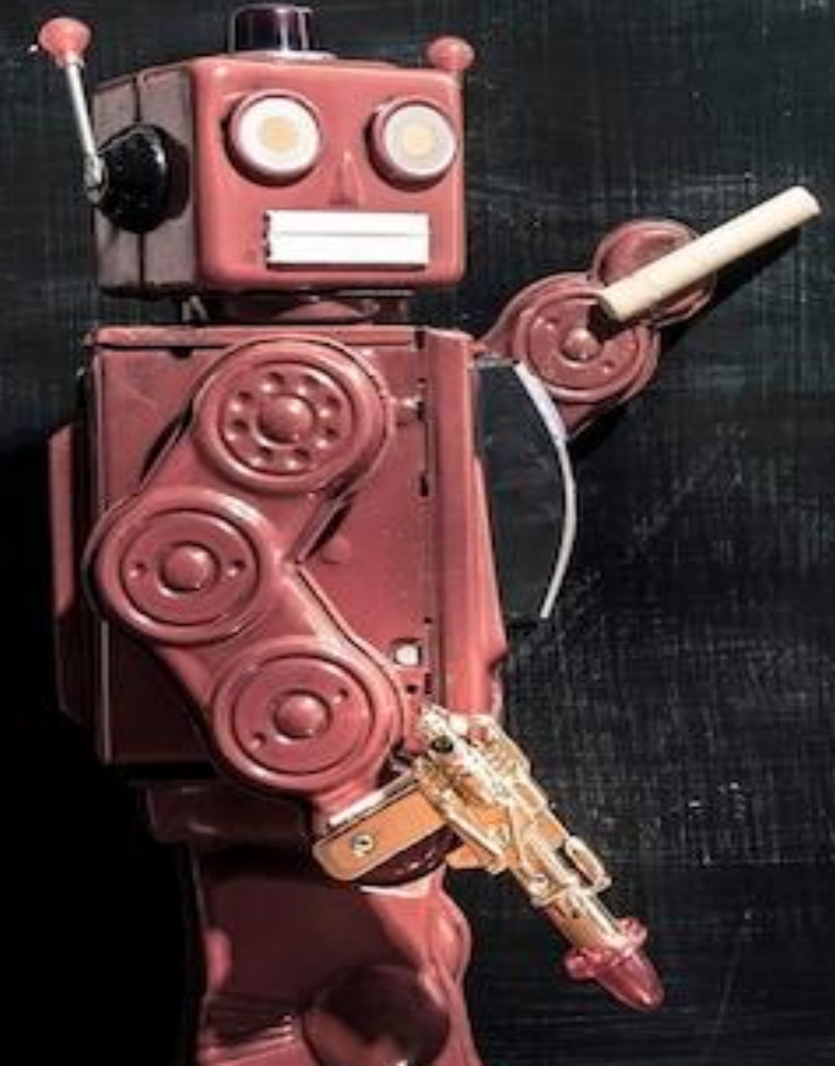


*videos from isaac sim documentation.

Thesis Objective

“Bridge simulation and real-world navigation via Reinforcement Learning (RL) algorithms”



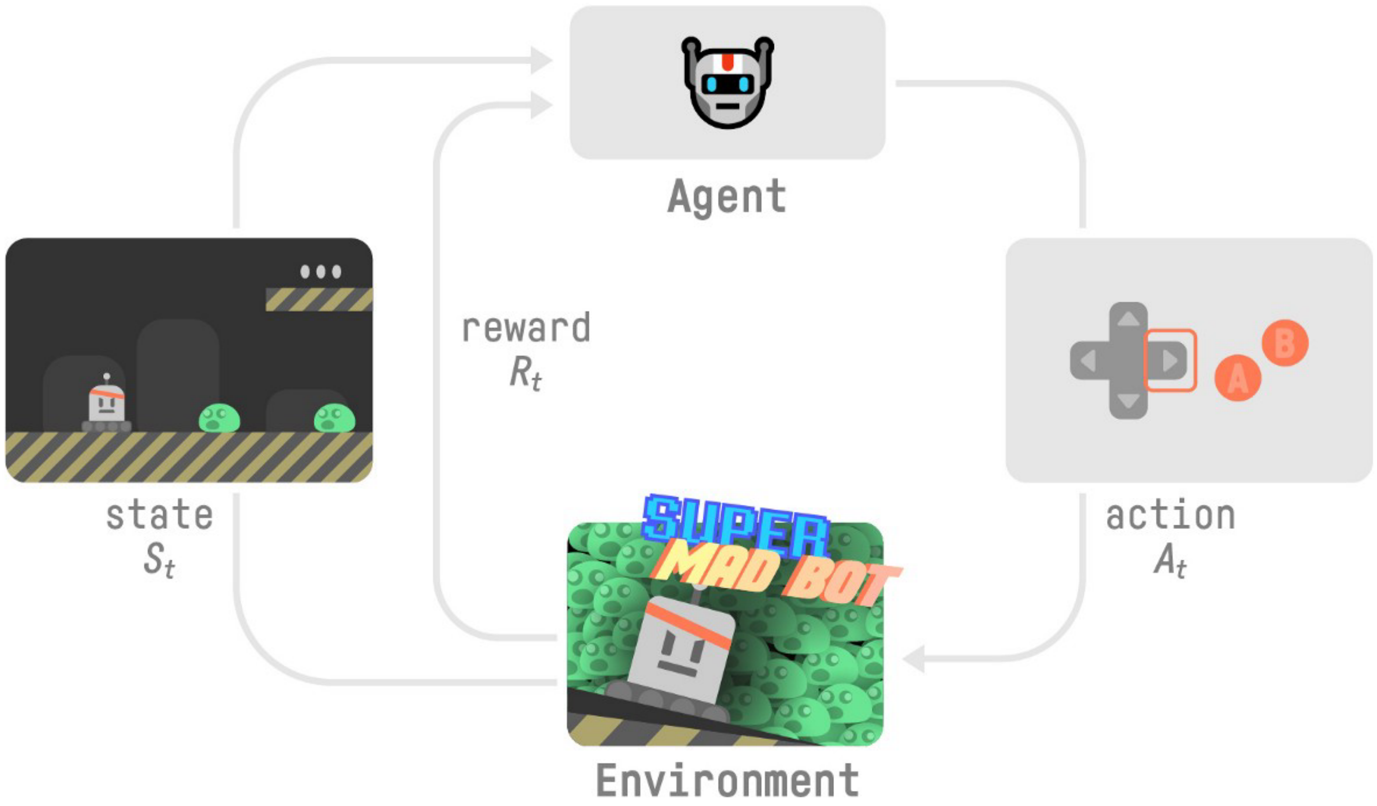


2. Theoretical framework

How do we study robotic navigation?

RL for Visual Semantic Navigation (VSN)

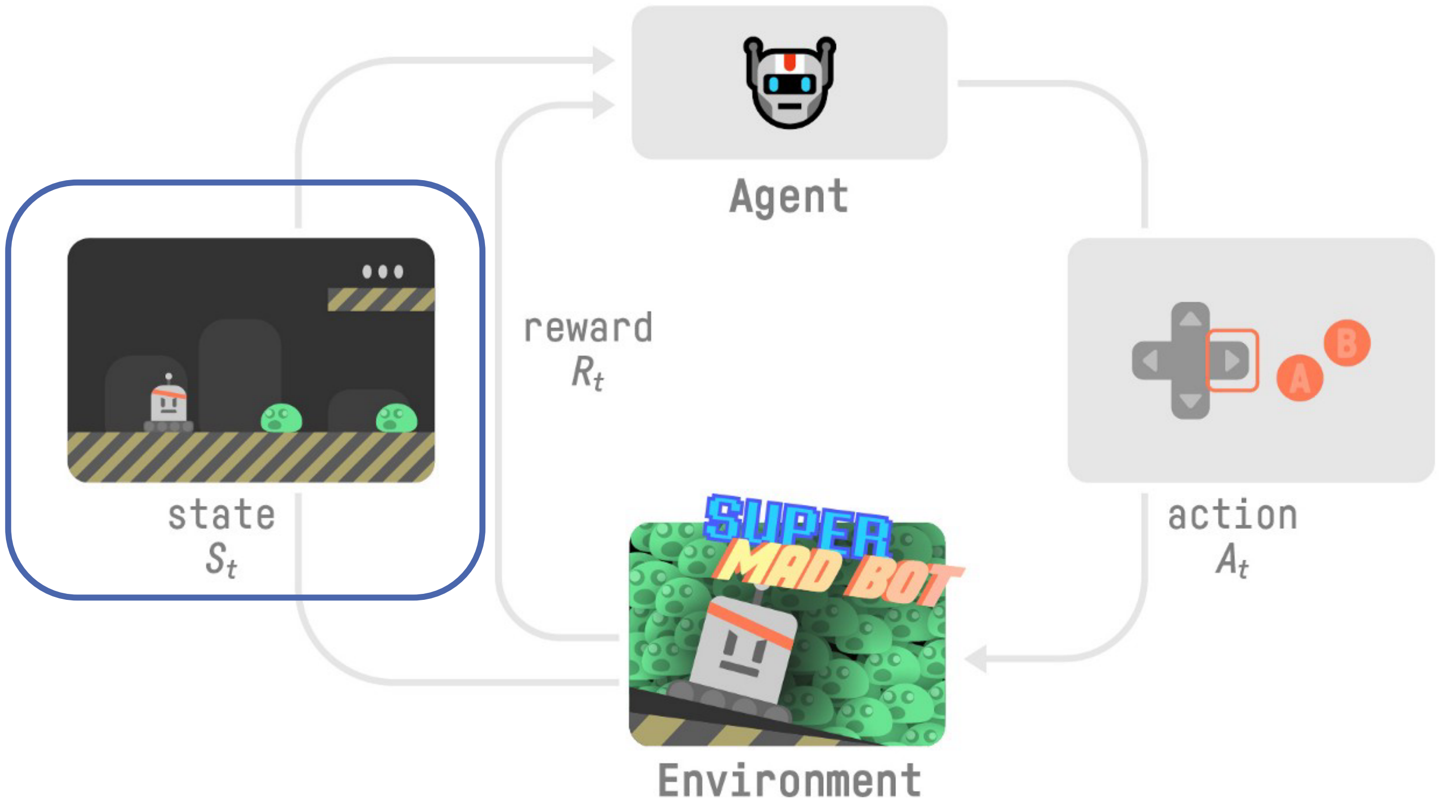
$$MDP = \{s_t, a_t, P_{a,t}, r_{a,t}\}$$



Universidad de Alcalá

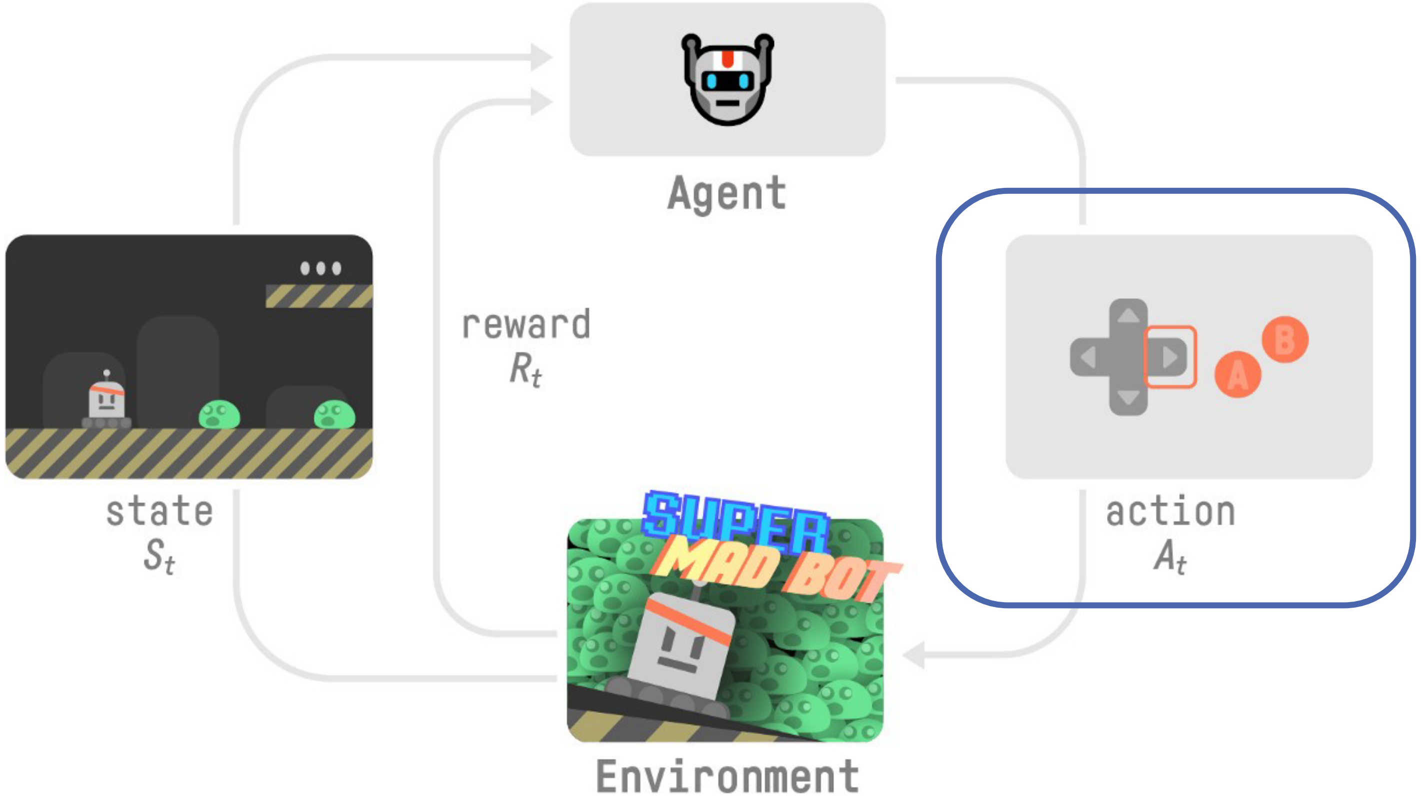
RL for Visual Semantic Navigation (VSN)

$$MDP = \boxed{\{s_t,} a_t, P_{a,t}, r_{a,t}\}$$



RL for Visual Semantic Navigation (VSN)

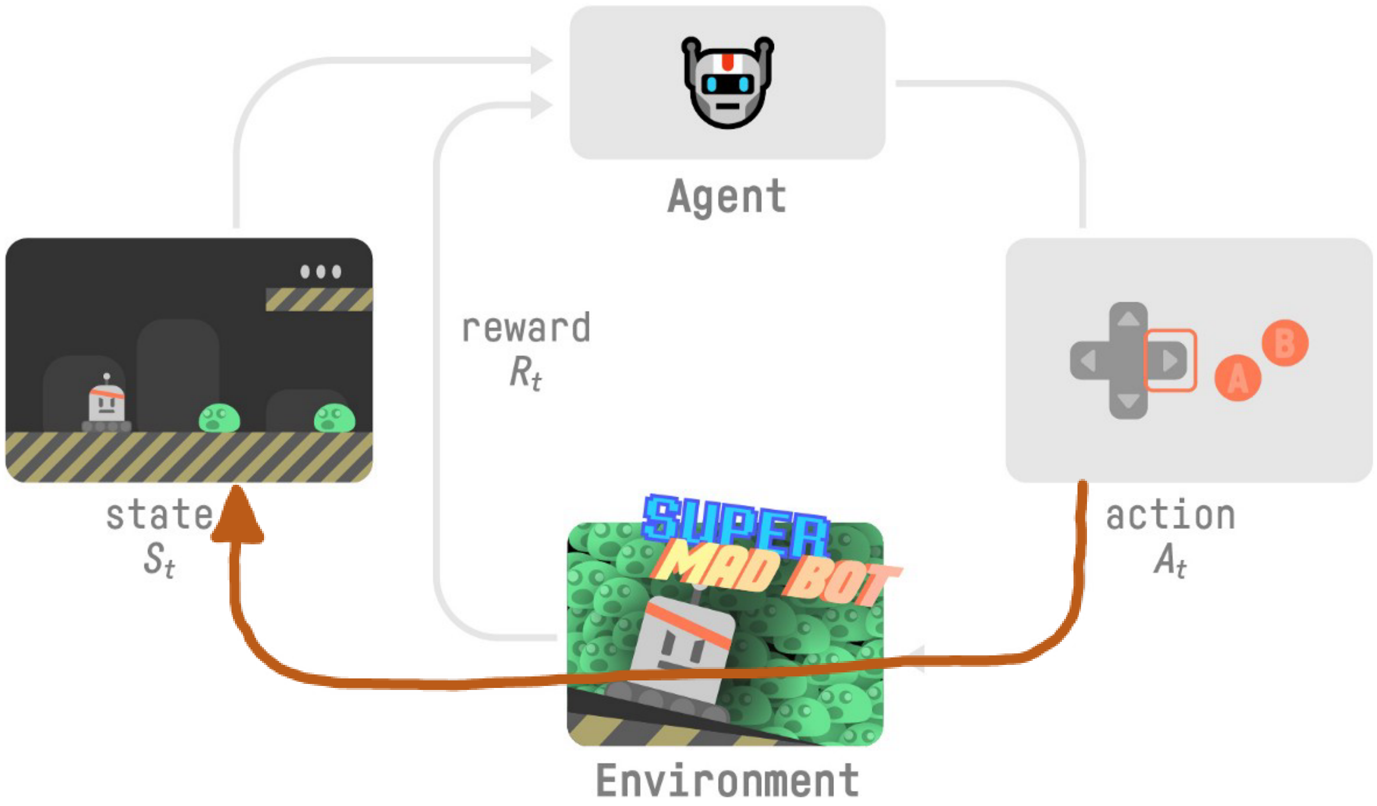
$$MDP = \{s_t, \boxed{a_t,} P_{a,t}, r_{a,t}\}$$



Universidad de Alcalá

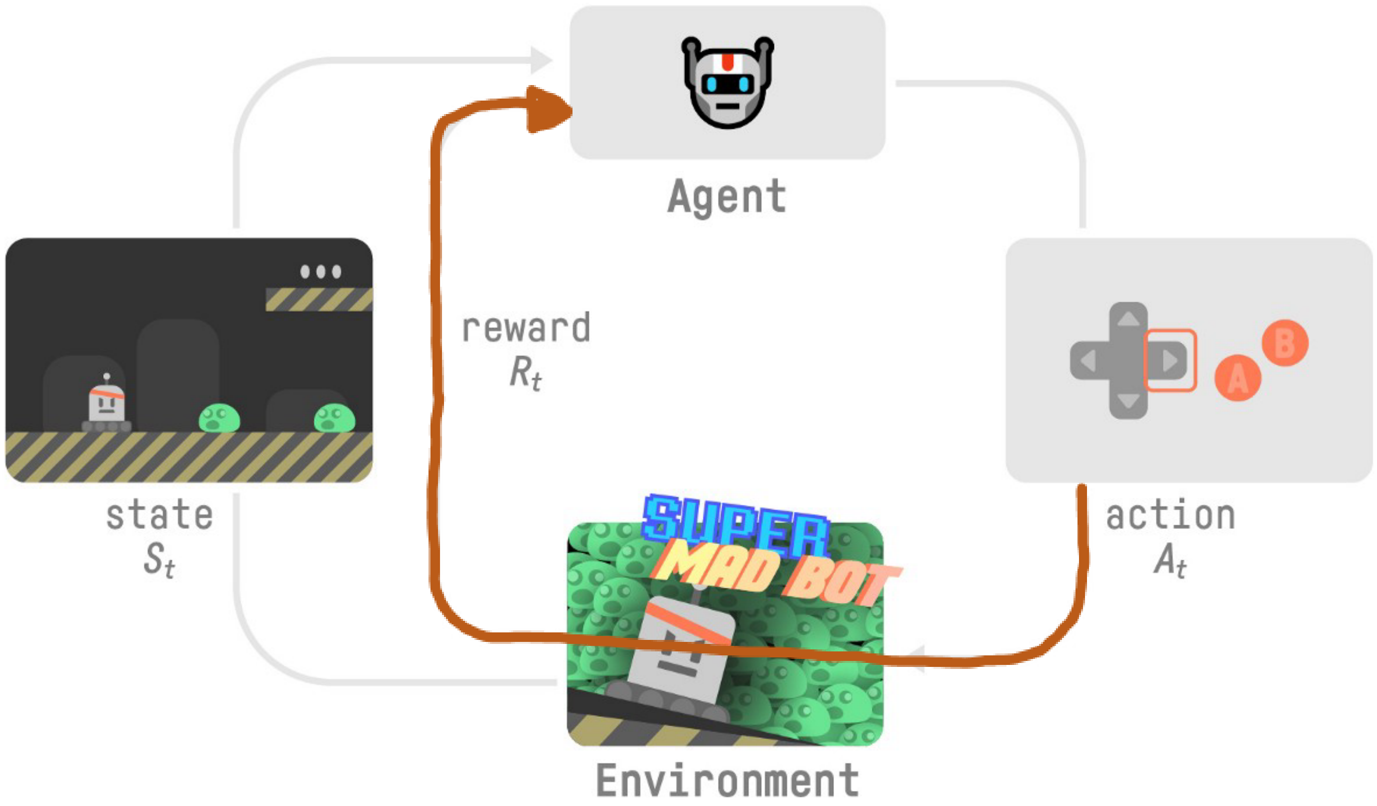
RL for Visual Semantic Navigation (VSN)

$$MDP = \{s_t, a_t, \boxed{P_{a,t},} r_{a,t}\}$$



RL for Visual Semantic Navigation (VSN)

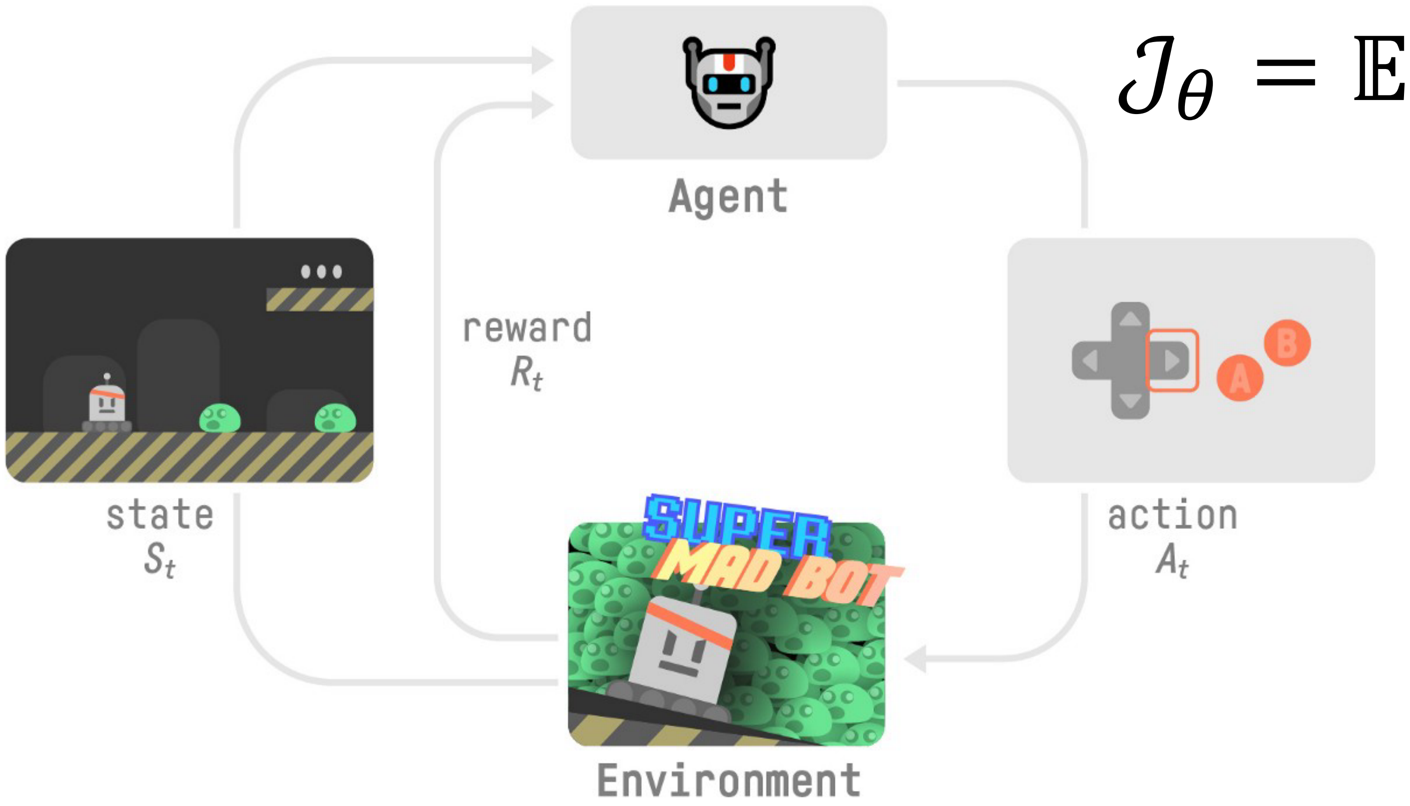
$$MDP = \{s_t, a_t, P_{a,t}, \boxed{r_{a,t}}\}$$



RL for Visual Semantic Navigation (VSN)

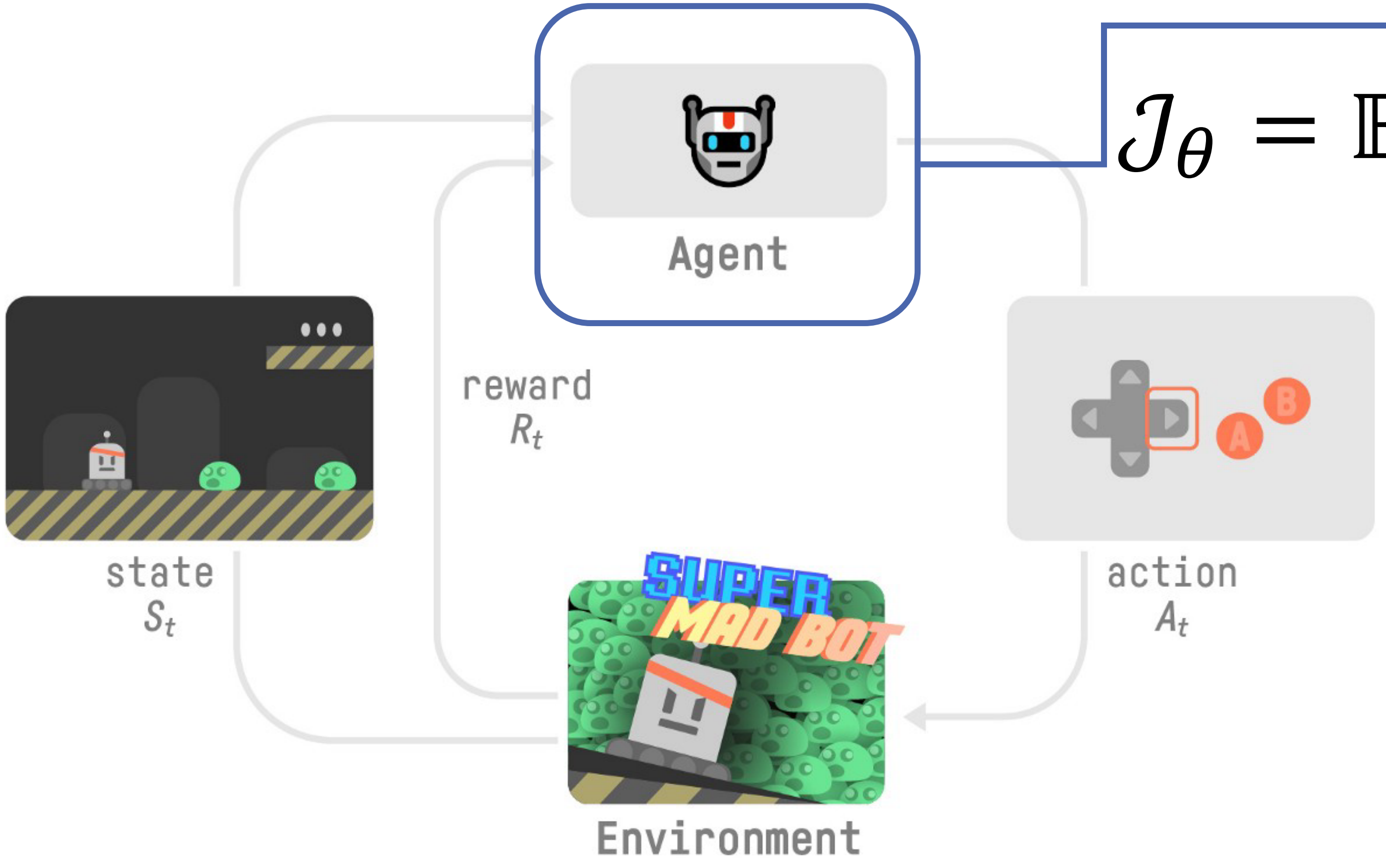
$$MDP = \{s_t, a_t, P_{a,t}, r_{a,t}\}$$

$$\mathcal{J}_\theta = \mathbb{E}_{a_t \sim \pi_\theta} \left\{ \sum_{t=0}^{\infty} r_{a_t}(s_t, s_{t+1}) \right\}$$



RL for Visual Semantic Navigation (VSN)

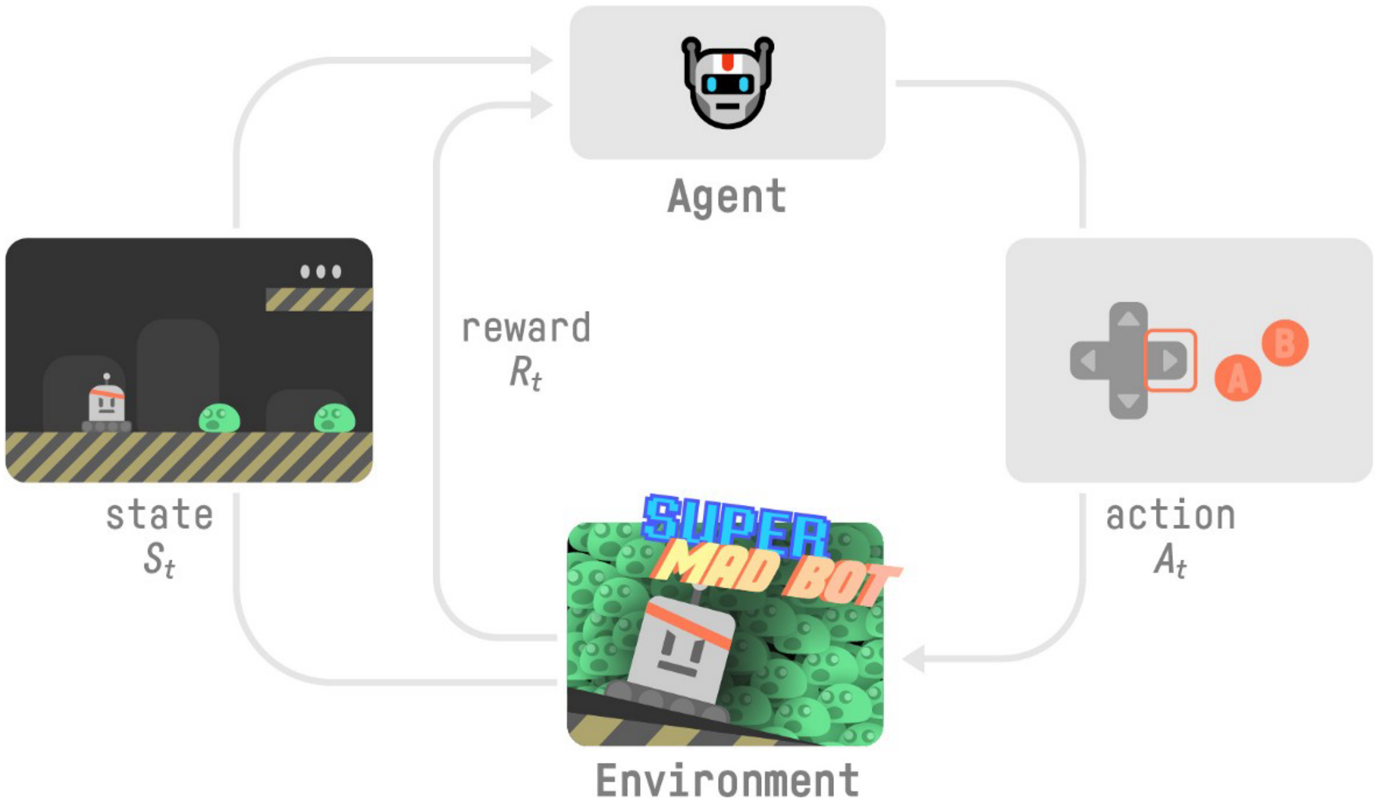
$$MDP = \{s_t, a_t, P_{a,t}, r_{a,t}\}$$



$$\mathcal{J}_\theta = \mathbb{E}_{a_t \sim \boxed{\pi_\theta}} \left\{ \sum_{t=0}^{\infty} r_{a_t}(s_t, s_{t+1}) \right\}$$

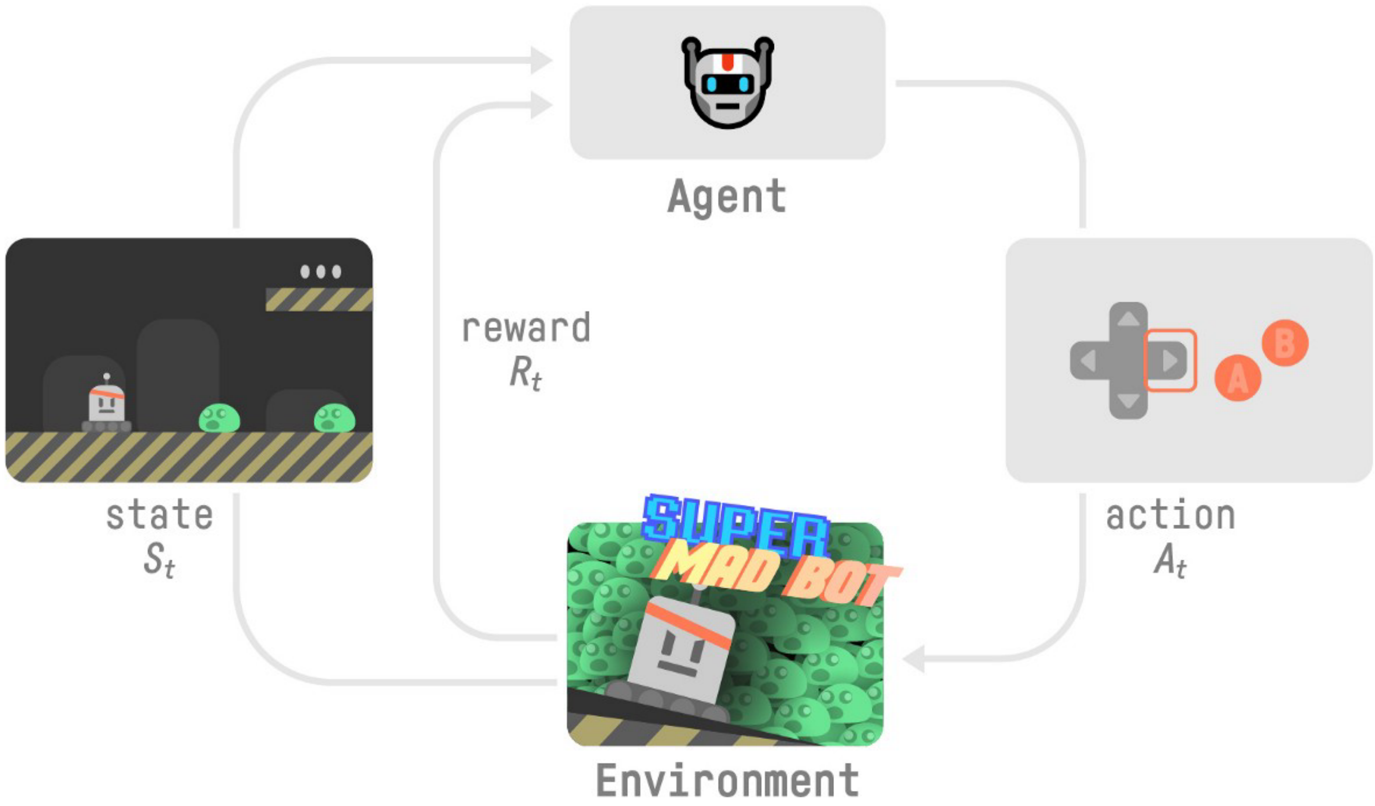
RL for Visual Semantic Navigation (VSN)

$$MDP = \{s_t, a_t, P_{a,t}, r_{a,t}\} \qquad POMDP = \{o_t, a_t, P_{a,t}, r_{a,t}\}$$



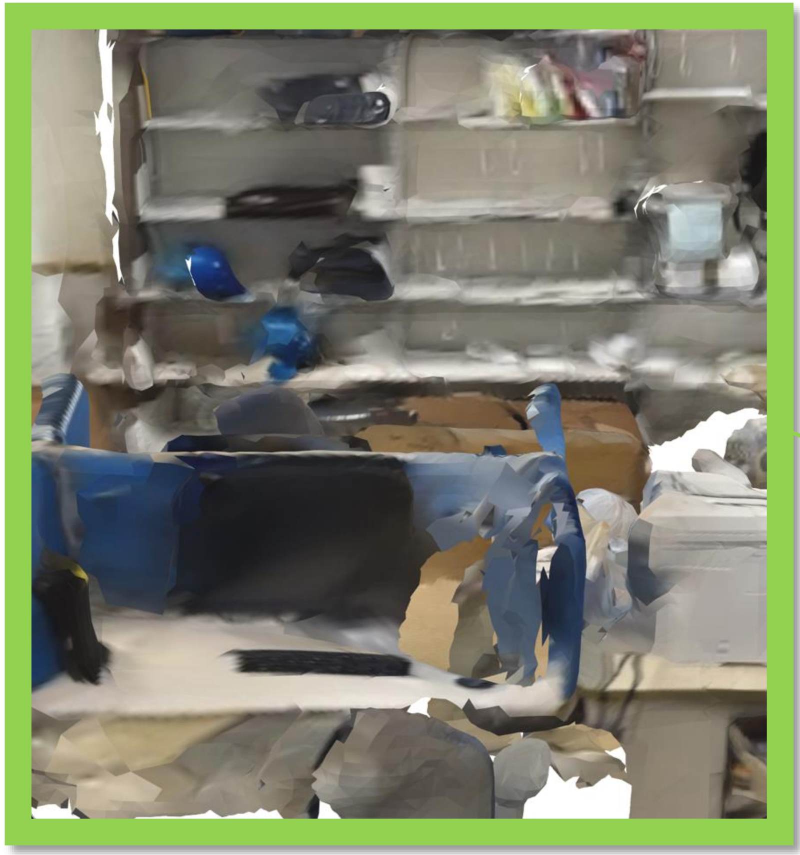
RL for Visual Semantic Navigation (VSN)

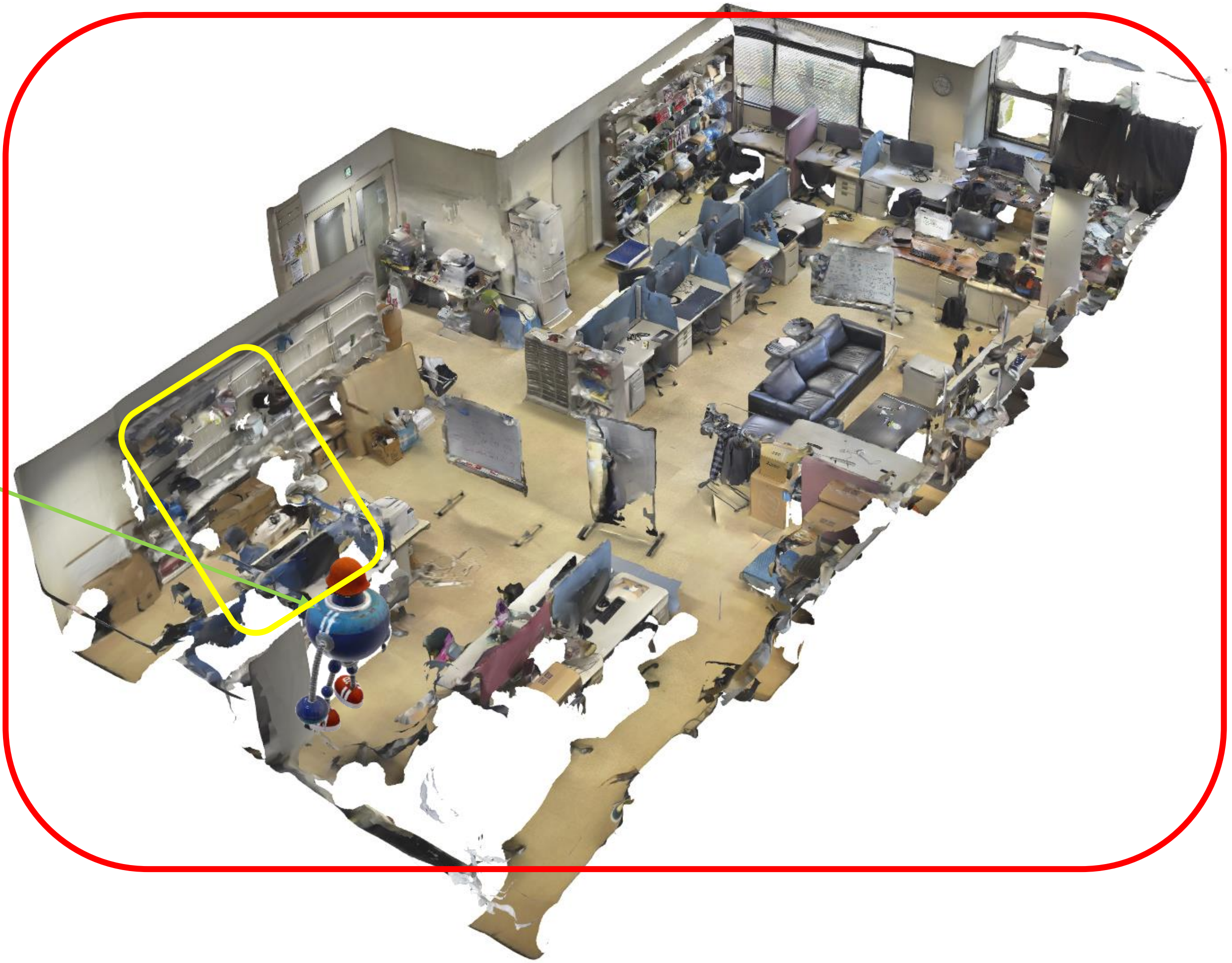
$$MDP = \boxed{\{s_t,} a_t, P_{a,t}, r_{a,t}\} \qquad POMDP = \boxed{\{o_t,} a_t, P_{a,t}, r_{a,t}\}$$



RL for Visual Semantic Navigation (VSN)

$$MDP = \boxed{\{s_t,} a_t, P_{a,t}, r_{a,t}\} \qquad POMDP = \boxed{\{o_t,} a_t, P_{a,t}, r_{a,t}\}$$





Three Families of VSN

1. Classical methods

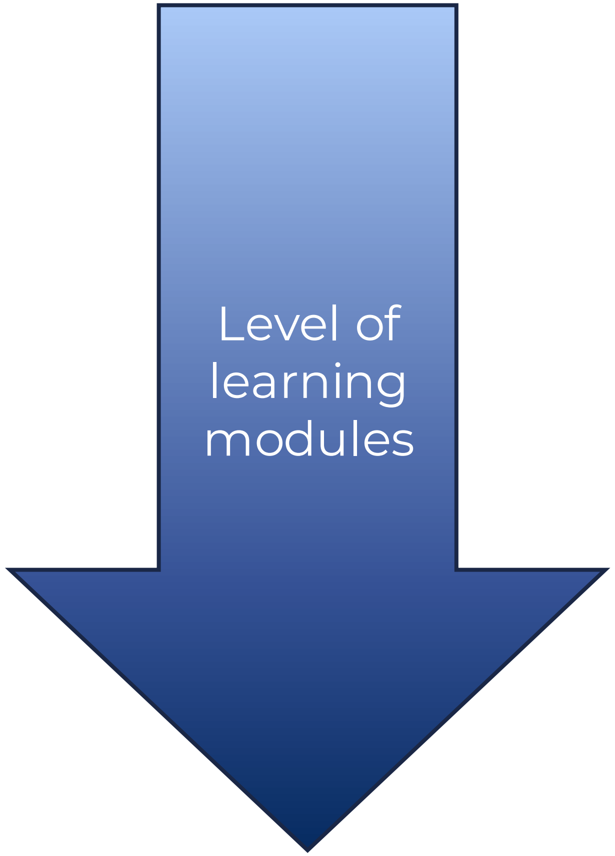
No learning components.

2. Modular learning methods

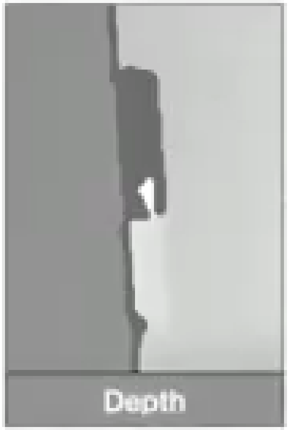
Mix between learning and non learning components.

3. End-to-end learning methods

Only learning components.

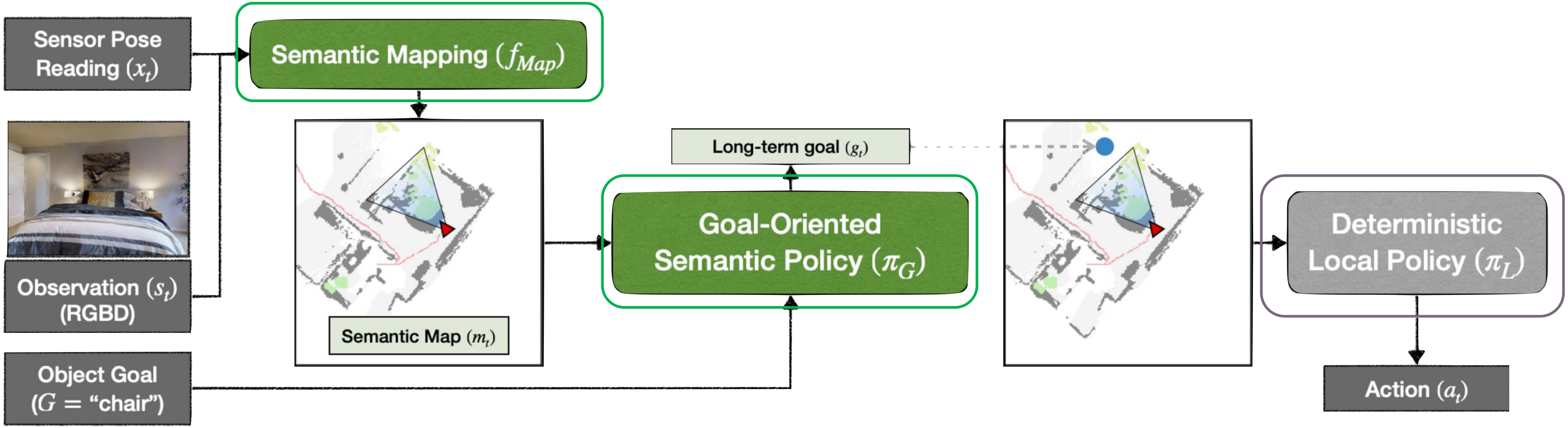


Classical methods



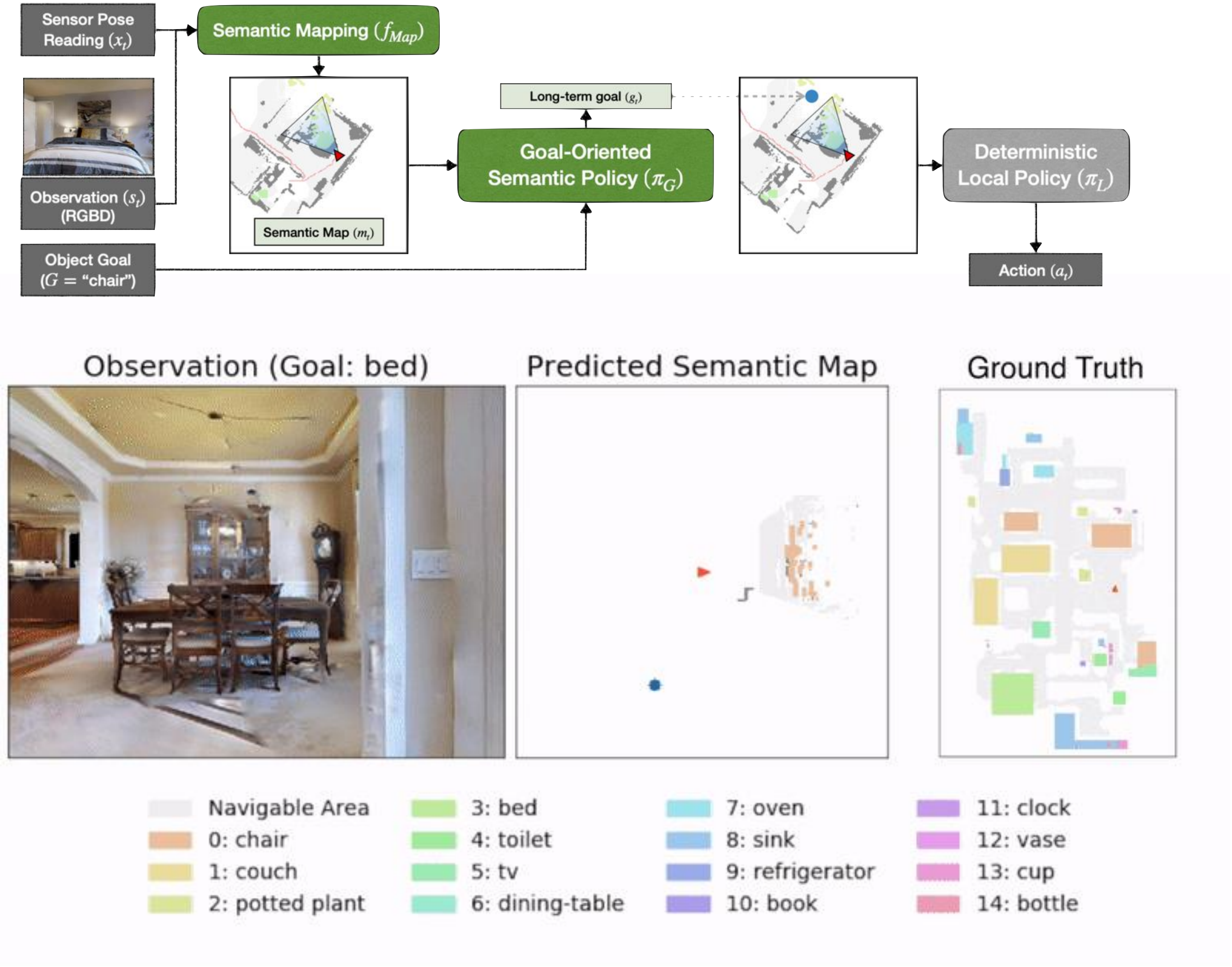
* Video grom Gervert et.al. 2023

Modular learning



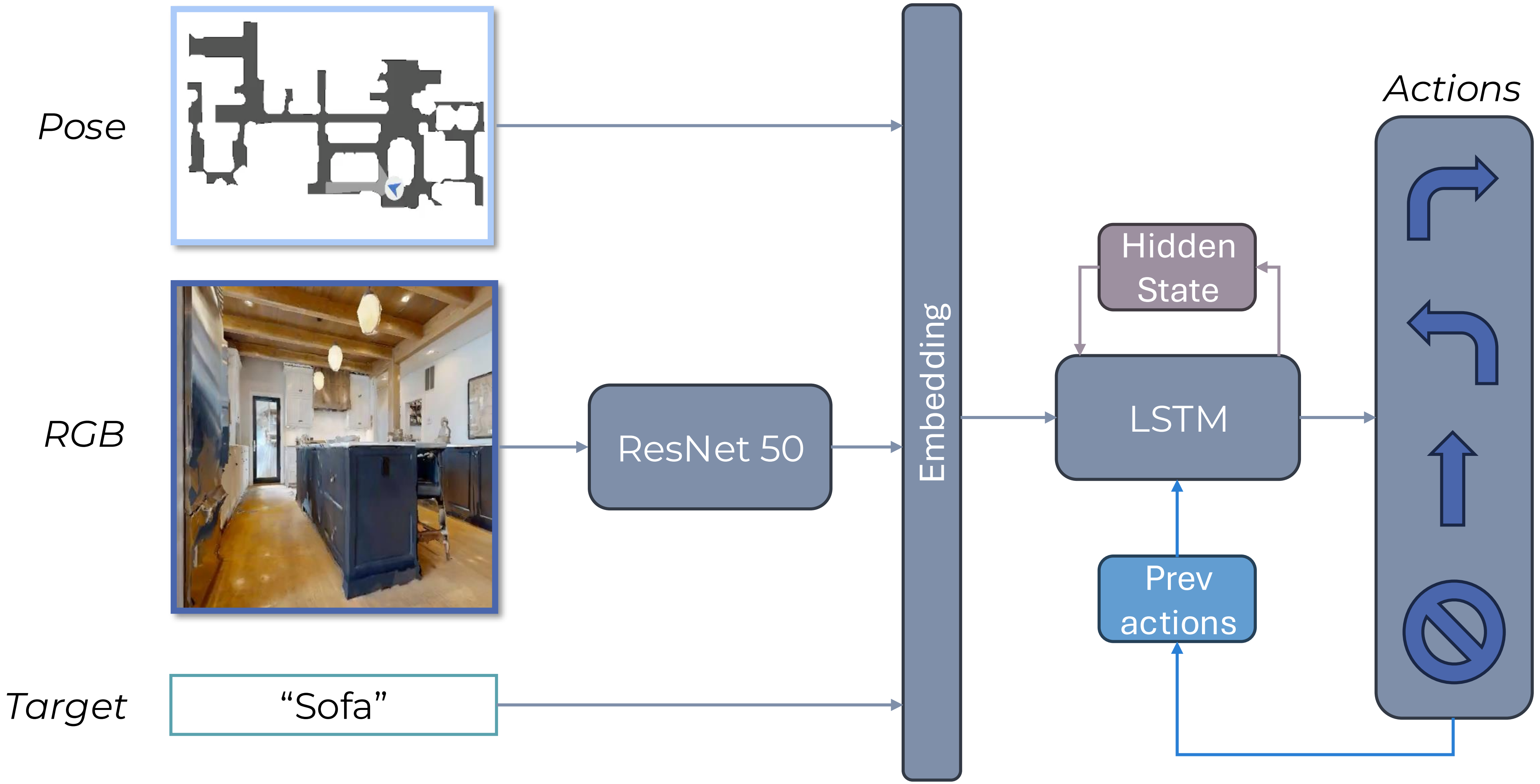
* Image grom Chaplot et.al. 2020

Modular learning



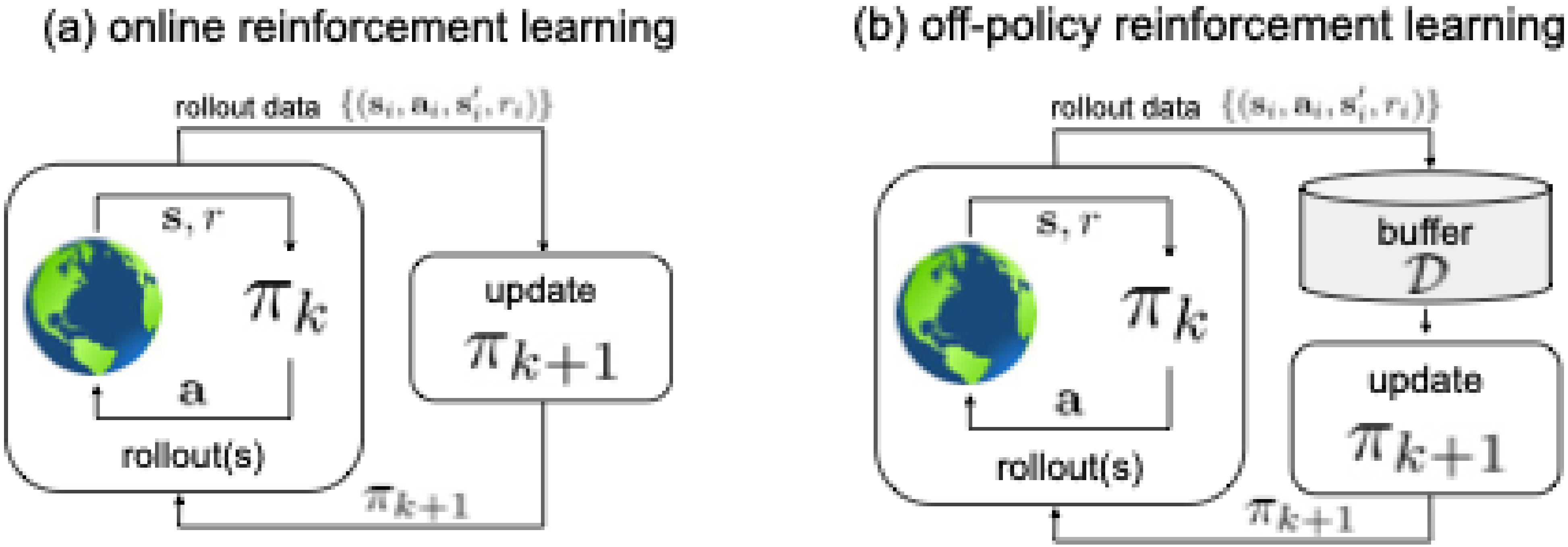
* Image grom Chaplot et.al. 2020

End-to-end learning



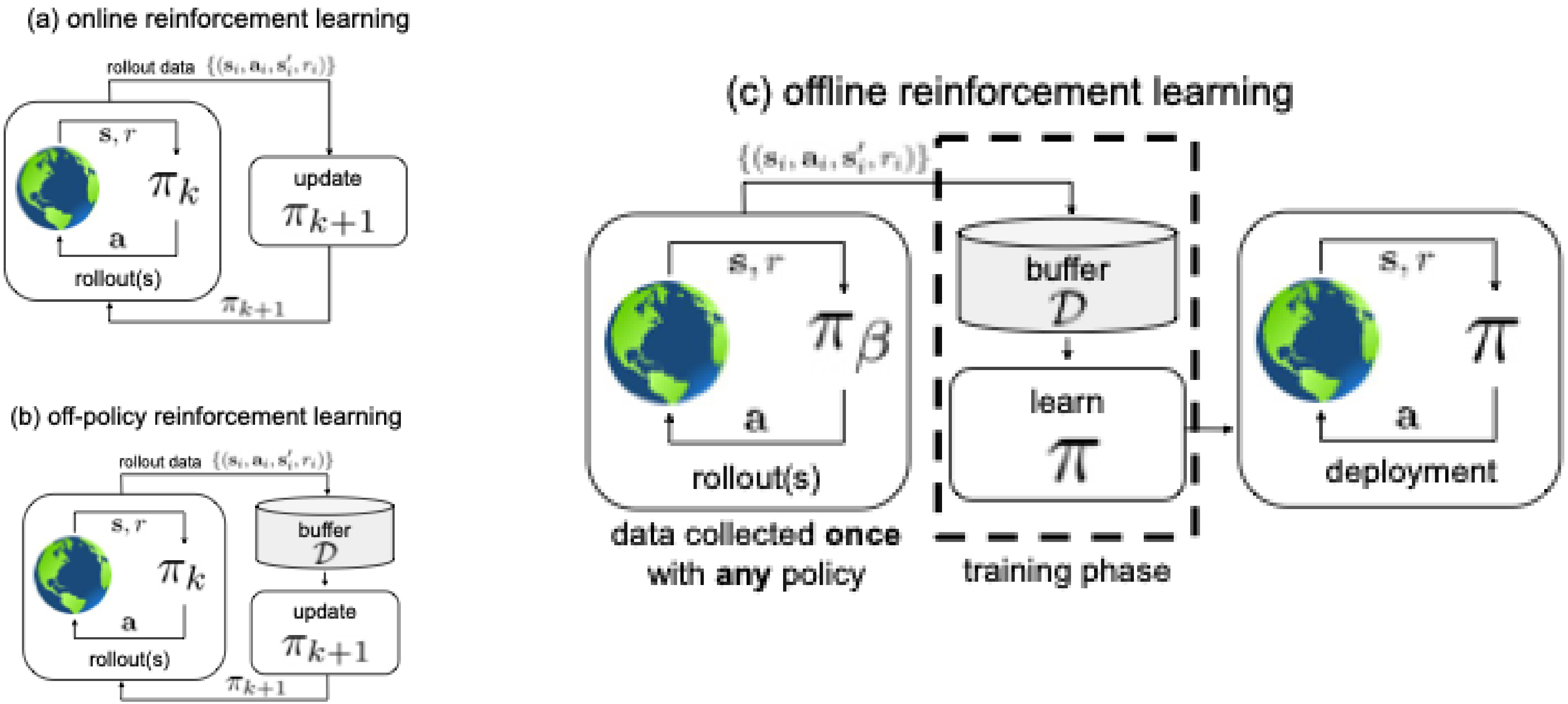
Universidad de Alcalá

Offline RL



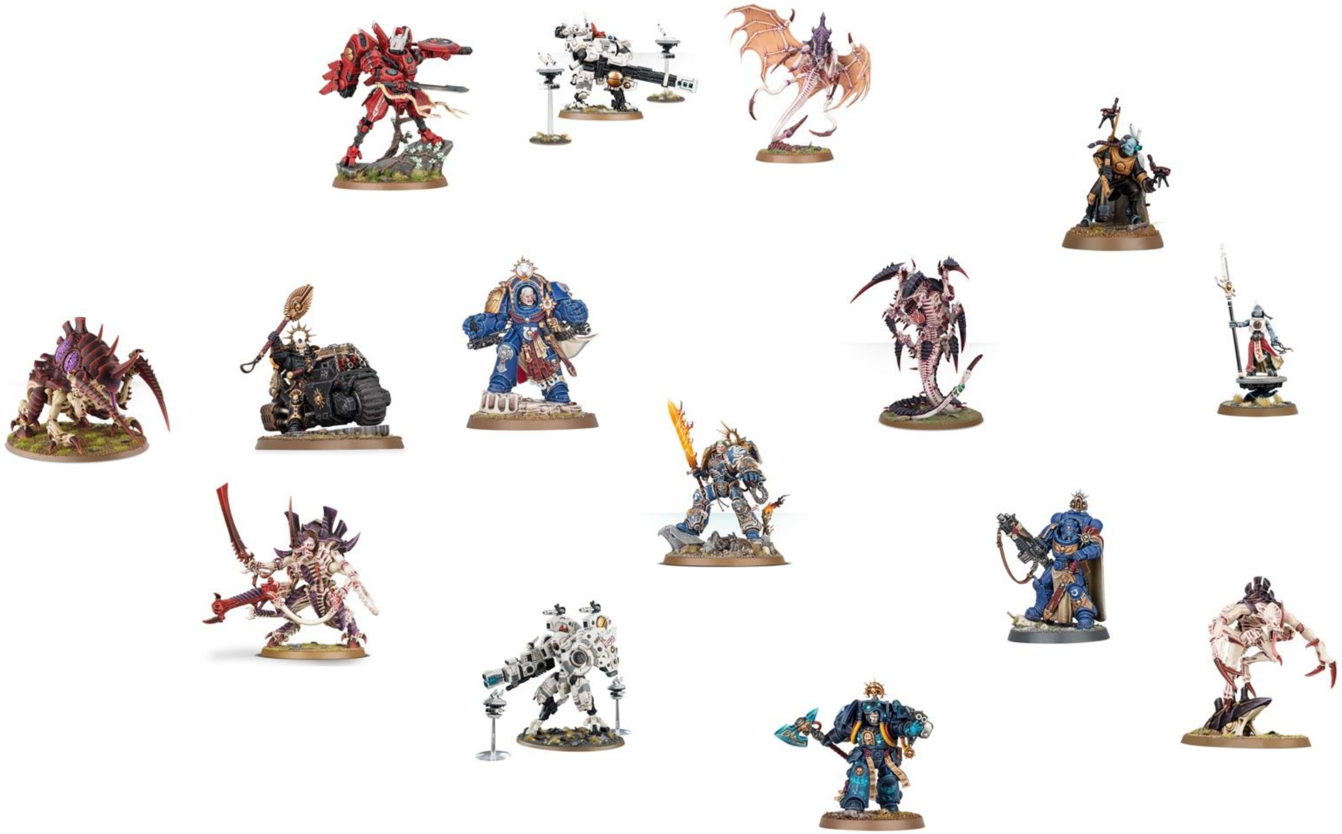
* Images grom Levine et.al. 2020

Offline RL


(a) online reinforcement learning
rollout data $\{(s_i, a_i, s'_i, r_i)\}$
s, r
π_k
a
rollout(s)
update
π_{k+1}
π_{k+1}
(b) off-policy reinforcement learning
rollout data $\{(s_i, a_i, s'_i, r_i)\}$
s, r
π_k
a
rollout(s)
buffer
$\mathcal{D}$
update
π_{k+1}
π_{k+1}
(c) offline reinforcement learning
$\{(s_i, a_i, s'_i, r_i)\}$
s, r
π_β
a
rollout(s)
buffer
$\mathcal{D}$
learn
π
s, r
π
a
deployment
data collected once
with any policy
training phase

Meta Learning

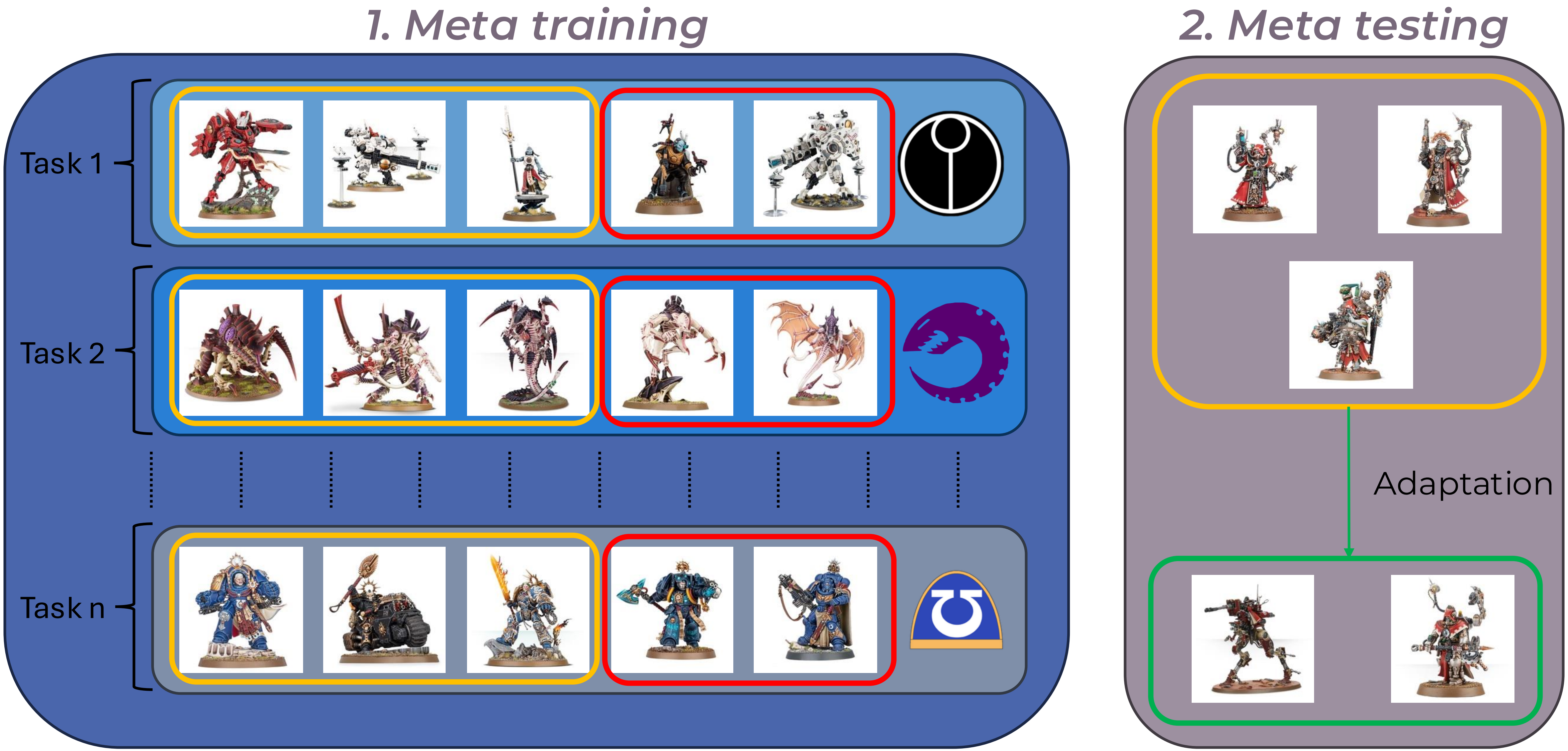
Meta Learning



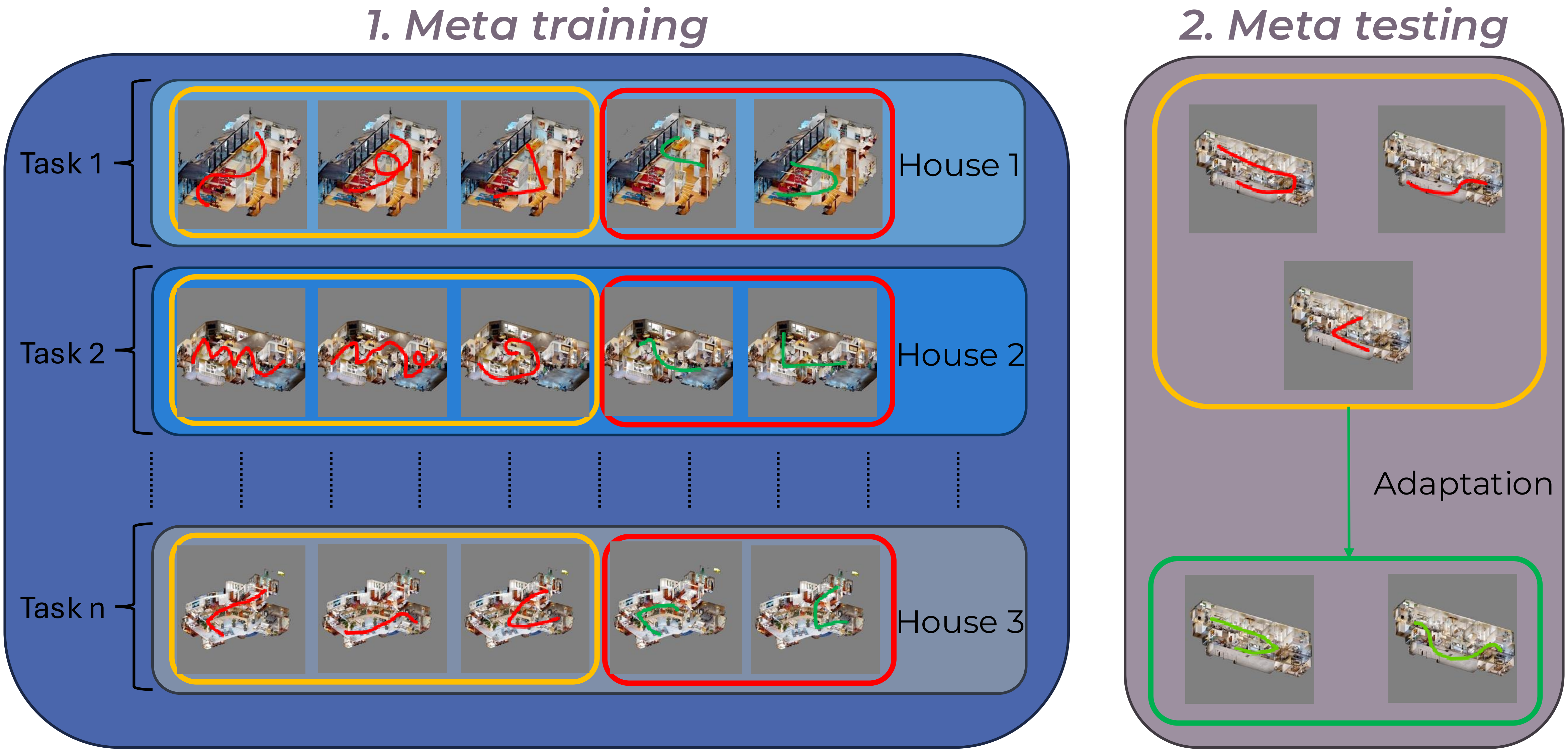
Meta Learning

1. Meta training

Meta Learning

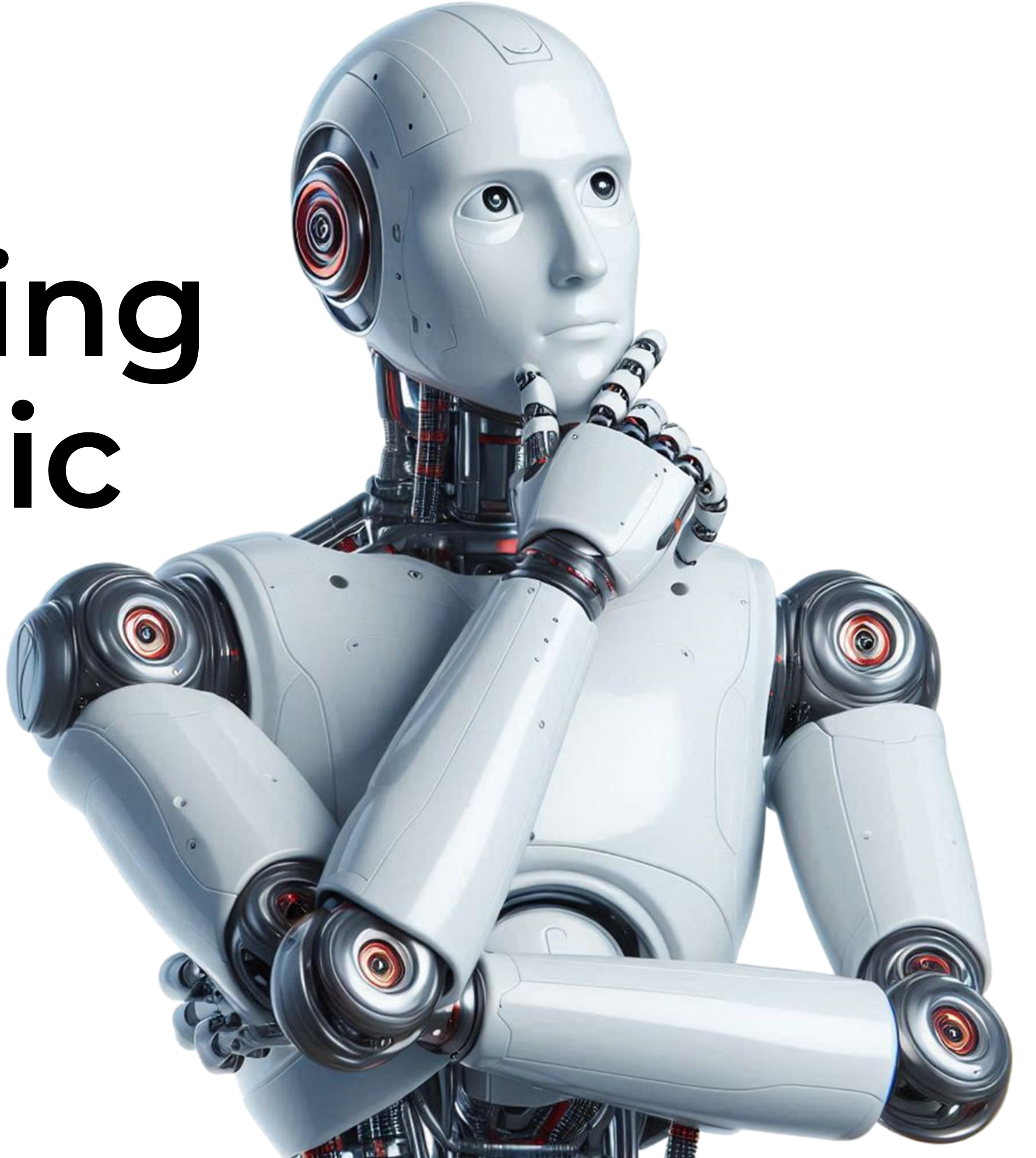


Meta Learning

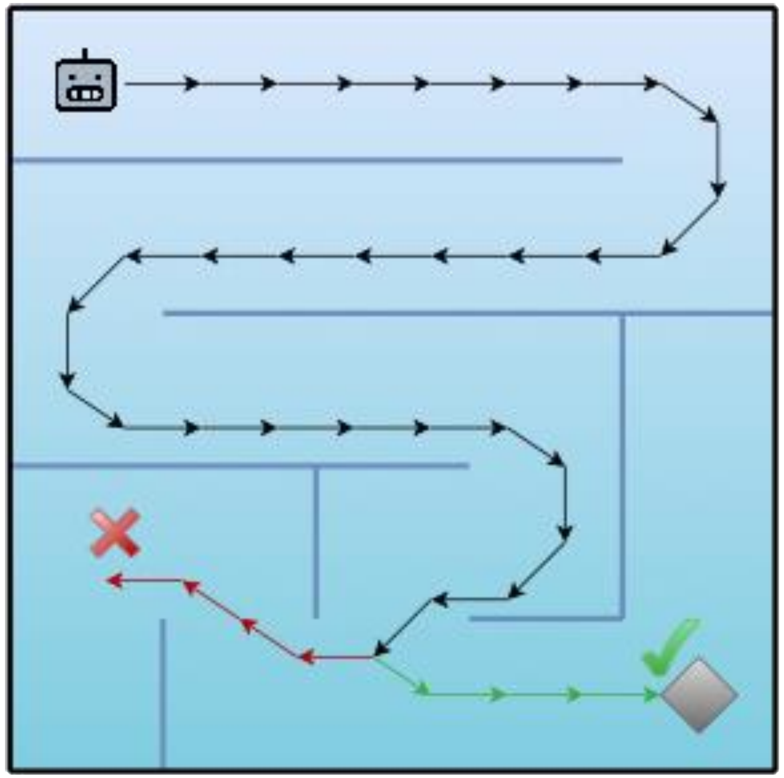


3. Understanding Visual Semantic Navigation

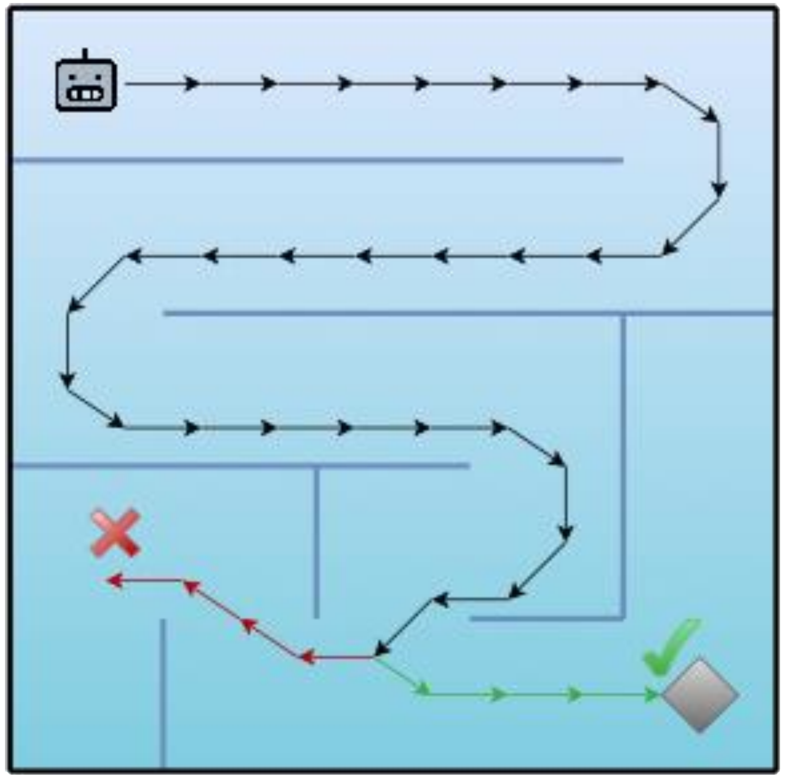
How do we train VSN agents using reinforcement learning



Motivation



Motivation

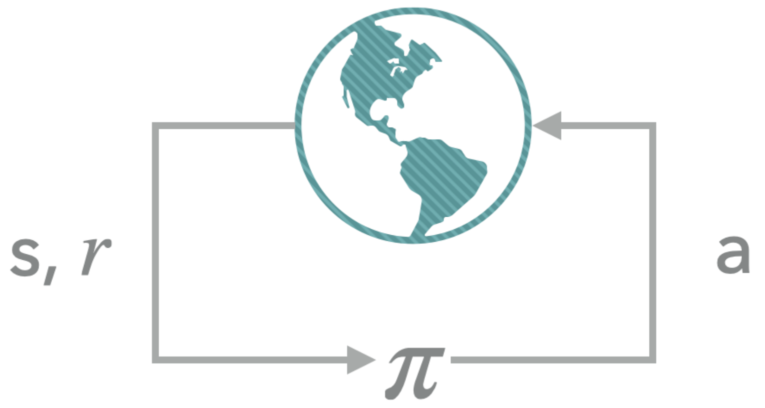


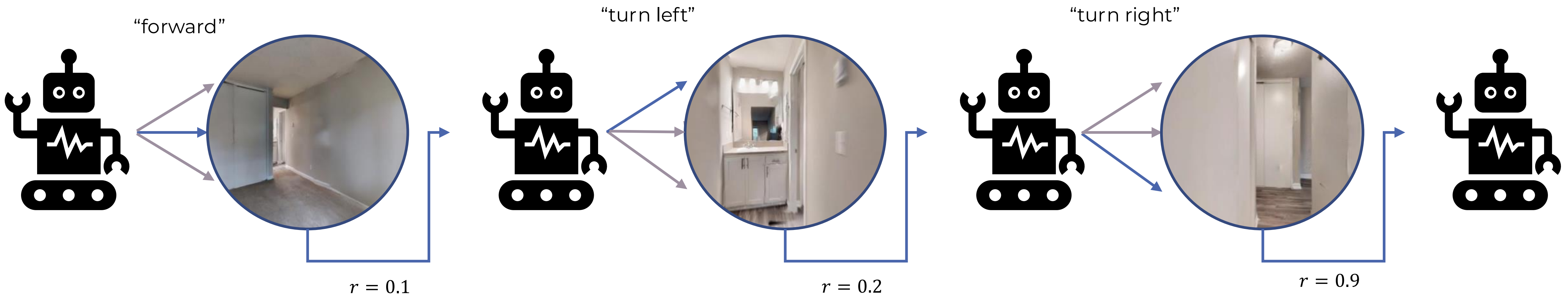
- Can an agent localize a target in an environment given just visual information?
- What are the main challenges a deep reinforcement learning agent has to overcome to successfully navigate to targets within a scene?
- First scientific problem of the thesis.

How to navigate

Reinforcement Learning with PPO

$$\pi_\theta^* = argmax_{\pi_\theta} \mathbb{E}_{\mathcal{T}\sim\pi_\theta}\left[\sum_{t=0}^{\mathrm{H}} r_{a_t}\gamma^{\mathrm{t}-1}\right]$$

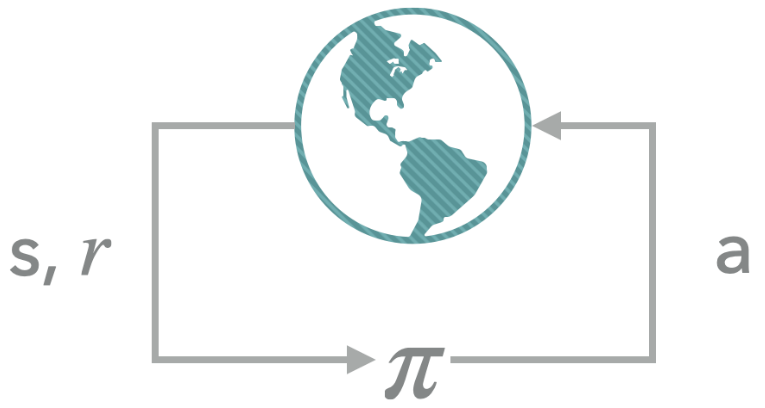




How to navigate

Reinforcement Learning with PPO

$$\pi_\theta^* = argmax_{\pi_\theta} \mathbb{E}_{\mathcal{T}\sim\pi_\theta}\left[\sum_{t=0}^{\mathrm{H}} r_{a_t}\gamma^{\mathrm{t}-1}\right]$$



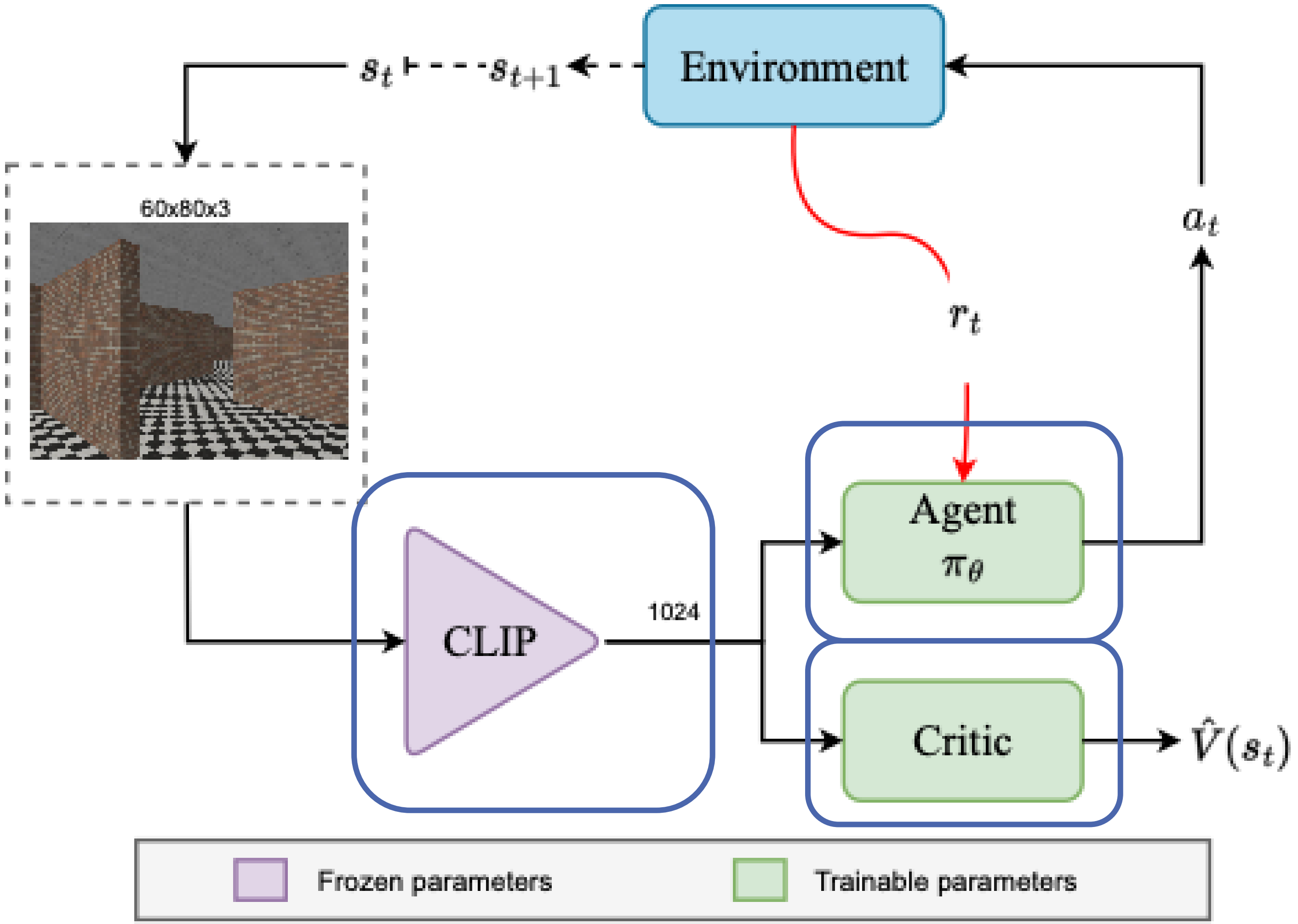
$$L_t^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t\left[\underbrace{L_t^{CLIP}(\theta)}_{\text{surrogate}} - c_1 \underbrace{L_t^{VF}(\theta)}_{\text{value loss}} + \underbrace{c_2 S[\pi_\theta](s_t)}_{\text{entropy loss}}\right]$$

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t\left[\min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\right] \qquad r_t(\theta) = \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\mathrm{old}}}(a_t \mid s_t)}$$

$$L_t^{VF} = (V_\theta(s_t) - V_t^{\mathrm{targ}})^2$$

Actor-critic:
Actor π_θ
Critic $\mathrm{V}_\theta(\mathrm{s}_\mathrm{t})$

How to navigate



Universidad de Alcalá

Problems of RL for navigation

1. How to choose the correct reward function

- Sparse rewards → almost no info for the agent in the environment.
- Dense rewards → gives more info to the agent but must be designed.

2. Trade off between exploration and exploitation

- Exploration is inefficient for navigation, but it has to be done in order to learn the environment.
- Exploitation let the agent use its previous knowledge of the environment to get to the target as quick as possible.

Contextualiza bien los problemas que vamos a resolver luego

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$

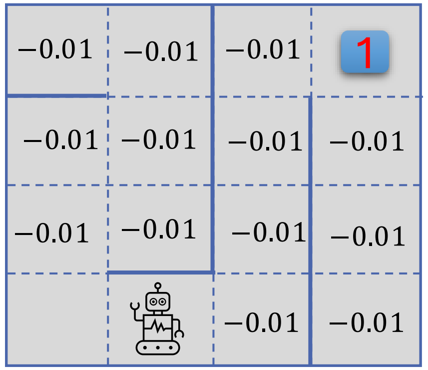
-0.01	-0.01	-0.01	1
-0.01	-0.01	-0.01	-0.01
-0.01	-0.01	-0.01	-0.01
	-0.01	-0.01	-0.01

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



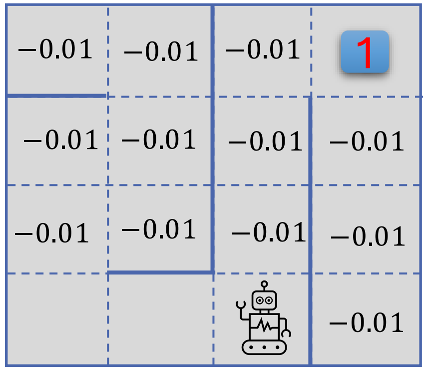
$$r_1 = -0{,}01$$

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



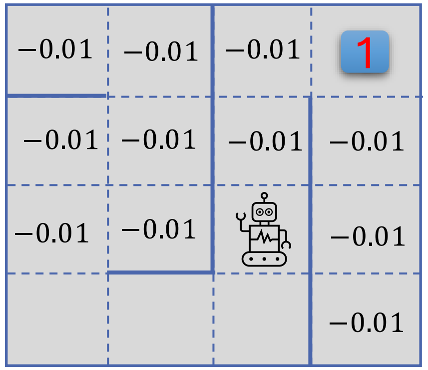
$$r_2 = -0{,}02$$

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



$$r_3 = -0{,}03$$

Universidad de Alcalá

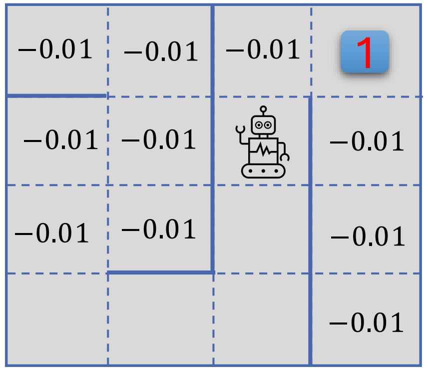
How to choose the correct reward

Sparse Reward

Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



$$r_4 = -0{,}04$$

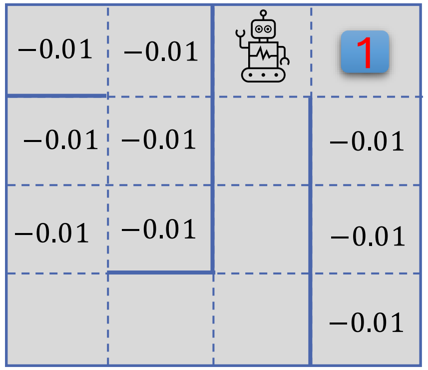
Universidad de Alcalá

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



$$r_5 = -0{,}05$$

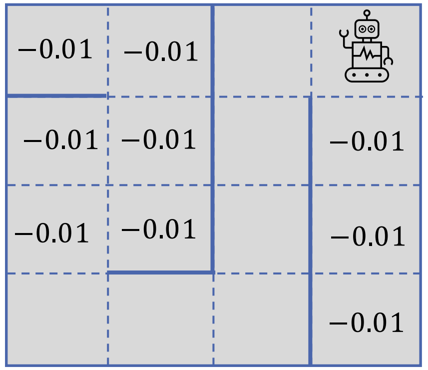
Universidad de Alcalá

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = -r_s + r_T$$



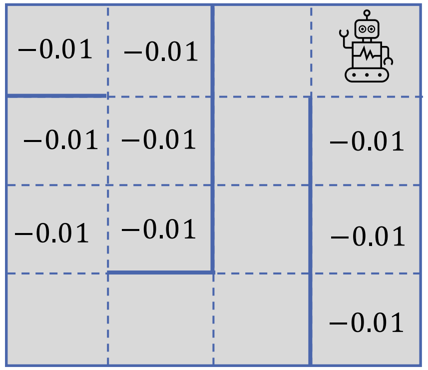
$$r_6 = 0{,}95$$

How to choose the correct reward

Sparse Reward → Rewards present in the environment are zero most of the time, except for when the agent reaches the target.

Navigation Reward

$$r_t = r_s + r_T$$

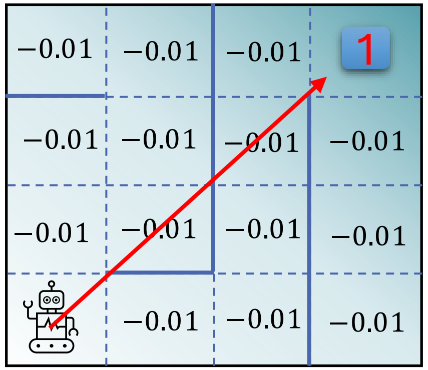


$$r_6 = 0{,}95$$

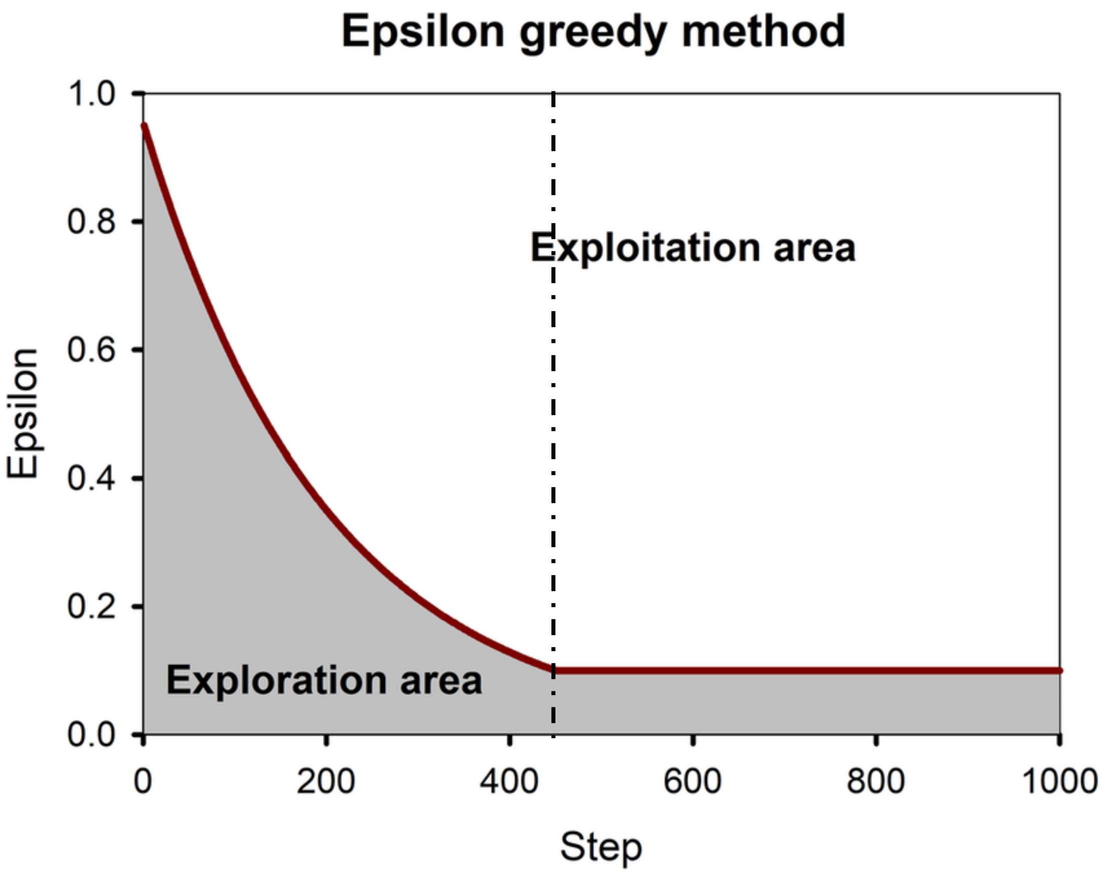
Reward Shaping

Distance Reward

$$r_t = \Delta d_{s_t} + r_s + r_T$$



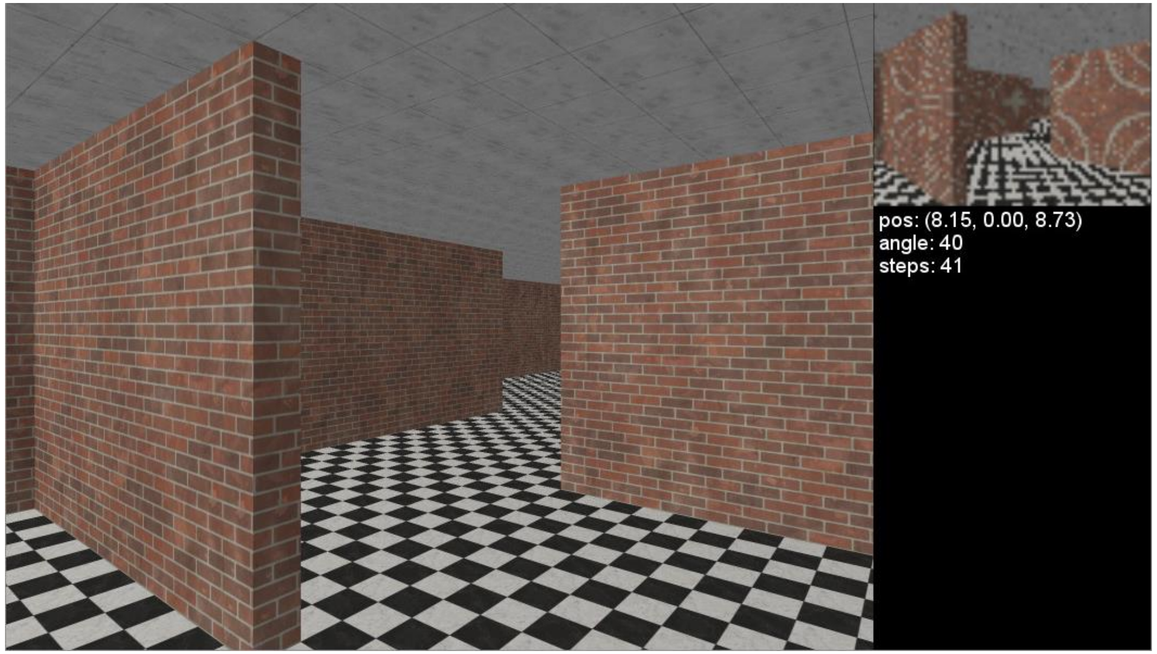
Exploration vs Exploitation

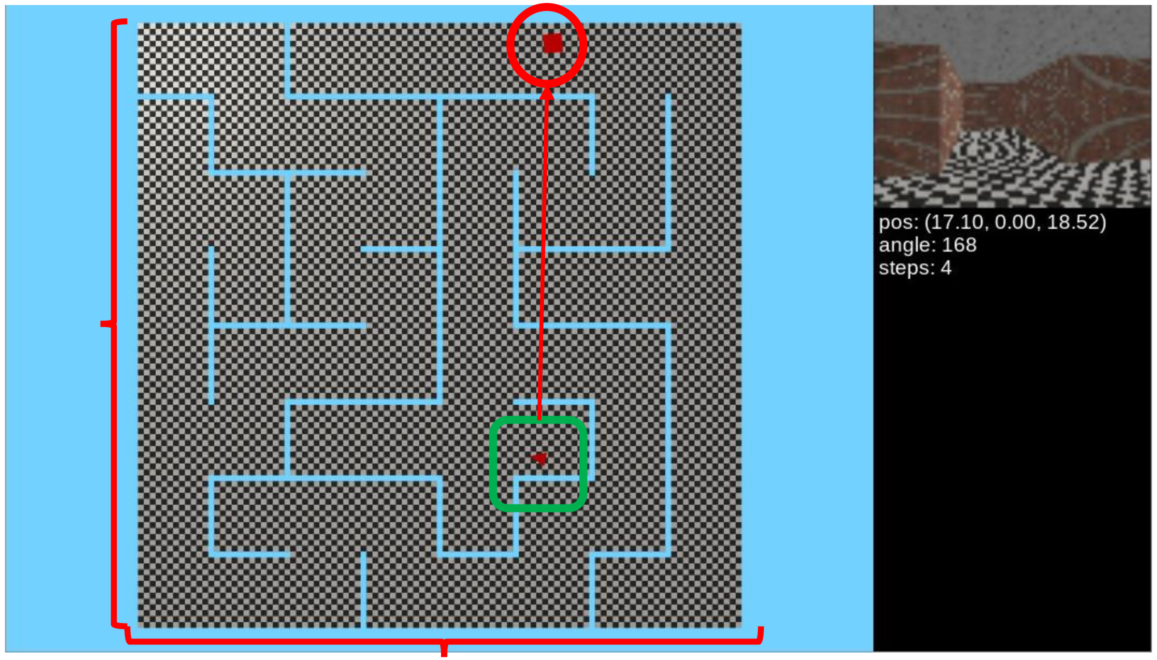


$$a_t = \begin{cases} argmax\ \pi_\vartheta & with\ probability\ 1 - \mathcal{E} \\ rand(a) \in \mathcal{A} & with\ probability\ \mathcal{E} \end{cases}$$

Experimental setup

Miniworld





We use two Maze sizes:

- **S3**: 3X3 tiling.
- **S5**: 5x5 tiling.

Experimental setup

Habitat

Hablar del 100 mazes ese

Experimental setup

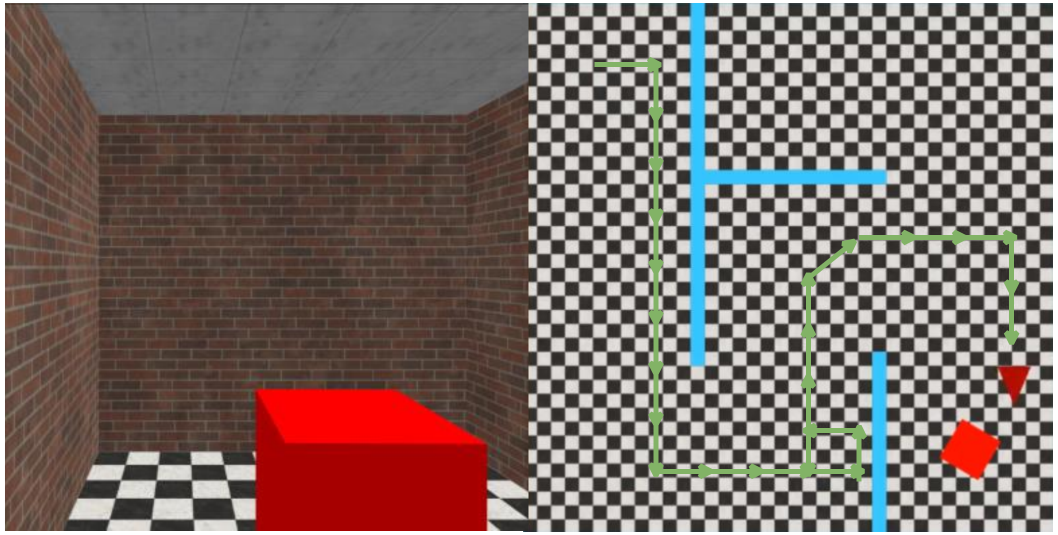
- **Simulators**: Miniworld-Maze and AI Habitat.
- **Task**: Find target in novel indoor environments.
- **Dataset**: HM3D for AI Habitat.
- **Action space**:
  - Move forward, turn left and turn right for Miniworld-Maze.
  - The previous ones plus look_up and look_down for AI Habitat.
- **Metrics**:
  - Success Rate (SR)
  - Steps Per Episode (SPE)
  - Shortest Path Length (SPL)
  - Distance To Goal (DTG)

Miniworld Maze results

Output type	Maze	SR	SPE	Reward
Ours + ϵ-*greedy*	$S3$	**0.75 ± 0.44**	**120.59 ± 111.85**	**6.80 ± 2.29**
	$S5$	**0.18 ± 0.38**	534.40 ± 130.20	**5.24 ± 5.73**
Ours + *stochastic*	$S3$	0.63 ± 0.49	127.42 ± 132.98	6.59 ± 2.41
	$S5$	0.17 ± 0.38	**521.39 ± 182.66**	5.14 ± 5.70
random	$S3$	0.18 ± 0.39	278.04 ± 51.55	0.37 ± 3.66
	$S5$	0.02 ± 0.14	596.07 ± 32.83	-2.09 ± 4.06

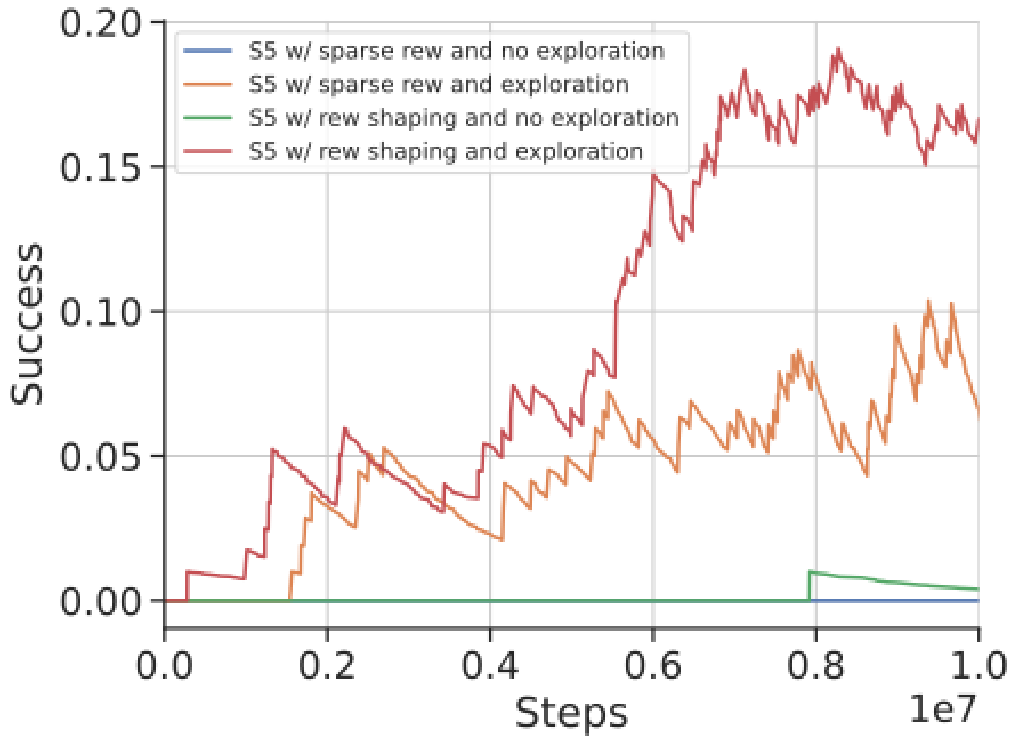
Miniworld Maze results

Output type	Maze	SR	SPE	Reward
Ours + ϵ-*greedy*	$S3$	**0.75** ± **0.44**	**120.59** ± **111.85**	**6.80** ± **2.29**
	$S5$	**0.18** ± **0.38**	534.40 ± 130.20	**5.24** ± **5.73**
Ours + *stochastic*	$S3$	0.63 ± 0.49	127.42 ± 132.98	6.59 ± 2.41
	$S5$	0.17 ± 0.38	**521.39** ± **182.66**	5.14 ± 5.70
random	$S3$	0.18 ± 0.39	278.04 ± 51.55	0.37 ± 3.66
	$S5$	0.02 ± 0.14	596.07 ± 32.83	-2.09 ± 4.06



Ablation study

Reward function	Exploration strategy	SR	SPE	Reward
distance reward	ϵ-*greedy*	**0.18 ± 0.38**	**534.40 ± 130.20**	**5.24 ± 5.73**
navigation reward	ϵ-*greedy*	0.09 ± 0.29	575.86 ± 91.94	0.08 ± 0.26
distance reward	No	0.02 ± 0.14	588.66 ± 79.78	-1.24 ± 4.18
navigation reward	No	0.00 ± 0.00	600.00 ± 0.00	0.00 ± 0.00



Habitat HM3D results

Output type	SR	SPL	DTG	SPE	Reward
Best agent + ϵ-*greedy*	**0.96** $\pm$ **0.19**	0.66 ± 0.25	0.25 ± 0.85	**189.99** $\pm$ **116.97**	**4.96** $\pm$ **1.99**
Best agent + *stochastic*	0.73 ± 0.45	0.58 ± 0.36	0.63 ± 1.17	231.23 ± 188.13	3.52 ± 3.90
random	0.05 ± 0.22	0.02 ± 0.10	4.49 ± 1.72	495.50 ± 26.96	-4.68 ± 2.16

Contar si hay una red un poco distinta pero en la red

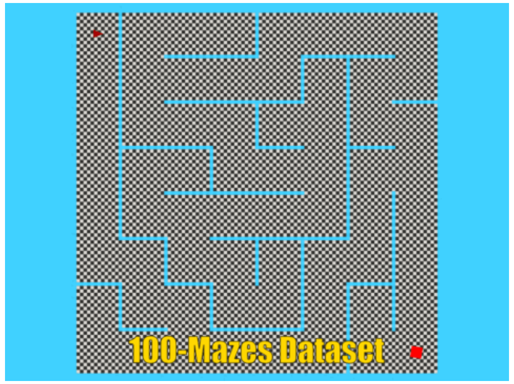
Habitat HM3D results

Output type	SR	SPL	DTG	SPE	Reward
Best agent + ϵ-*greedy*	**0.96 ± 0.19**	0.66 ± 0.25	0.25 ± 0.85	**189.99 ± 116.97**	**4.96 ± 1.99**
Best agent + *stochastic*	0.73 ± 0.45	0.58 ± 0.36	0.63 ± 1.17	231.23 ± 188.13	3.52 ± 3.90
random	0.05 ± 0.22	0.02 ± 0.10	4.49 ± 1.72	495.50 ± 26.96	-4.68 ± 2.16

Mirara a ve
Si tengo
Mas videos

Conclusions

- First paper on VSN and RL.
- Developed a state-of-the-art VSN that can navigate in different environments.
- Release of a collection of 100 mazes dataset.



- Code available in github.

Associated paper:

Towards Clear Evaluation of Robotic Visual Semantic Navigation, 2023

Gutiérrez-Alvarez C., Hernández-García S., Nasri N., Cuesta-Infante Alfredo., López-Sastre RJ.

4. Real World VSN

How actual VSN algorithms behave in the real world

Motivation

Can a robotic agent navigate and interact in the real world as in simulation?

* Images from Ramrakhya et.al. 2023

Motivation

Can a robotic agent navigate and interact in the real world as in simulation?

ROS4VSN
ROS library

+

ANY
VSN model

+

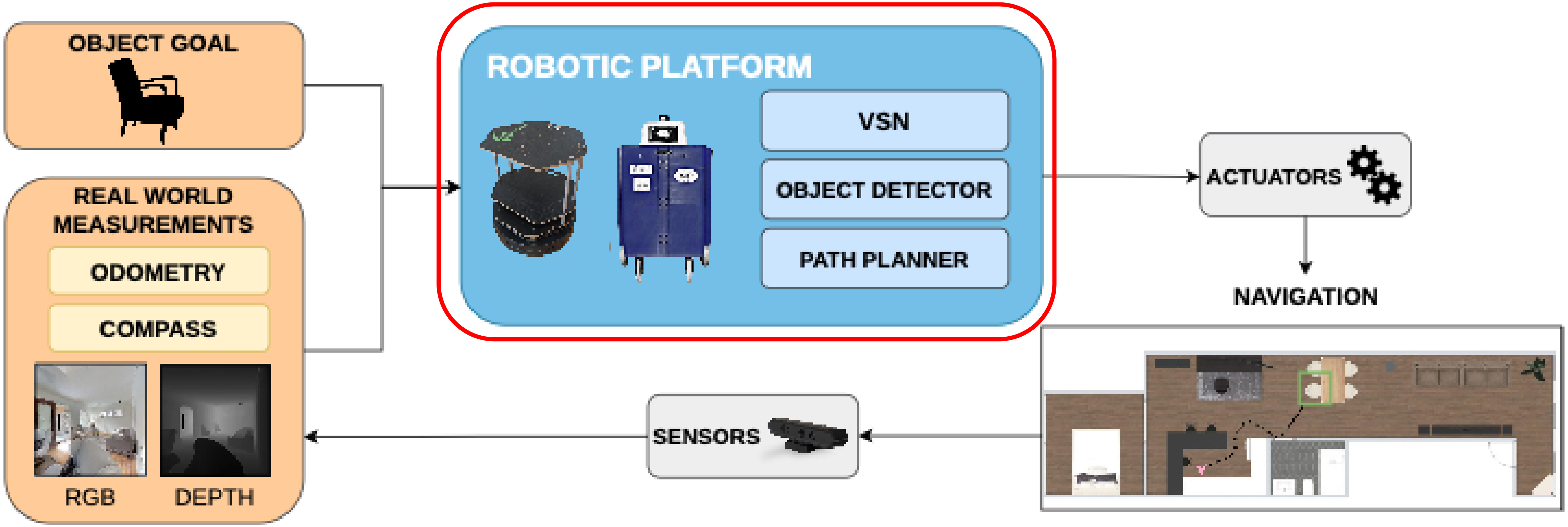
=

Real World VSN

Universidad de Alcalá

Real World VSN with ROS4VSN

Novel ROS library to study how VSN algorithms behave in the real world



The core problem

* Images from Ramrakhya et.al. 2023

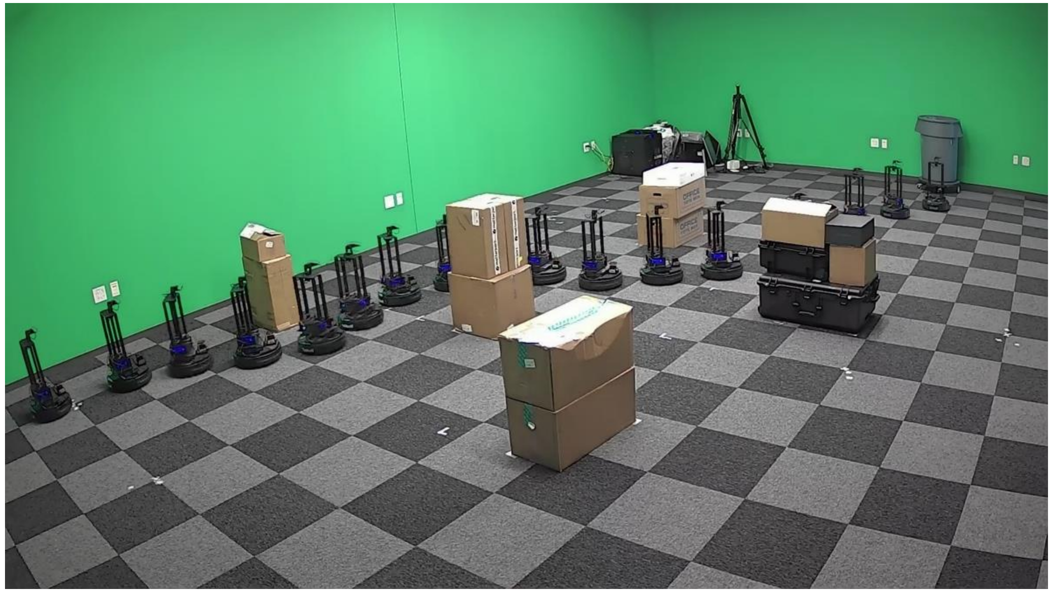
The core problem

* Images grom Ramrakhya et.al. 2023

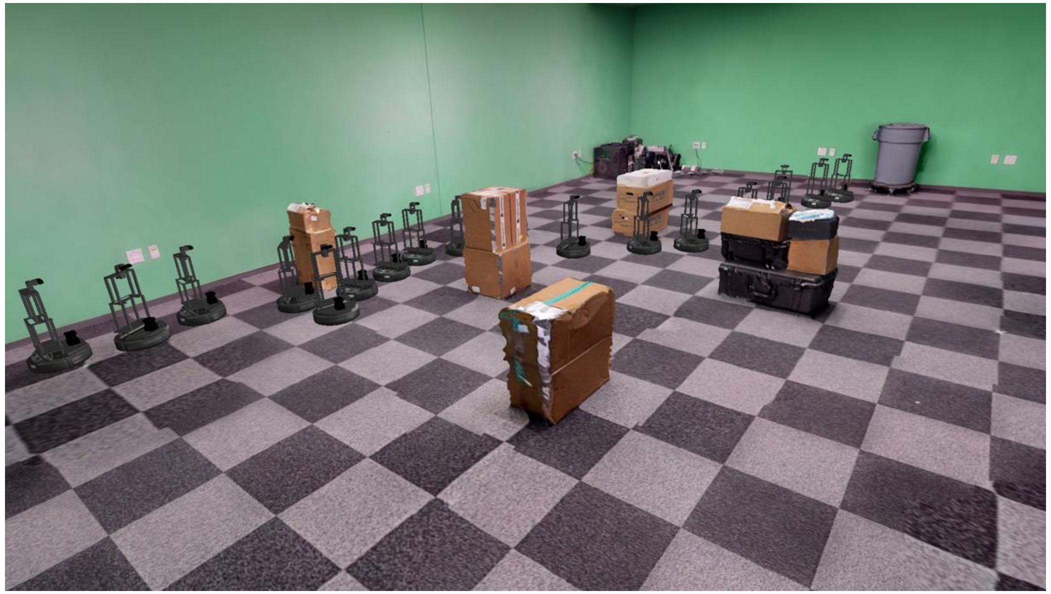
Why simulation is not enough

RGB Domain Gap

Real world



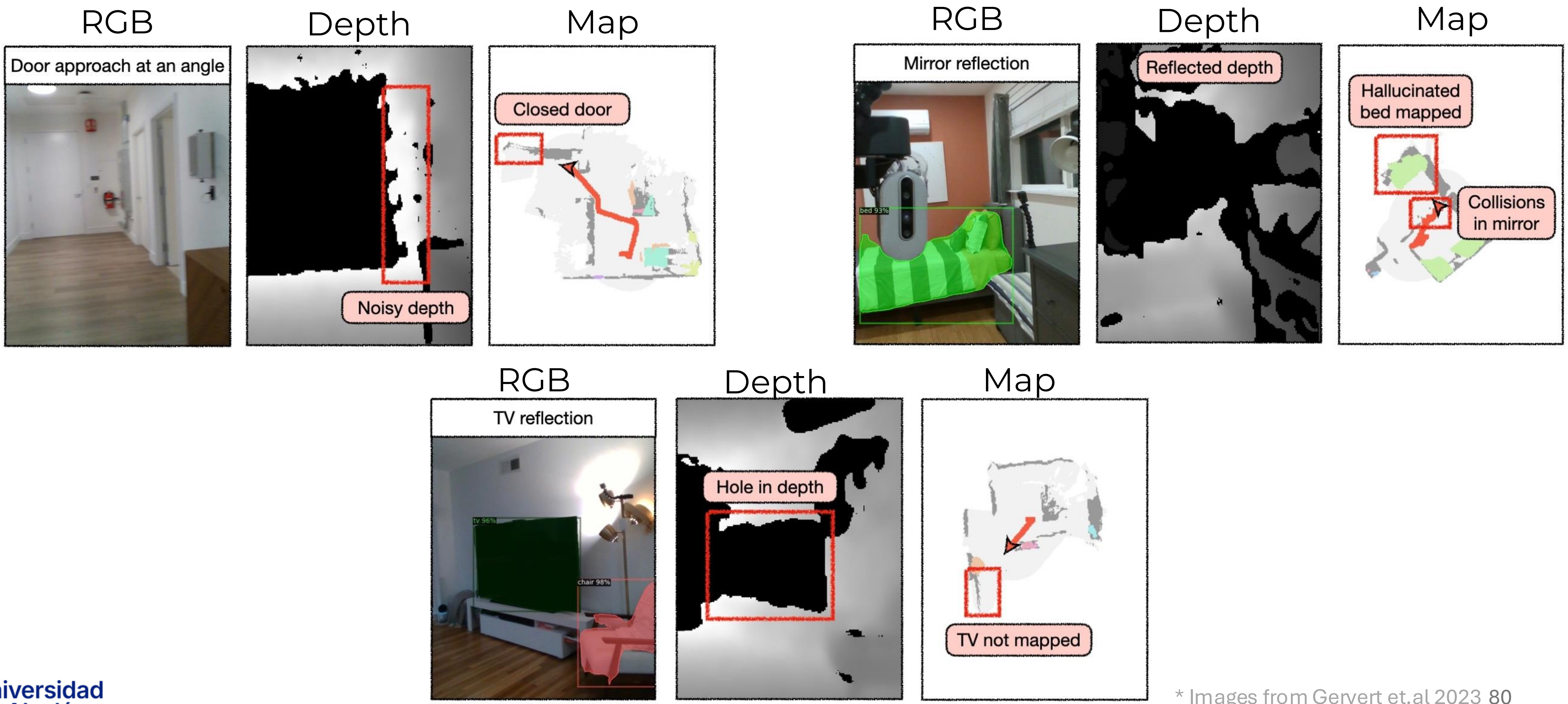
Simulation



* Images from Gervert et.al 2023

Why simulation is not enough

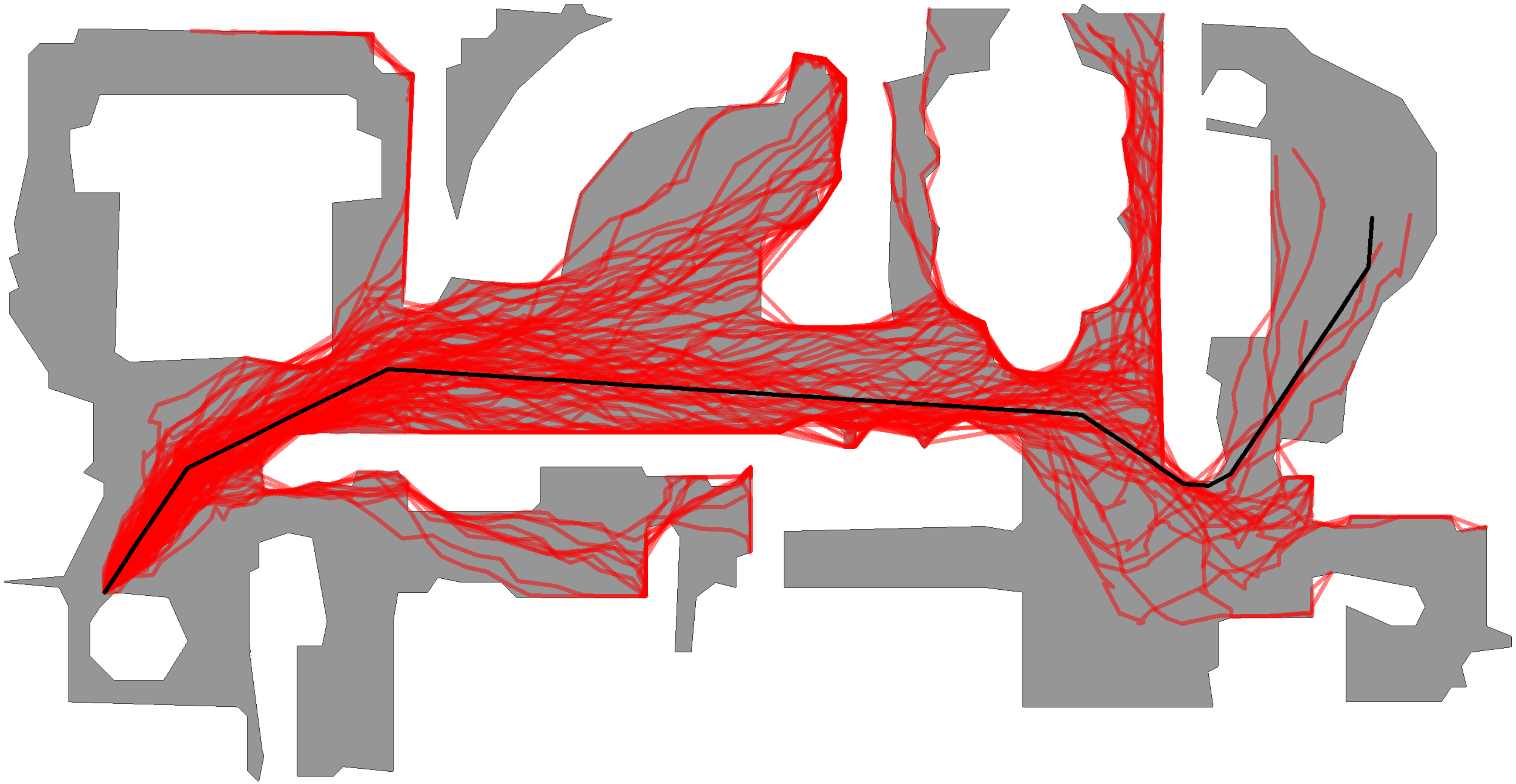
Depth Domain Gap



* Images from Gervert et.al 2023

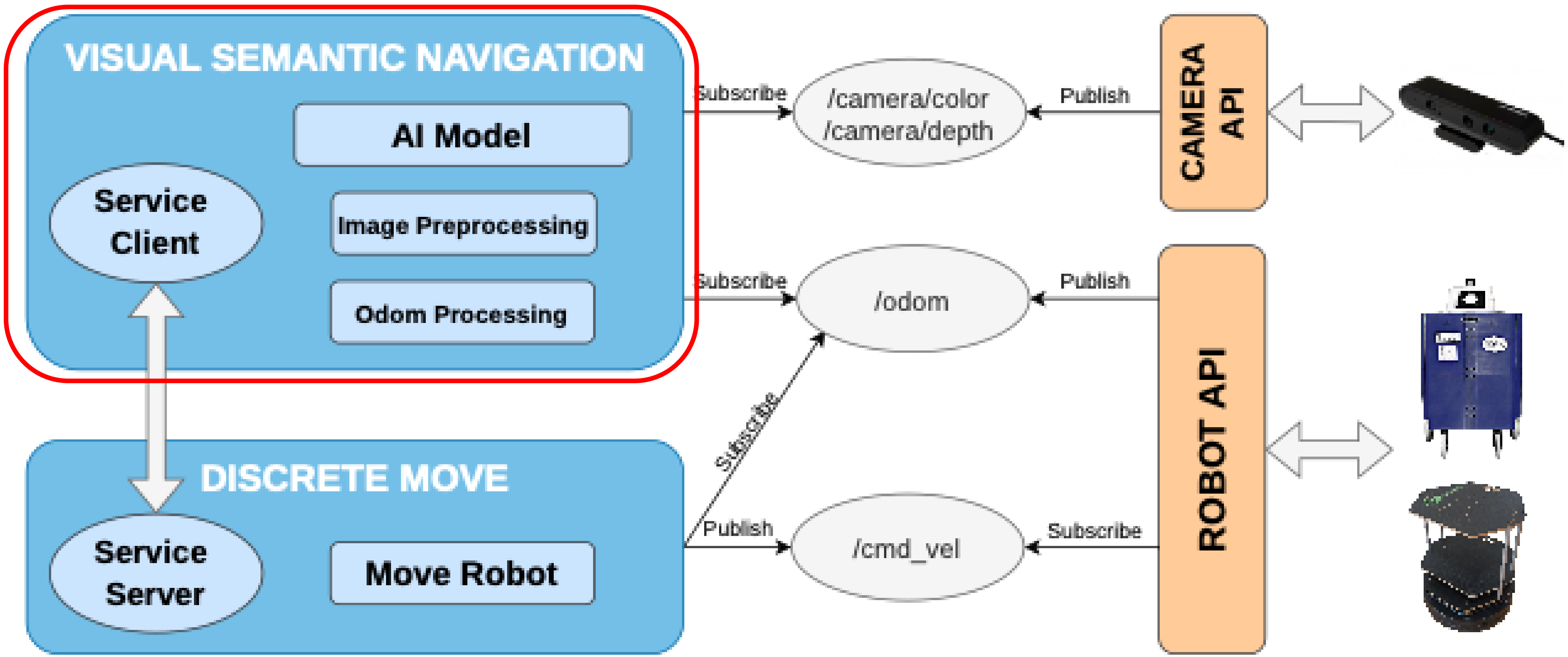
Why simulation is not enough

Actuators Domain Gap



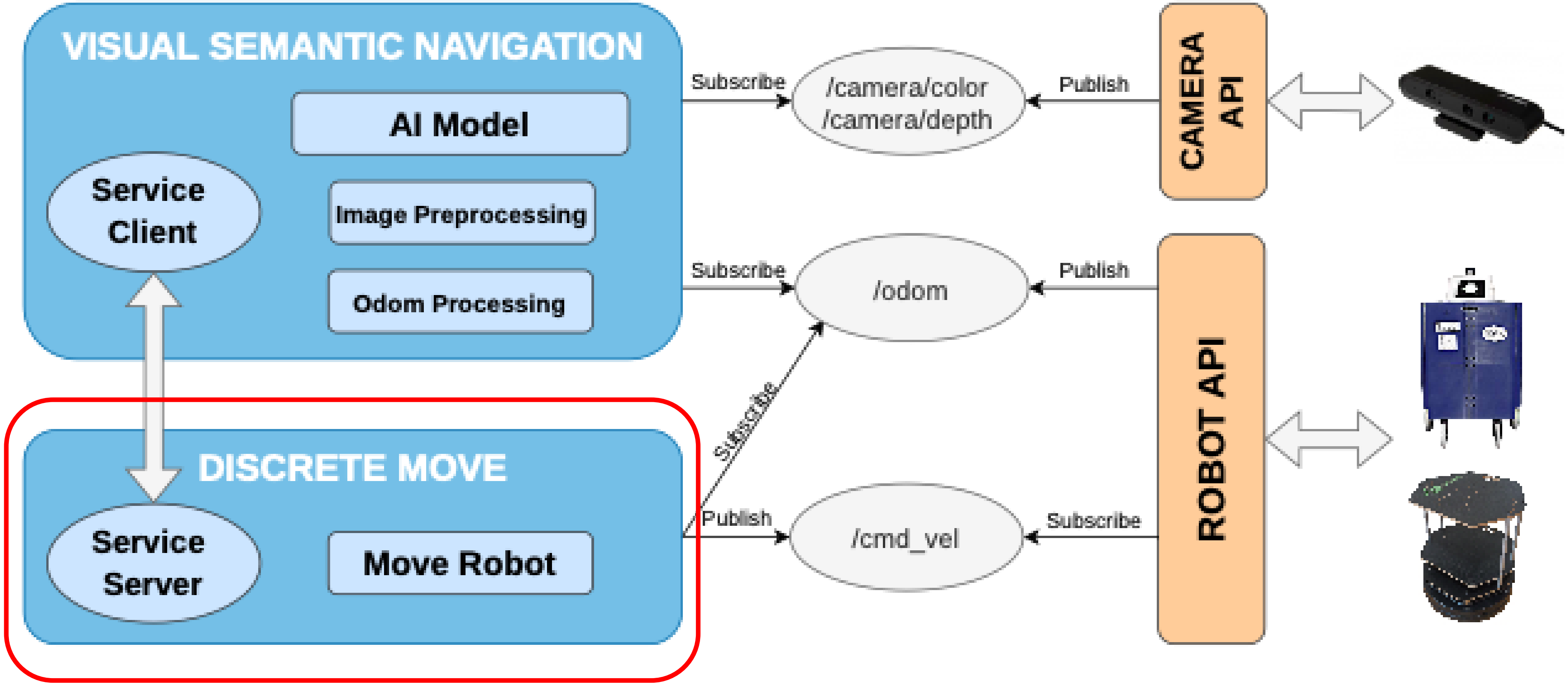
* Images from Kadian et.al 2020

ROS4VSN: System architecture

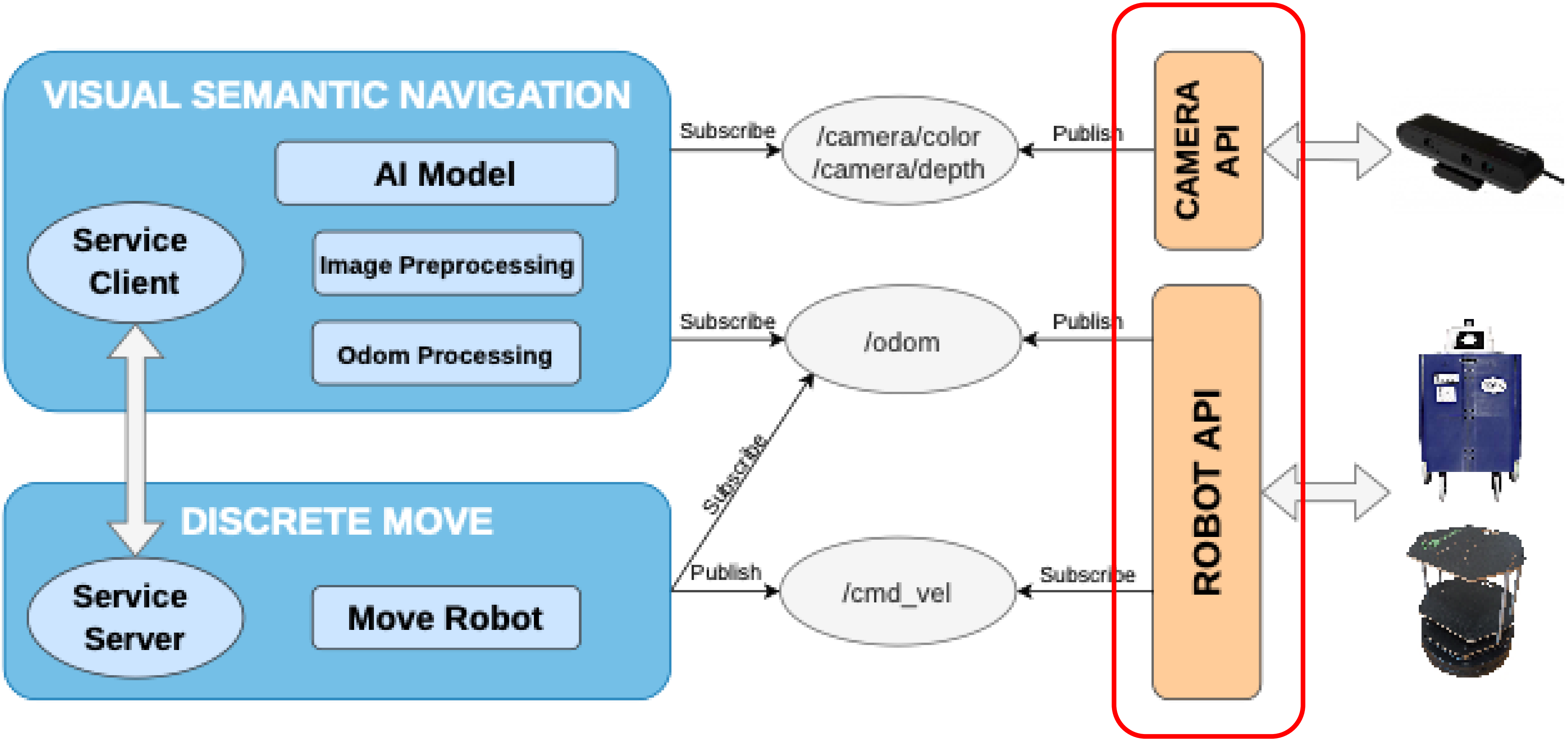


Universidad de Alcalá

ROS4VSN: System architecture

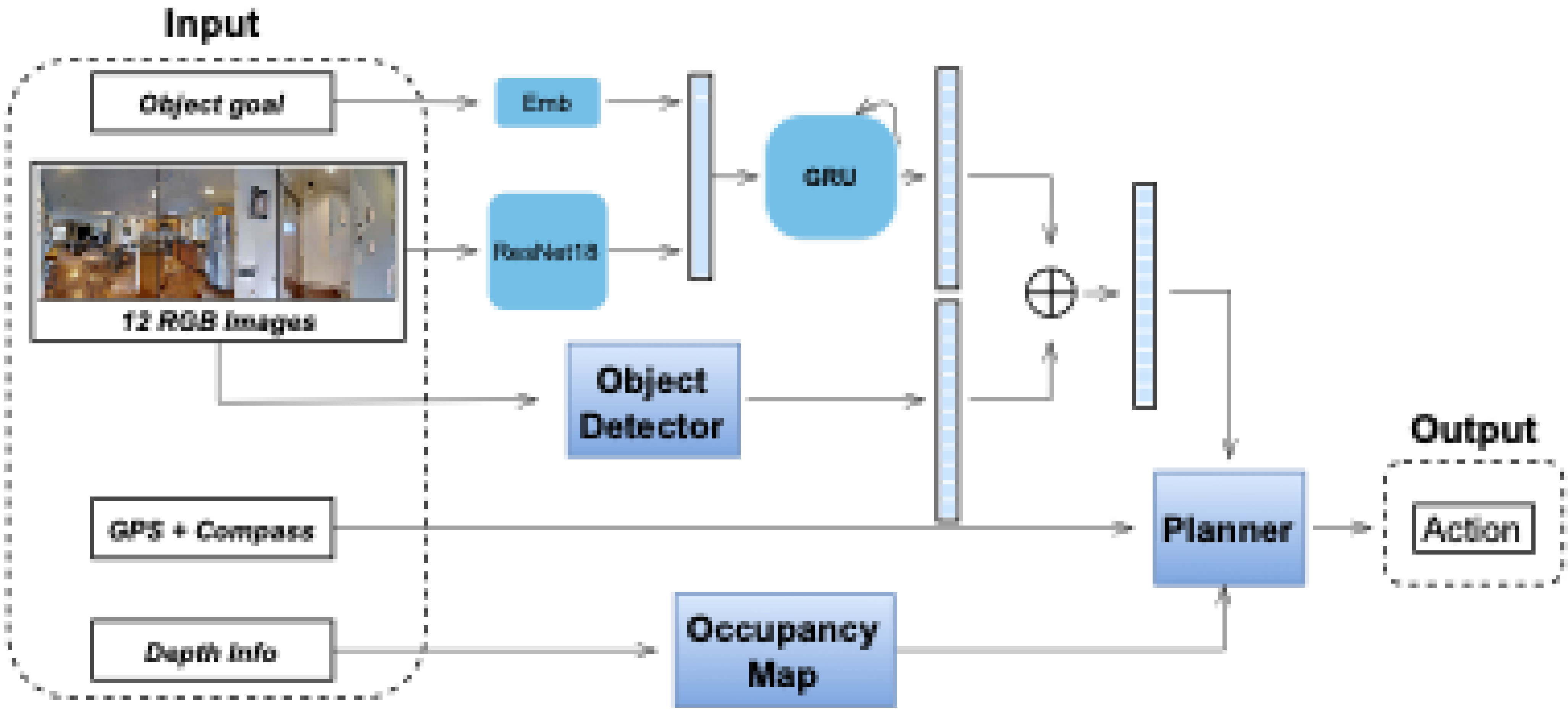


ROS4VSN: System architecture



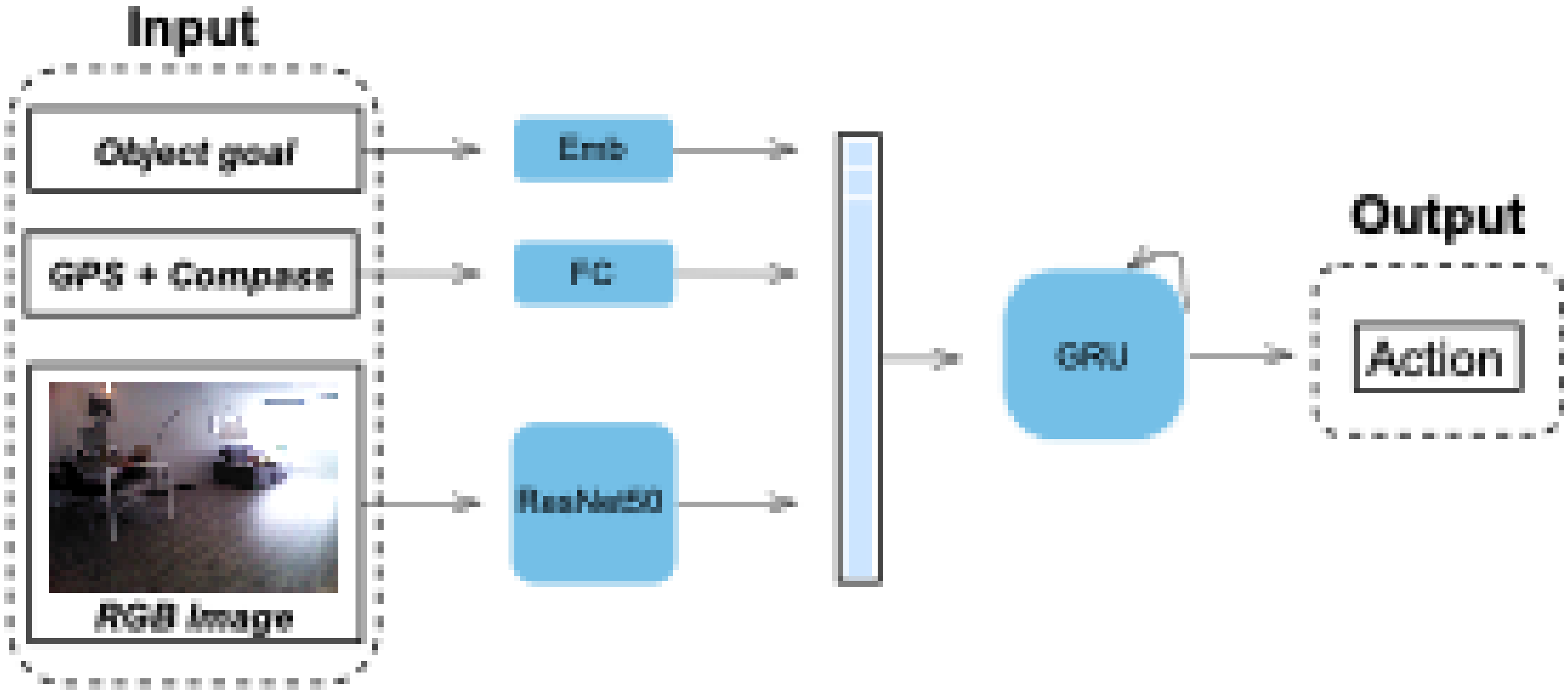
VSN Models Integrated

VLV – Modular learning – Chang et.al 2020



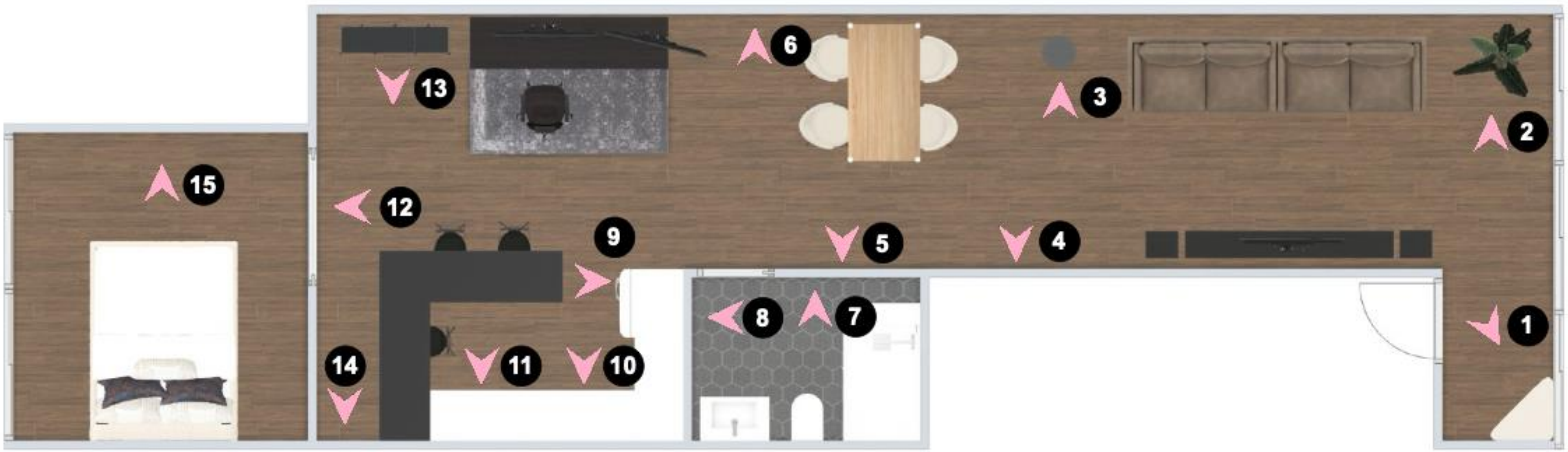
VSN Models Integrated

PIRLNAV – End-to-end learning – Ramrakhya et.al 2023



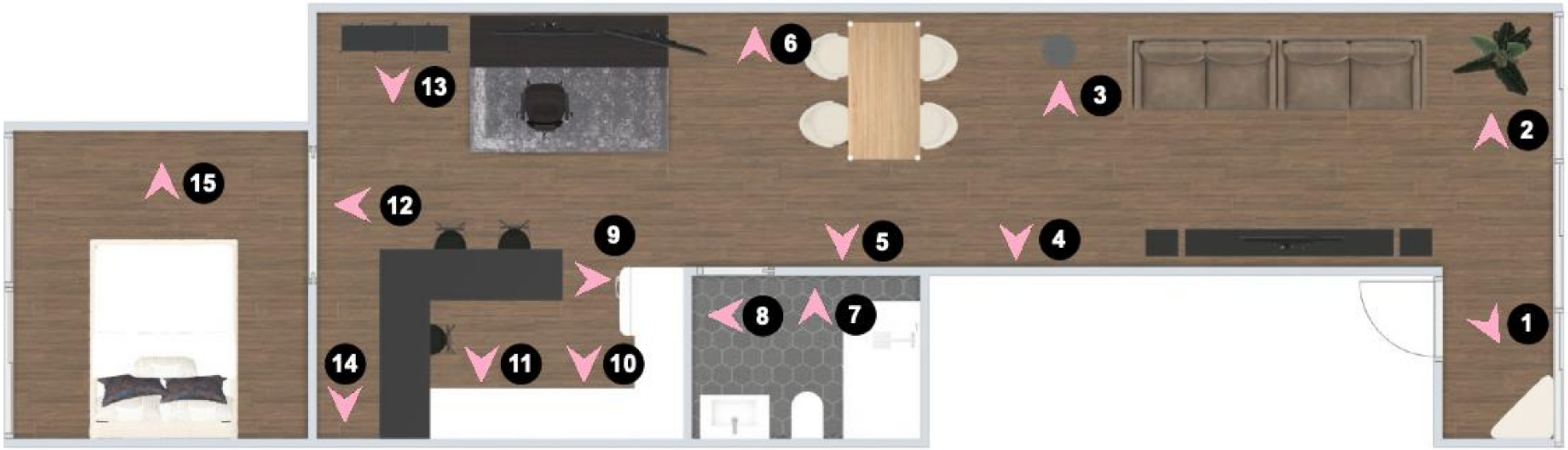
Poner imagen con diagrama que ambos se entrenan

Real world experimental setup



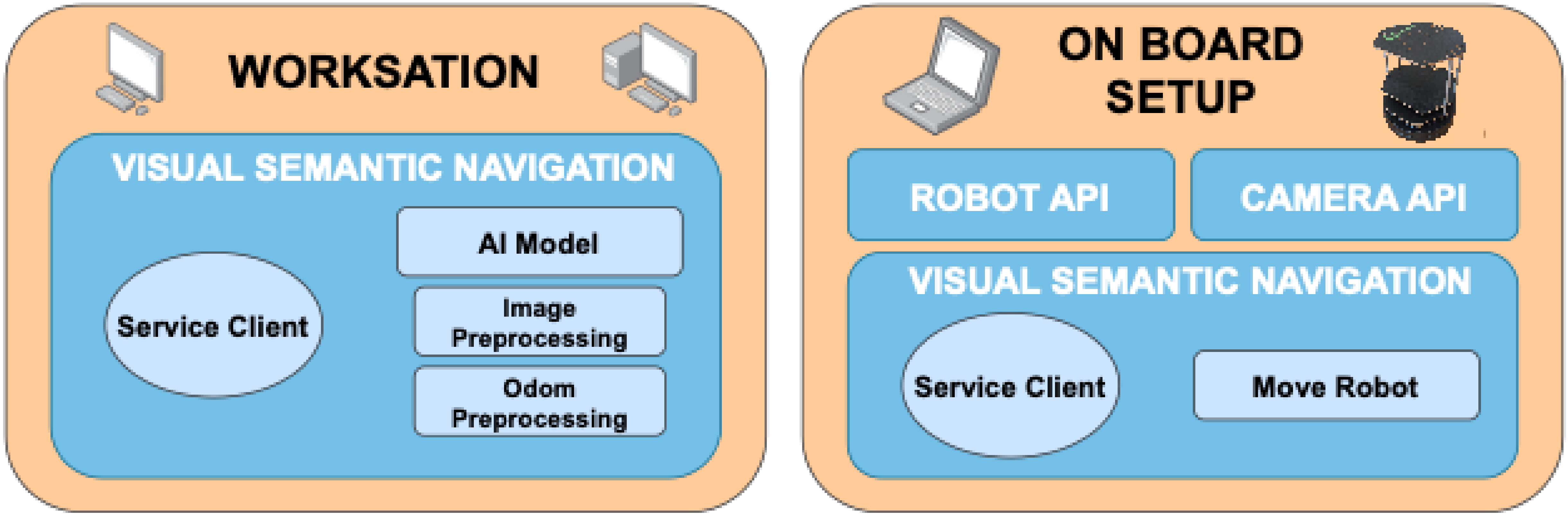
Object Goal
Chair
Sofa
Table
Bed
Toilet

Real world experimental setup



Object Goal
Chair
Sofa
Table
Bed
Toilet

Real world experimental setup



VLV real world results

Universidad de Alcalá

VLV real world results

Object Goal	*Successful episodes*	SR	*Avg. number of actions*
Chair	6/15	40%	30
Sofa	6/15	40%	65
Table	6/15	40%	42
Bed	3/15	20%	39
Toilet	1/15	6,67%	42

PIRLNav real world results

PIRLNav real world results

Object Goal	*Successful episodes*	SR	*Avg. number of actions*
Chair	5/15	33,33%	49
Monitor	5/15	33,33%	91
Sofa	5/15	33,33%	70
Bed	3/15	20,00%	97
Toilet	1/15	6,67%	61
Plant	0/15	0,00%	82

The big numbers

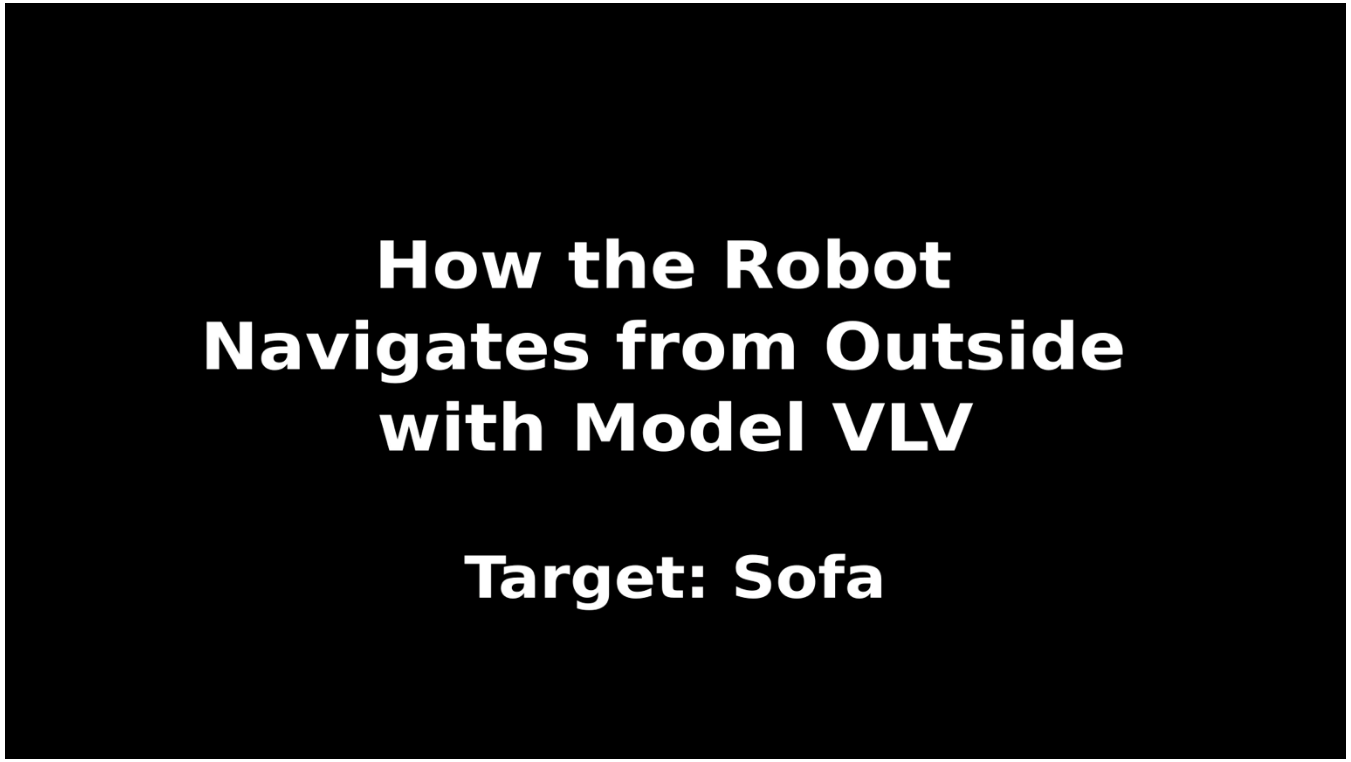
The success rate of end-to-end learning is greater in sim,
but it suffers a larger performance drop in the real world

Models	SR (Real World)	SR (Virtual Environment)
VLV [31]	29.33%	39%
PIRLNAV [45]	21.11%	65%

The big numbers

The success rate of end-to-end learning is greater in sim,
but it suffers a larger performance drop in the real world

Models	SR (Real World)	SR (Virtual Environment)
VLV [31]	29.33%	39%
PIRLNAV [45]	21.11%	65%



Universidad de Alcalá

Conclusions

- Developed a new ROS robotic framework for deploying VSN algorithms in the real world in any robot.
- The ROS4VSN library is very stable with more than 38h and 5km of operation.
- Modular learning wins end-to-end learning in real-world.
- There is still a lot of room for improvement on VSN algorithms to work in the real world.
- Code available in github.

Associated publicaitons:

Springer
Applied Intelligence

Visual Semantic Navigation with Real Robots, 2025

Gutiérrez-Alvarez C., Ríos-Navarro P., Flor-Rodríguez-Rabadán R., Avecedo-Rodríguez FJ., López-Sastre RJ.

2023 iROS

IROS late braking results

Evaluation of Visual Semantic Navigation Models in Real Robots, 2023

Gutiérrez-Alvarez C., Ríos-Navarro P., Flor-Rodríguez-Rabadán R., Avecedo-Rodríguez FJ., López-Sastre RJ.

5. Bridging the gap

Strategies to go easier from simulation to the real world

How to bridge the gap

1. How to do RL with real world data

- Can we use offline RL to train policies that are able to navigate?

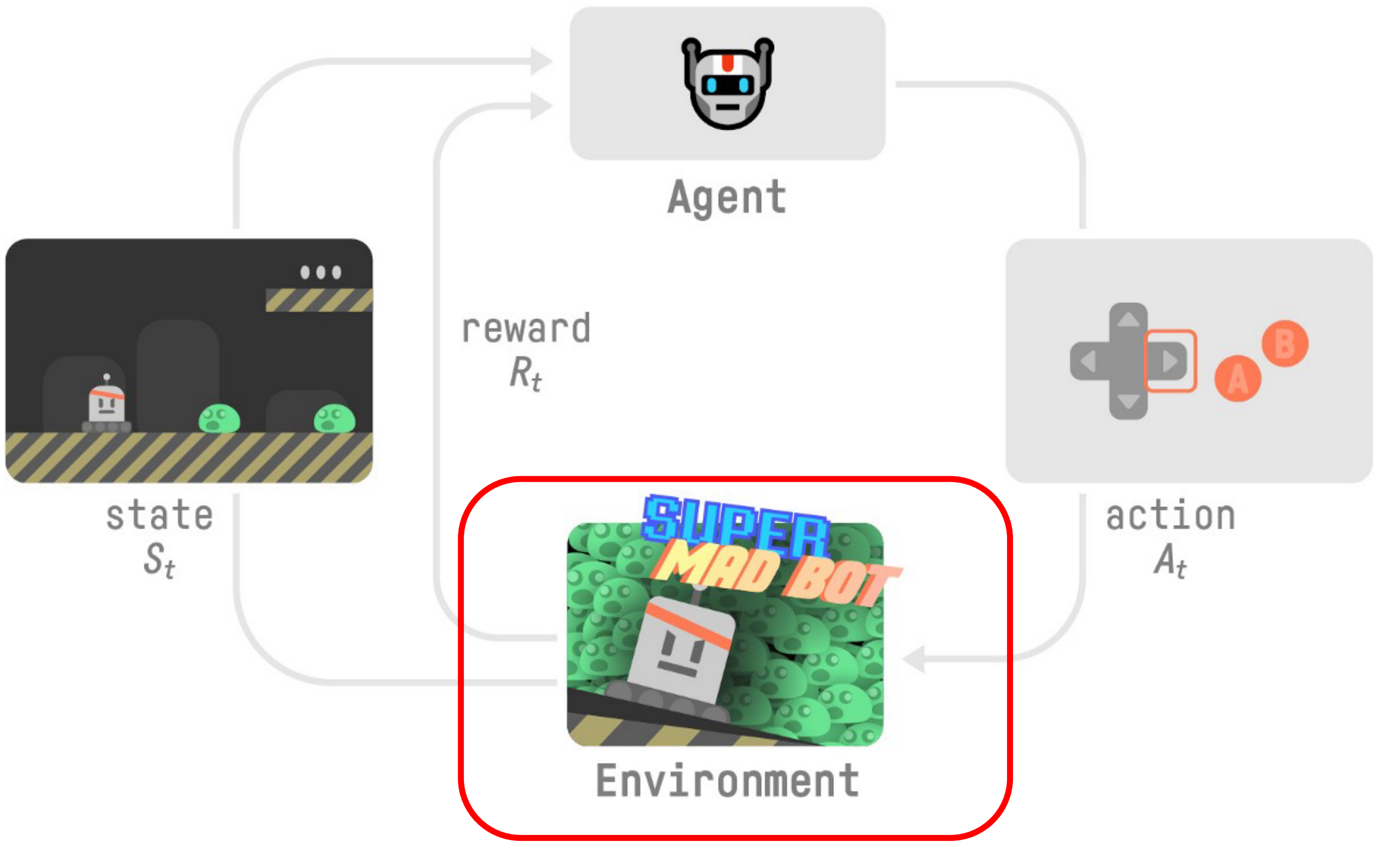
2. How to learn to navigate from a few examples

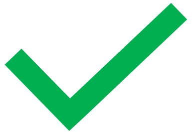
- Can we train meta-algorithms capable of navigate in new environments with few navigation trajectories?

Universidad de Alcalá

Why standard RL is not enough

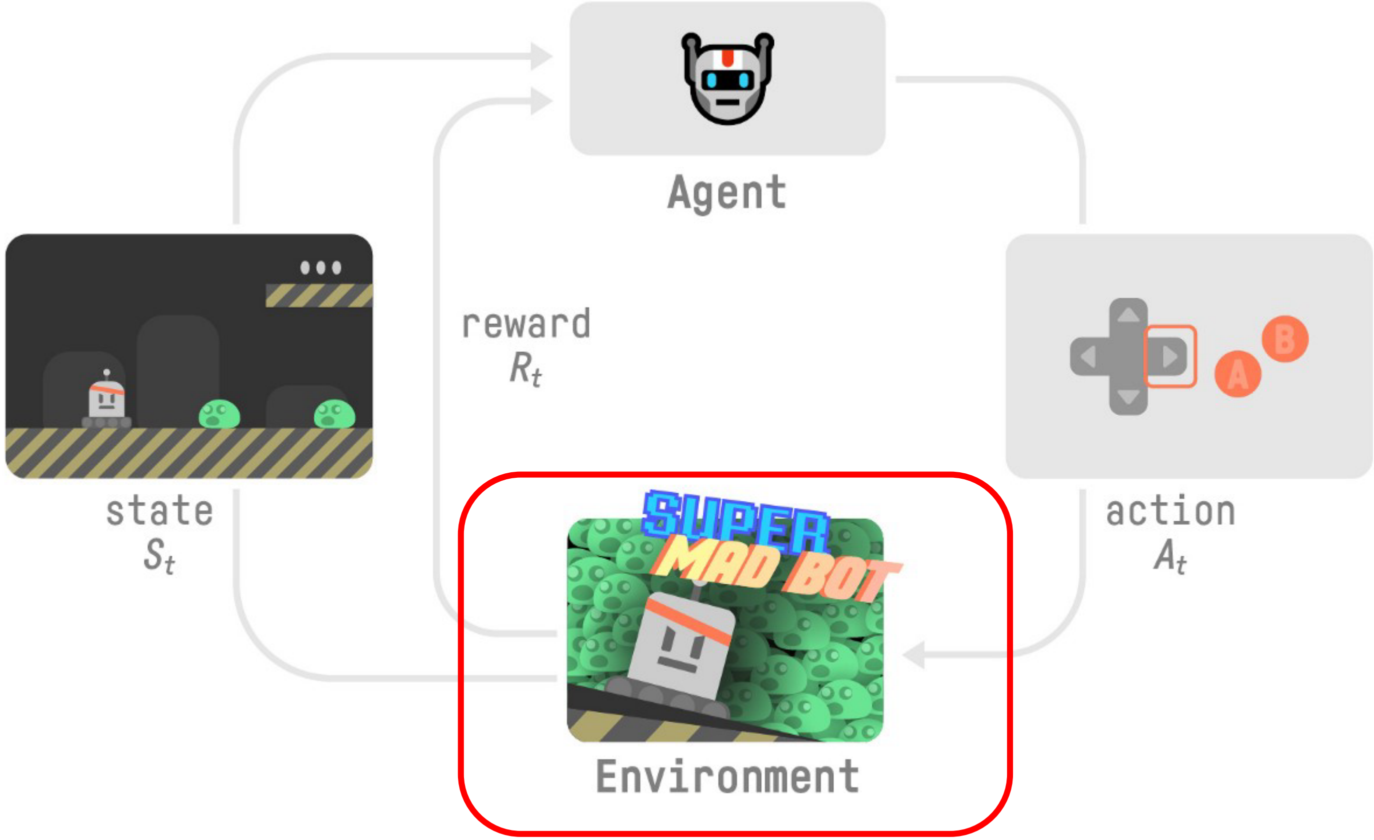
Why standard RL is not enough





Simulation

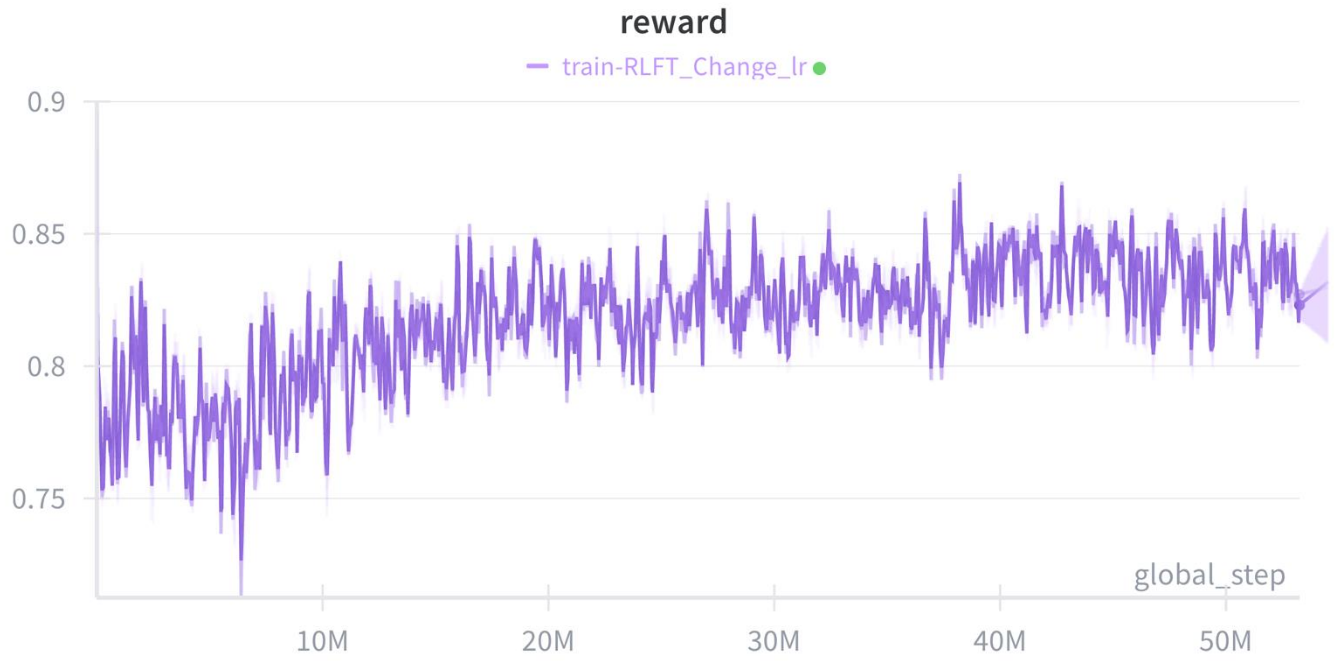
Why standard RL is not enough



Simulation

Real World

Why standard RL is not enough



- 50M steps took 50h.
- Trained on a 4GPU compute node at 170fps.
- Suppose a real robot can perform 1 action per second:

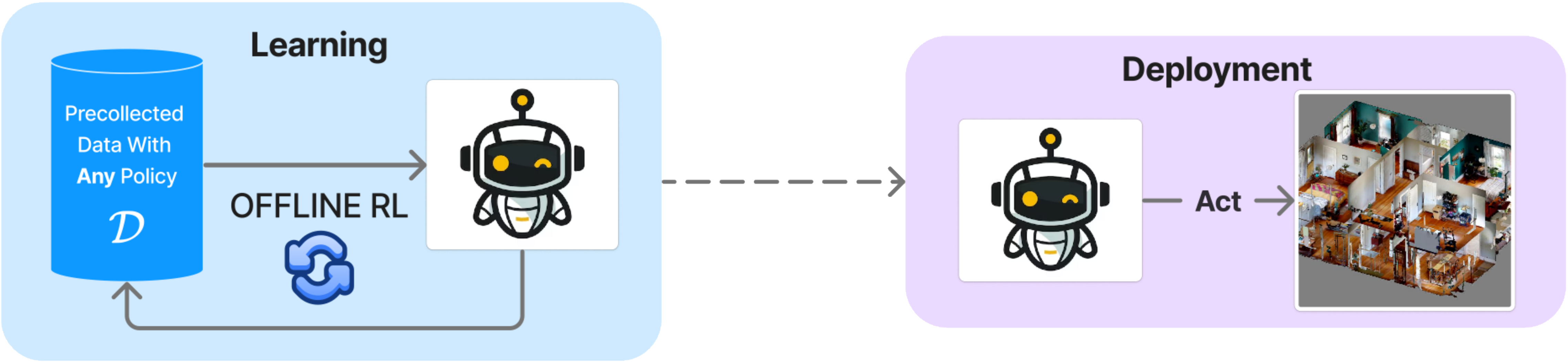
50M interaction steps would take a whole year in the real world!

Why standard RL is not enough

What if we could use precollected datasets?

Offline Reinforcement Learning

Offline RL consists of learning from a fixed dataset of trajectories without ever querying the environment.



OffNav: offline RL without extrapolation

- OffNav is an offline RL framework for visual semantic navigation.
- It is based in Implicit Q-Learning algorithm [1] adapted to work with habitat simulator.

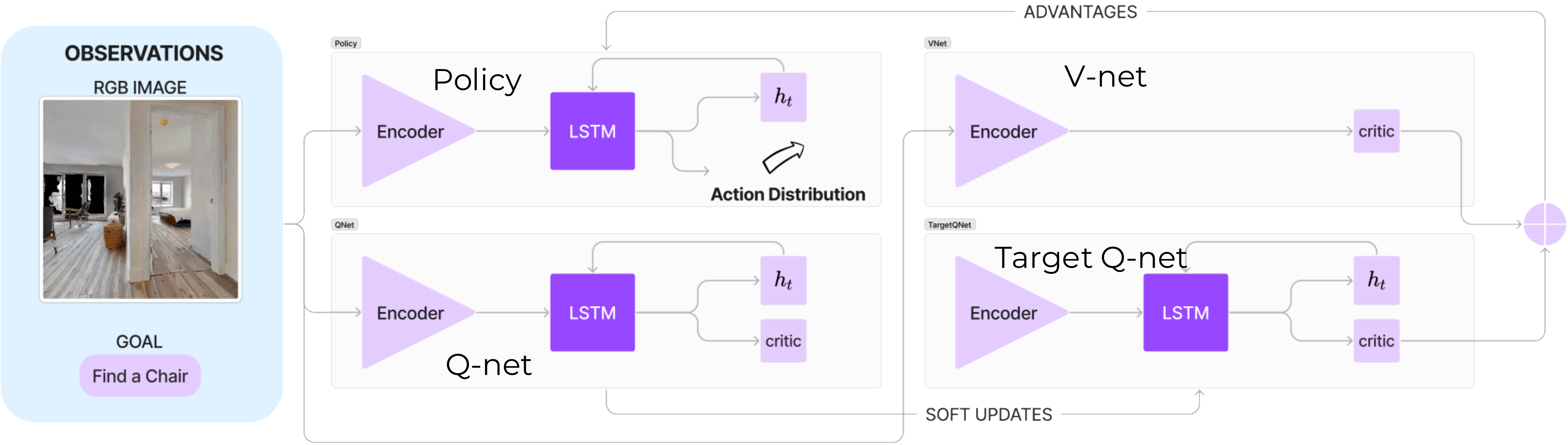
$$L_V(\psi) = \mathbb{E}_{(s,a)\sim\mathcal{D}}\left[\overbrace{L_2^\tau}^{\text{expectile regression}}\left(\underbrace{Q_{\hat{\theta}}(s,a)}_{\text{in-distribution}} - V_\psi(s)\right)\right]$$

$$L_Q(\theta) = \mathbb{E}_{(s,a,s')\sim\mathcal{D}}\left[\left(r(s,a) + \gamma V_\psi(s') - Q_\theta(s,a)\right)^2\right]$$

$$L_\pi(\phi) = \mathbb{E}_{(s,a)\sim\mathcal{D}}\left[\exp\left(\boxed{\beta}\underbrace{\left(Q_{\hat{\theta}}(s,a) - V_\psi(s)\right)}_{\text{maximum of Q values}}\right)\underbrace{\log \pi_\phi(a \mid s)}_{\text{behavior cloning}}\right]$$

[1] Kostrikov et.al 2021

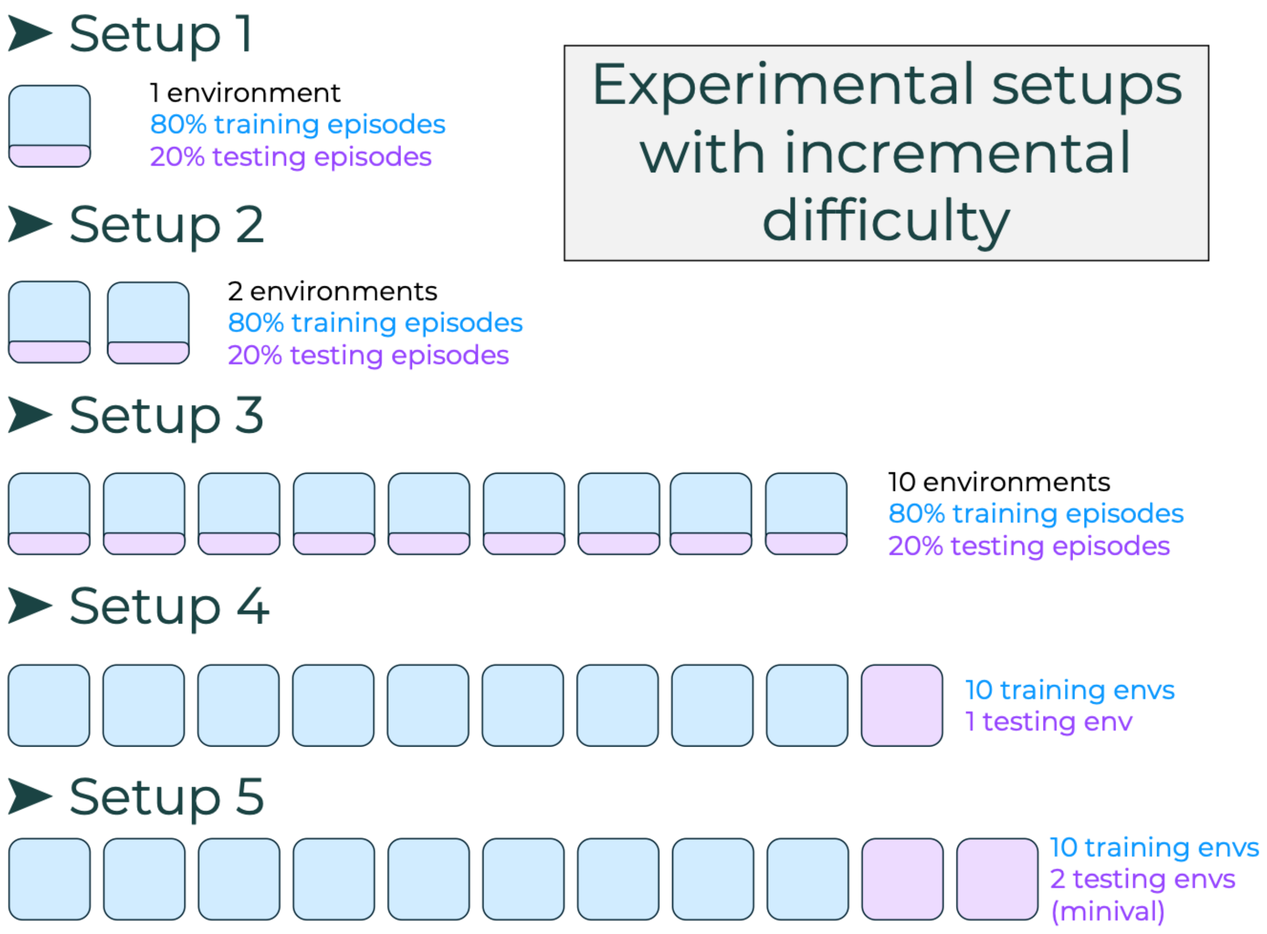
OffNav: offline RL without extrapolation



[1] Kostrikov et.al 2021

Experimental setups

- The model implemented is very heavy, consuming up to 80GB of VRAM for 8 envs.
- That's why this work uses an incremental experimental setup.
- A normal habitat HM3D experimental setup consists of 80 training scenes and 20 validation environments.



Experimental results

Success rate agains behavior cloning baseline (PirlNav)

Experimental Setup	*OffNav*	*PirlNav*
Setup 1	100%	100%
Setup 2	**79.31%**	72.50%
Setup 3	75.78%	**77.63%**
Setup 4	25.00%	**27.27%**
Setup 5	**34.78%**	26.09%

How to bridge the gap

1. How to do RL with real world data

- Can we use offline RL to train policies that are able to navigate?

2. How to learn to navigate from a few examples

- Can we train meta-algorithms capable of navigate in new environments with few navigation trajectories?

Real data collection problems

- OffNav algorithm was trained with 77k human recorded trajectories in habitat simulator.
- On chapter 4, the robots spent 38h operating to achieve a total of 150 trajectories.

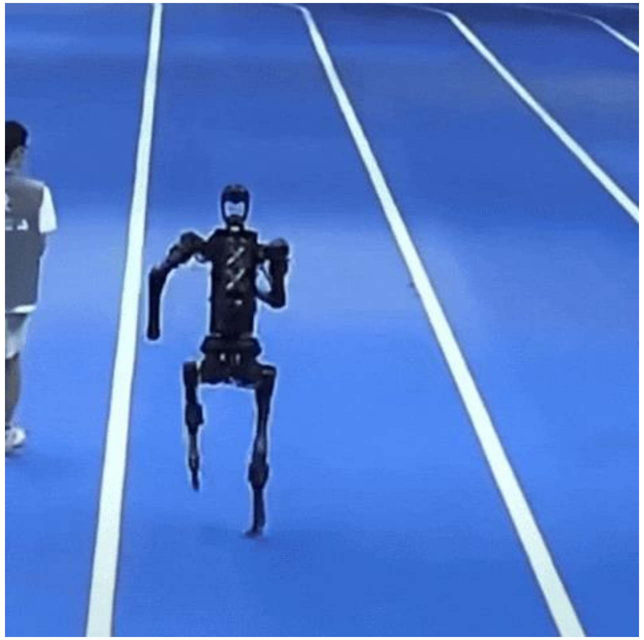
Collecting 77k trajectories would take more than two years in the real world!

Universidad de Alcalá

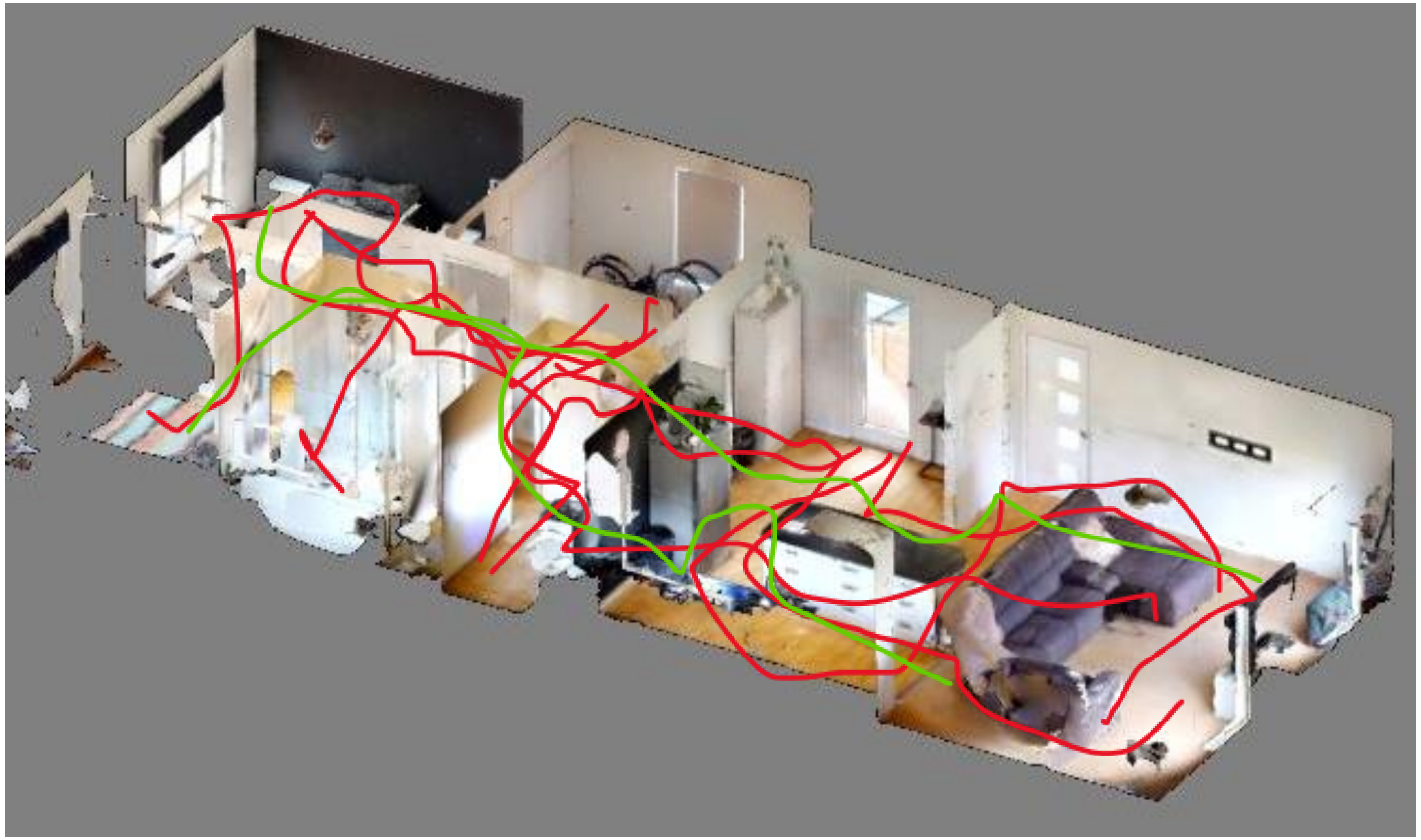
* Images from Gervert et.al 2023

Real data collection problems

Data collection can be risky!



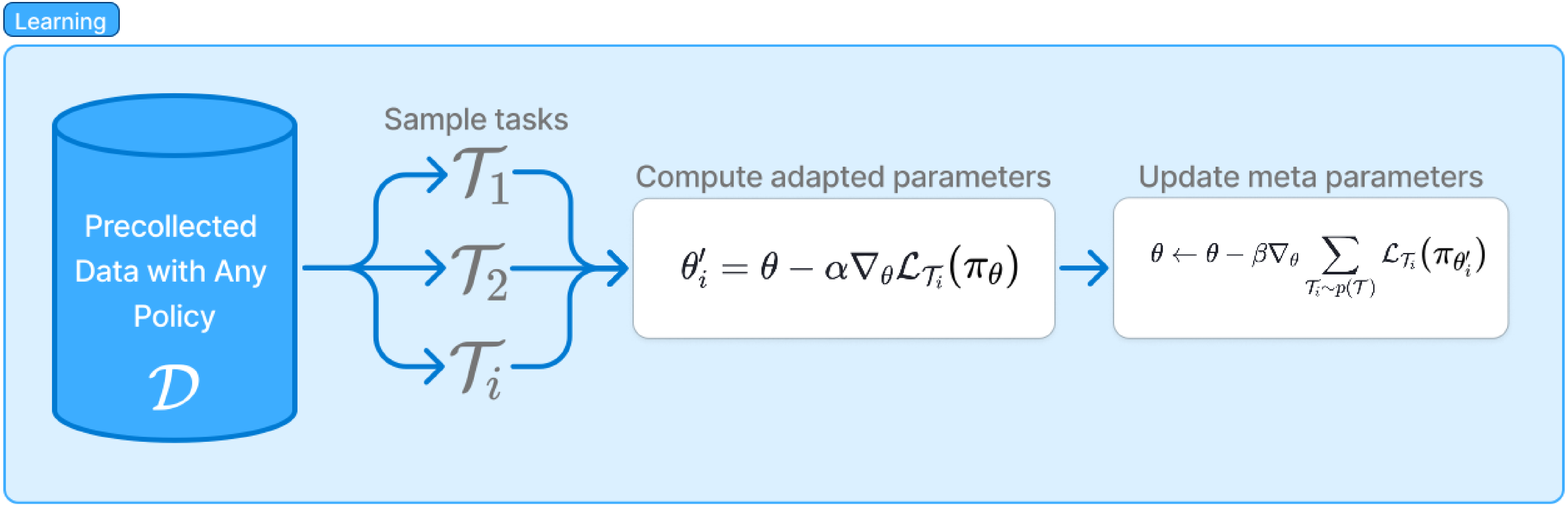
Why meta-imitation learning?



Why meta-imitation learning?

- Few demonstrations.
- Fast adaptation.
- Better generalization.

MetaNav: Learning to adapt



Universidad de Alcalá

MetaNav: Learning to adapt

Deployment

Simulation $\mathcal{T}_1$

Compute adapted parameters

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(\pi_\theta)$$

Model parameters adapted to new **simulation** task

Simulation $\mathcal{T}_2$

Compute adapted parameters

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(\pi_\theta)$$

Model parameters adapted to new **simulation** task

Universidad de Alcalá

MetaNav: evaluation

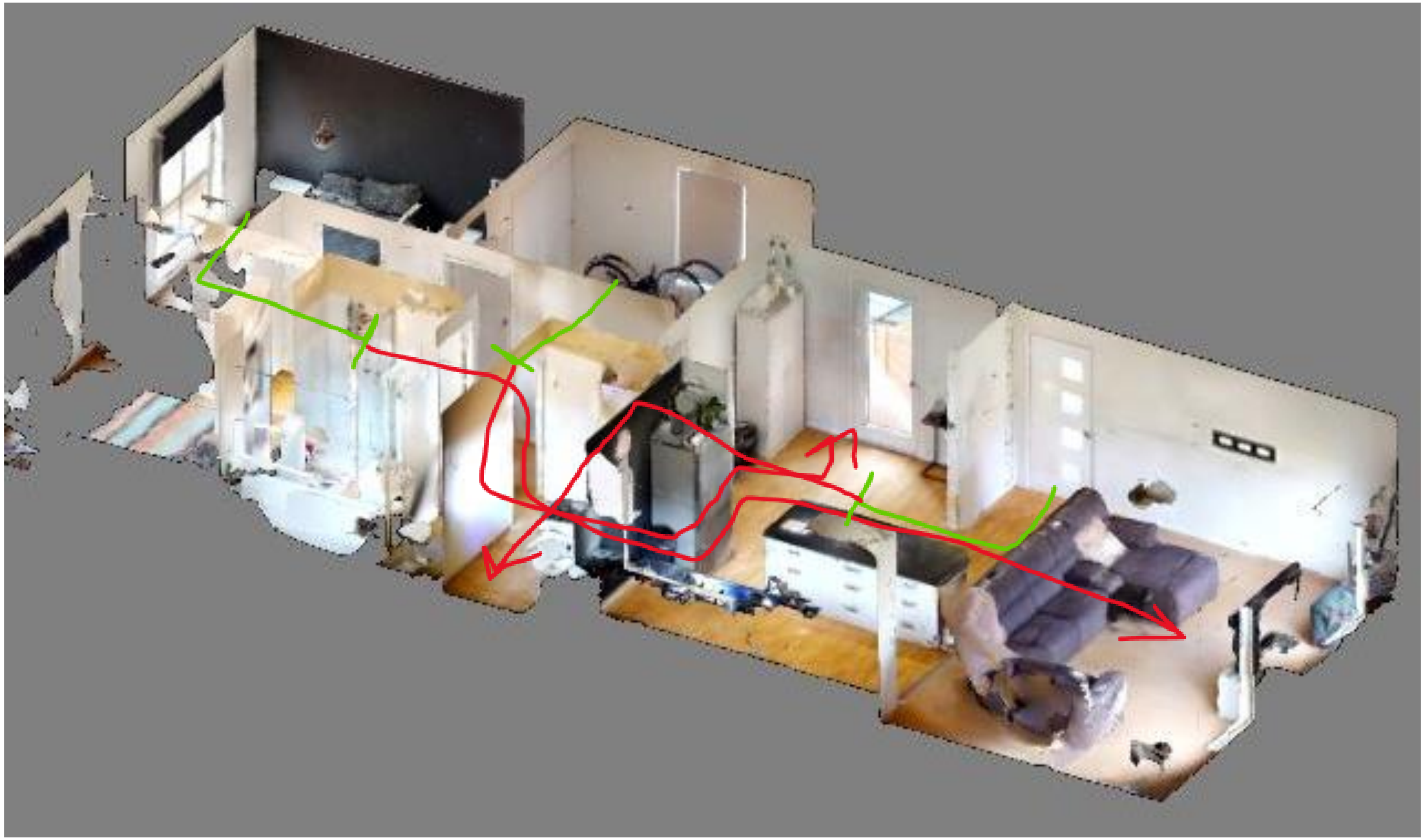
Continuous evaluation

Experience

Evaluation

MetaNav: evaluation

Per-episode evaluation



Experience

Evaluation

MetaNav: experimental results

Continuous evaluation

Setup	SR (↑)	SPL (↑)	Distance to Goal (↓)
1	89.18%	40.04%	0.29
2	76.10%	33.92%	0.97
3	64.19%	33.11%	1.99
4	23.07%	11.87%	12.23
5	21.74%	9.38%	7.99

Per-episode evaluation

Setup	SR (↑)	SPL (↑)	Distance to Goal (↓)
1	83.33%	40.03%	0.29
2	60.78%	26.58%	1.74
3	55.19%	26.21%	2.54
4	16.67%	4.84%	12.72
5	25.00%	9.31%	8.19

Final results

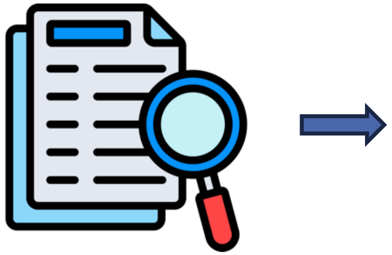
Experimental Setup	*OffNav*	*PirlNav*	*MetaNav*
SETUP 1	**100%**	**100%**	89.18%
SETUP 2	**79.31%**	72.50%	76.10%
SETUP 3	75.78%	**77.63%**	64.19%
SETUP 4	25.00%	**27.27%**	23.07%
SETUP 5	**34.78%**	26.09%	25.00%

Meta-training +25M parameters ✗ ⟶ Meta-training task aware encoders ✓

Conclusions

- Both OffNav and MetaNav are novel approaches to robot navigation that have demonstrated capable of navigating.
- OffNav is able to perform better that the behavior cloning baseline in some scenarios.
- While MetaNav is not able to perform better than the baseline or OffNav, it is able to navigate and the philosophy of navigating on novel environments with a few trajectories is promising.
- However, the results are not strong enough and suggest that further research has to be delivered to make this methods viable.

Associated publicaiton:



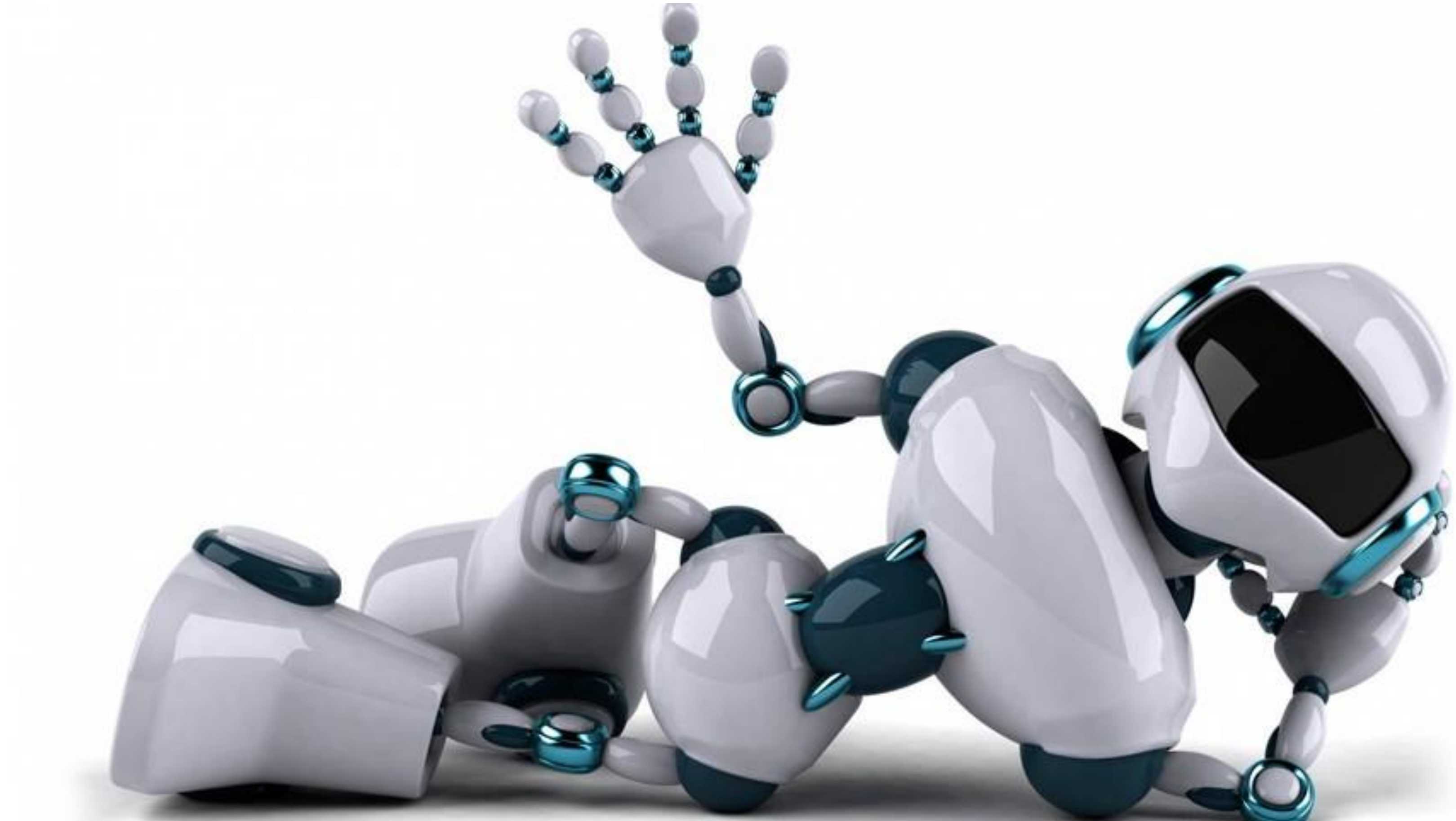
HARL workshop
ICRA 2025

Offnav: Offline Reinforcement Learning for Visual Semantic Navigation

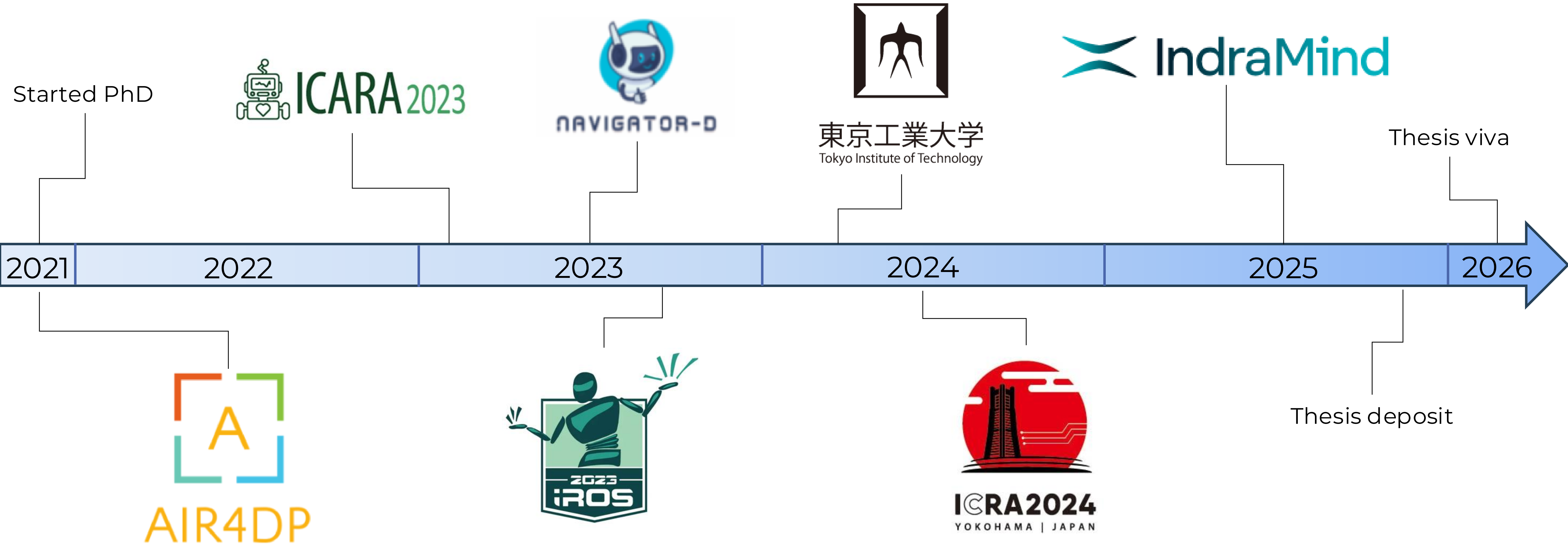
Gutiérrez-Alvarez C., Flor-Rodríguez-Rabadán R., Avecedo-Rodríguez FJ., López-Sastre RJ., Kanezaki A.

6. Final closure

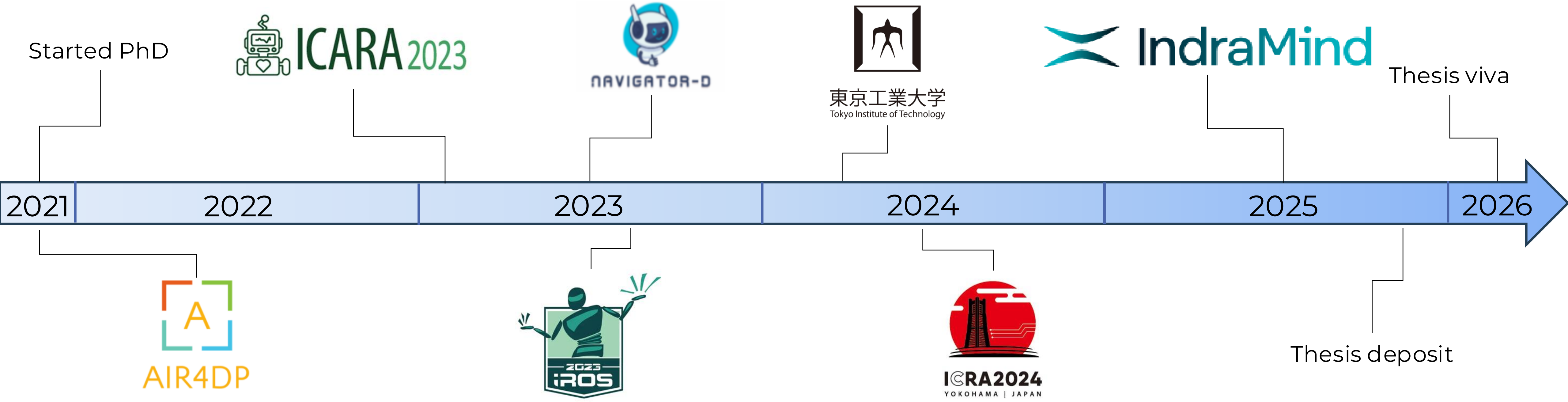
Scientific trajectory, impact and final conclusions



My Phd journey at a glance



My Phd journey at a glance



9 Papers

3 International Conferences

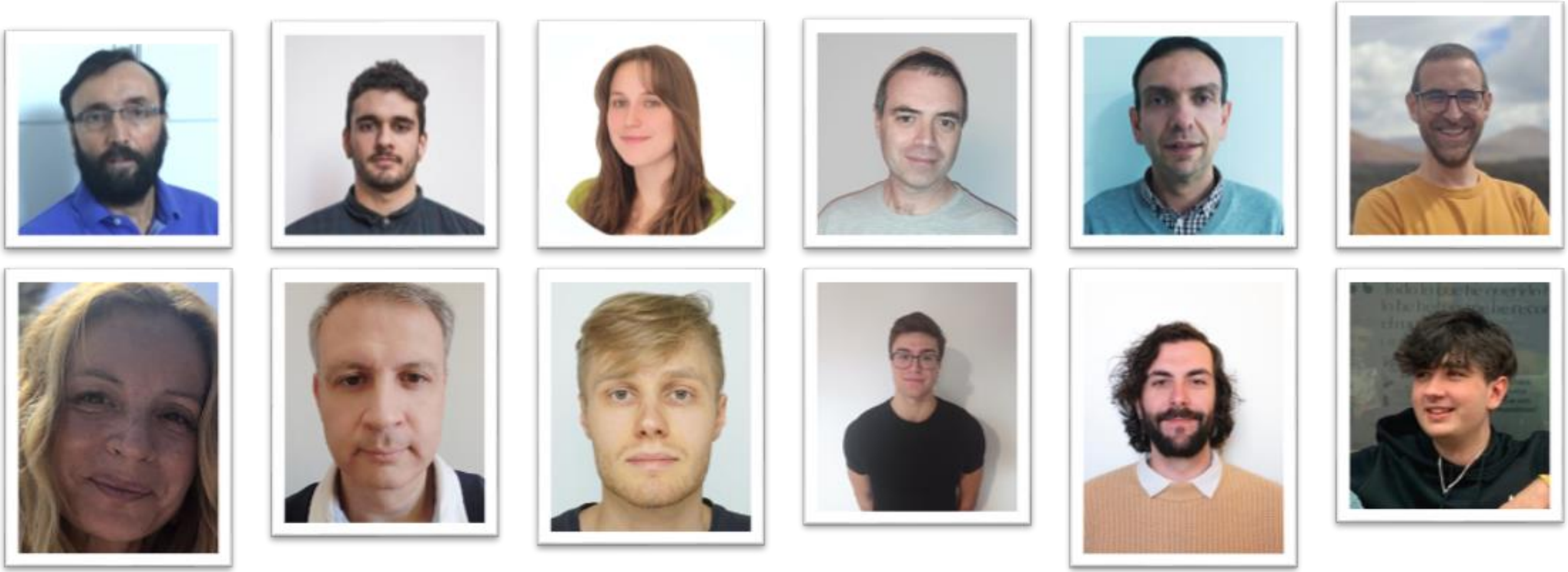
2 Research Projects

2 Courses

34 Citations

5 Supervised Undergrad Thesis

My lab



International research experience

March 2024 – Sep 2024

Asako Kanezaki
Associate Professor
Tokyo Institute of Technology

東京科学大学　金﨑研究室
Automation & Knowledge Laboratory

東京工業大学
Tokyo Institute of Technology

Associated publicaiton:

HARL workshop
ICRA 2025

Offnav: Offline Reinforcement Learning for Visual Semantic Navigation

Gutiérrez-Alvarez C., Flor-Rodríguez-Rabadán R., Avecedo-Rodríguez FJ., López-Sastre RJ., Kanezaki A.

Attended: ICRA2024 YOKOHAMA | JAPAN

Scholarships:

- FPI scholarship from Spanish Ministry of Science: 5780€.
- Mobility scholarship from UAH: 3000€.

International research experience

Scientific publications

Publications directly related to the thesis

1. **Gutiérrez-Alvarez C.**, Ríos-Navarro P., Flor-Rodríguez-Rabadán R., Acevedo-Rodríguez F.J., López-Sastre R.J., *Visual Semantic Navigation with Real Robots*, in Applied Intelligence, 2025. 5 citations, JCR Q2
2. **Gutiérrez-Alvarez C.**, Acevedo-Rodríguez F.J., López-Sastre R.J., Kanezaki A., OffNav: *Offline Reinforcement Learning for Visual Semantic Navigation*, in ICRA Human-aligned Reinforcement Learning for Autonomous Agents and Robots Workshop, 2024. 0 citations
3. **Gutiérrez-Alvarez C.**, Ríos-Navarro P., Flor-Rodríguez-Rabadán R., Acevedo-Rodríguez F.J., López-Sastre R.J., *Evaluation of Visual Semantic Navigation Models in Real Robots*, in IROS Late Breaking Results, 2023. 0 citations
4. **Gutiérrez-Alvarez C.**, Hernández-García S, Nasri N, Cuesta-Infante Alfredo, López-Sastre RJ, *Towards Clear Evaluation of Robotic Visual Semantic Navigation*, in ICARA, 2023. 0 citations

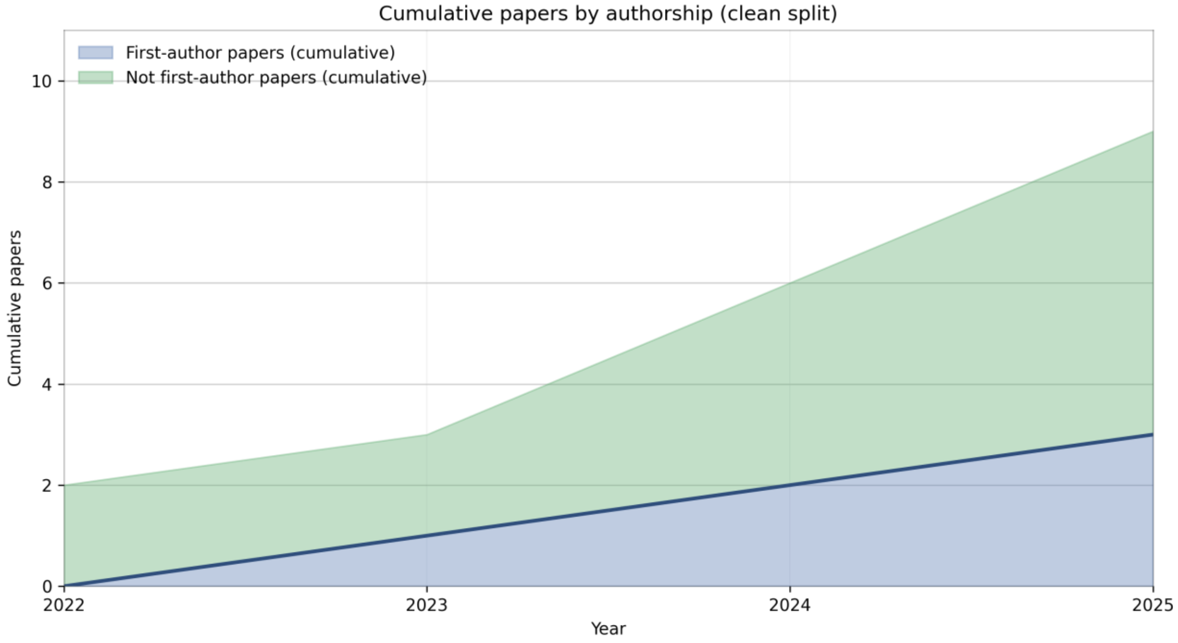
Scientific publications

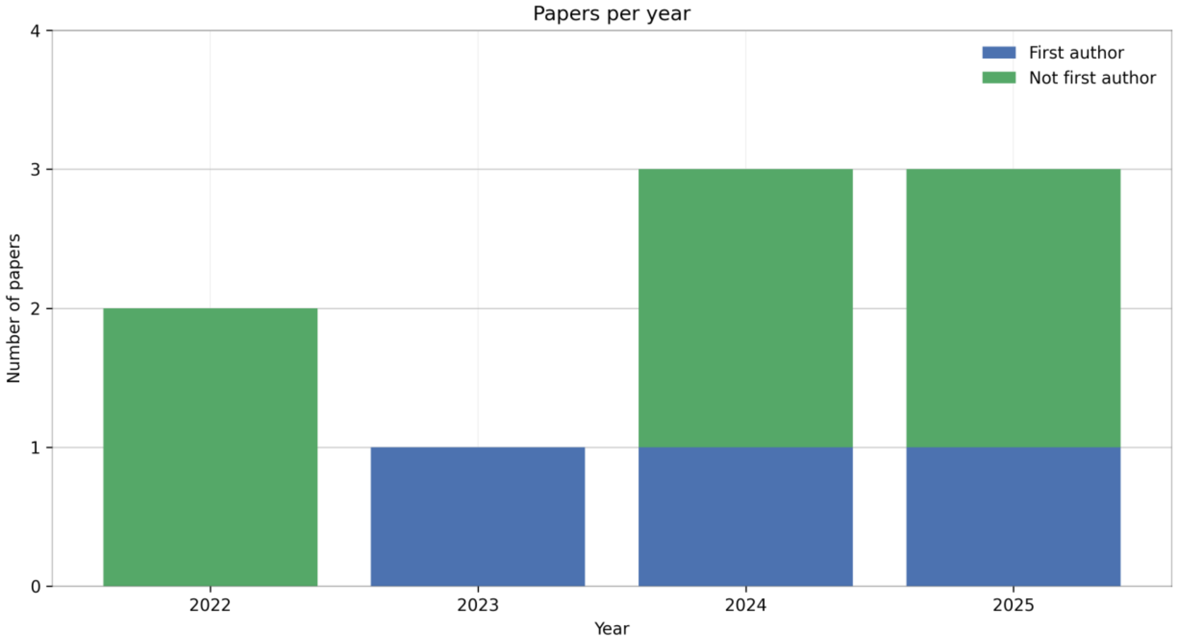
Side publications

1. Flor-Rodríguez-Rabadán R., **Gutiérrez-Álvarez C.**, Acevedo-Rodríguez, F.J., Lafuente-Arroyo S., López-Sastre R.J., *SEMNAV: A Semantic Segmentation-Driven Approach to Visual Semantic Navigation*, in ArXiv, 2025. 0 citations
2. Blanco-Fernández E., **Gutiérrez-Alvarez C.**, Nasri N., Maldonado-Bascón, S., López-Sastre R.J., *Live Video Captioning*, in Multimedia Tools and Applications, 2025. 4 citations, JCR Q2
3. Nasri N, **Gutiérrez-Álvarez C.**, López-Sastre RJ, Lafuente-Arroyo S., Maldonado-Bascón S. *Realistic Continual Learning Approach using Pretrained Models*, in ArXiv 2024. 0 citations
4. Lafuente-Arroyo S., Maldonado-Bascón S., Delgado-Mena D., **Gutiérrez-Alvarez C.**, Acevedo-Rodríguez F.J., *Multisensory Integration for Topological Indoor Localization of Mobile Robots in Complex Symmetrical Environments*, in Expert Systems with Applications, 2023. 7 citations, JCR Q1
5. Nasri N, López-Sastre RJ, Pacheco-da-Costa S, Fernández-Munilla I, **Gutiérrez-Álvarez C.**, Pousada-García T, Acevedo-Rodríguez FJ, Maldonado-Bascón S. *Assistive Robot with an AI-Based Application for the Reinforcement of Activities of Daily Living: Technical Validation with Users Affected by Neurodevelopmental Disorders*, in Applied Sciences, 2022. 18 citations, JCR Q2

Bibliometric impact

Papers



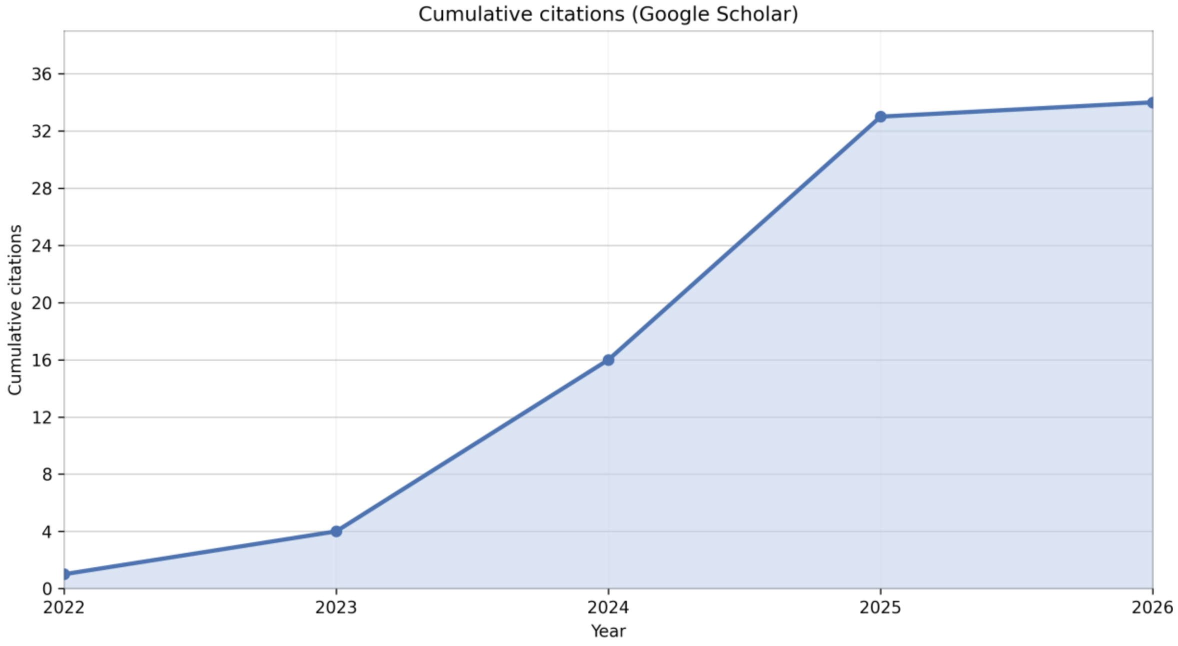


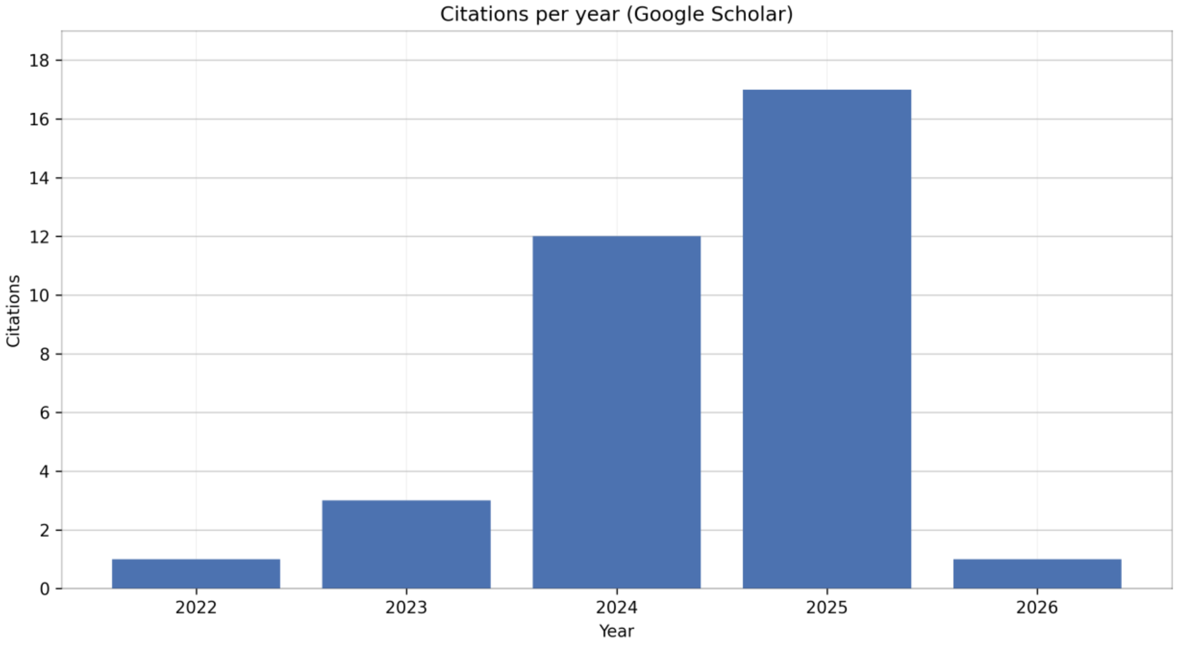
- Total papers: 9
- First author: 3
- Not first author: 6

Universidad de Alcalá

Bibliometric impact

Citations





- Total citations: 34
- H-index: 3
- I10-index: 1

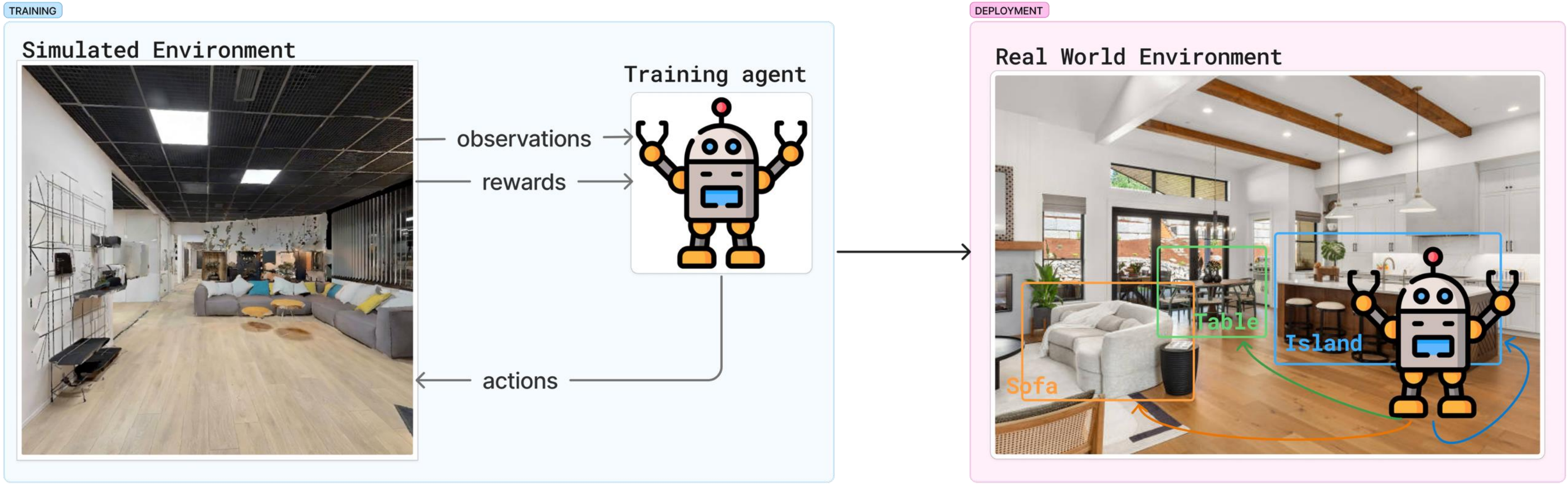
Universidad de Alcalá

Limitations & future work

- Add more types of multimodal sensor to make the navigation closer to that of humans:
  - Audio sensors.
  - Tactile sensor.
- Explore more complex tasks: not only navigating to an object, but rearranging room objects or following complex instructions via text.
- Try new meta learning approaches that do not heavily modify the subjacent algorithm: the method used in chapter 5 meta adapts the whole parameters of the model, which can hurt performance. It could be more promising to use meta learning approaches that do not modify the parameters and could for example represent the task information into an encoder.

Global Scientific Conclusions

- High performance in simulation does **not** guarantee real-world robustness.
- Modular architectures remain **more reliable** for real robotic deployment.
- Data-efficient learning is **essential** for scalable embodied intelligence.




The end
Thank you!
Universidad de Alcalá