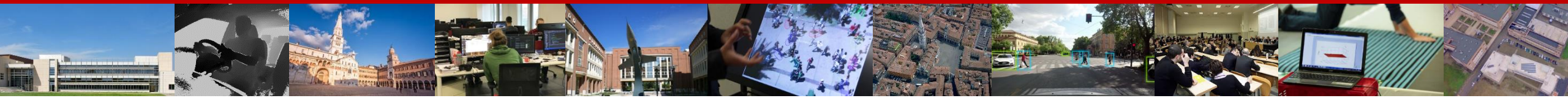
CAIP 2021

Out of the Box:
Embodied Navigation in the Real World



Roberto Bigazzi, Federico Landi, Marcella Cornia, Silvia Cascianelli, Lorenzo Baraldi, Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab

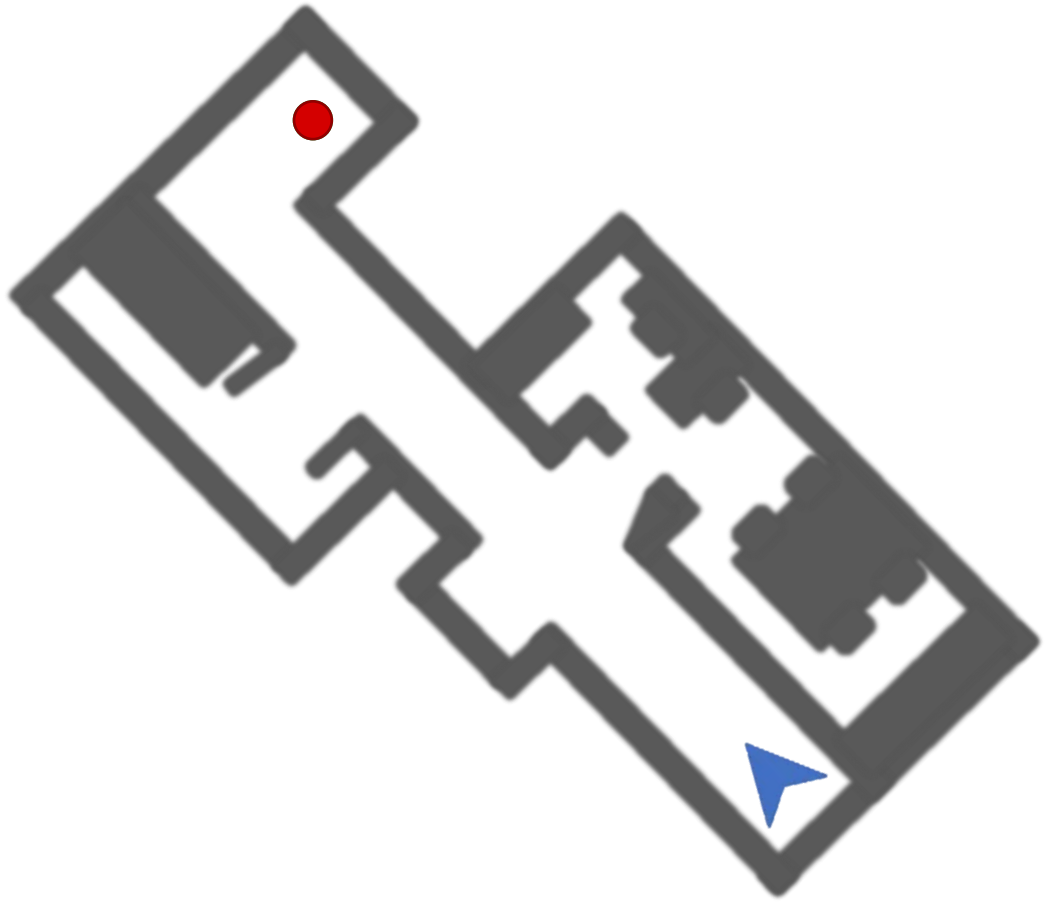
POINTGOAL NAVIGATION IN THE REAL WORLD

AImage Lab

Embodied agent for pointgoal navigation in the real world.

Differences between simulation and real-world lead to lower performance using the real robot.

Mitigate this problem aligning the training in simulation with the real-world setting.



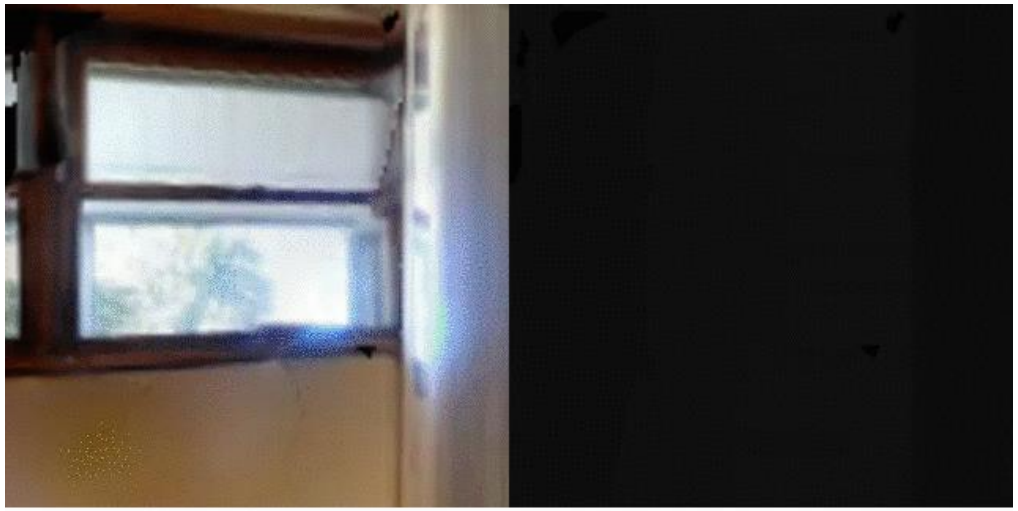
NAVIGATION IN SIMULATED ENVIRONMENTS

AImage Lab

Embodied AI agents are trained using simulators with millions of frames.

Simulation is more convenient in terms of costs, safety and time.

Trained model is finally deployed on the real robot



An episode of pointgoal navigation with Habitat Simulator.

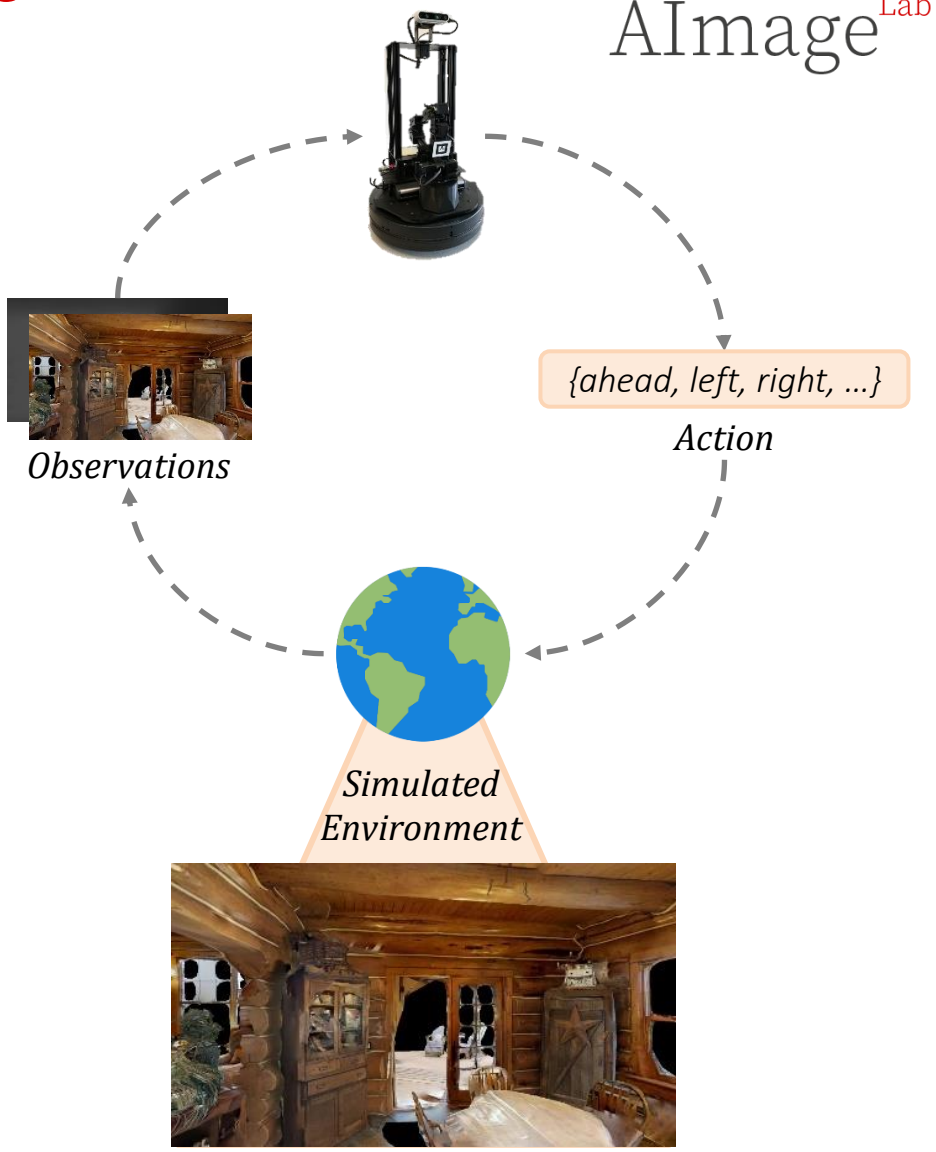
NAVIGATION IN SIMULATED ENVIRONMENTS

The agent is deployed in a simulated environment.

The simulator extracts observations for the agent.

The agent samples the atomic actions.

New observations are extracted.

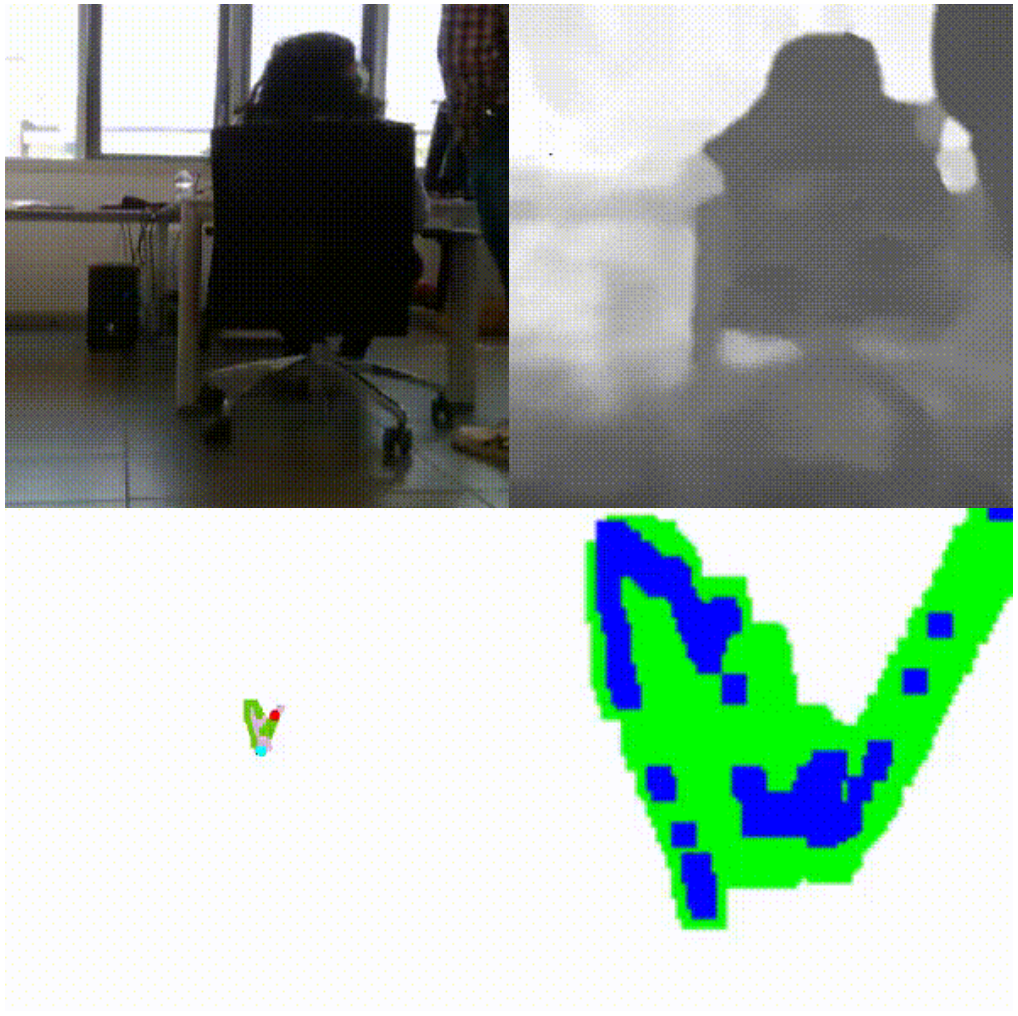


NAVIGATION IN REAL WORLD ENVIRONMENTS

The model deployed on the real robot suffers the discrepancies between training (simulated) and evaluation (real world).

Main differences:

- Hyperparameters not matching;
- Agent could exploit tricks when using the simulator;
- Sensor and actuation noise.



An episode of pointgoal navigation in the real world.

ADAPTING FOR REAL WORLD

AImage Lab

Hyperparameters not matching:

- Real robot parameters

Tricks in simulation:

- Avoid sliding

Sensor and actuation noise:

- Proper noise models
- Depth restoration
- Pose computation exploits odometry sensor

: during training : during evaluation on the real robot

HYPERPARAMETER DIFFERENCES

AImage Lab

During the training, we should consider the characteristics of the real robot that differ from the default parameters of the simulator.

The main differences to be addressed are the height of the robot, the field of view of the cameras, and the range of the depth camera.

	Height	RGB FoV	Depth FoV	Depth Range
Default for Simulation	1.25	**H**: 90, **V**: 90	**H**: 90, **V**: 90	[0.0, 10.0]
LoCoNav (ours)	0.60	**H**: 70, **V**: 90	**H**: 57, **V**: 86	[0.0, 5.00]

DEPTH OBSERVATION RESTORATION

AImage Lab

Depth is restored using a hole filling algorithm followed by the application of a median filter.

DEPTH OBSERVATION RESTORATION

AImage Lab

Depth is restored using a hole filling algorithm followed by the application of a median filter.

COMPUTATION OF THE POSE $\tilde{\chi}_t$

AImage Lab

We adapt the odometry sensor of the robot to compute position and orientation:

Position: $\tilde{\mathbf{x}}_t = (\tilde{x}_t, \tilde{y}_t, 1)$

$$\tilde{\mathbf{x}}_t = \mathbf{A}^{-1}\mathbf{x}_t$$

$$\mathbf{x}_t = (x_t, y_t, 1)$$

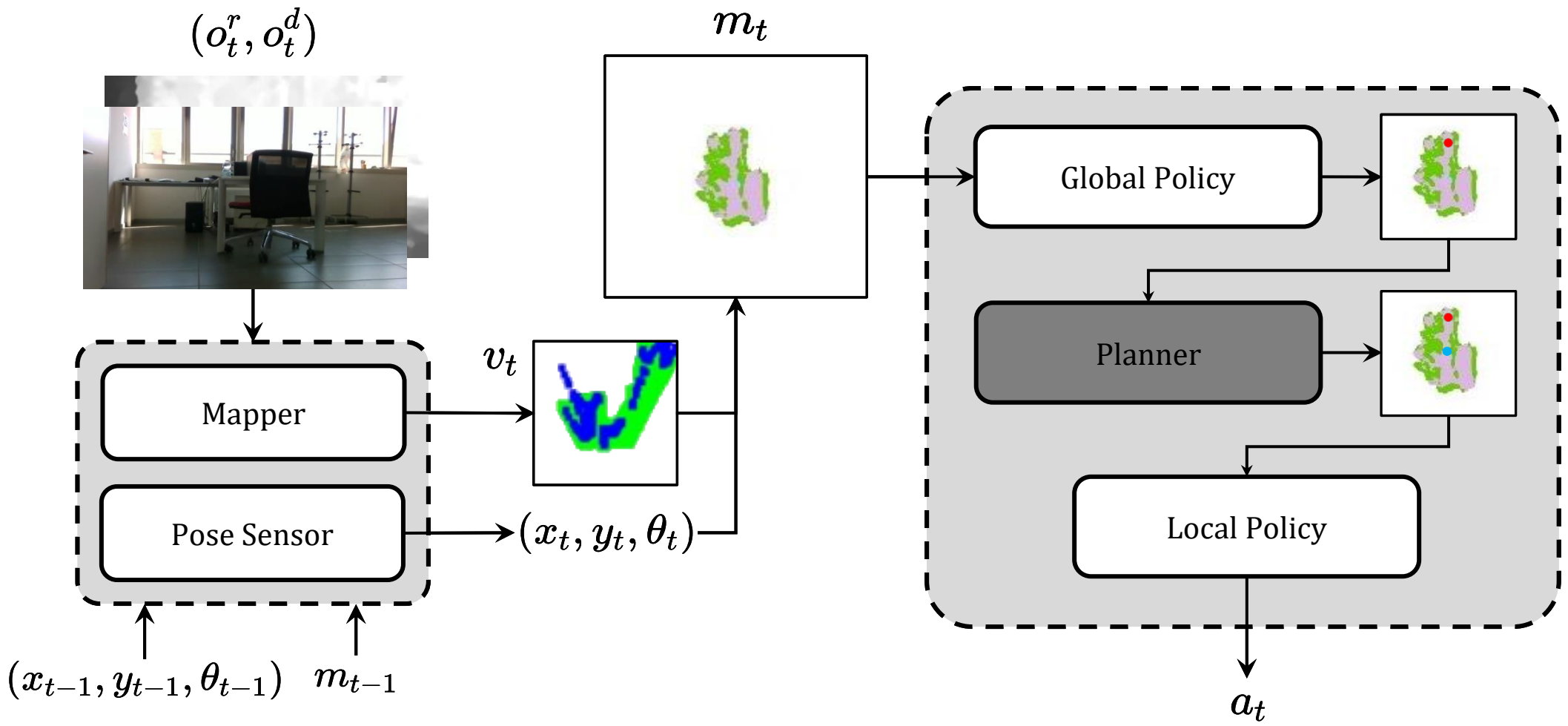
$$\mathbf{A} = \begin{pmatrix} \mathbf{R}_0 & \mathbf{t}_0 \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & -\sin\theta_0 & x_0 \\ \sin\theta_0 & \cos\theta_0 & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\chi_0 = (x_0, y_0, \theta_0)$$

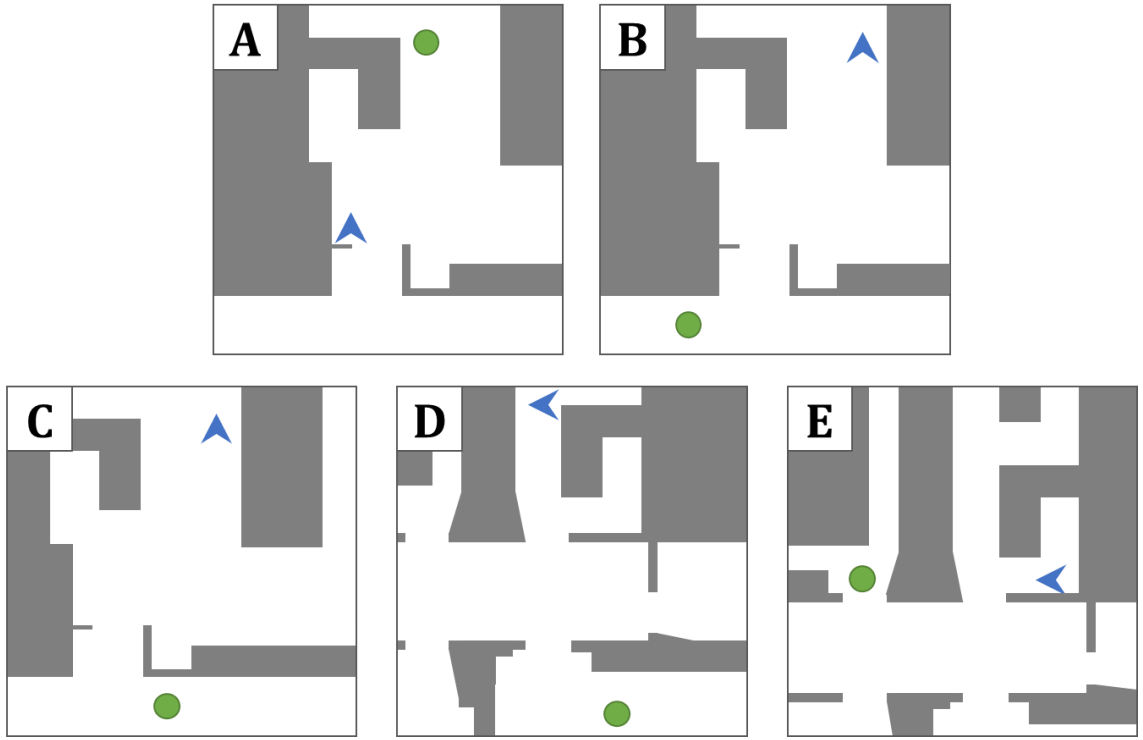
Orientation: $\tilde{\theta}_t = \theta_t - \theta_0$

$$\tilde{\chi}_t = (\tilde{x}_t, \tilde{y}_t, \tilde{\theta}_t)$$

LOCONAV ARCHITECTURE



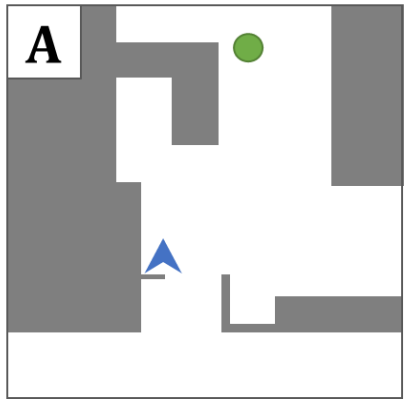
REAL WORLD POINTGOAL NAVIGATION



Path	Length(m)	Time(s)	# Step
A	3.80	124	23
B	6.75	239	45
C	5.95	223	43
D	6.55	217	42
E	4.20	227	33

NAVIGATION RESULTS

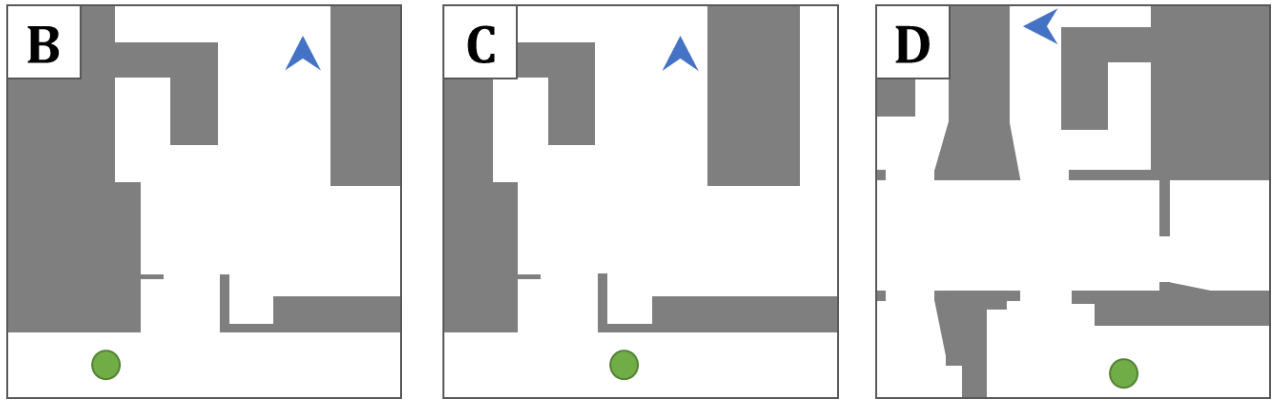
If the goal is in the same room of the agent, results are optimal.



Path	SR ↑	SPL ↑	HFR ↓	BR ↓	Abs. Steps	Norm. Steps ↑	Abs. Time	Norm. Time ↑
A	1.0	0.718	0.0	0.30	32.70±1.73	0.717±0.033	176.11±10.39	0.718±0.031
B	0.8	0.711	0.10	0.22	51.67±1.72	0.880±0.027	273.70±8.24	0.879±0.030
C	0.5	0.205	0.10	0.78	123.44±10.66	0.374±0.034	631.15±50.09	0.372±0.036
D	0.5	0.318	0.10	0.89	65.67±3.90	0.645±0.037	344.00±20.08	0.657±0.038
E	0.2	0.060	0.40	1.00	135.17±29.97	0.290±0.049	722.76±162.01	0.38±0.066
Overall	0.6	0.402	0.14	0.60	-	0.608±0.036	-	0.617±0.034

NAVIGATION RESULTS

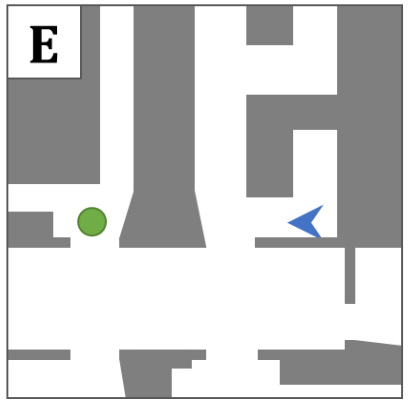
When the agent needs to change room, it is still able to successfully terminate the majority of the experiments.



Path	SR ↑	SPL ↑	HFR ↓	BR ↓	Abs. Steps	Norm. Steps ↑	Abs. Time	Norm. Time ↑
A	1.0	0.718	0.0	0.30	32.70±1.73	0.717±0.033	176.11±10.39	0.718±0.031
B	0.8	0.711	0.10	0.22	51.67±1.72	0.880±0.027	273.70±8.24	0.879±0.030
C	0.5	0.205	0.10	0.78	123.44±10.66	0.374±0.034	631.15±50.09	0.372±0.036
D	0.5	0.318	0.10	0.89	65.67±3.90	0.645±0.037	344.00±20.08	0.657±0.038
E	0.2	0.060	0.40	1.00	135.17±29.97	0.290±0.049	722.76±162.01	0.38±0.066
Overall	0.6	0.402	0.14	0.60	-	0.608±0.036	-	0.617±0.034

NAVIGATION RESULTS

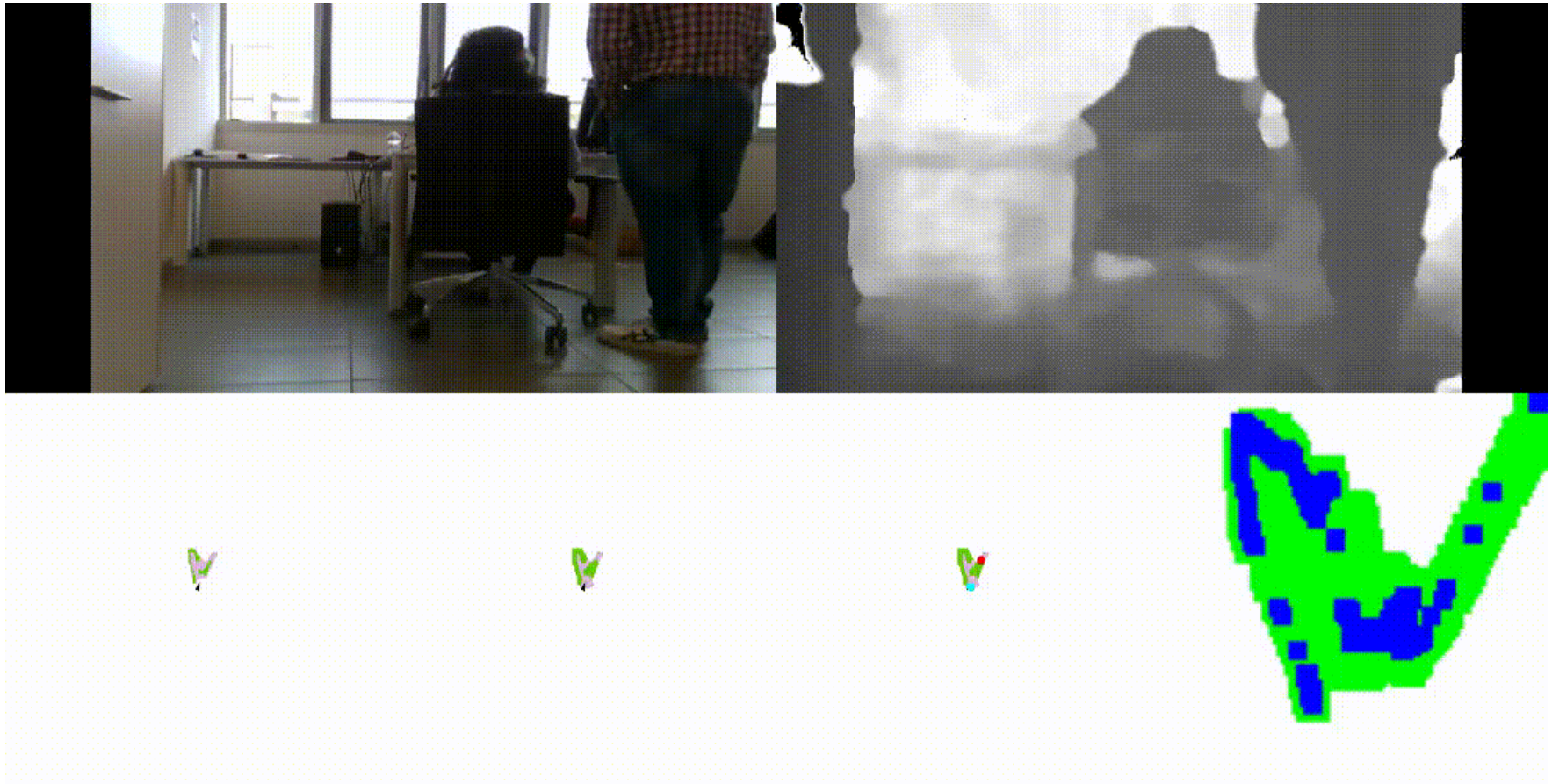
The episode that involves getting out of a room and enter a door immediately after represented the most challenging scenario due to the difficulty in identifying the second door while approaching it with a high angle.



Path	SR ↑	SPL ↑	HFR ↓	BR ↓	Abs. Steps	Norm. Steps ↑	Abs. Time	Norm. Time ↑
A	1.0	0.718	0.0	0.30	32.70±1.73	0.717±0.033	176.11±10.39	0.718±0.031
B	0.8	0.711	0.10	0.22	51.67±1.72	0.880±0.027	273.70±8.24	0.879±0.030
C	0.5	0.205	0.10	0.78	123.44±10.66	0.374±0.034	631.15±50.09	0.372±0.036
D	0.5	0.318	0.10	0.89	65.67±3.90	0.645±0.037	344.00±20.08	0.657±0.038
E	0.2	0.060	0.40	1.00	135.17±29.97	0.290±0.049	722.76±162.01	0.38±0.066
Overall	0.6	0.402	0.14	0.60	-	0.608±0.036	-	0.617±0.034

NAVIGATION EXAMPLES

AImage Lab



NAVIGATION EXAMPLES

AImage Lab

CAIP 2021

Thank you for your attention

Out of the Box: *Embodied Navigation in the Real World*


Roberto Bigazzi


Federico Landi


Marcella Cornia


Silvia Cascianelli


Lorenzo Baraldi


Rita Cucchiara

{name.surname}@unimore.it

University of Modena and Reggio Emilia, Italy

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

AImage Lab