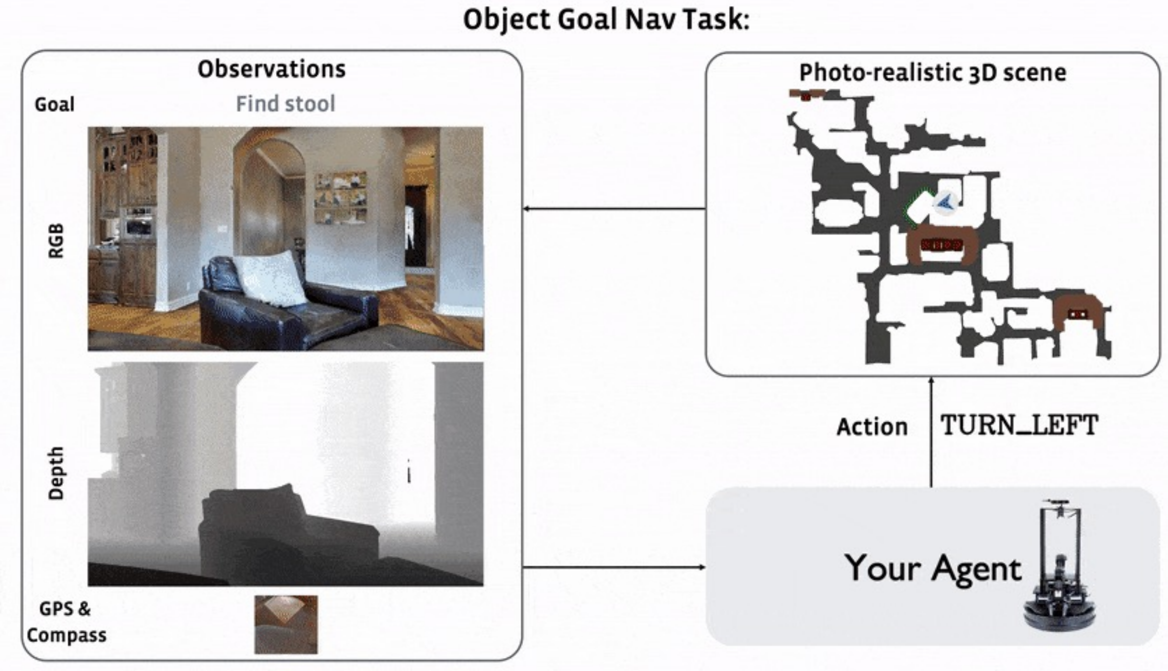
Embodied Navigation Tasks

Speaker: Liu Dai

Collaborator: Fanpeng Meng

Advisor: Jiazhao Zhang, He Wang

Sep. 8, 2022



CONTENT

- **Task Introduction**
  - Task Definition
  - Habitat Challenge
- **Selected Papers**
  - Map-Based Methods
  - Vision-Based Methods

Task Introduction

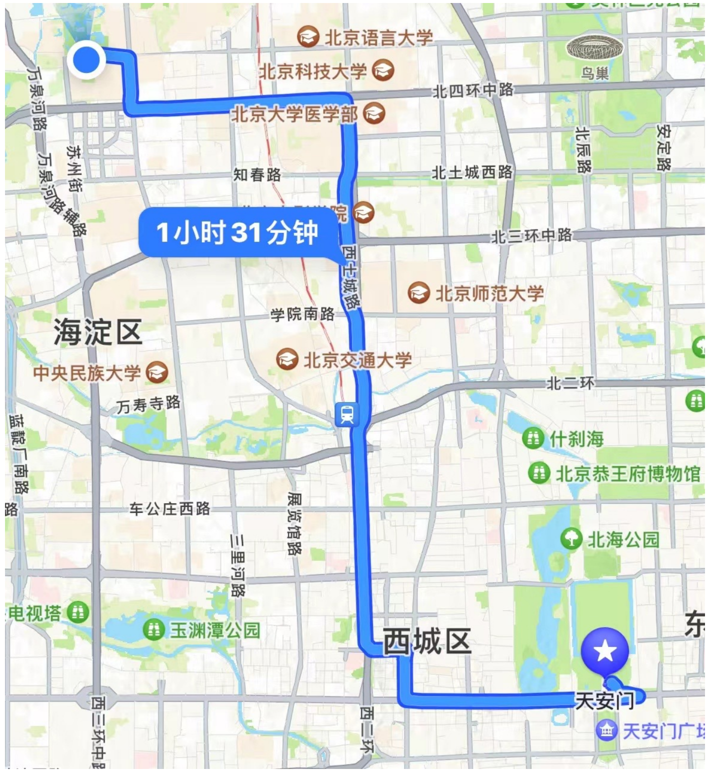
What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Outdoor Navigation



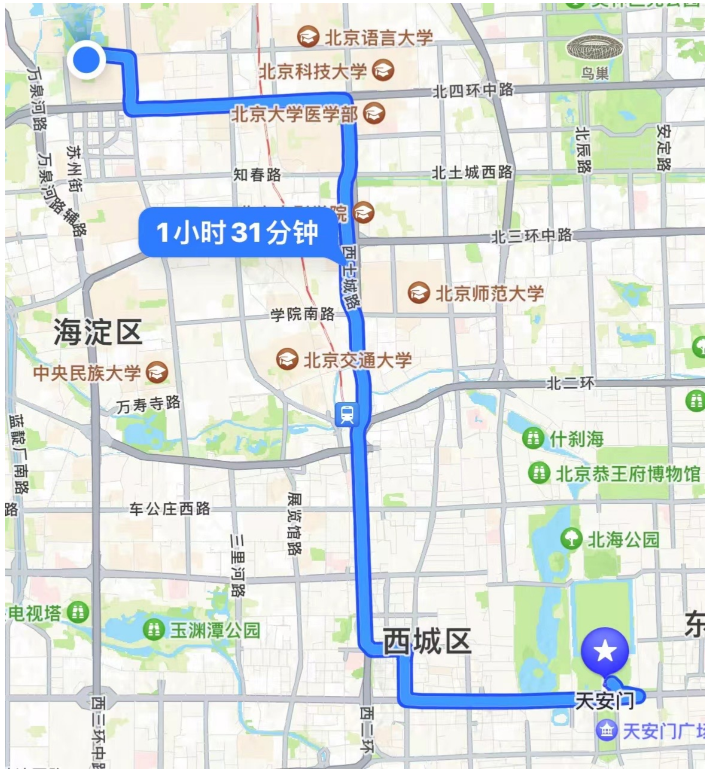
VS

Indoor Navigation

What is Navigation?

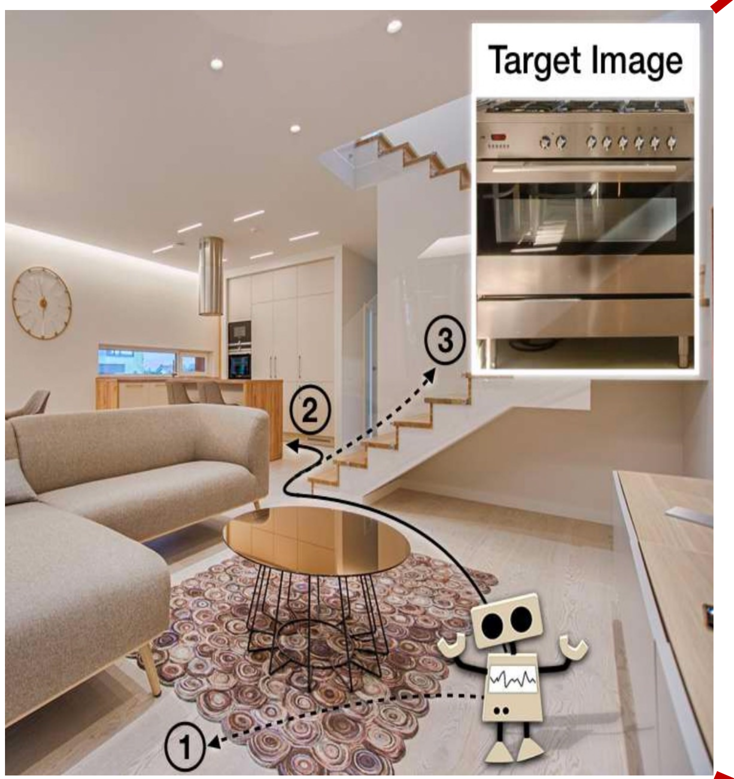
- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Outdoor Navigation



VS

Indoor Navigation



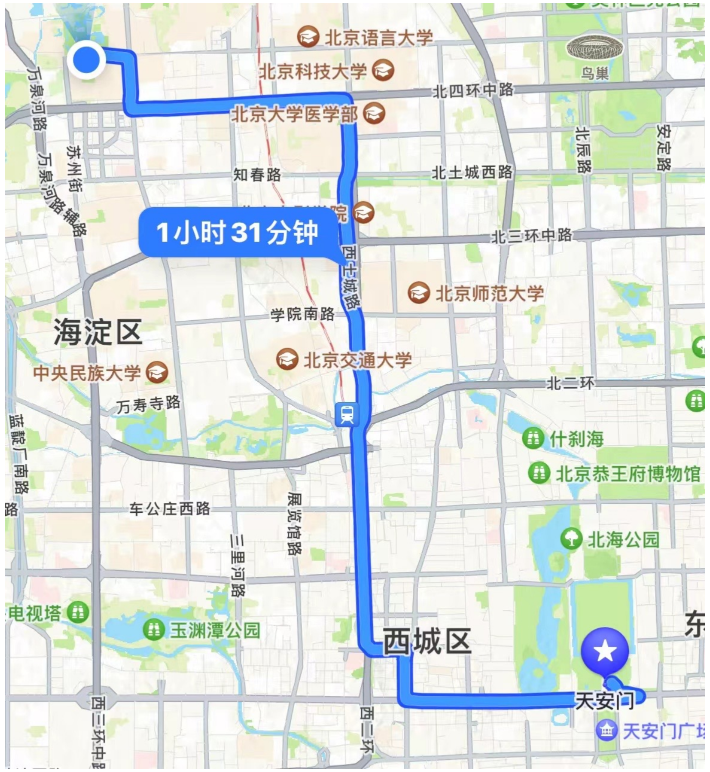
Challenges:

a) No HD Map
b) Noisy Indoor Positioning
c) Real 3D Environment

What is Navigation?

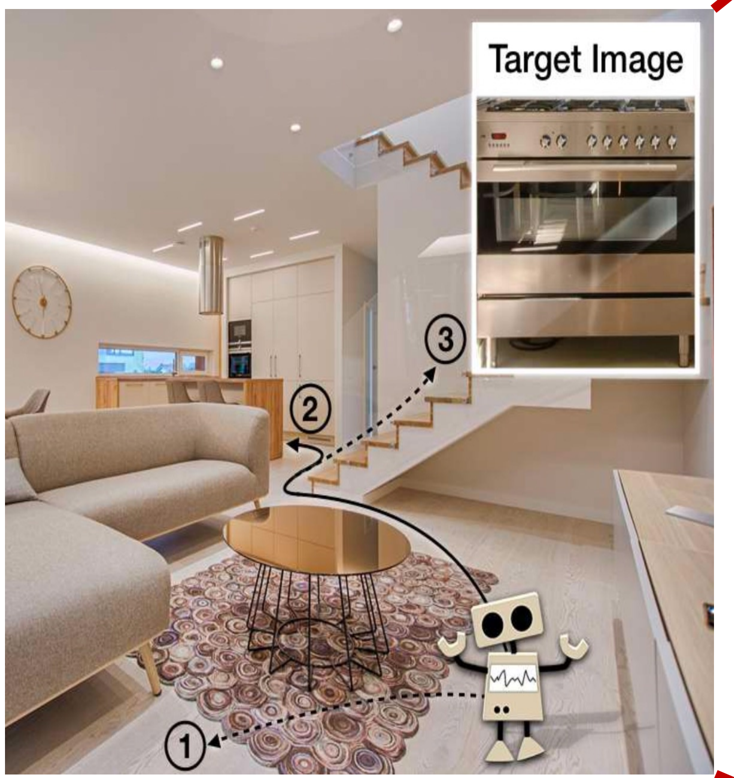
- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Outdoor Navigation



VS

Indoor Navigation



Challenges:

a) No HD Map
b) Noisy Indoor Positioning
c) Real 3D Environment

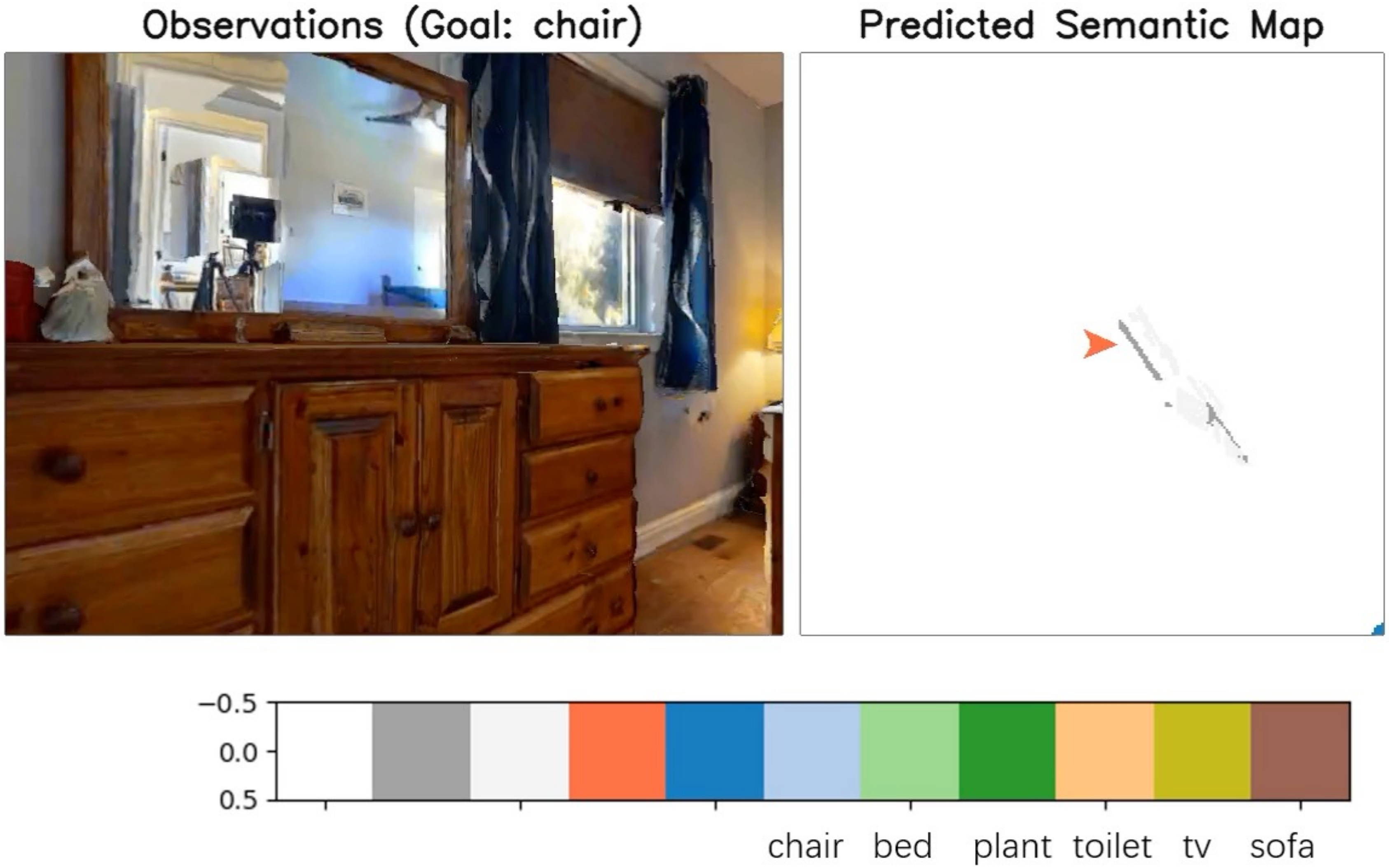
Values:

a) Necessary for Embodied AI
b) Home Robot
c) Robot for Special Tasks

What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Demo:

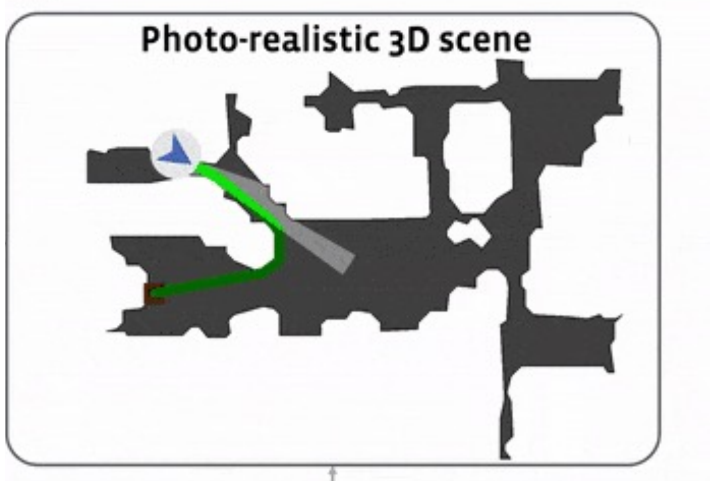


What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Explicit Target Locations:

Point Goal



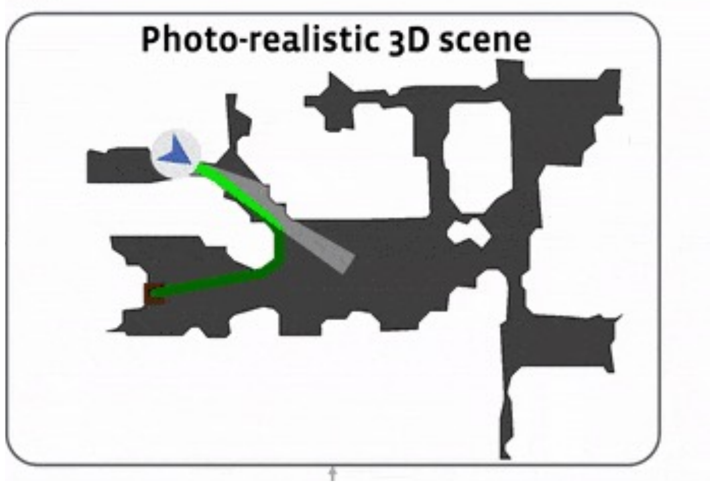
Goal：
Go 5m south, 3m west relative to start.

What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Explicit or ***Implicit*** Target Locations:

Point Goal



Goal：
Go 5m south, 3m west relative to start.

Image Goal

Goal：
Go where the photo was taken.

What is Navigation?

- **Definition** of **Navigation**: Direct or find a way ***from one place to another***.

Explicit or ***Implicit*** Target Locations:

Point Goal

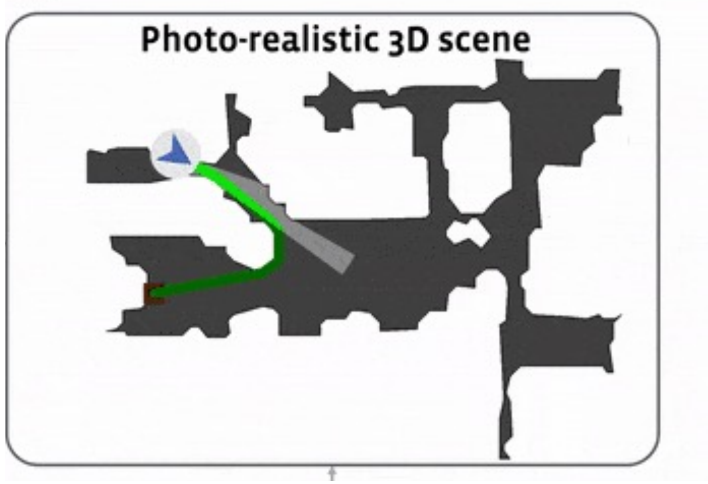
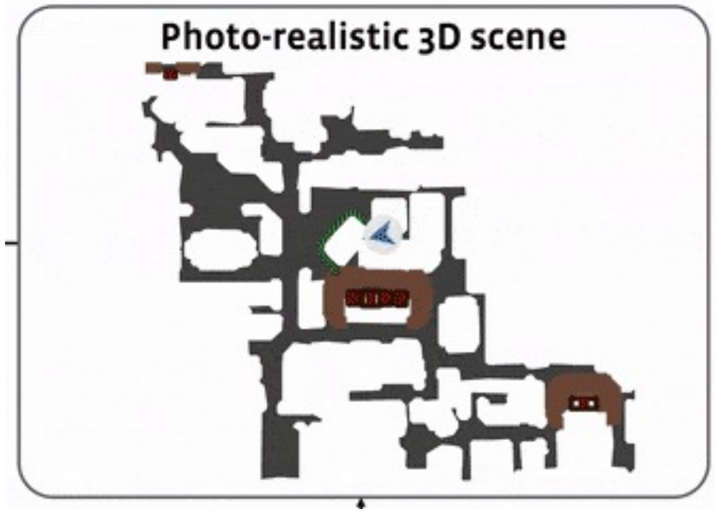


Image Goal

Object Goal



Goal：
Go 5m south, 3m west relative to start.

Goal：
Go where the photo was taken.

Goal：
Go find a sofa.

What is Navigation?

- **How to Evaluate?**

(1) Success Rate

(2) SPL (Success Weighted by Path Planning)

$$\text{SPL} = \frac{1}{N}\sum_{i=1}^{N} S_i \frac{l_i}{\max(p_i, l_i)}$$

S_i : binary for success
l_i : shortest path length
p_i : actual path length

(3) Soft SPL

$$SoftSPL = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{d_{T_i}}{d_{init_i}}\right)\left(\frac{l_i}{max(p_i, l_i)}\right)$$

d_{Ti} : distance to target location d_{initi} : distance to start location

(4) Distance to Goal

What is Navigation?

- **How to Evaluate?**

(1) Success Rate

(2) SPL (Success Weighted by Path Planning)

$$\text{SPL} = \frac{1}{N}\sum_{i=1}^{N} S_i \frac{l_i}{\max(p_i, l_i)}$$

S_i : binary for success
l_i : shortest path length
p_i : actual path length

(3) Soft SPL

$$SoftSPL = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{d_{T_i}}{d_{init_i}}\right)\left(\frac{l_i}{max(p_i, l_i)}\right)$$

d_{Ti} : distance to target location d_{initi} : distance to start location

(4) Distance to Goal

- **Datasets:**

(1) Matterport 3D(MP3D)
Proposed in 2017

(2) Gibson 3D
Proposed in CVPR 2018

(3) Habitat Matterport 3D
Proposed in NIPS 2021

Larger

More Realistic

More Information

Why do we care about Navigation?

- Basic Module of Embodied AI: **Navigation + Manipulation**

Point Goal

Image Goal

Object Goal

……

$\left.\begin{array}{l}\textbf{Point Goal}\\\textbf{Image Goal}\\\textbf{Object Goal}\\\textbf{……}\end{array}\right\}$ **Navigation**

Why do we care about Navigation?

- **Basic Module of Embodied AI: Navigation + Manipulation**

Point Goal

Image Goal

Object Goal

……

Navigation

Rearrangement

Goal：
Rearrange the apple on the desk to the sofa.

Language+

Goal：
“Hey, bobo. Where is my wallet? Give it to me.”

Why do we care about Navigation?

- **Basic Module of Embodied AI: Navigation + Manipulation**

Point Goal

Image Goal

Object Goal

……

Navigation

Rearrangement

Goal：
Rearrange the apple on the desk to the sofa.

Language+

Goal：
"Hey, bobo. Where is my wallet? Give it to me."

Habitat Challenge

Habitat

PointGoal Nav
2019 - 2021

ObjectGoal Nav
2020 -

Rearrangement
2022 -

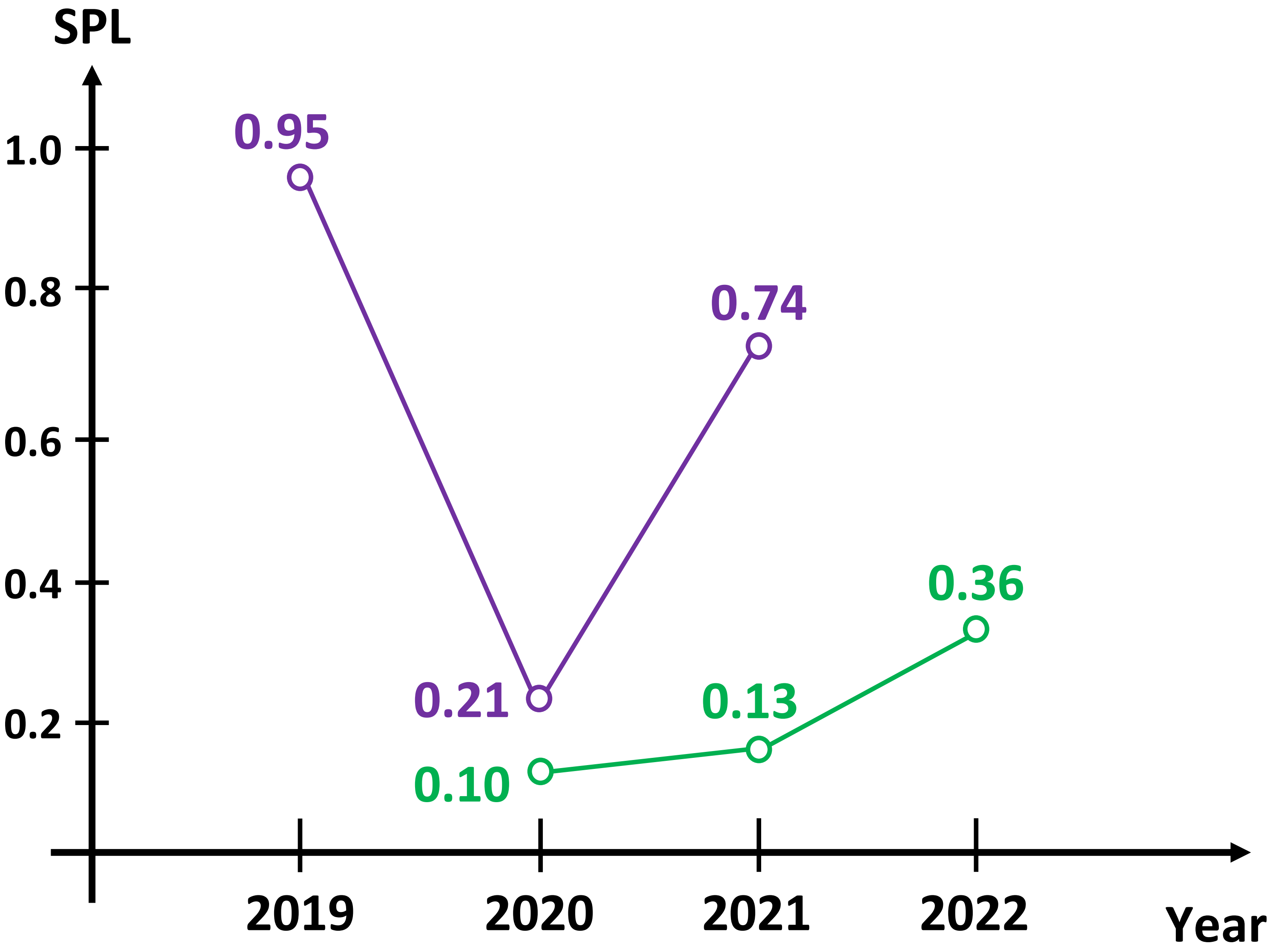
Habitat Challenge

Habitat

PointGoal Nav
2019 - 2021

ObjectGoal Nav
2020 -

Rearrangement
2022 -



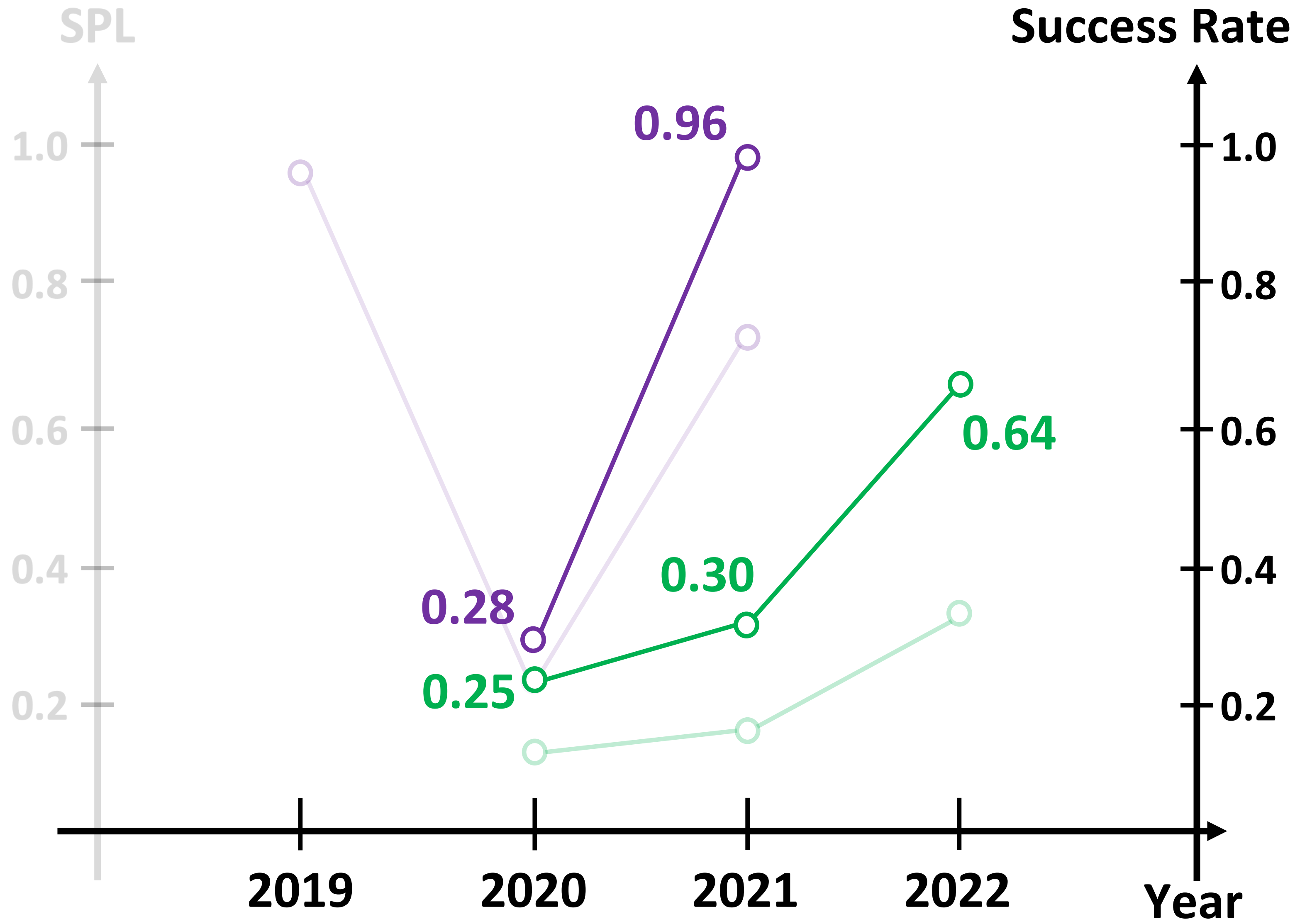
Habitat Challenge

Habitat

PointGoal Nav
2019 - 2021

ObjectGoal Nav
2020 -

Rearrangement
2022 -



Selected Papers

Selected Papers

- Map-Based
- Vision-Based

Selected Papers

Map-Based

Vision-Based

Learning to Explore Using Active Neural SLAM

Devendra Singh Chaplot[1 †], **Dhiraj Gandhi**[2], **Saurabh Gupta**[3 *],
Abhinav Gupta [1,2 *], **Ruslan Salakhutdinov**[1 *]
[1]Carnegie Mellon University, [2]Facebook AI Research, [3]UIUC

(ICLR 2020)

Learning to Explore Using Active Neural SLAM

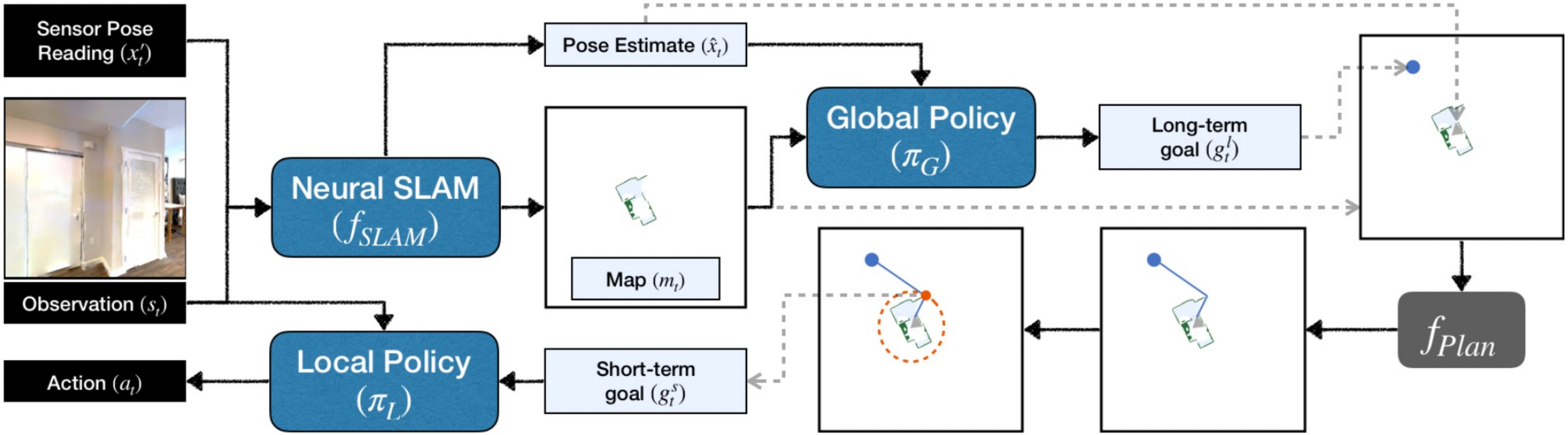
Devendra Singh Chaplot[1 †], **Dhiraj Gandhi**[2], **Saurabh Gupta**[3*],
Abhinav Gupta [1,2*], **Ruslan Salakhutdinov**[1*]
[1]Carnegie Mellon University, [2]Facebook AI Research, [3]UIUC

(ICLR 2020)

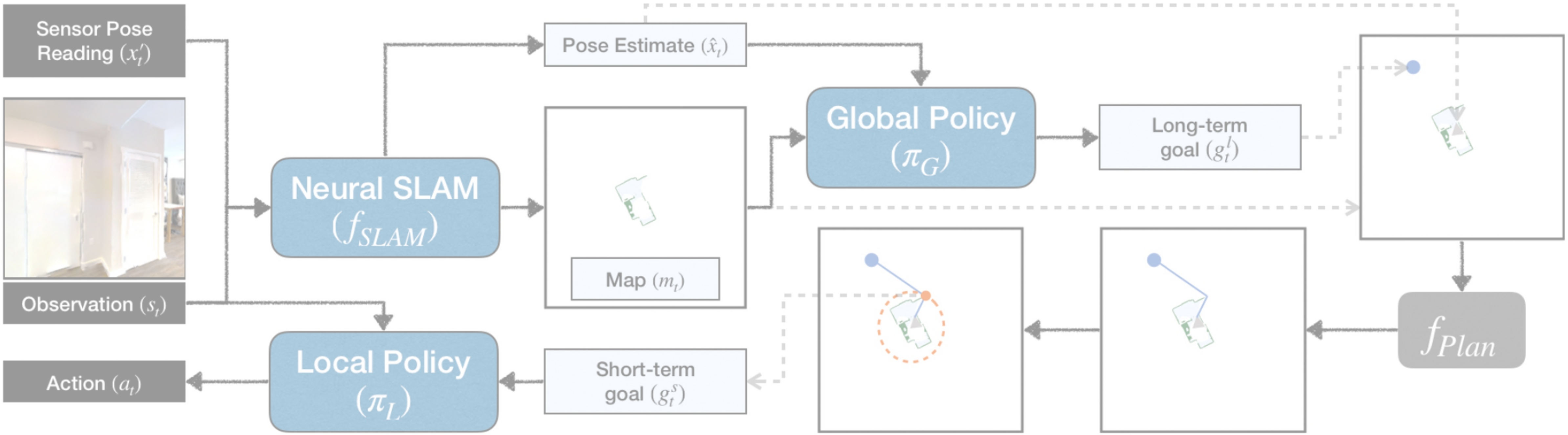
Goal: **Maximize Exploration Coverage in a fixed time budget**

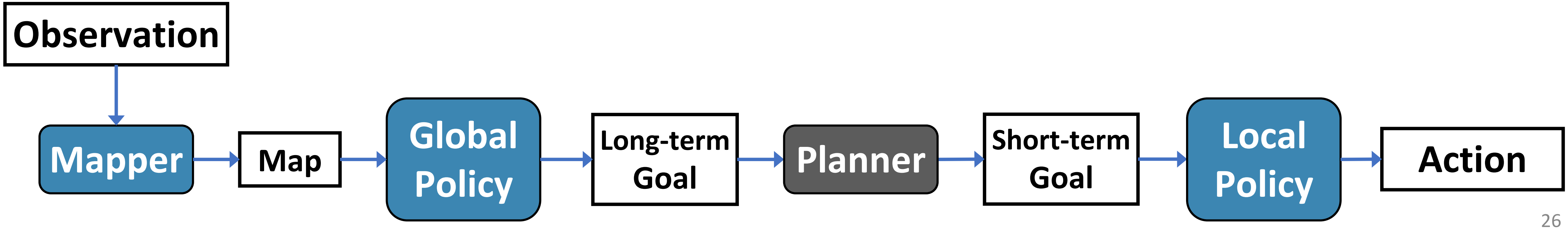
Input: **RGB Image, Pose Sensor with Noise**

Active Neural SLAM

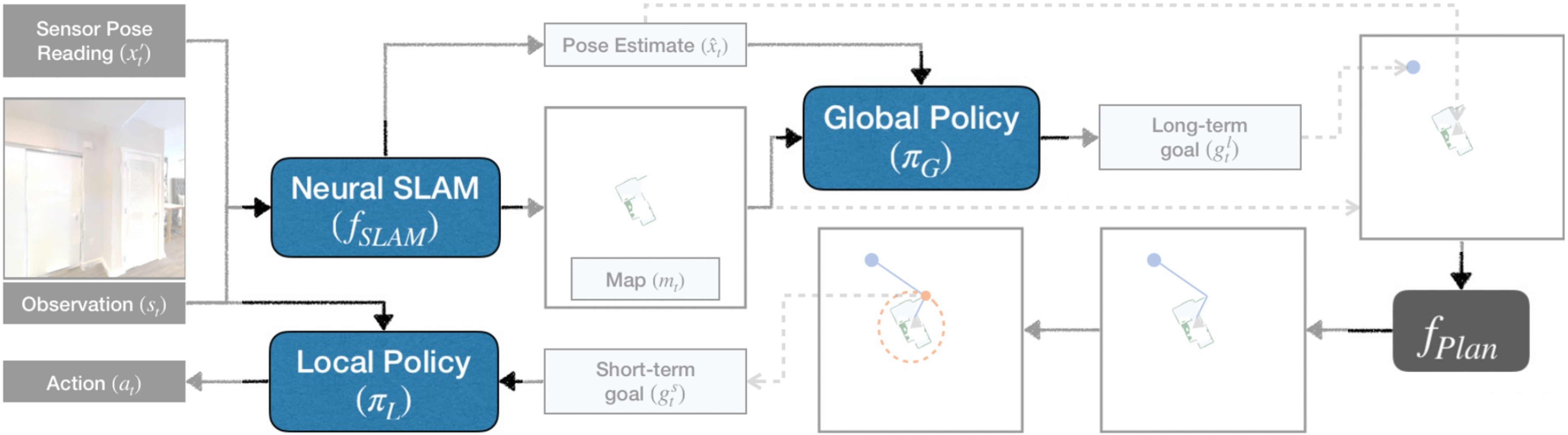


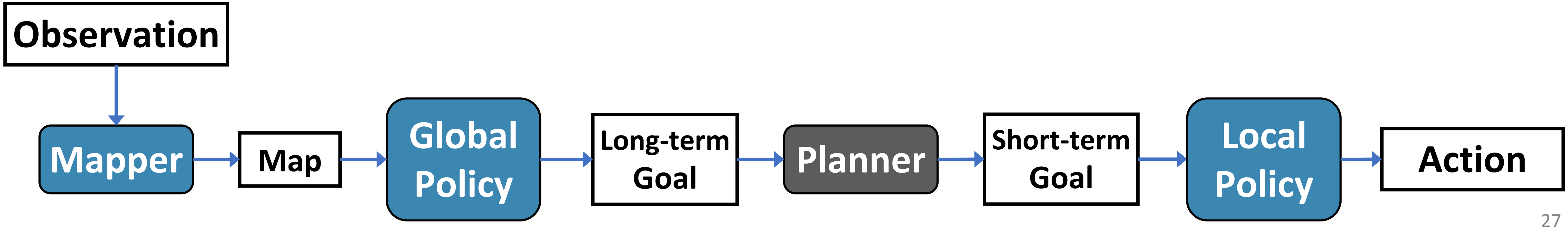
Active Neural SLAM



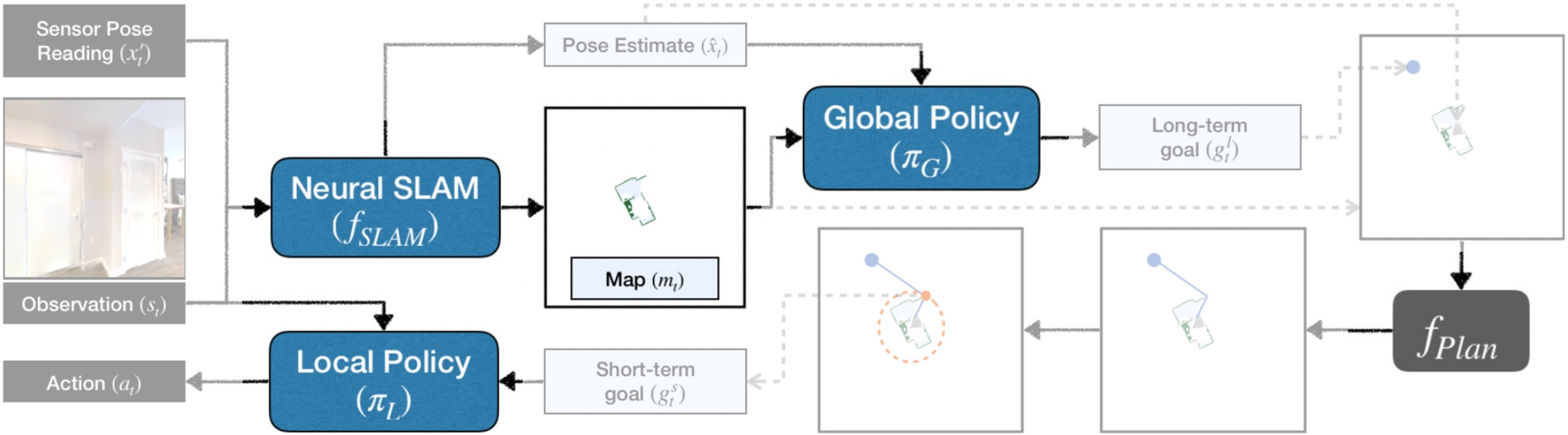


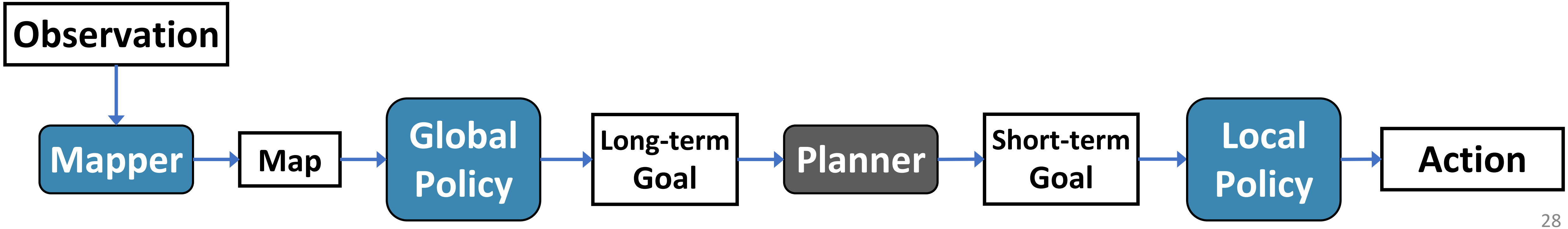
Active Neural SLAM



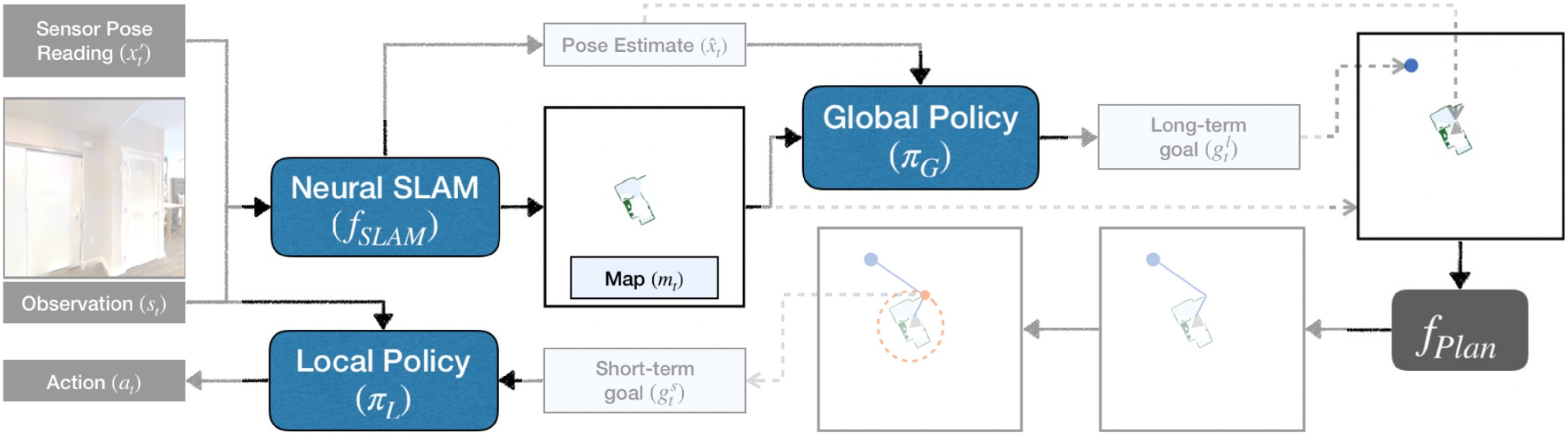


Active Neural SLAM



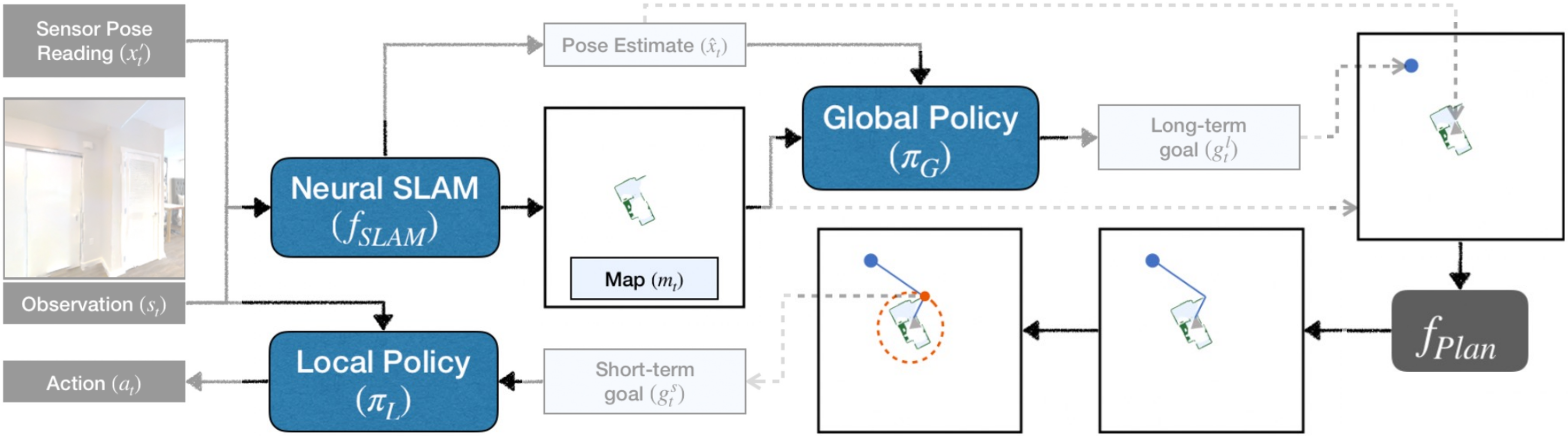


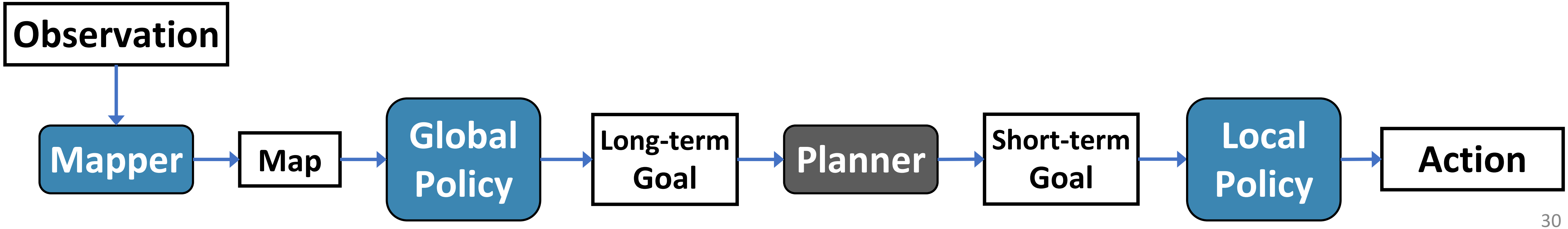
Active Neural SLAM



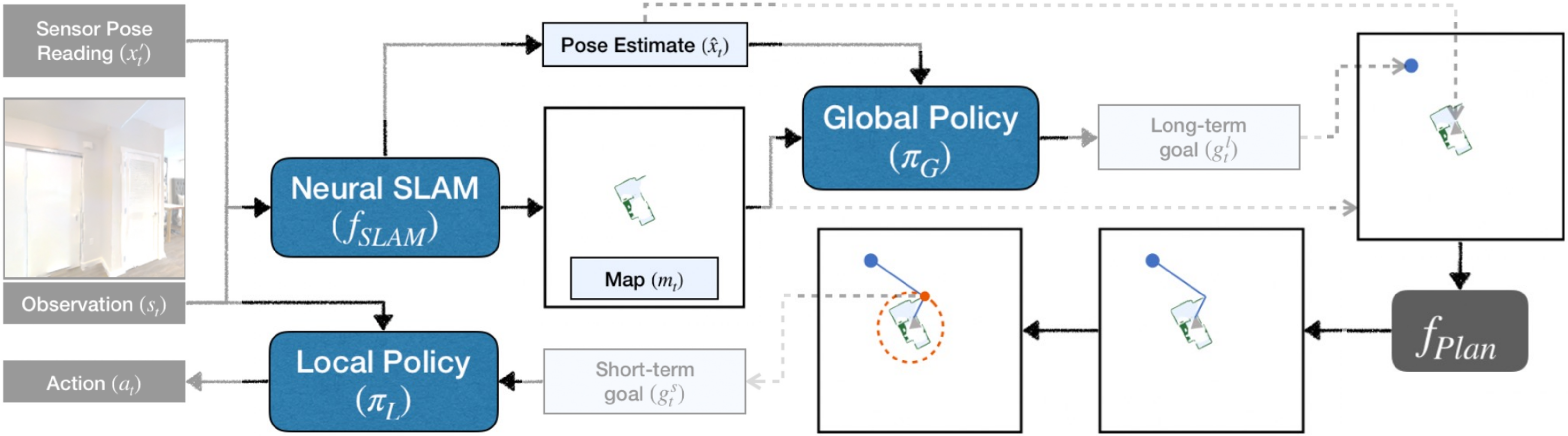
Observation → Mapper → Map → Global Policy → Long-term Goal → Planner → Short-term Goal → Local Policy → Action

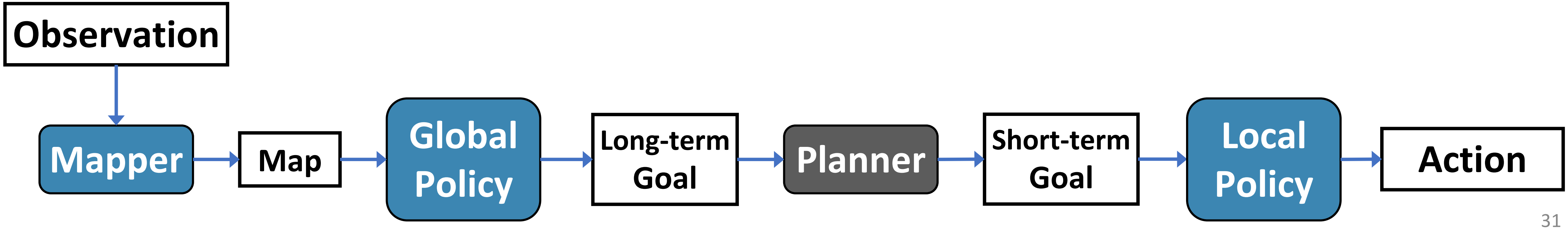
Active Neural SLAM



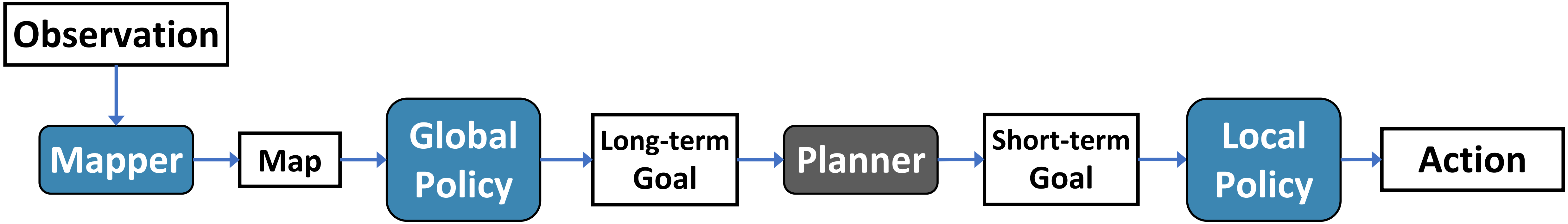


Active Neural SLAM

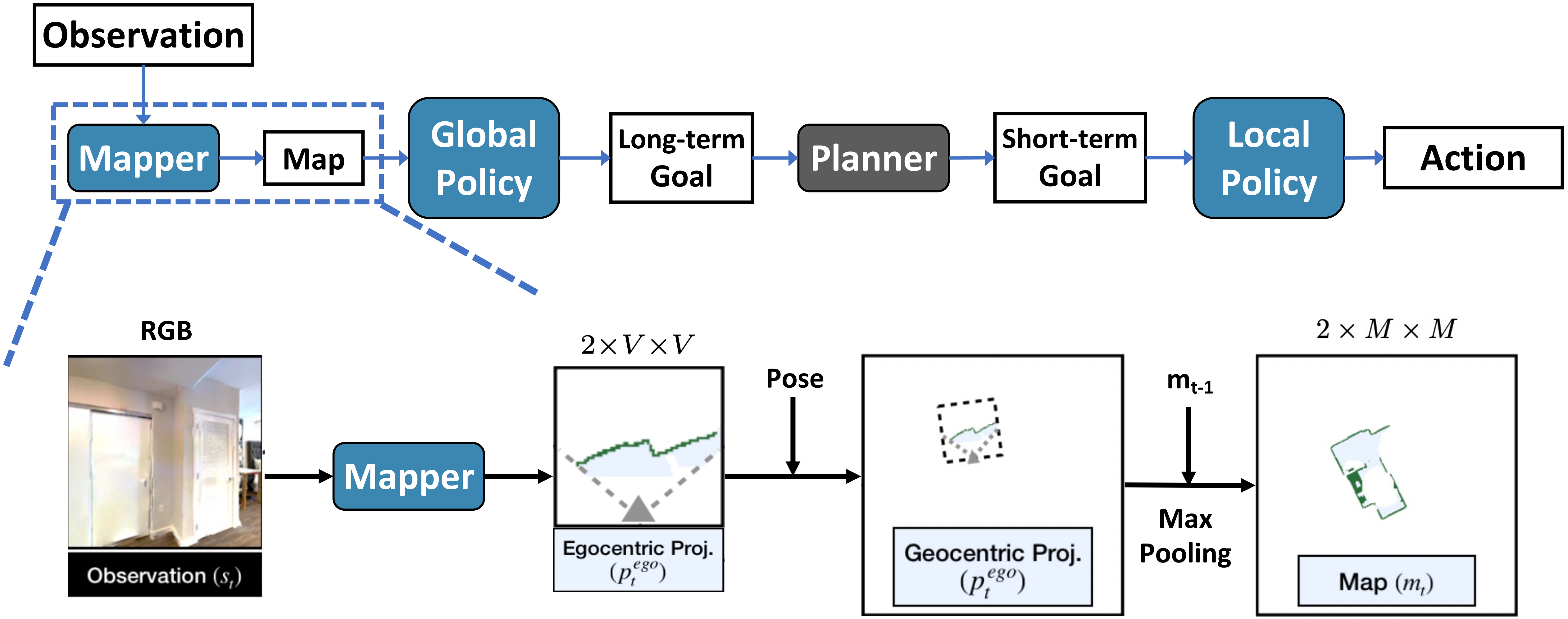




Active Neural SLAM

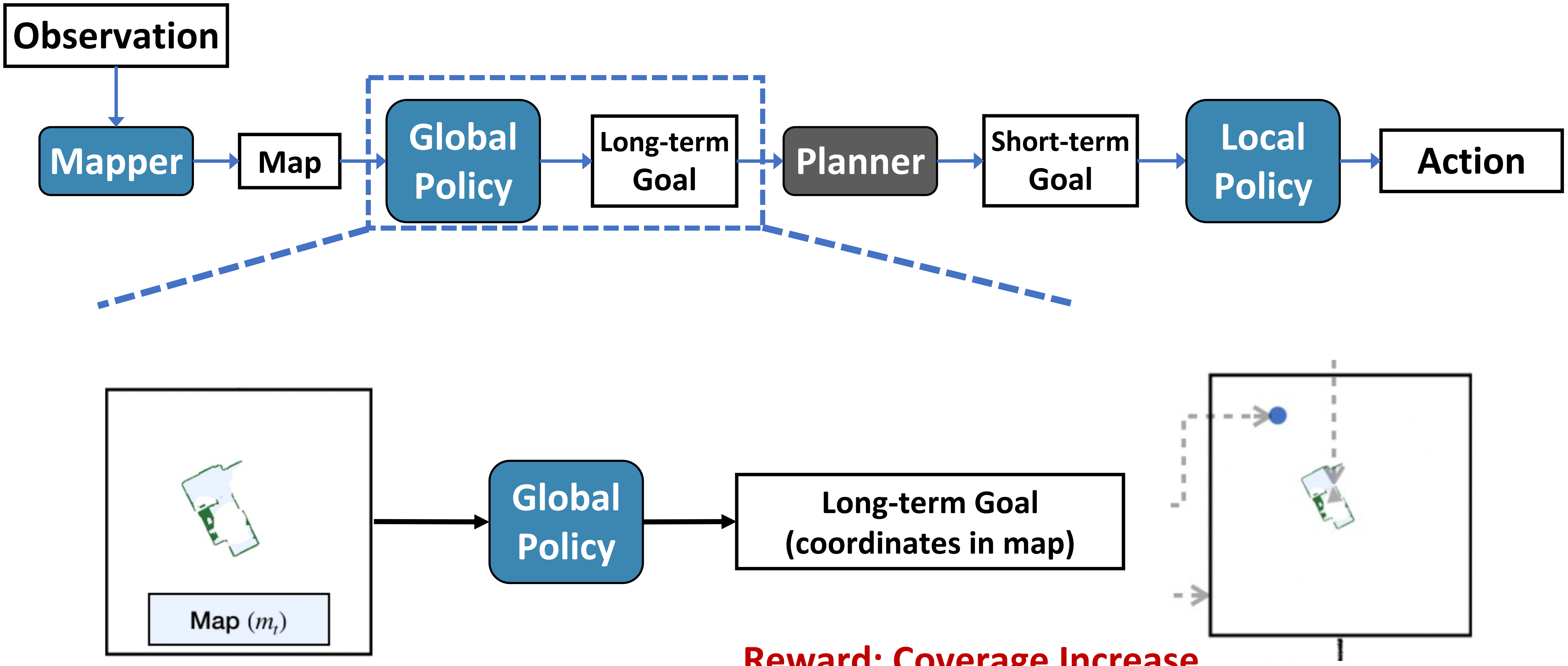


Active Neural SLAM



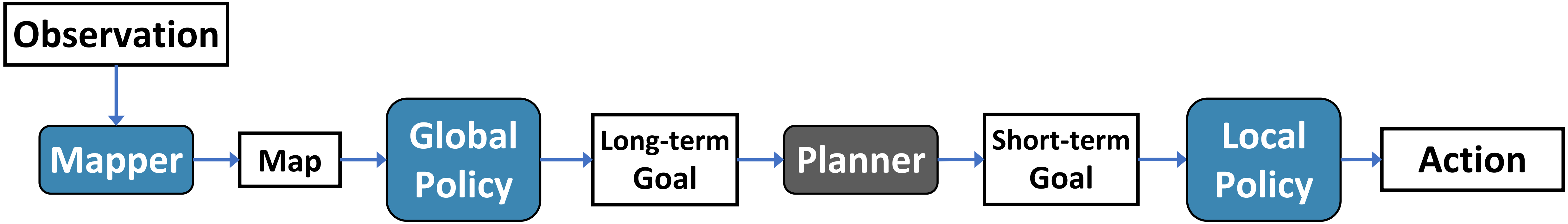
2 Channels of Occupancy Map: A Location being Occupied & Explored

Active Neural SLAM



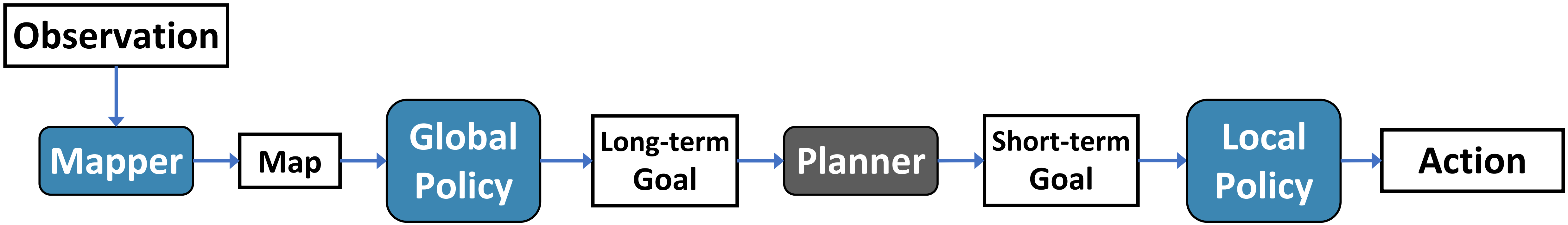
Reward: Coverage Increase

Active Neural SLAM



	Gibson Val		Domain Generalization MP3D Test	
Method	% Cov.	Cov. (m2)	% Cov.	Cov. (m2)
RL + 3LConv [1]	0.737	22.838	0.332	47.758
RL + Res18	0.747	23.188	0.341	49.175
RL + Res18 + AuxDepth [2]	0.779	24.467	0.356	51.959
RL + Res18 + ProjDepth [3]	0.789	24.863	0.378	54.775
Active Neural SLAM (ANS)	**0.948**	**32.701**	**0.521**	**73.281**

PointNav in Habitat Challenge 2019

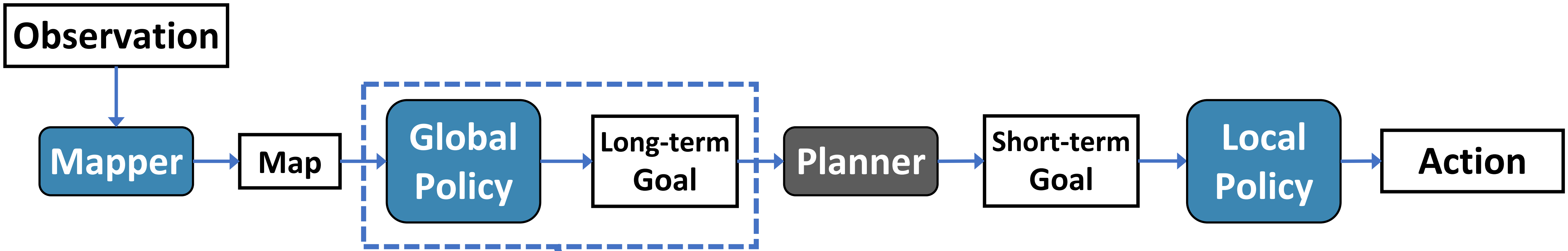


How to transfer to
PointNav in Habitat Challenge 2019

Goal: Go 5m south, 3m west.

Input: RGB/RGB-D + Perfect Pose Sensor

PointNav in Habitat Challenge 2019



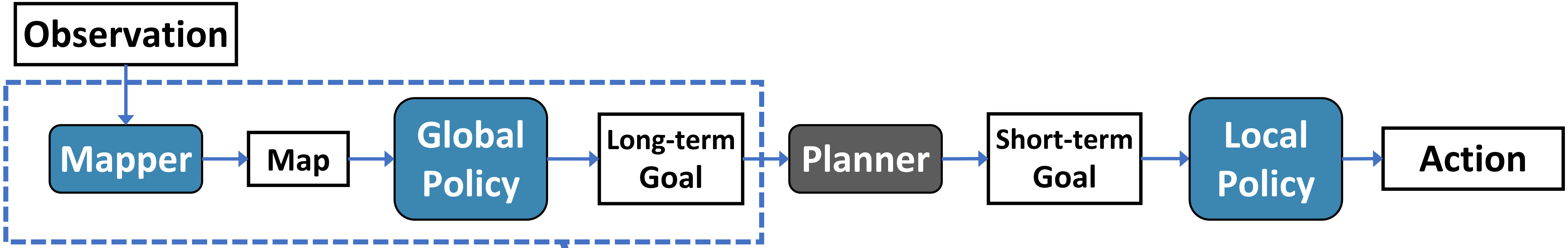
How to transfer to
PointNav in Habitat Challenge 2019

Goal: Go 5m south, 3m west.

Input: RGB/RGB-D + Perfect Pose Sensor

RGB Track: Fix the target point as the long-term goal

PointNav in Habitat Challenge 2019



How to transfer to
PointNav in Habitat Challenge 2019

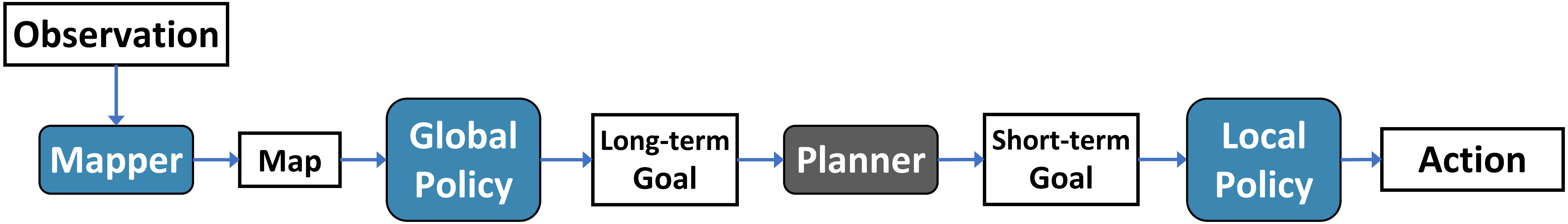
Goal: Go 5m south, 3m west.

Input: RGB/RGB-D + Perfect Pose Sensor

RGB Track: Fix the target point as the long-term goal

RGB-D Track: Fix the target point as the long-term goal + Projecting Mapper

PointNav in Habitat Challenge 2019



RGB Leaderboard

Rank	Team	SPL
1	Arnold	0.805
1	Mid-level-Features	0.800
3	CHROMA	0.712
4	ARF-RL	0.699
5	MTank	0.260
6	Policy-Police	0.260

RGB-D Leaderboard

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

Goal: Go 5m south, 3m west.

Input: RGB/RGB-D + Perfect Pose Sensor

PointNav in Habitat Challenge 2020

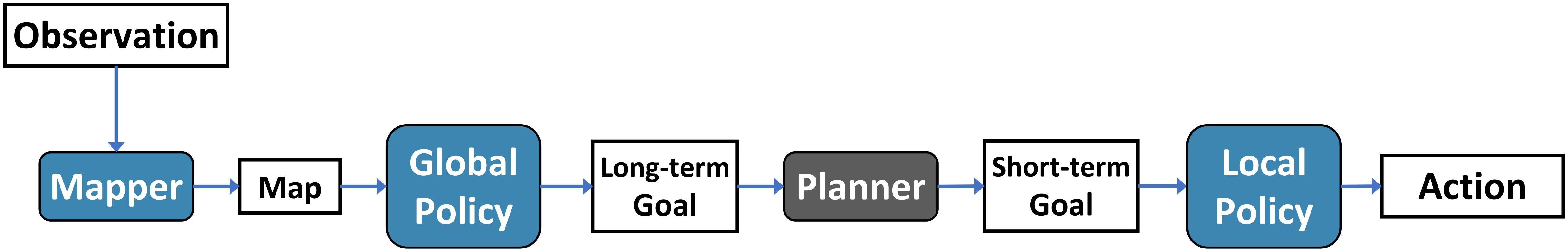
What’s new in PointNav of Habitat Challenge 2020:

Input: Only Noisy RGB-D Image

PointNav in Habitat Challenge 2020

What's new in PointNav of Habitat Challenge 2020:

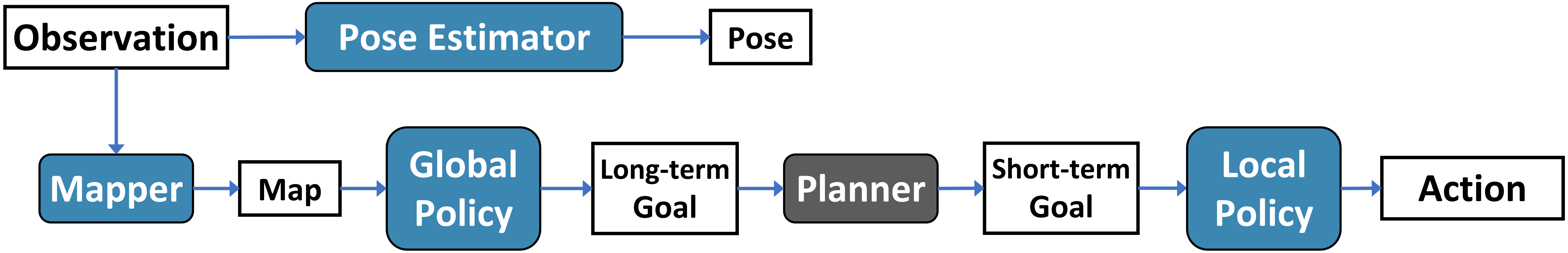
Input: Only Noisy RGB-D Image



PointNav in Habitat Challenge 2020

What's new in PointNav of Habitat Challenge 2020:

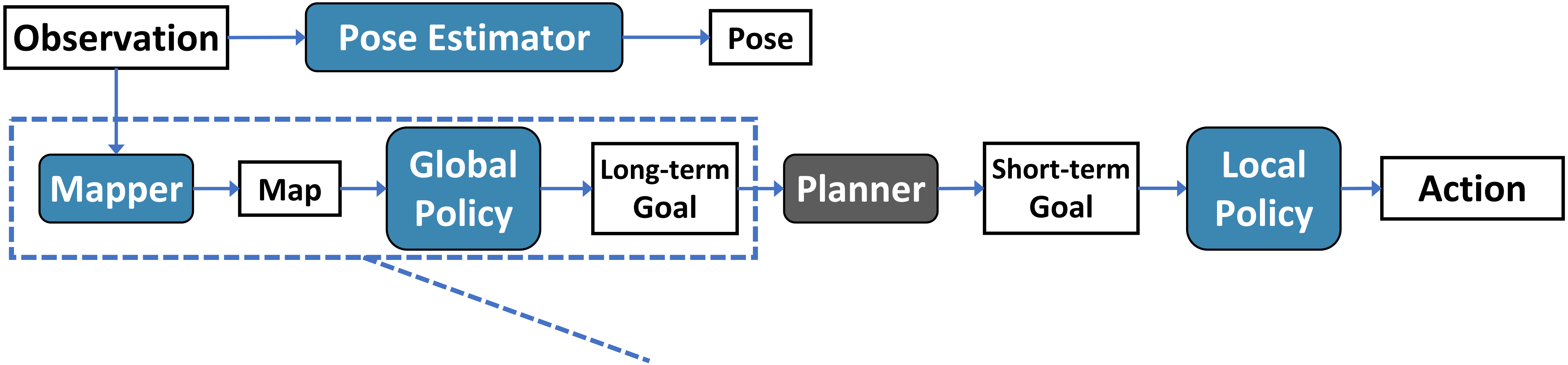
Input: Only Noisy RGB-D Image



PointNav in Habitat Challenge 2020

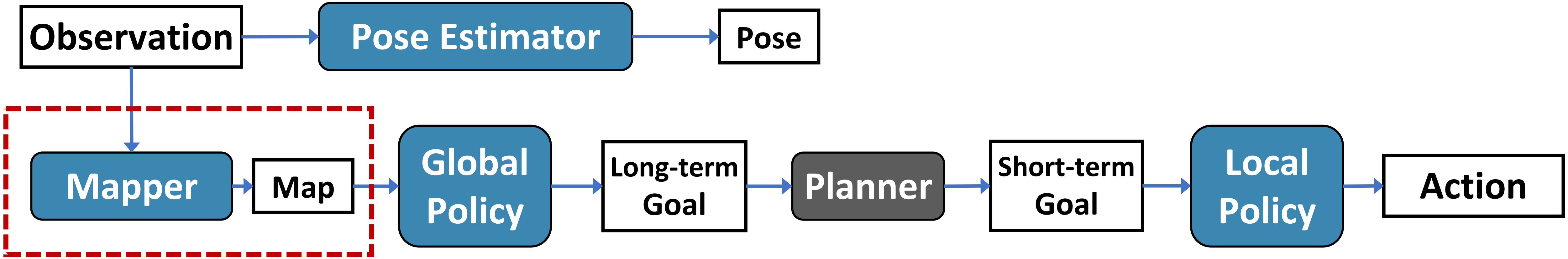
What's new in PointNav of Habitat Challenge 2020:

Input: Only Noisy RGB-D Image



Fix the target point as the long-term goal + Projecting Mapper

PointNav in Habitat Challenge 2020



2020 Winner!

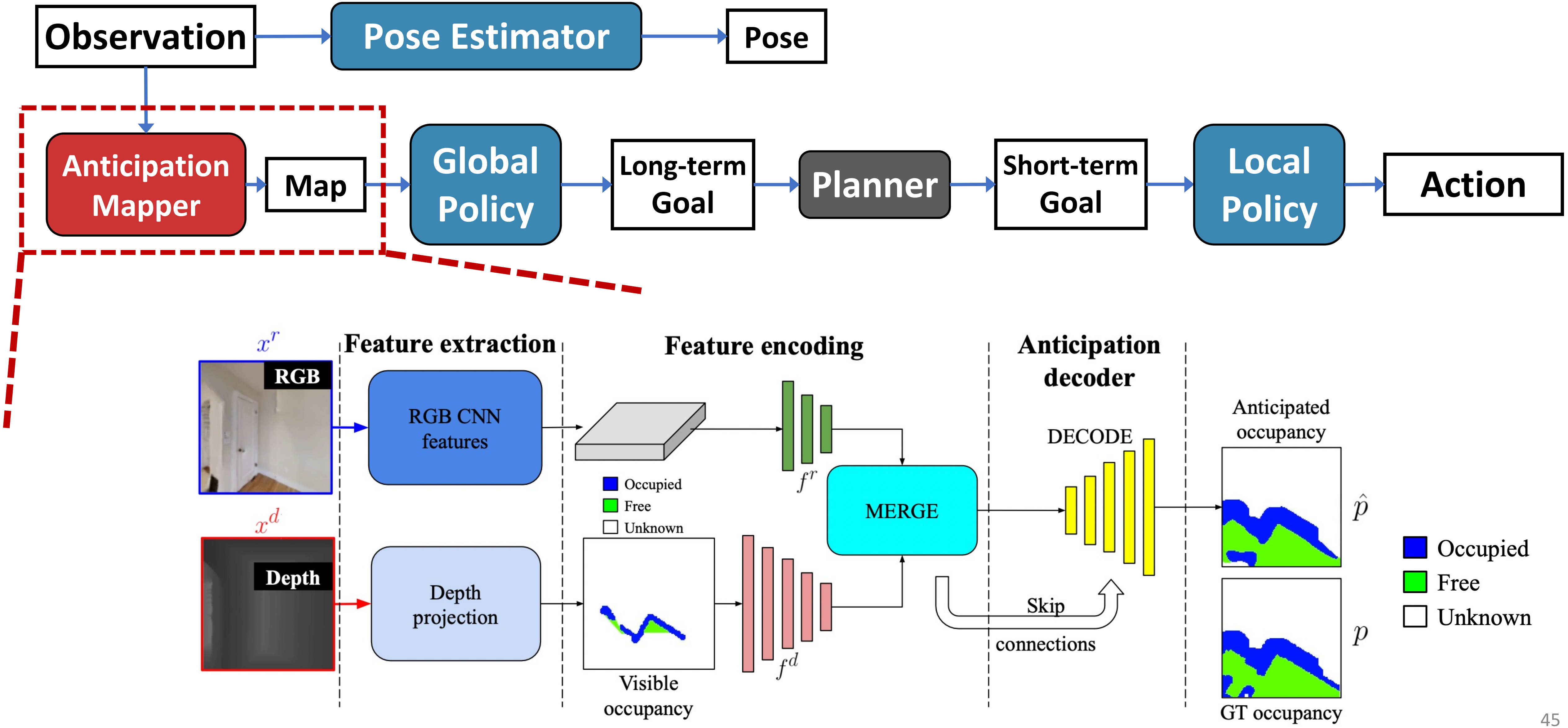
Occupancy Anticipation for Efficient Exploration and Navigation

(ECCV 2020)

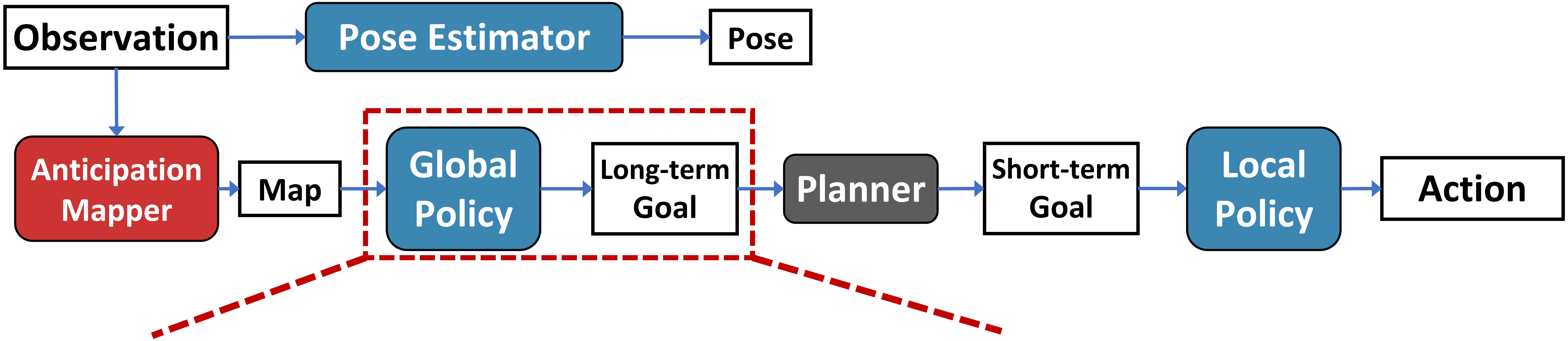
Santhosh K. Ramakrishnan[1,2], Ziad Al-Halah[1], and Kristen Grauman[1,2]

[1] The University of Texas at Austin, Austin TX 78712, USA
[2] Facebook AI Research, Austin TX 78701, USA

PointNav in Habitat Challenge 2020



PointNav in Habitat Challenge 2020

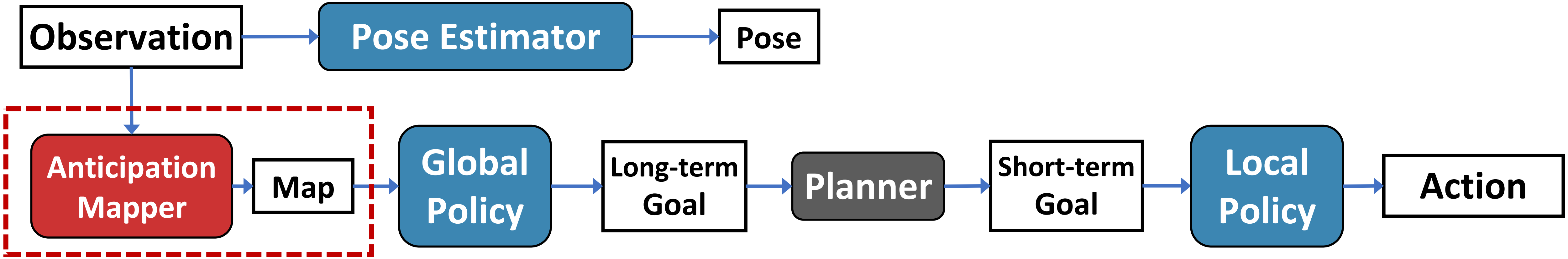


For Exploration: $$\text{Accuracy}(\hat{m}, m) = \sum^{G^2} \sum^{2} \mathbb{1}[\hat{m}_{ij} = m_{ij}], \quad m : \textit{Global Map}$$

$$R_t^{anticp} = \text{Accuracy}(\hat{m}_t, m) - \text{Accuracy}(\hat{m}_{t-1}, m).$$

For PointNav: **Fix the target location as global goal.**

PointNav in Habitat Challenge 2020



Occupancy Anticipation

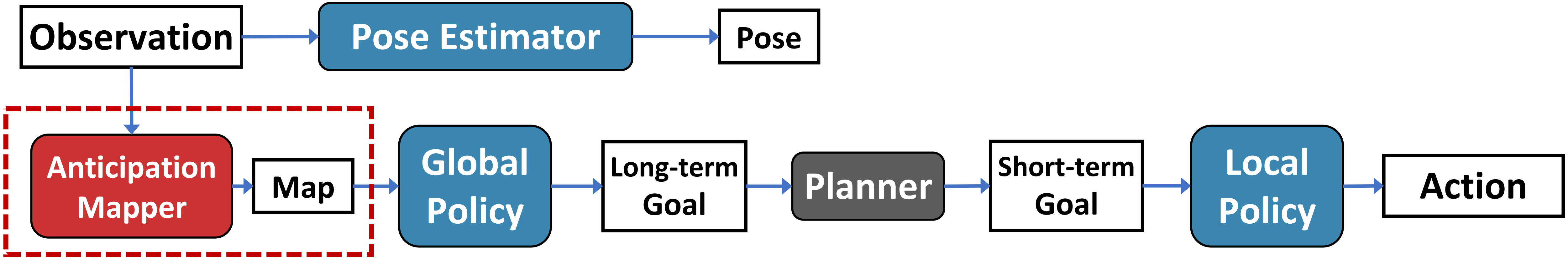
Rank	Team	SPL	SOFT_SPL	DISTANCE_TO_GOAL	SUCCESS
1	OccupancyAnticipation	0.21	0.50	2.29	0.28
2	ego-localization	0.15	0.60	1.82	0.19
3	DAN	0.13	0.24	4.00	0.25
4	Information Bottleneck	0.06	0.43	2.72	0.09
5	cogmodel_team	0.01	0.33	4.27	0.01
6	UCULab	0.001	0.11	5.97	0.002

2020 Winner!

Goal: Go 5m south, 3m west.

Input: RGB-D Image

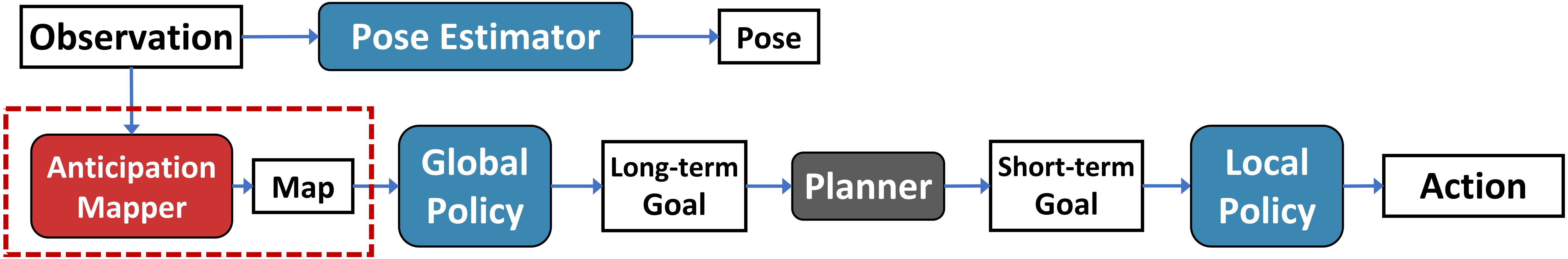
PointNav in Habitat Challenge 2020



Q: Why Occupancy Anticipation works for PointNav?

A: Better modeling the navigable spaces.

PointNav in Habitat Challenge 2020



Q: Why Occupancy Anticipation works for PointNav?

A: Better modeling the navigable spaces.

Q: Why the performance drops so much if not given a pose sensor?

A: Bad pose estimation becomes a drag on map-based methods.

ObjectNav in Habitat Challenge 2020

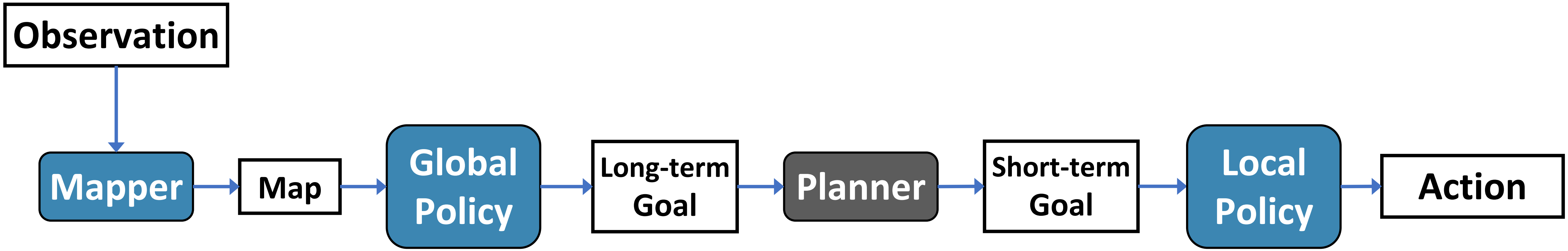
Input: **RGB-D Input + Oracle Pose Sensor**

Goal: ***Find a chair***

ObjectNav in Habitat Challenge 2020

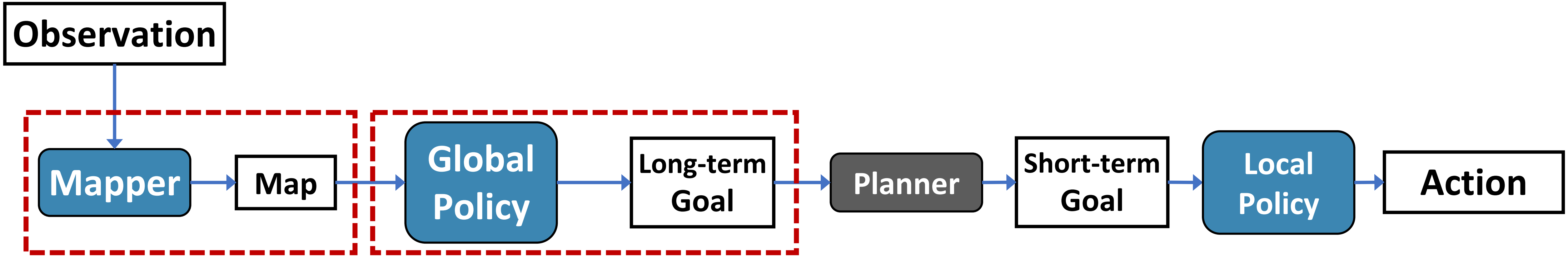
Input: **RGB-D Input + Oracle Pose Sensor**

Goal: ***Find a chair***



How to transfer the pipeline to Object Goal Navigation?

ObjectNav in Habitat Challenge 2020



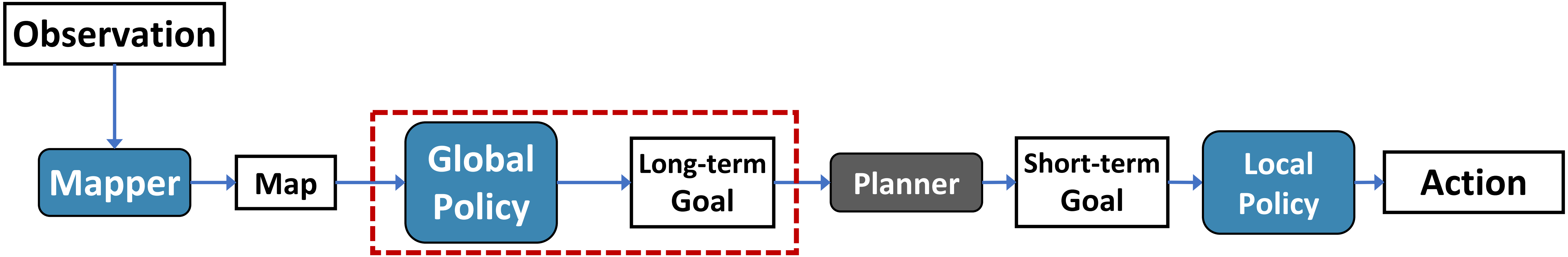
2020 Winner!

Object Goal Navigation using Goal-Oriented Semantic Exploration

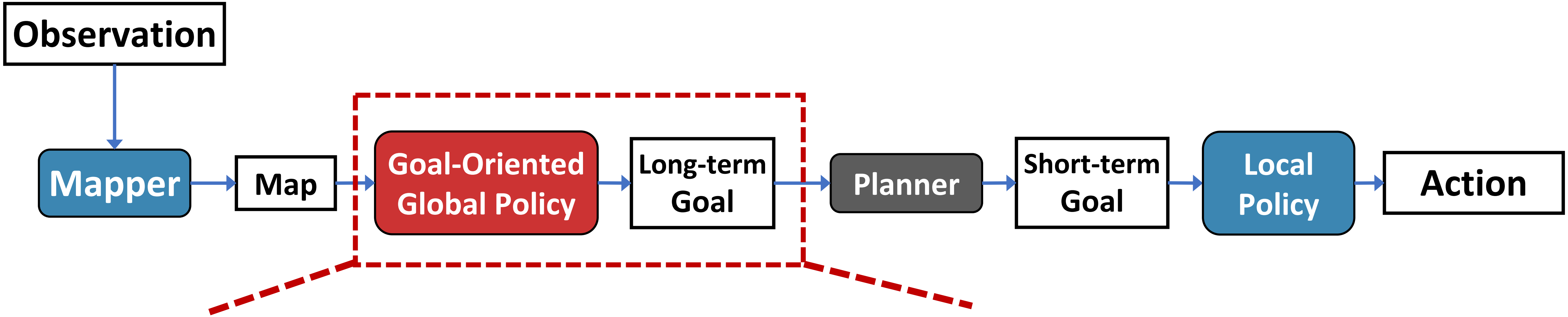
(NeurIPS 2020)

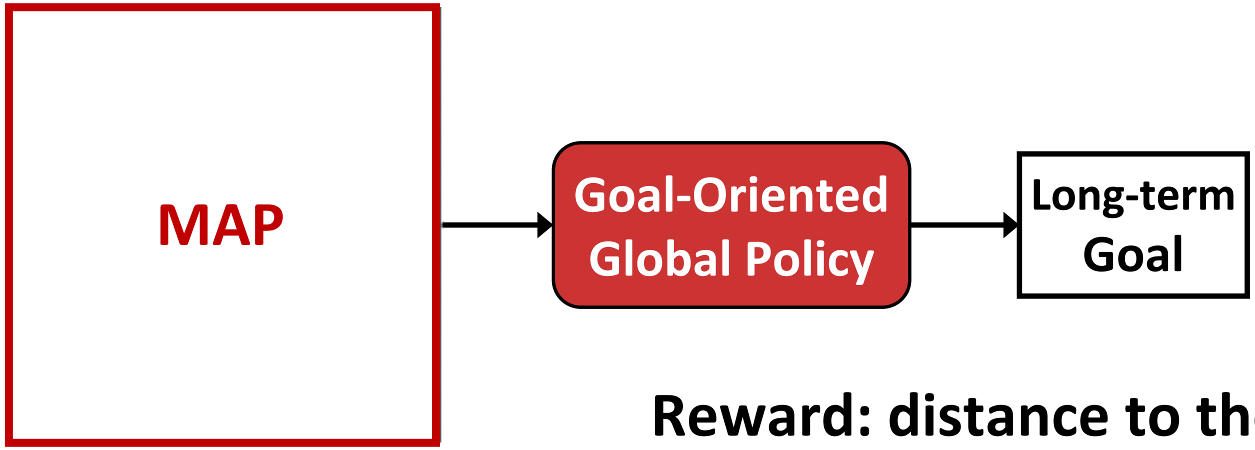
Devendra Singh Chaplot[1†], **Dhiraj Gandhi**[2], **Abhinav Gupta**[1,2*], **Ruslan Salakhutdinov**[1*]
[1]Carnegie Mellon University, [2]Facebook AI Research

ObjectNav in Habitat Challenge 2020



ObjectNav in Habitat Challenge 2020

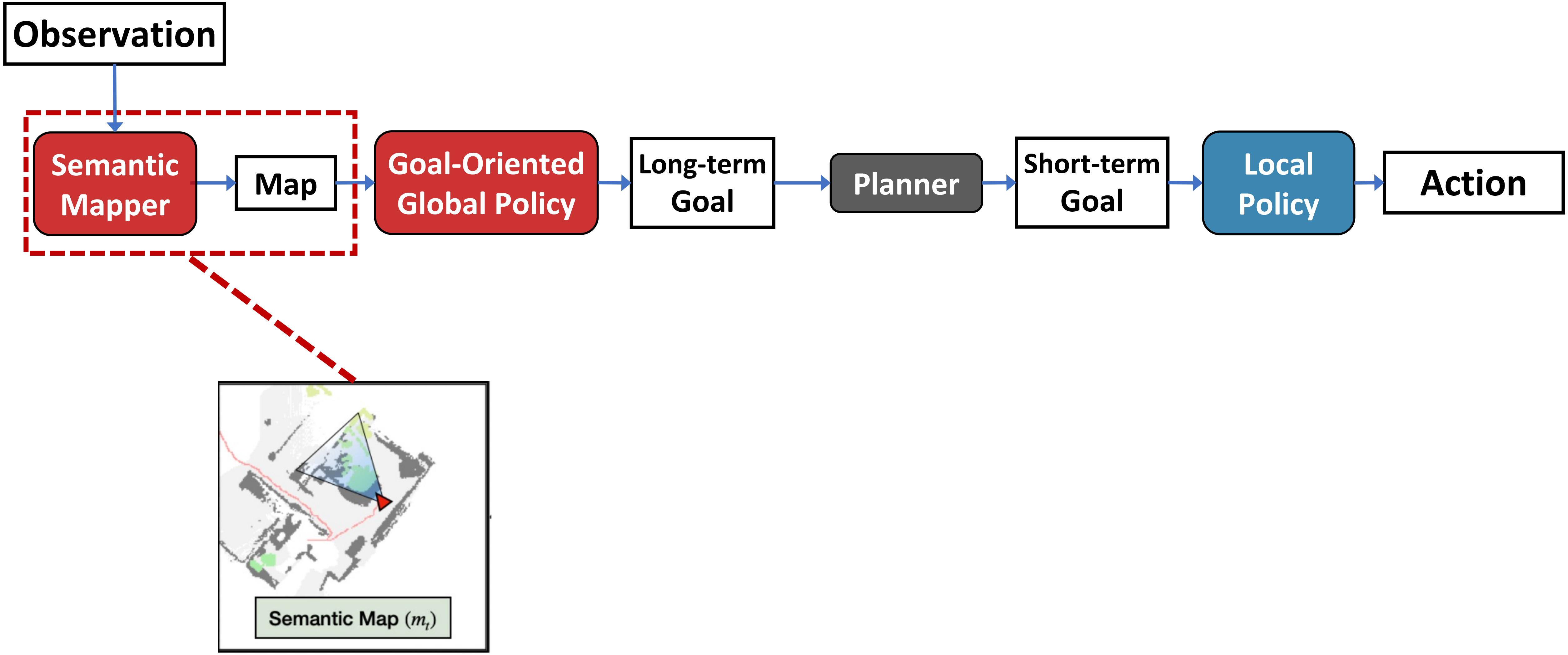




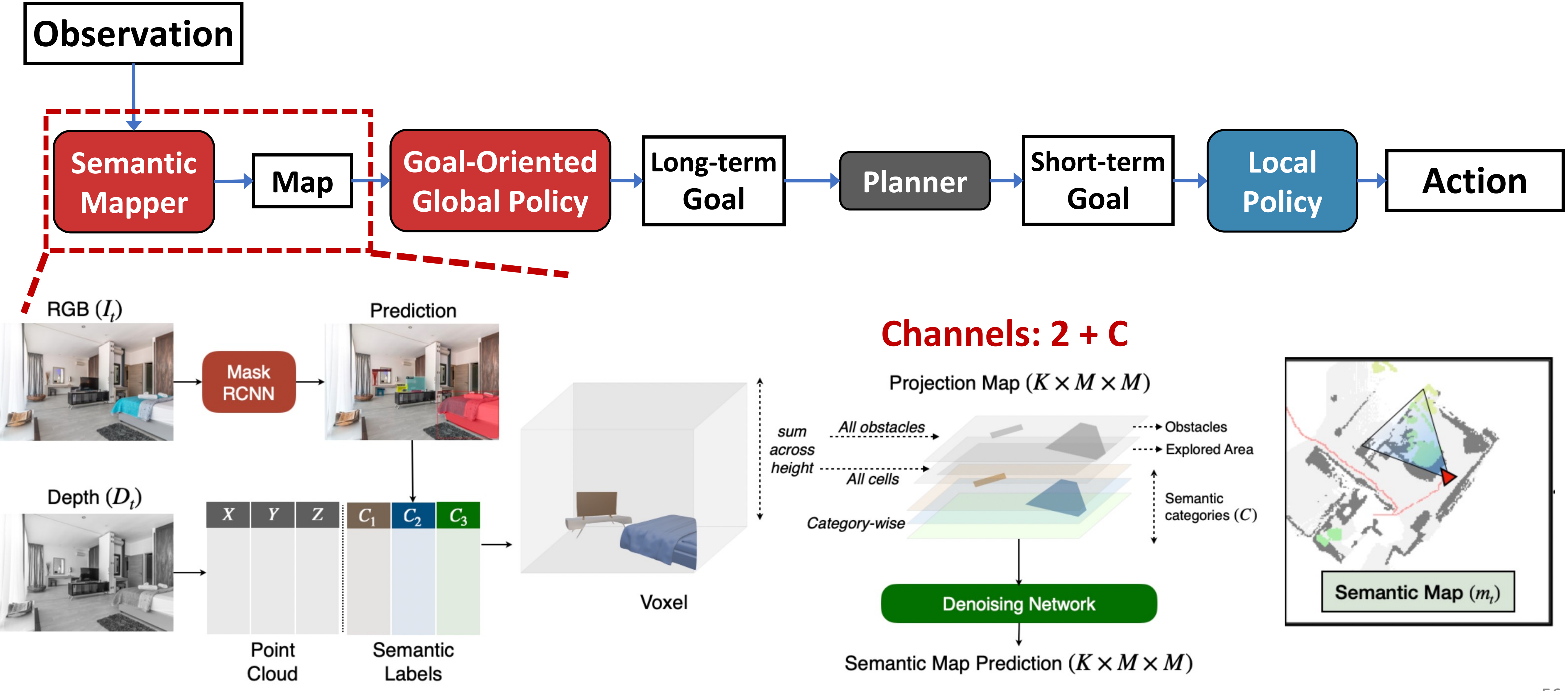
Where to check to reach the target object?

Reward: distance to the nearest goal

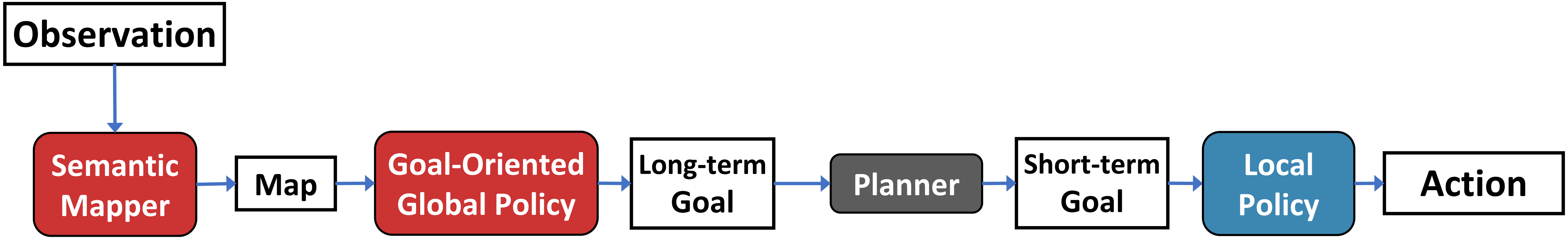
ObjectNav in Habitat Challenge 2020



ObjectNav in Habitat Challenge 2020



ObjectNav in Habitat Challenge 2020



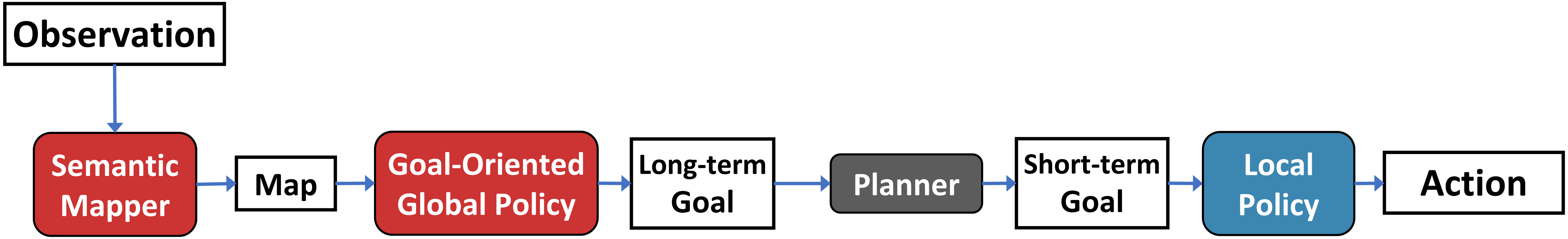
2020 Winner!

Rank	Team	SPL	SOFT_SPL	DISTANCE_TO_GOAL	SUCCESS
1	Arnold	0.10	0.18	6.33	0.25
2	SRCB-Robot-Sudoer	0.10	0.22	6.91	0.19
3	Active Exploration	0.05	0.17	7.34	0.13
4	Black Sheep	0.03	0.16	7.03	0.10
5	Blue Ox	0.02	0.14	7.23	0.07
6	UCULab	0.001	0.11	5.97	0.002

Goal: ***Go find a chair.***

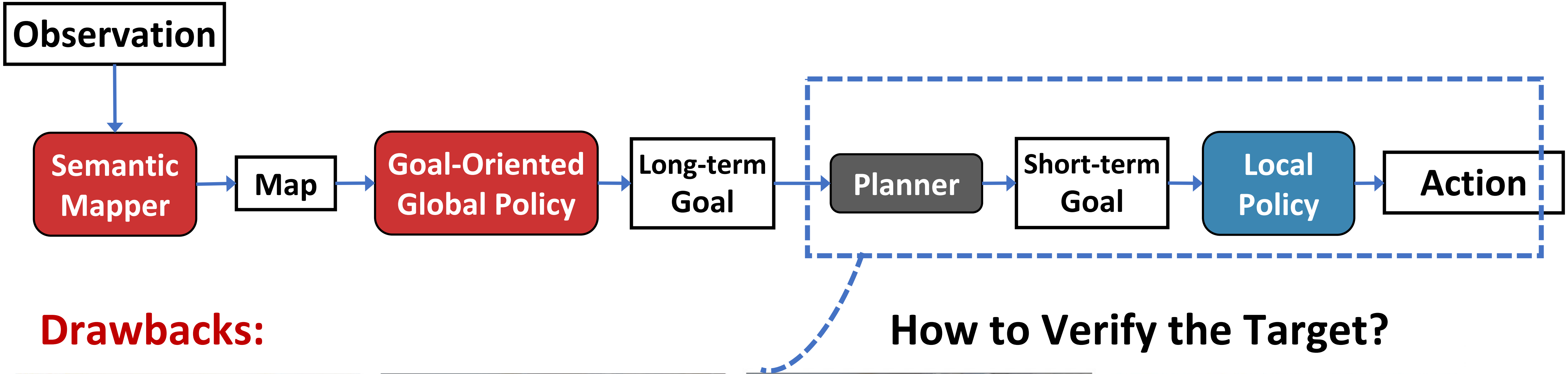
Input: **RGB-D Image + Oracle Pose**

ObjectNav in Habitat Challenge 2020



Drawbacks:

ObjectNav in Habitat Challenge 2020



Drawbacks:

How to Verify the Target?

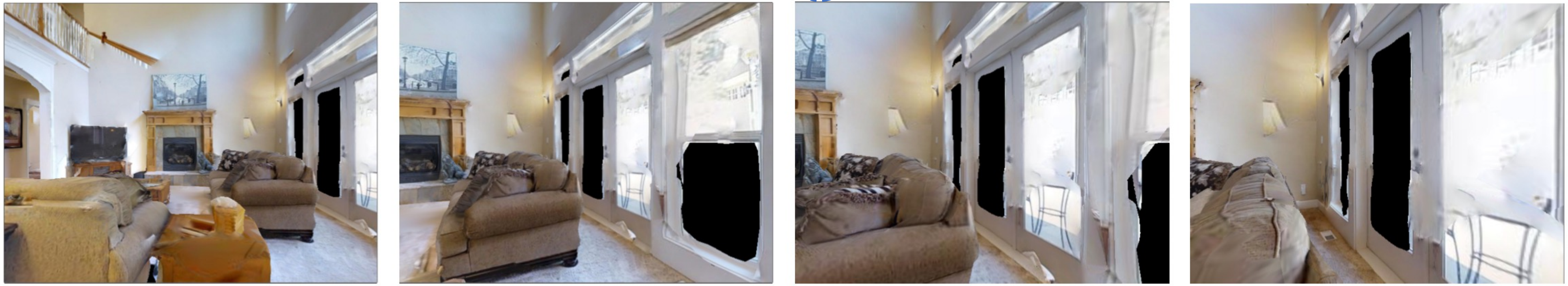


Fig. 4: Tasked to go to a bed, the agent mistakes the sofa as a bed in the last frame and stops. Such a false detection results in failure.

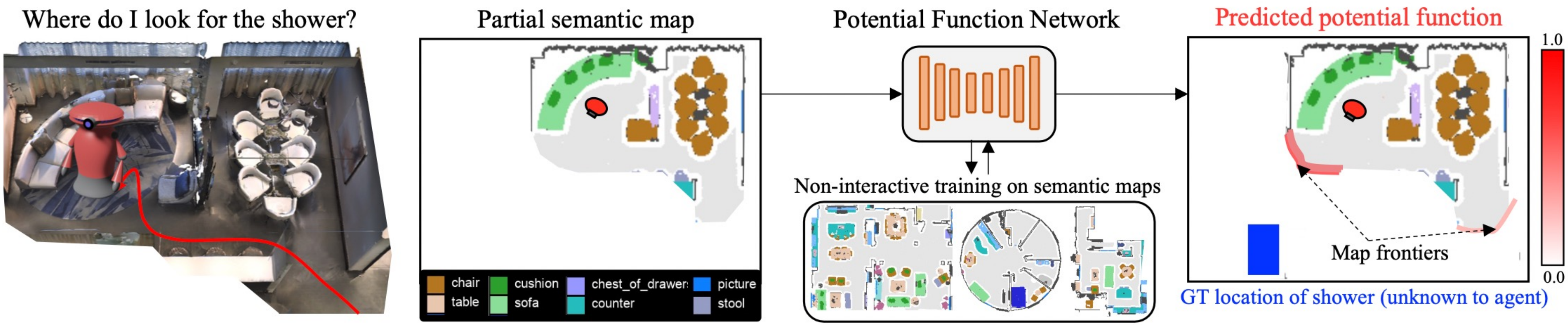
Fig. 4 comes from ***Stubborn: A Strong Baseline for Indoor Object Navigation*, Hankuan Luo et al.**

Other Map-Based Methods for ObjecNav

PONI: Potential Functions for ObjectGoal Navigation with Interaction-free Learning

(CVPR 2022, Oral)

Santhosh Kumar Ramakrishnan[1,2], Devendra Singh Chaplot[1], Ziad Al-Halah[2], Jitendra Malik[1,3], Kristen Grauman[1,2]
[1]Meta AI [2]UT Austin [3]UC Berkeley



Selected Papers

Map-Based

Vision-Based

PointNav in Habitat Challenge 2021

2021 Runner-up for PointNav!

Robust Visual Odometry for Realistic Point-Goal Navigation

Ruslan Partsey[1], Oleksandr Maksymets[2], and Oles Dobosevych[1]

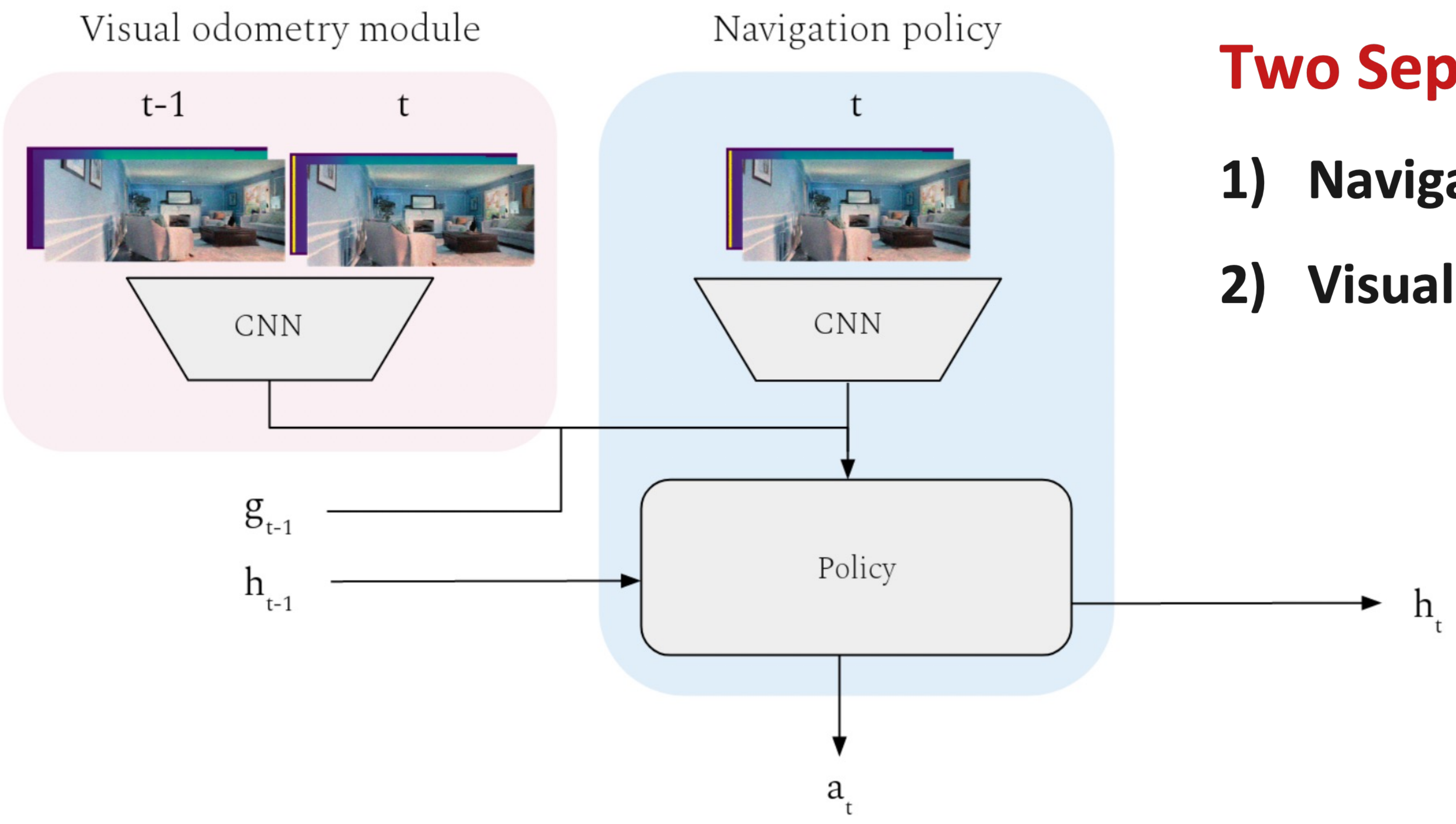
[1] Ukrainian Catholic University
[2] Facebook AI Research

Goal: **Go 5m south, 3m west relative to start.**

Input: **Only RGB-D Image with noise.**

PointNav in Habitat Challenge 2021

Overview

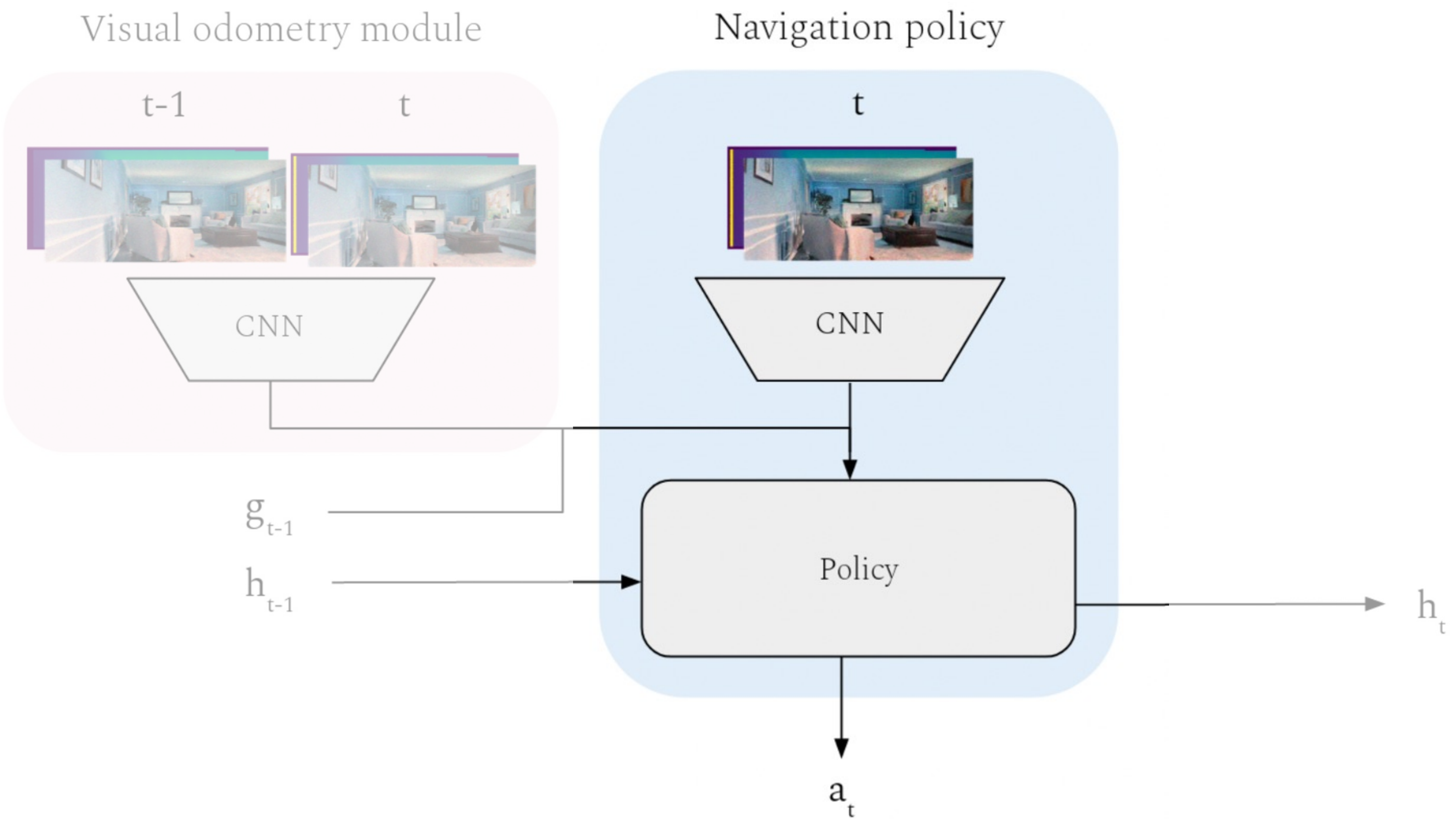


Two Separately-Trained Modules:

1) **Navigation Policy (blue)**
2) **Visual Odometry (pink)**

PointNav in Habitat Challenge 2021

Overview



Navigation Policy:

Architecture: LSTM + ResNet18

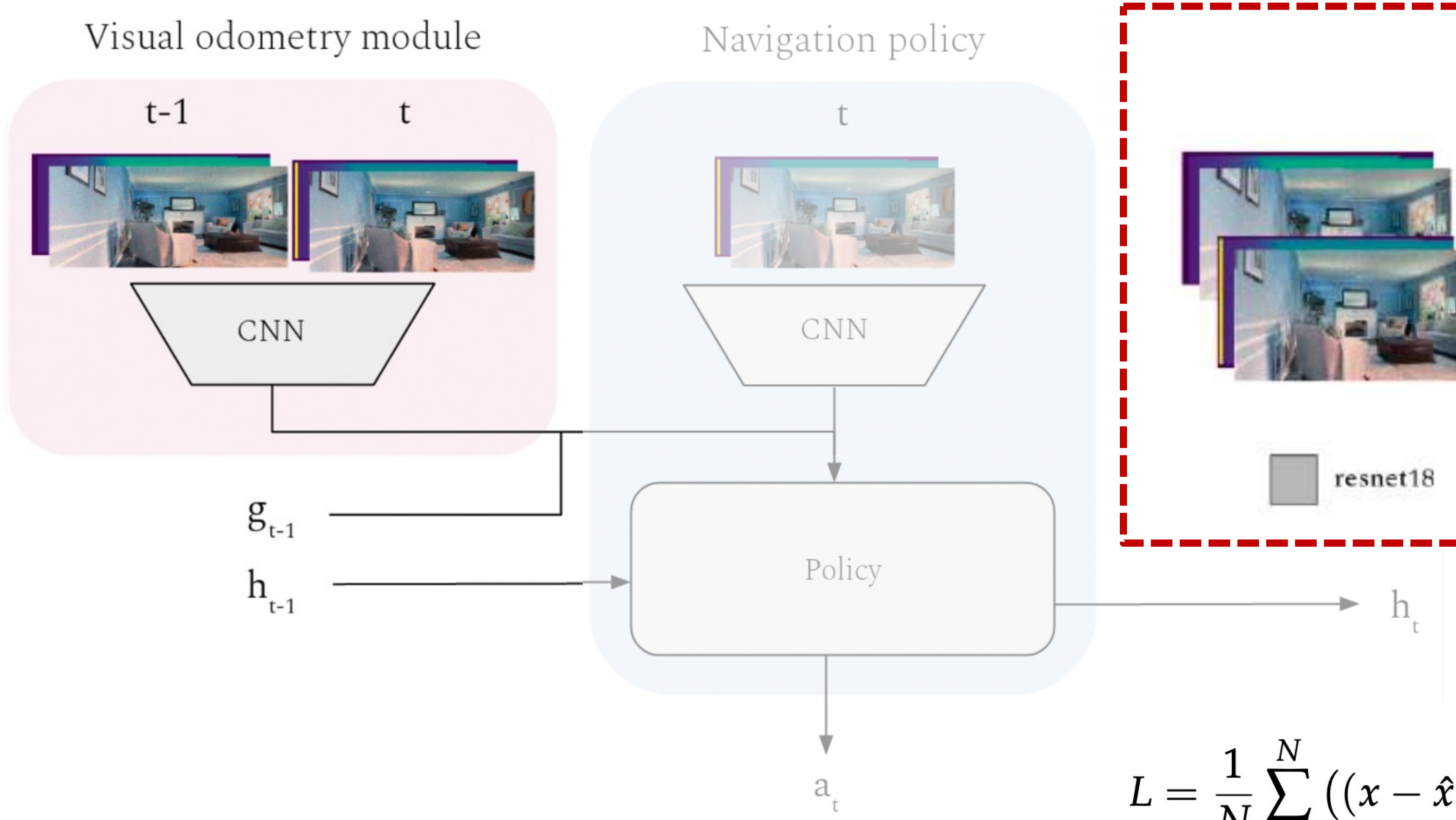
Training: PPO

Input: Depth + GT Pose + …
(Replace GT Pose with Visual Odometry in evaluation)

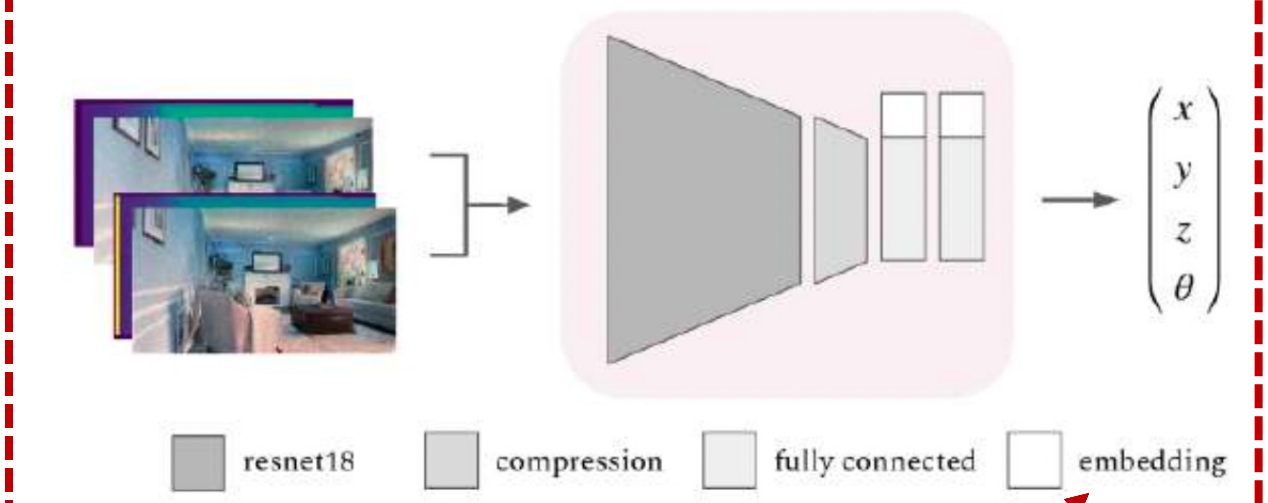
Output: Action, …

PointNav in Habitat Challenge 2021

Overview



Visual Odometry



meta-information (action taken/collision)

$$L = \frac{1}{N}\sum_{i=1}^{N}\left((x-\hat{x})^2 + (y-\hat{y})^2 + (z-\hat{z})^2\right) + \frac{1}{N}\sum_{i=1}^{N}(\theta-\hat{\theta})^2$$

PointNav in Habitat Challenge 2021

Tricks for Visual Odometry

1) How to embed meta-info?
2) Whether to crop out interested area?
3) Data Augmentation?

TABLE 4.7: Visual odometry metrics (subject to 1e+2 multiplication).

Experiment name	Epoch	Translation MAE Total	Forward	Left	Right	Rotation MAE Total	Forward	Left	Right
bs	20	4.61	5.41	3.48	3.67	2.37	1.58	3.61	3.17
	40	4.10	4.53	3.47	3.65	1.85	1.29	2.65	2.49
bs + col_emb	20	5.51	6.99	3.48	3.73	2.98	2.15	3.90	4.18
	40	4.65	5.45	3.56	3.67	2.40	1.64	3.28	3.45
bs + act_emb	20	3.13	2.86	3.45	3.48	1.32	0.99	1.74	1.76
	40	2.89	2.39	3.43	3.65	1.15	0.78	1.62	1.65
bs + col_emb + act_emb	20	3.10	2.80	3.37	3.62	1.37	1.02	1.79	1.87
	40	3.00	2.56	3.48	3.65	1.26	0.84	1.74	1.85
bs + act_emb 2fc	20	3.00	2.67	3.39	3.47	1.25	0.96	1.58	1.68
	40	2.89	2.43	3.41	3.58	1.16	0.82	1.59	1.63
bs + col_emb + act_emb 2fc	20	3.08	2.80	3.38	3.49	1.24	0.94	1.59	1.65
	40	2.87	2.43	3.31	3.57	1.17	0.85	1.55	1.60
bs + act_emb 2fc + vflip	20	2.92	2.54	3.37	3.44	1.14	0.85	1.51	1.53
	40	2.73	2.30	3.23	3.32	1.03	0.73	1.41	1.41
bs + act_emb 2fc + vflip + inv_rot	20	2.89	2.62	3.20	3.30	1.06	0.82	1.37	1.37
	40	2.76	2.42	3.17	3.23	0.96	0.69	1.28	1.32
bs + act_emb 2fc + 320x450->160x225	20	3.28	3.41	3.03	3.19	1.24	1.09	1.40	1.49
	40	3.14	3.18	3.01	3.16	1.15	0.95	1.38	1.44
bs + act_emb 2fc + 320x450->180x320	20	3.12	3.15	3.02	3.16	1.26	1.02	1.57	1.57
	40	2.94	2.89	3.00	3.00	1.18	1.00	1.41	1.42
bs + act_emb 2fc + 320x350->180x320	20	3.17	3.58	2.59	2.67	1.18	1.14	1.19	1.27
	40	2.76	3.05	2.36	2.41	0.98	0.95	1.02	1.00
sepact_right	20	3.49	NA	NA	3.49	1.73	NA	NA	1.73
	40	3.46	NA	NA	3.46	1.71	NA	NA	1.71
sepact_left	20	3.34	NA	3.34	NA	1.76	NA	1.76	NA
	40	3.27	NA	3.27	NA	1.55	NA	1.55	NA
sepact_fwd	20	2.75	2.75	NA	NA	0.97	0.97	NA	NA
	40	2.27	2.27	NA	NA	0.77	0.77	NA	NA
bs + act_emb 2fc + lscale	20	2.37	1.88	2.97	3.03	0.82	0.53	1.22	1.18
	40	2.21	1.65	2.91	2.95	0.77	0.48	1.15	1.15

PointNav in Habitat Challenge 2021

Tricks for Visual Odometry

1) How to embed meta-info?
2) Whether to crop out interested area?
3) Data Augmentation?

Questions:

1) Why a simple LSTM + a tricky Visual Odometry work so well for PointNav?

2) With a good Visual Odometry, will map-based methods also work well for the task?

TABLE 4.7: Visual odometry metrics (subject to 1e+2 multiplication).

Experiment name	Epoch	Translation MAE				Rotation MAE			
		Total	Forward	Left	Right	Total	Forward	Left	Right
bs	20	4.61	5.41	3.48	3.67	2.37	1.58	3.61	3.17
	40	4.10	4.53	3.47	3.65	1.85	1.29	2.65	2.49
bs + col_emb	20	5.51	6.99	3.48	3.73	2.98	2.15	3.90	4.18
	40	4.65	5.45	3.56	3.67	2.40	1.64	3.28	3.45
bs + act_emb	20	3.13	2.86	3.45	3.48	1.32	0.99	1.74	1.76
	40	2.89	2.39	3.43	3.65	1.15	0.78	1.62	1.65
bs + col_emb + act_emb	20	3.10	2.80	3.37	3.62	1.37	1.02	1.79	1.87
	40	3.00	2.56	3.48	3.65	1.26	0.84	1.74	1.85
bs + act_emb 2fc	20	3.00	2.67	3.39	3.47	1.25	0.96	1.58	1.68
	40	2.89	2.43	3.41	3.58	1.16	0.82	1.59	1.63
bs + col_emb + act_emb 2fc	20	3.08	2.80	3.38	3.49	1.24	0.94	1.59	1.65
	40	2.87	2.43	3.31	3.57	1.17	0.85	1.55	1.60
bs + act_emb 2fc + vflip	20	2.92	2.54	3.37	3.44	1.14	0.85	1.51	1.53
	40	2.73	2.30	3.23	3.32	1.03	0.73	1.41	1.41
bs + act_emb 2fc + vflip + inv_rot	20	2.89	2.62	3.20	3.30	1.06	0.82	1.37	1.37
	40	2.76	2.42	3.17	3.23	0.96	0.69	1.28	1.32
bs + act_emb 2fc + 320x450->160x225	20	3.28	3.41	3.03	3.19	1.24	1.09	1.40	1.49
	40	3.14	3.18	3.01	3.16	1.15	0.95	1.38	1.44
bs + act_emb 2fc + 320x450->180x320	20	3.12	3.15	3.02	3.16	1.26	1.02	1.57	1.57
	40	2.94	2.89	3.00	3.00	1.18	1.00	1.41	1.42
bs + act_emb 2fc + 320x350->180x320	20	3.17	3.58	2.59	2.67	1.18	1.14	1.19	1.27
	40	2.76	3.05	2.36	2.41	0.98	0.95	1.02	1.00
sepact_right	20	3.49	NA	NA	3.49	1.73	NA	NA	1.73
	40	3.46	NA	NA	3.46	1.71	NA	NA	1.71
sepact_left	20	3.34	NA	3.34	NA	1.76	NA	1.76	NA
	40	3.27	NA	3.27	NA	1.55	NA	1.55	NA
sepact_fwd	20	2.75	2.75	NA	NA	0.97	0.97	NA	NA
	40	2.27	2.27	NA	NA	0.77	0.77	NA	NA
bs + act_emb 2fc + lscale	20	2.37	1.88	2.97	3.03	0.82	0.53	1.22	1.18
	40	2.21	1.65	2.91	2.95	0.77	0.48	1.15	1.15

ObjectNav in Habitat Challenge 2021

2021 Winner for ObjectNav!

Auxiliary Tasks and Exploration Enable ObjectGoal Navigation

(ICCV 2021)

Joel Ye[1*] Dhruv Batra[1,2] Abhishek Das[2] Erik Wijmans[1]

[1]Georgia Institute of Technology [2] Facebook AI Research

Goal: ***Go find a chair.***

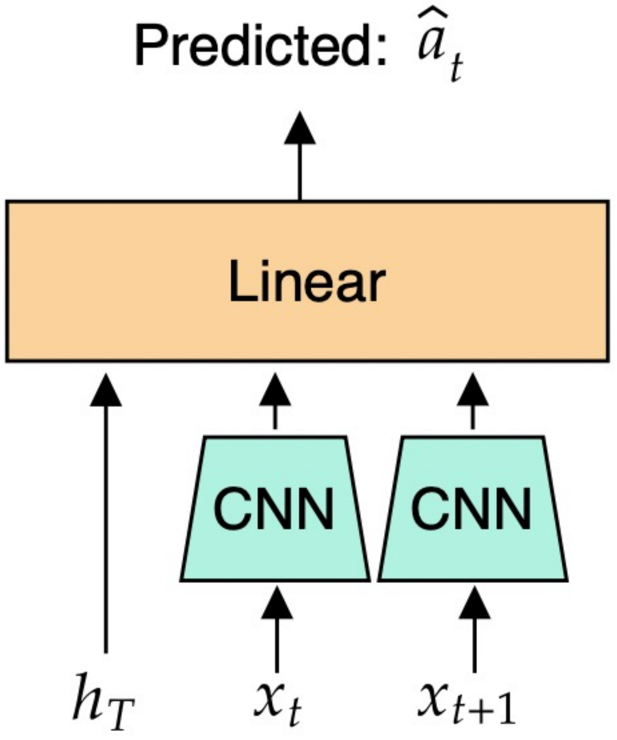
Input: **RGB-D Image with noise + Perfect Pose Sensor**

ObjectNav in Habitat Challenge 2021

Auxiliary Tasks:

(1) Inverse Dynamics (ID)

Decoding action taken from two successive observations.



(a) Inverse Dynamics

$$L_{ID} = \sum_{i=1}^{T-1} L_{CE}(I(\phi_i, \phi_{i+1}, h_T), a_i)$$

ϕ_t : visual embeddings

ObjectNav in Habitat Challenge 2021

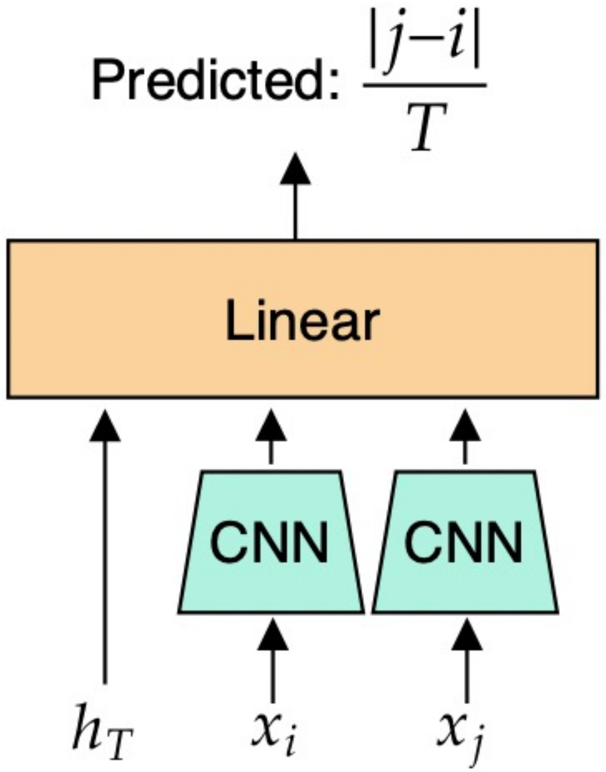
Auxiliary Tasks:

(1) Inverse Dynamics (ID)

Decoding action taken from two successive observations.

(2) Temporal Distance (TD)

Decoding the timestep difference between two observations.



(b) Temporal Distance

$$L_{TD} = \frac{1}{2}((i-j) - \mathcal{T}(\phi_i, \phi_j, h_T))^2$$

ObjectNav in Habitat Challenge 2021

Auxiliary Tasks:

(1) Inverse Dynamics (ID)

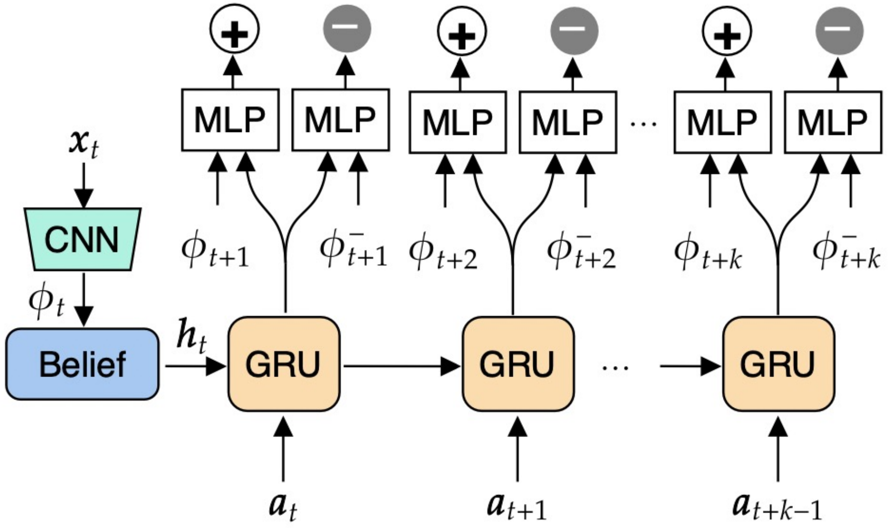
Decoding action taken from two successive observations.

(2) Temporal Distance (TD)

Decoding the timestep difference between two observations.

(3) Action-Conditional Contrastive Predictive Coding (CPC|A)

Decoding future visual embeddings at every timestep from other visual embeddings using a secondary GRU.



(c) CPCIA

ObjectNav in Habitat Challenge 2021

Auxiliary Tasks:

(1) Inverse Dynamics (ID)

Decoding action taken from two successive observations.

(2) Temporal Distance (TD)

Decoding the timestep difference between two observations.

(3) Action-Conditional Contrastive Predictive Coding (CPC|A)

Decoding future observation embeddings at every timestep from other observation embeddings using a secondary GRU.

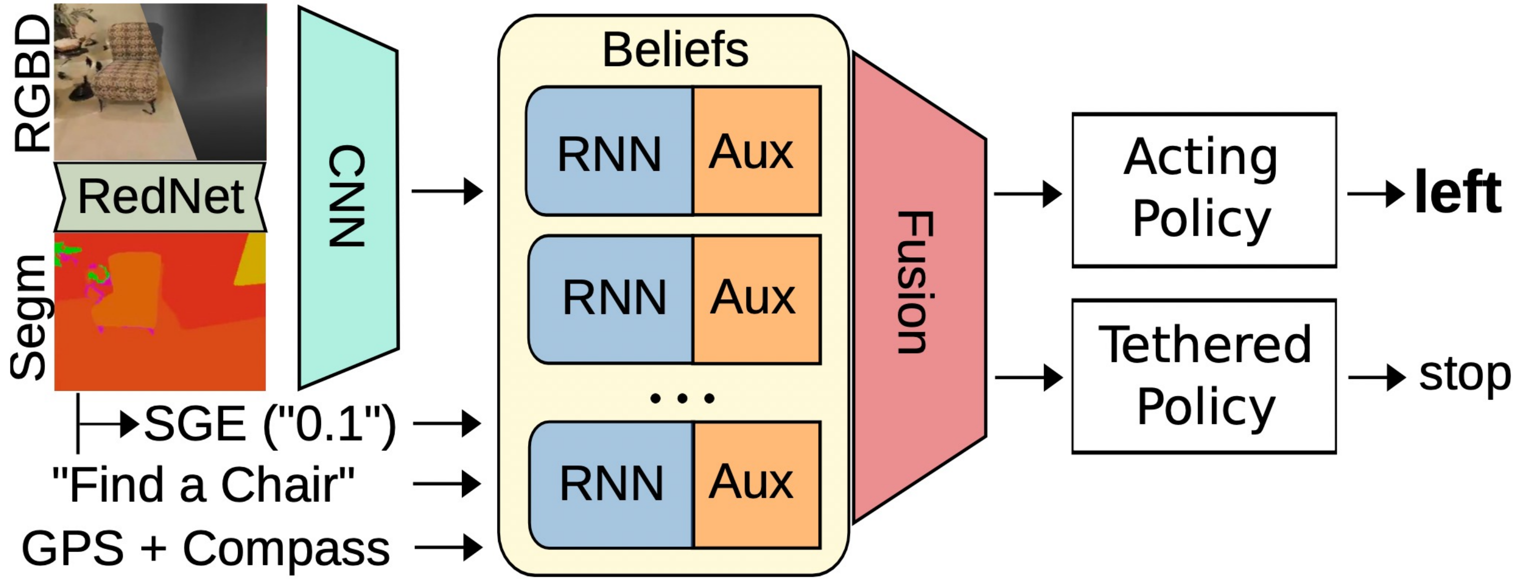
(4) Action Distribution Prediction (ADP)

(5) Generalized Inverse Dynamics (GID)

(6) Coverage Prediction (CP)

ObjectNav in Habitat Challenge 2021

Fuse Auxiliary Tasks for ObjectNav:



SGE : the fraction of the frame occupied by the goal object. (**S**emantic **G**oal **E**xists)

Beliefs : Output cell States from all individual GRUs for Aux Tasks.

Fusion : An attention layer conditioned on the observation embedding.

$$L_{\text{total}}(\theta_m;\theta_a) = L_{\text{RL}}(\theta_m) - \alpha H_{action}(\theta) + L_{\text{Aux}}(\theta_m;\theta_a)$$

H_{action} : entropy across action distribution

$$L_{\text{Aux}}(\theta_m;\theta_a) = \sum_{i=1}^{n_{\text{Aux}}} \beta^i L^i_{\text{Aux}}(\theta_m;\theta^i_a) - \mu H_{attn}(\theta_m)$$

H_{attn} : entropy across attention distribution over aux tasks

Comments & Discussion

Topic 1: Map-Based VS Vision-Based

Comment: Map-Based Methods with higher Interpretability & Expansibility

Topic 2: More Data VS More Elegant Methods

Background:

(1) **PointNav** solved in 2021, achieving 99.6% success rate, with 2.5 billion training frames and a simple RNN network.

(2) Boost performance in **GoalNav**, with auxiliary data(human annotated /synthetic/human demonstrations…)

(3) **CLIP**

Question: How should we regard those who achieve nearly perfect performance if simply with big data?

References

Talks & Websites:

(1) **Devendra Chaplot's Ph.D Thesis Defense** https://www.youtube.com/watch?v=rJ7tGT5cHtU

(2) **Habitat Challenge** https://aihabitat.org/challenge/2022/

Papers:

(1) ***Learning to Explore Using Active SLAM***, Chaplot et al, ICLR2020

(2) ***Occupancy Anticipation for Efficient Exploration and Navigation***, Santhosh et al, ECCV2020

(3) ***Object Goal Navigation using Goal-Oriented Semantic Exploration***, Chaplot et al, NIPS2020

(4) ***PONI: Potential Functions for ObjectGaol Navigation with Interaction-Free Learning***, Santhosh et al, CVPR22

(5) ***Robust Visual Odometry for Realistic Point-Goal Navigation***, Partsey et al

(6) ***Auxiliary Tasks and Exploration Enable ObjectGoal Navigation***, Joel Ye et al, ICCV2021

Acknowledgments

Advisors: ***Jiazhao Zhang, He Wang***

Thank you for your invaluable advices!

Embodied Navigation Tasks

Thanks for Listening!

Any Questions?