Topological Semantic Graph Memory for Image-Goal Navigation

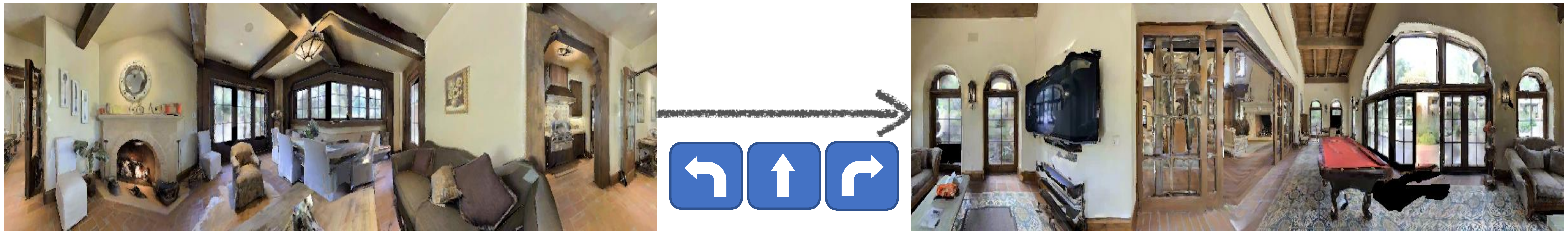
CoRL 2022 (oral)

Nuri Kim, Obin Kwon, Hwiyeon Yoo, Yunho Choi, Jeongho Park, and Songhwai Oh

Department of Electrical and Computer Engineering, Seoul National University

Image Goal Navigation

Source Image **Goal Image**

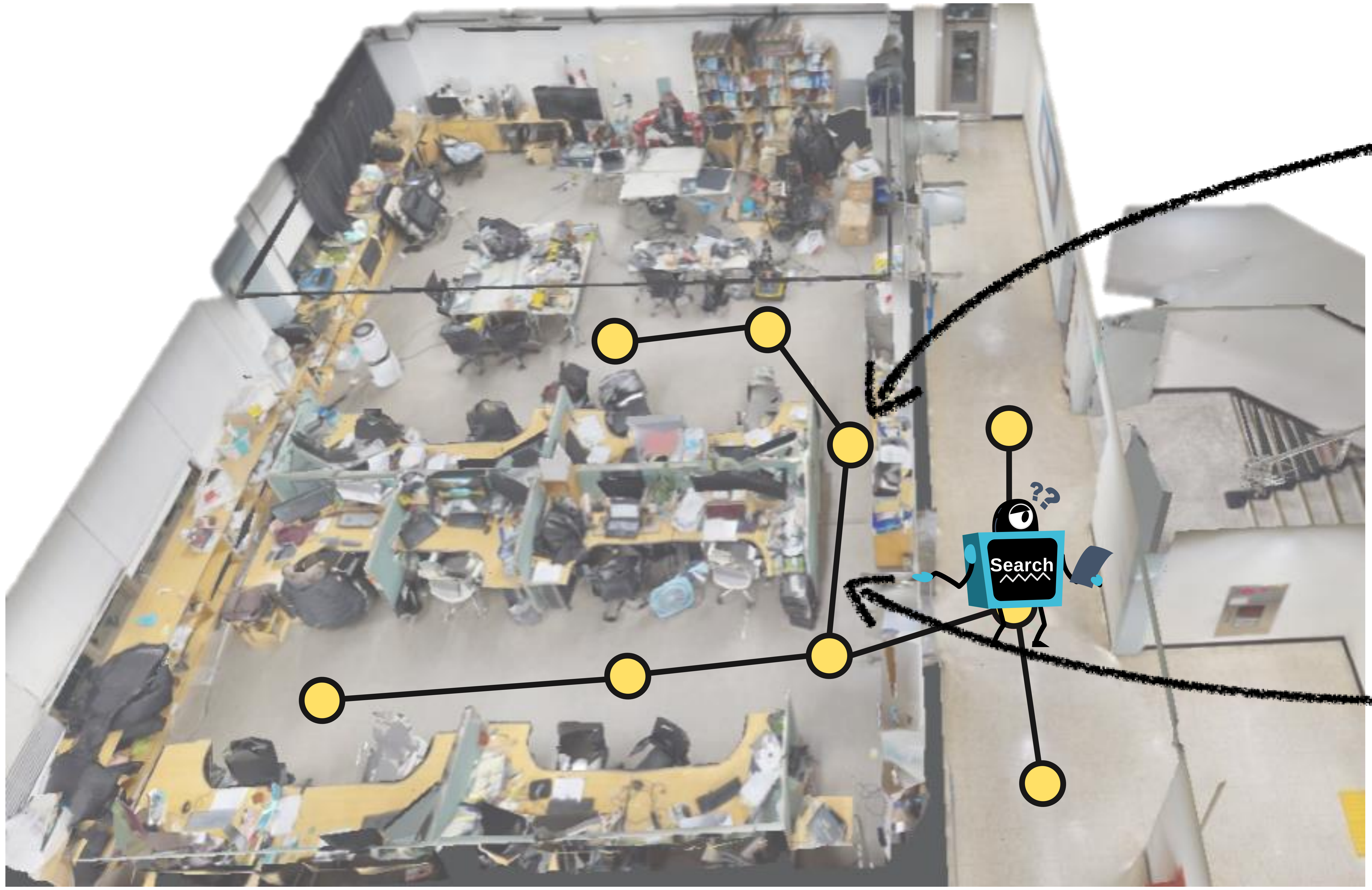


- Agent observations are panoramic images
- Take actions to navigate to the goal location
- Take the **stop** action at the goal location

Slide from Chaplot, et al. "Neural Topological SLAM for Visual Navigation." CVPR 2020

Graph Memory

RLLAB
Robot Learning Laboratory



A vertex represents an area in the environment

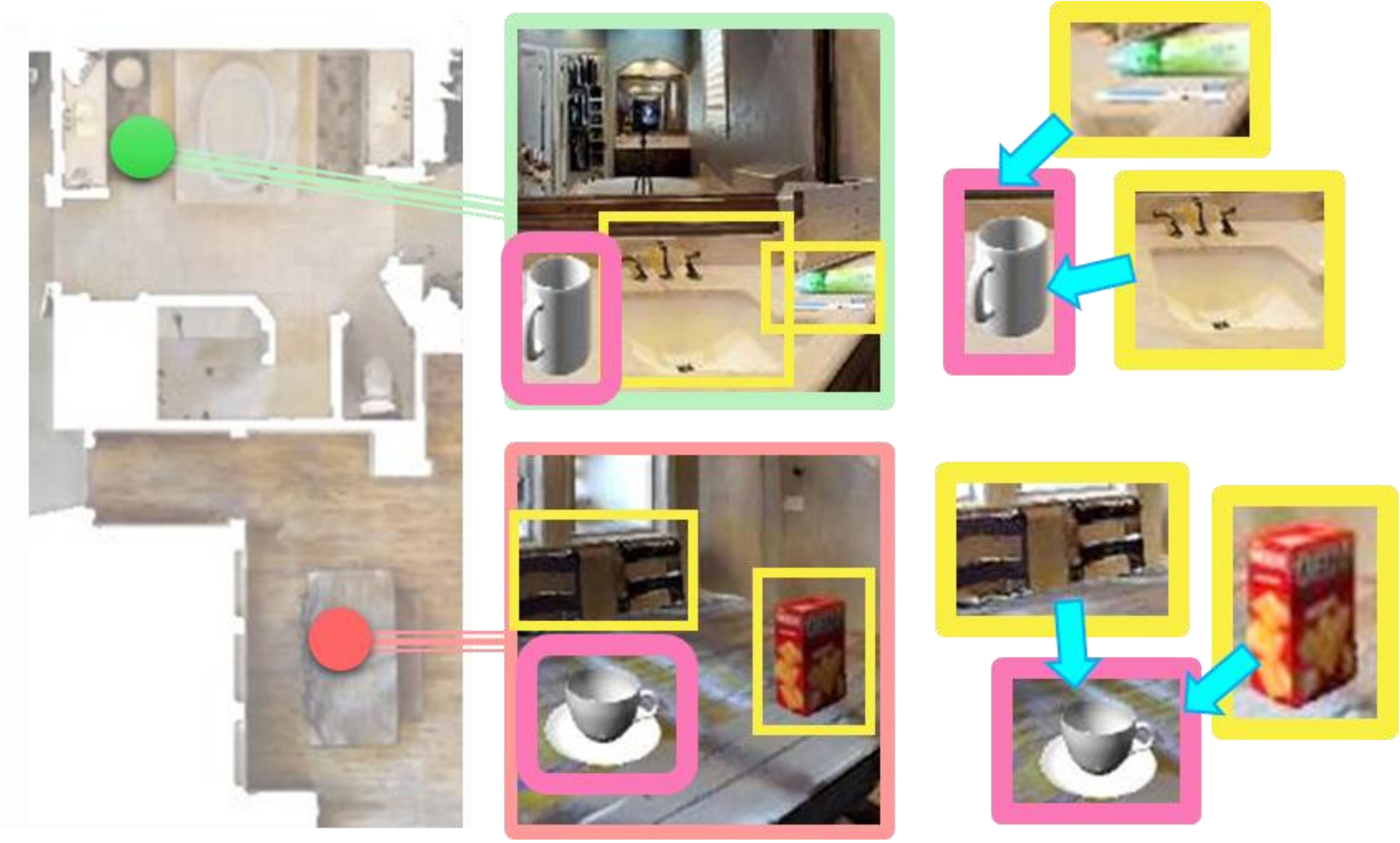
An edge represents the relationship between two vertices, such as reachability and proximity

Semantic Contexts

RLLAB Robot Learning Laboratory

Why landmark knowledges are better integrated into graph memory?

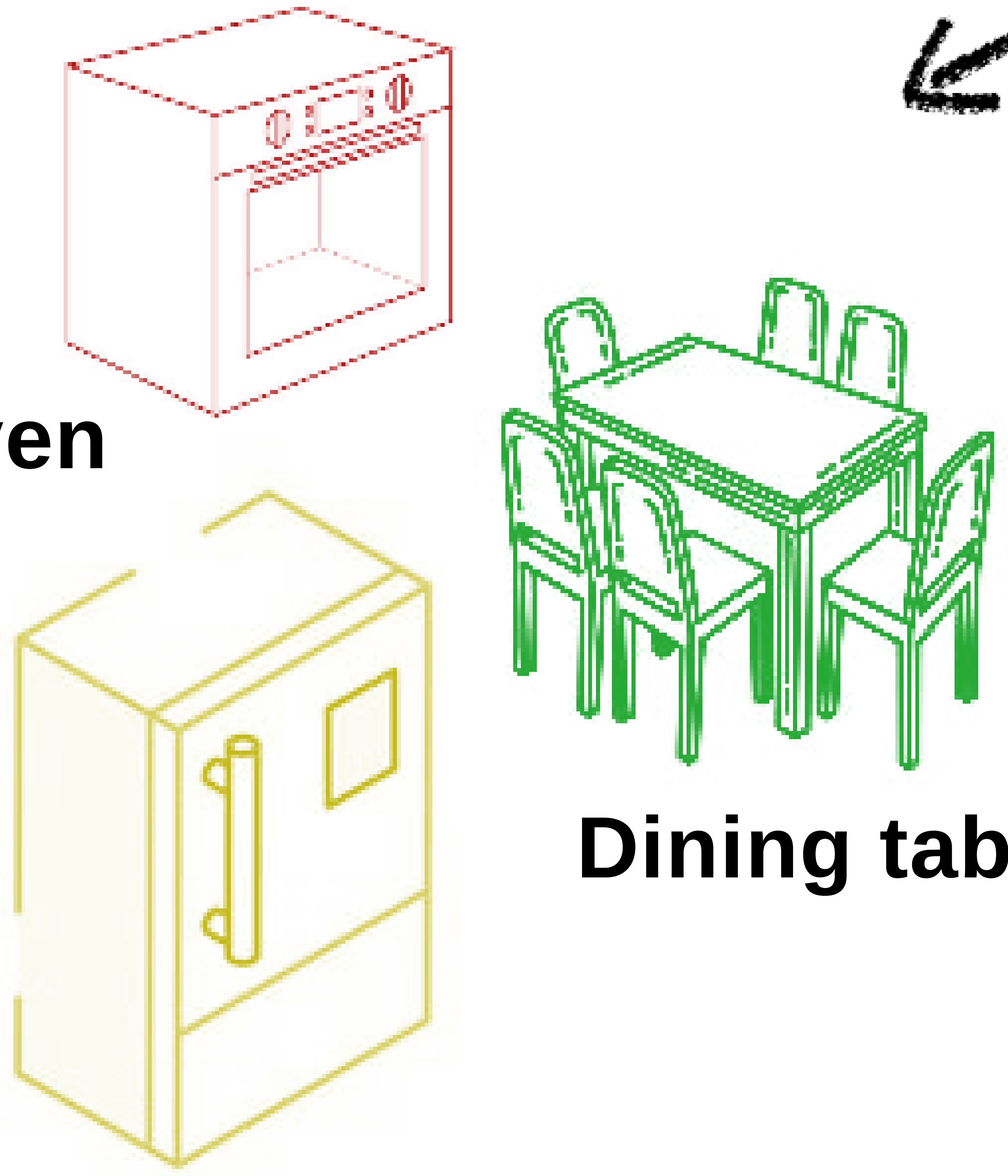
Semantic Contexts: Object Context



Neighboring objects make an object unique

Semantic Contexts: Place-Object Context

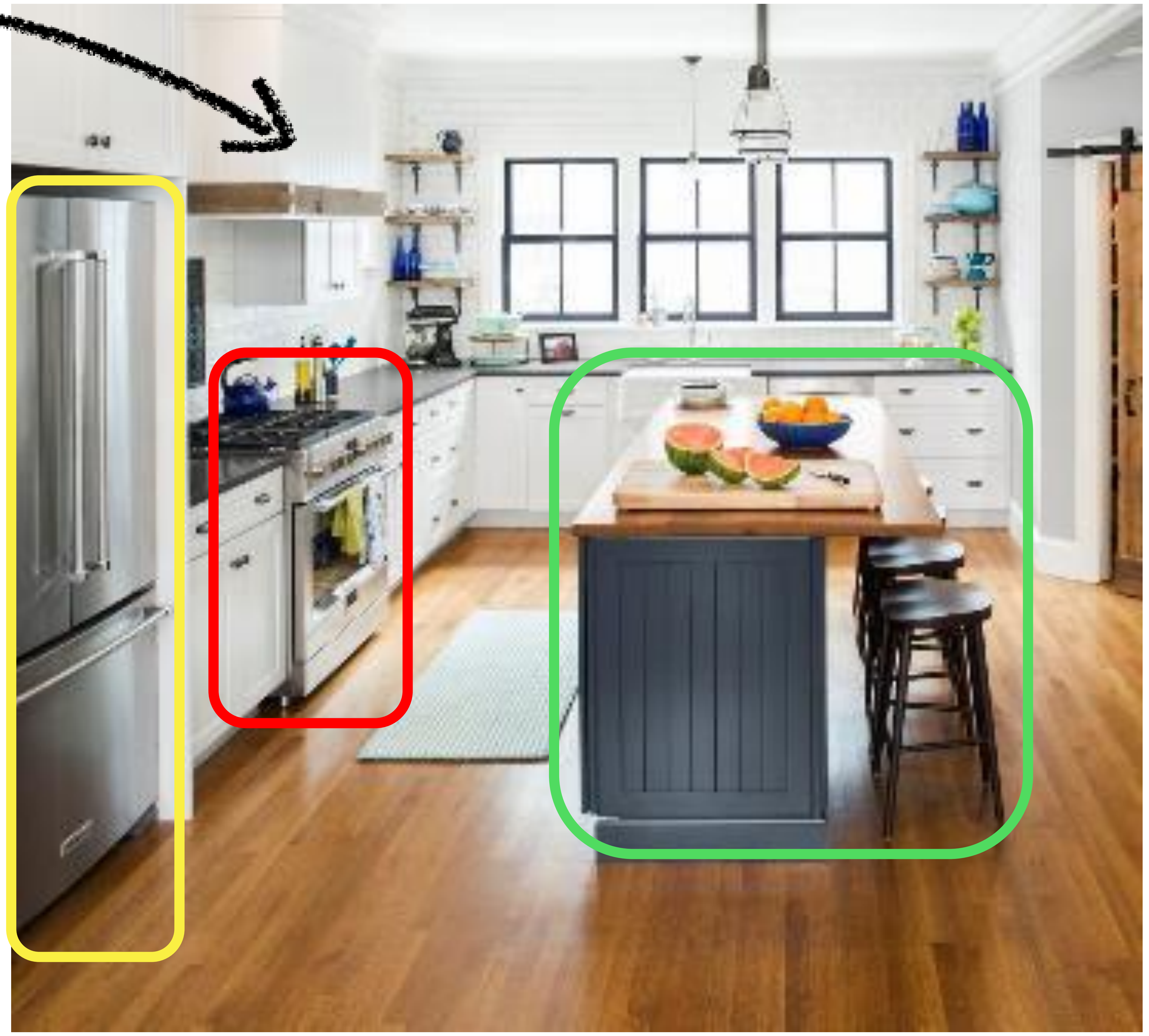
Place and Objects are highly related



Oven

Dining table

Refrigerator



Kitchen

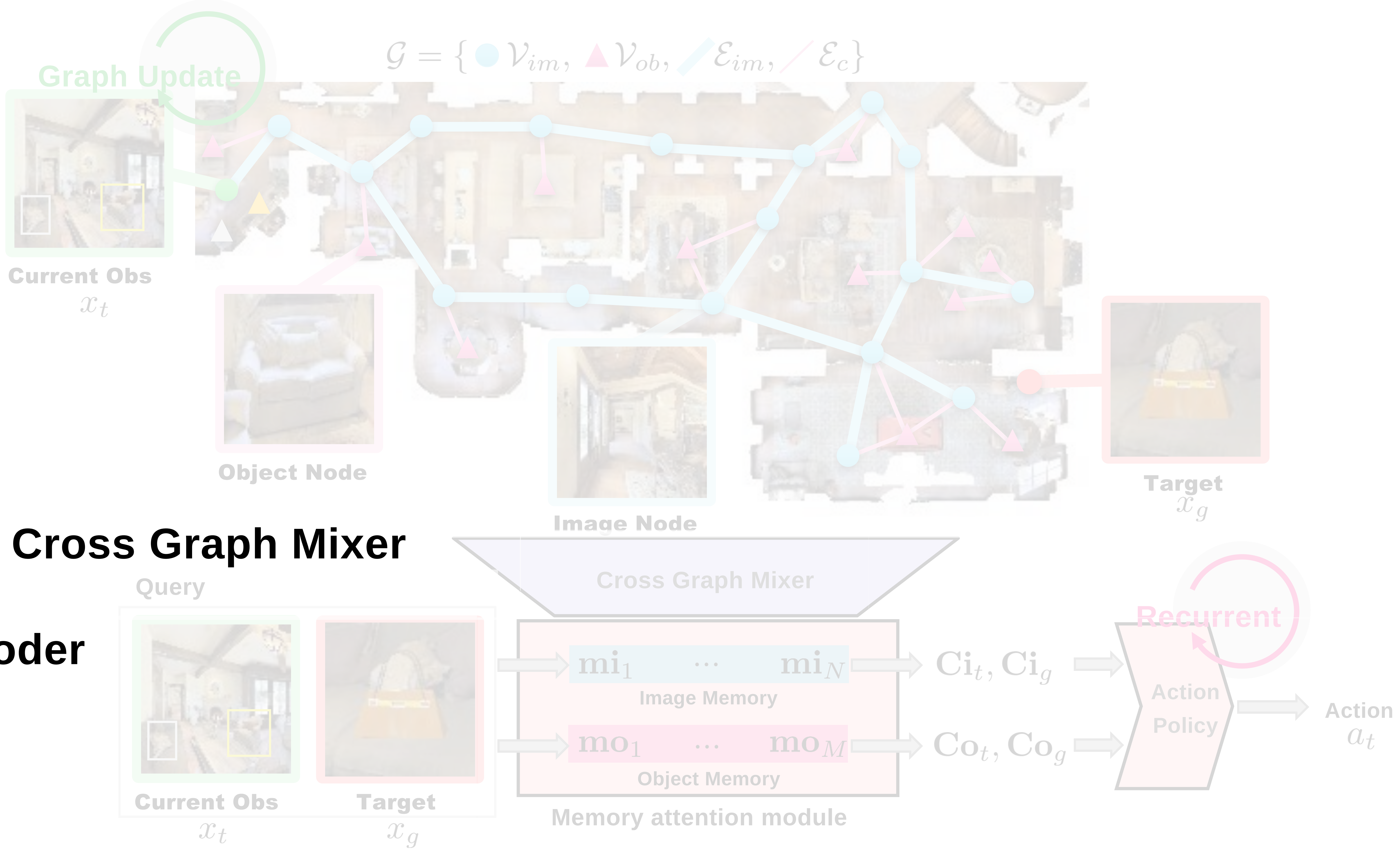
How to embed landmark knowledge into topological graph memory?

Topological Semantic Graph Memory

Graph Builder

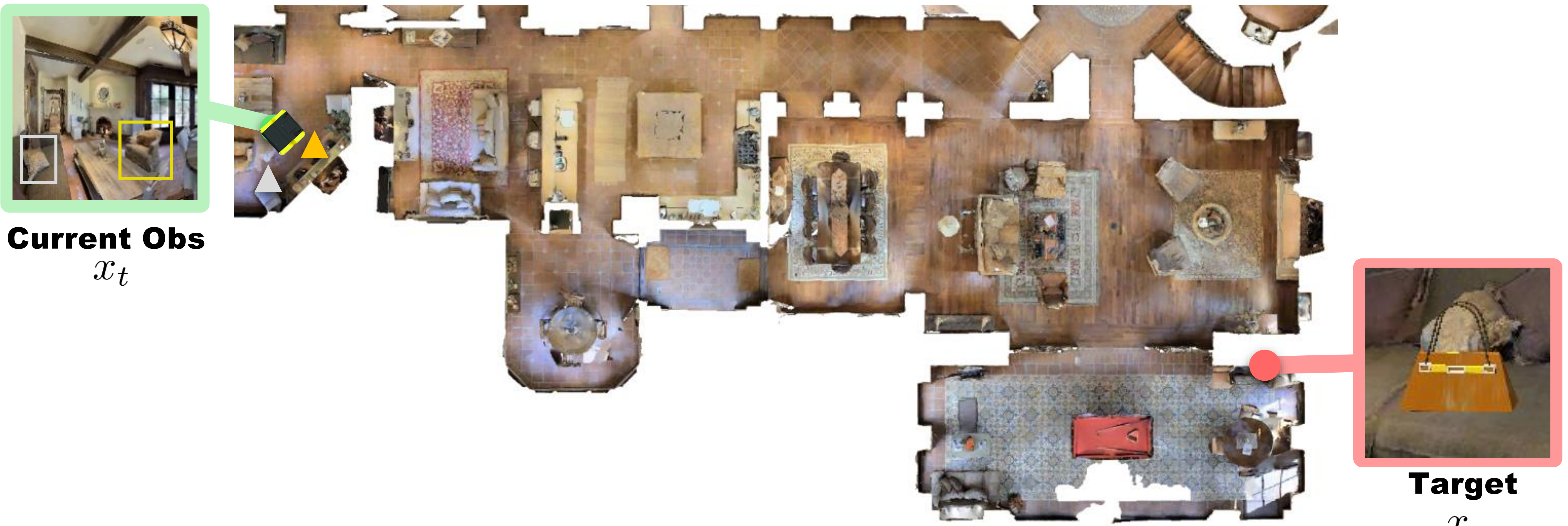
Cross Graph Mixer

Memory Decoder



Graph Builder

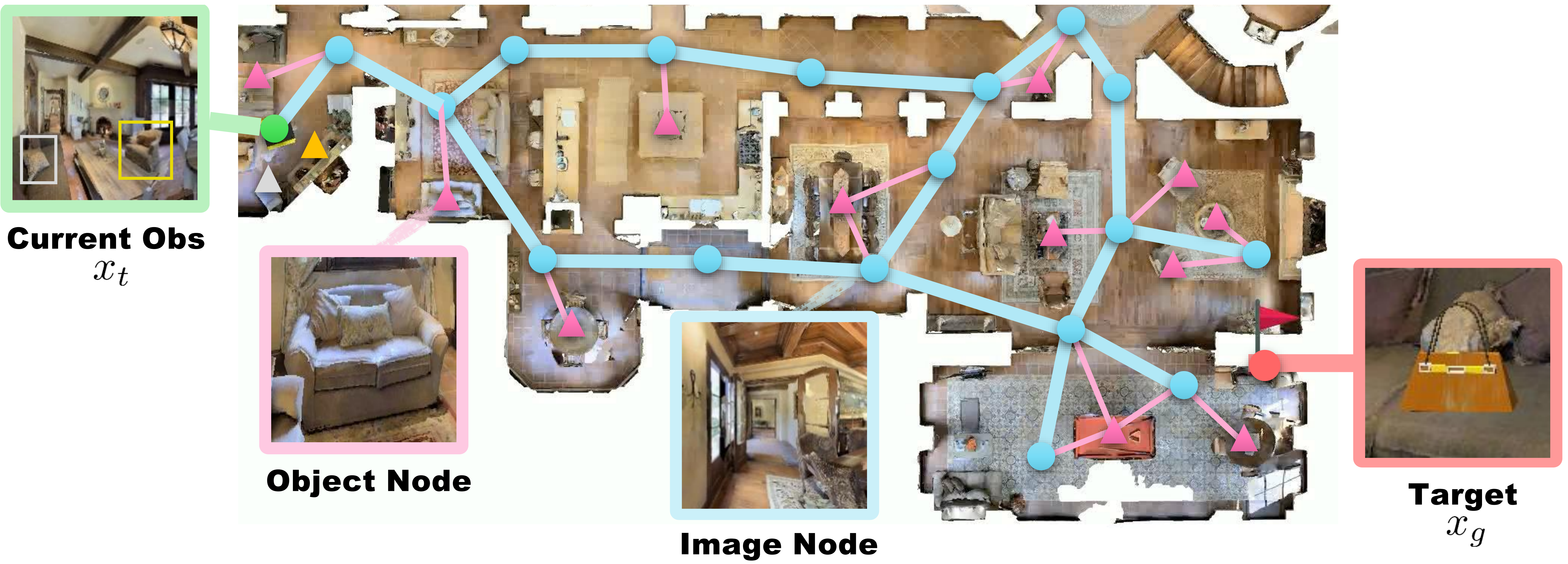
$$\mathcal{G} = \{ \mathcal{V}_{im}, \mathcal{V}_{ob}, \mathcal{E}_{im}, \mathcal{E}_c \}$$



✱ Note that floorplan and node positions are only used for illustration and not given as input to agent

Graph Builder

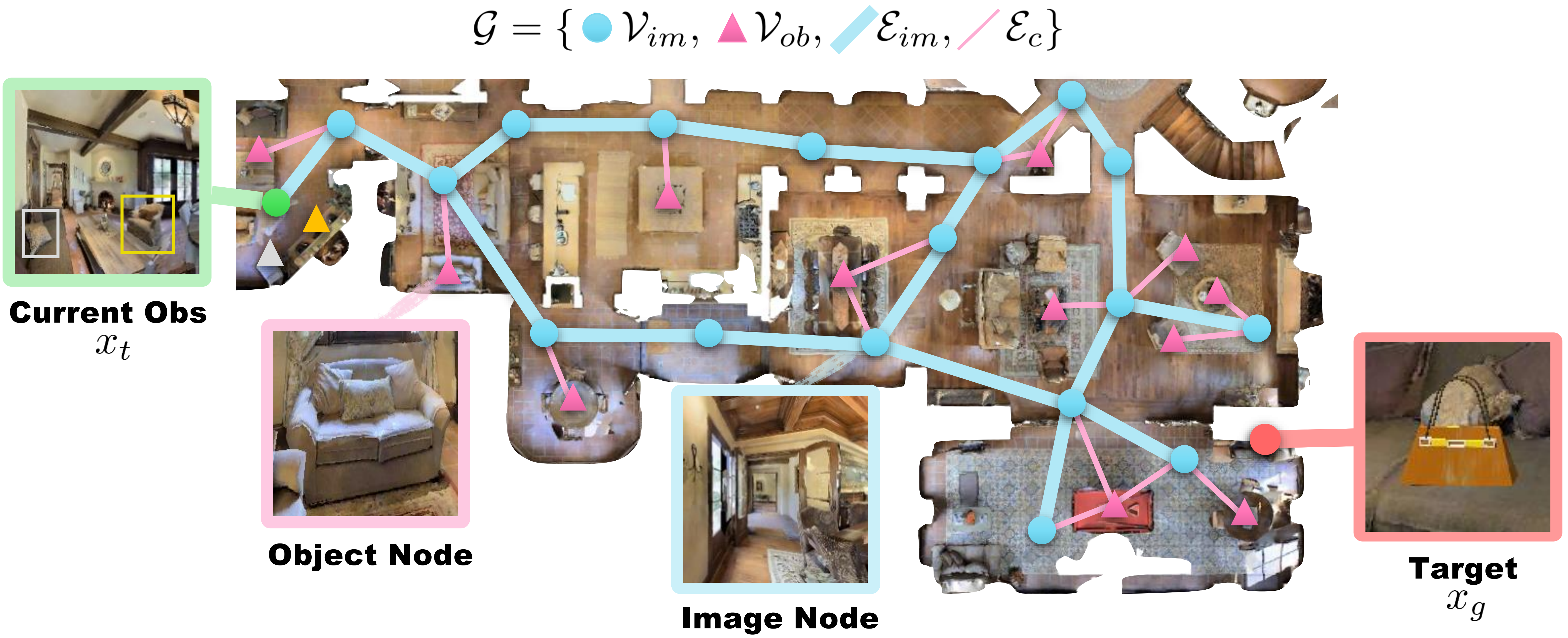
RLLAB
Robot Learning Laboratory



✱ Note that floorplan and node positions are only used for illustration and not given as input to agent

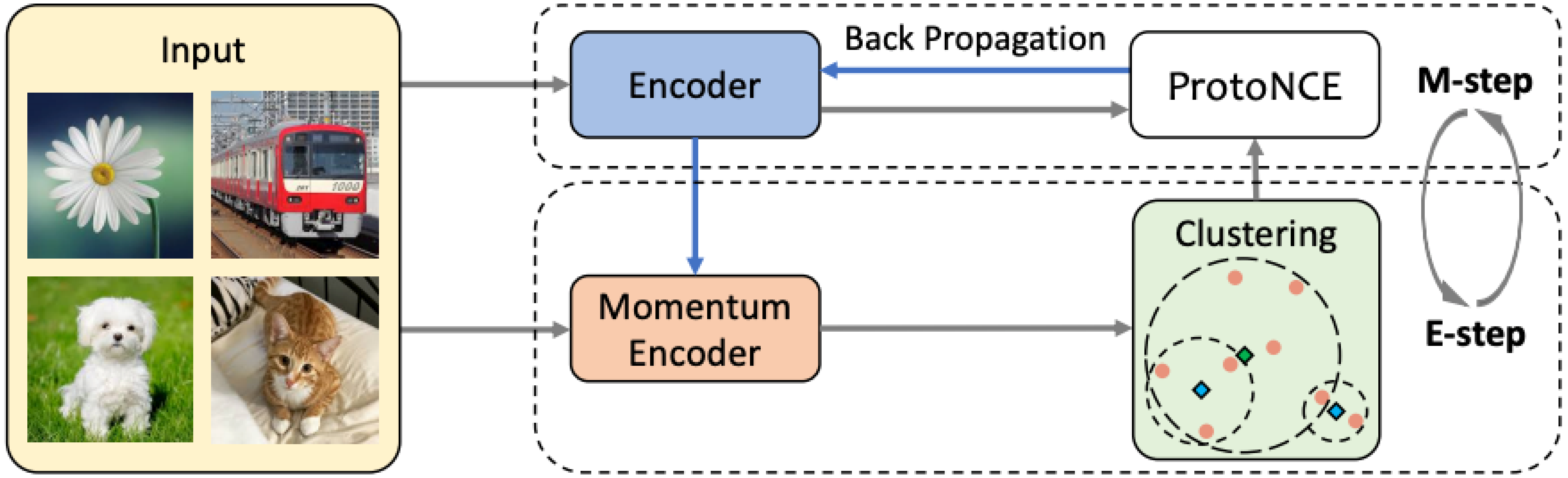
Graph Builder: Image Graph

RLLAB Robot Learning Laboratory

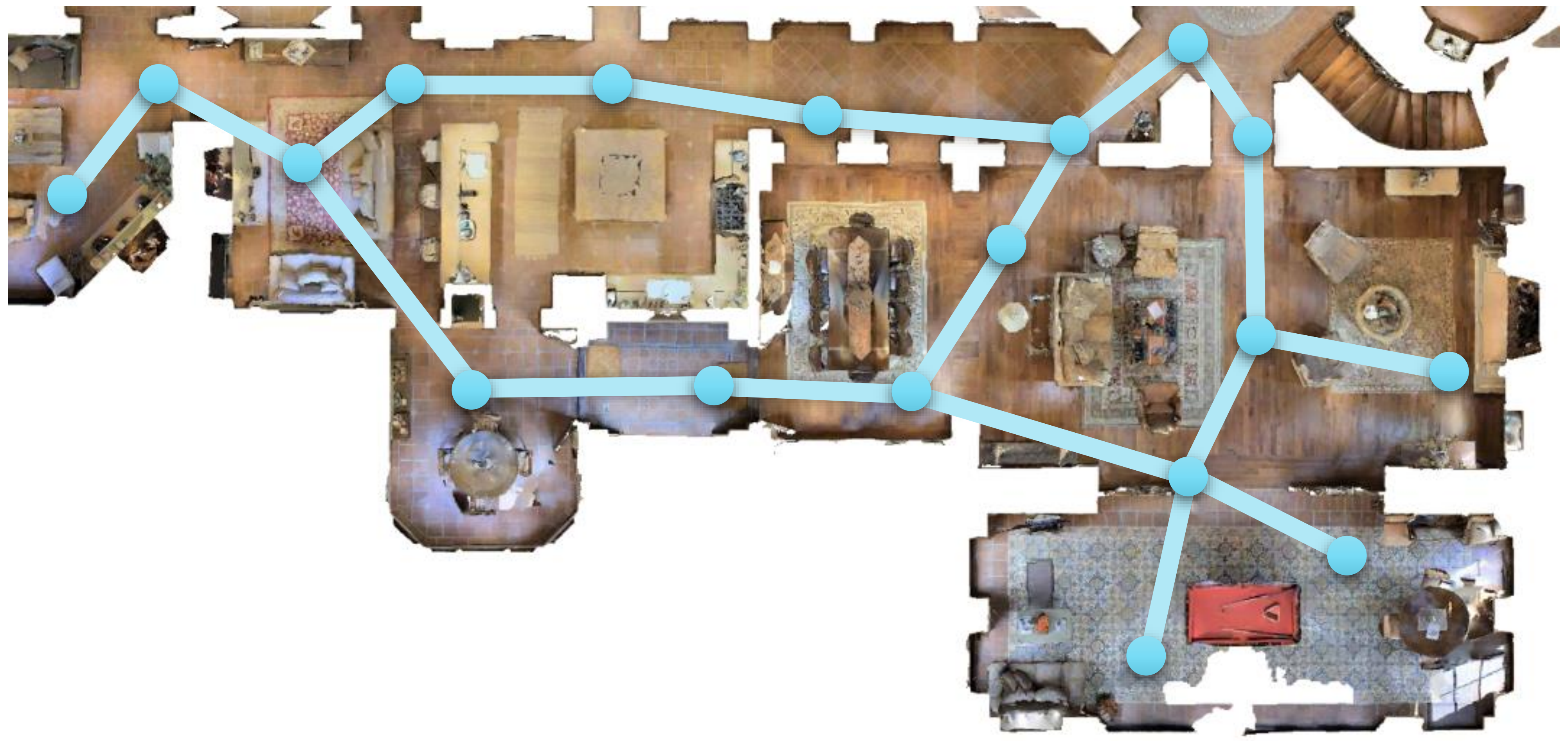


✱ Note that floorplan and node positions are only used for illustration and not given as input to agent

Graph Builder: Image Graph



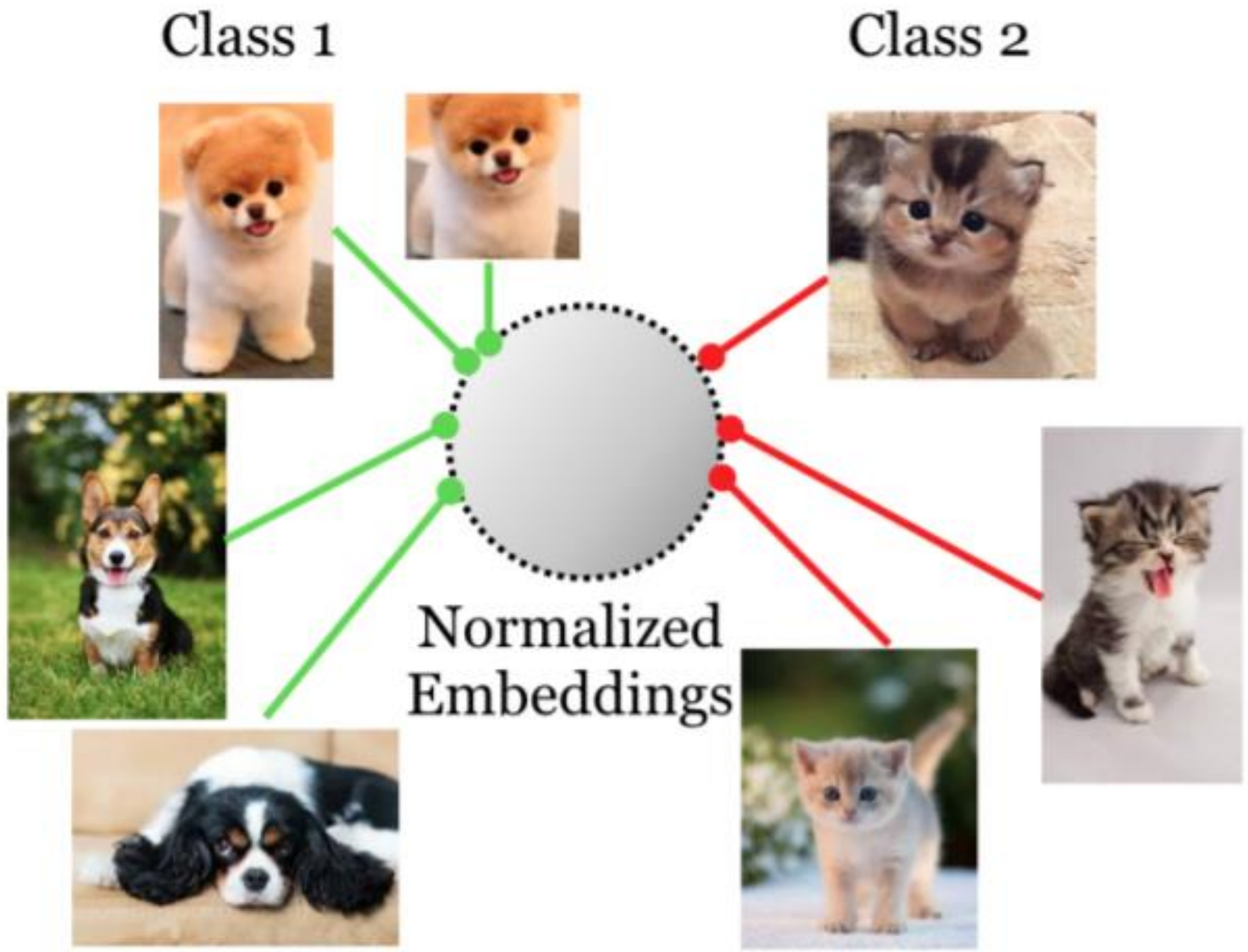
Prototypical Contrastive Learning [1]



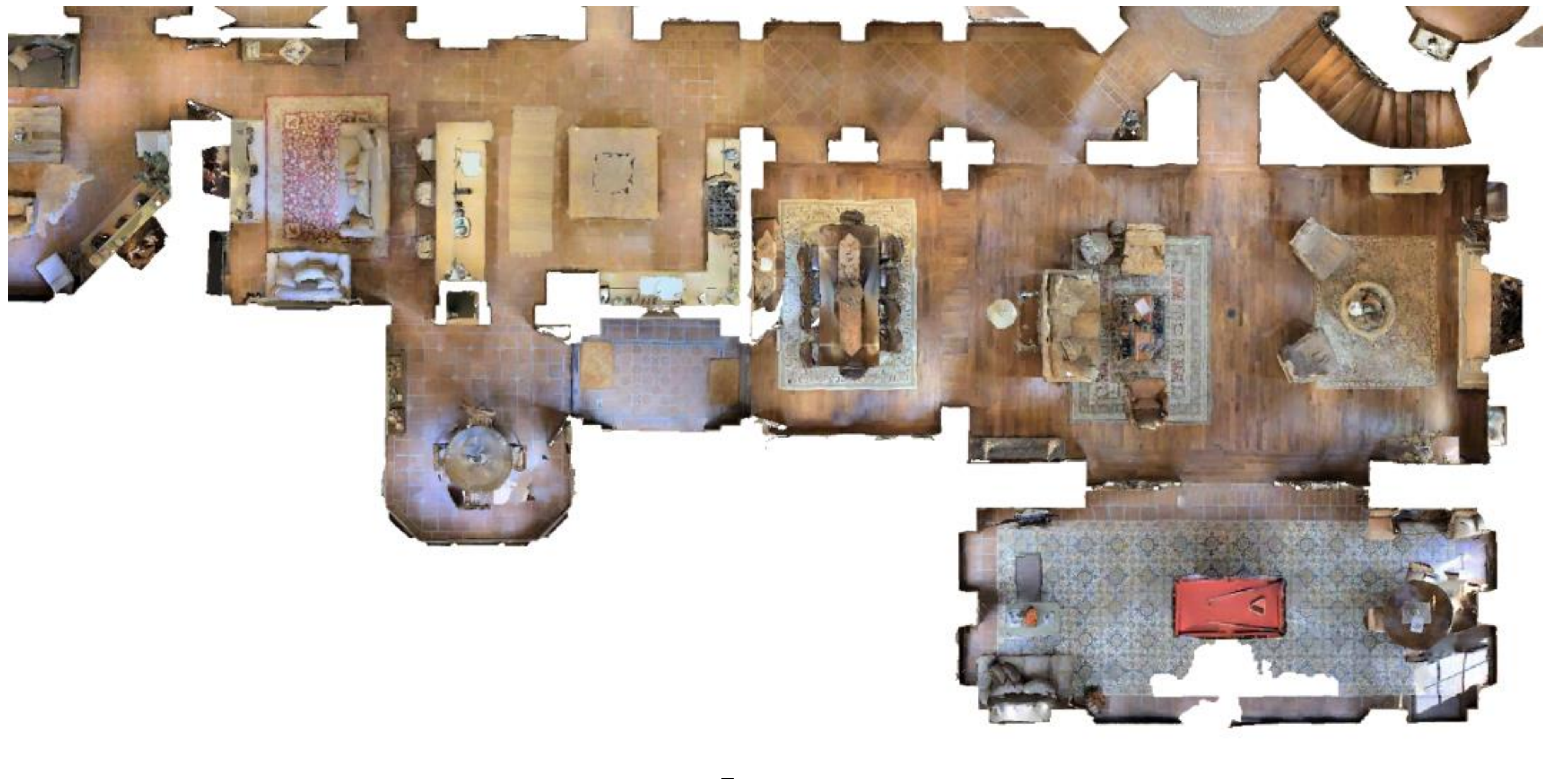
VGM [2]

[1] Li, Junnan, et al. "Prototypical contrastive learning of unsupervised representations." ICLR 2021
[2] Obin Kwon, et al. "Visual graph memory with unsupervised representation for visual navigation." ICCV 2021.

Graph Builder: Object Graph

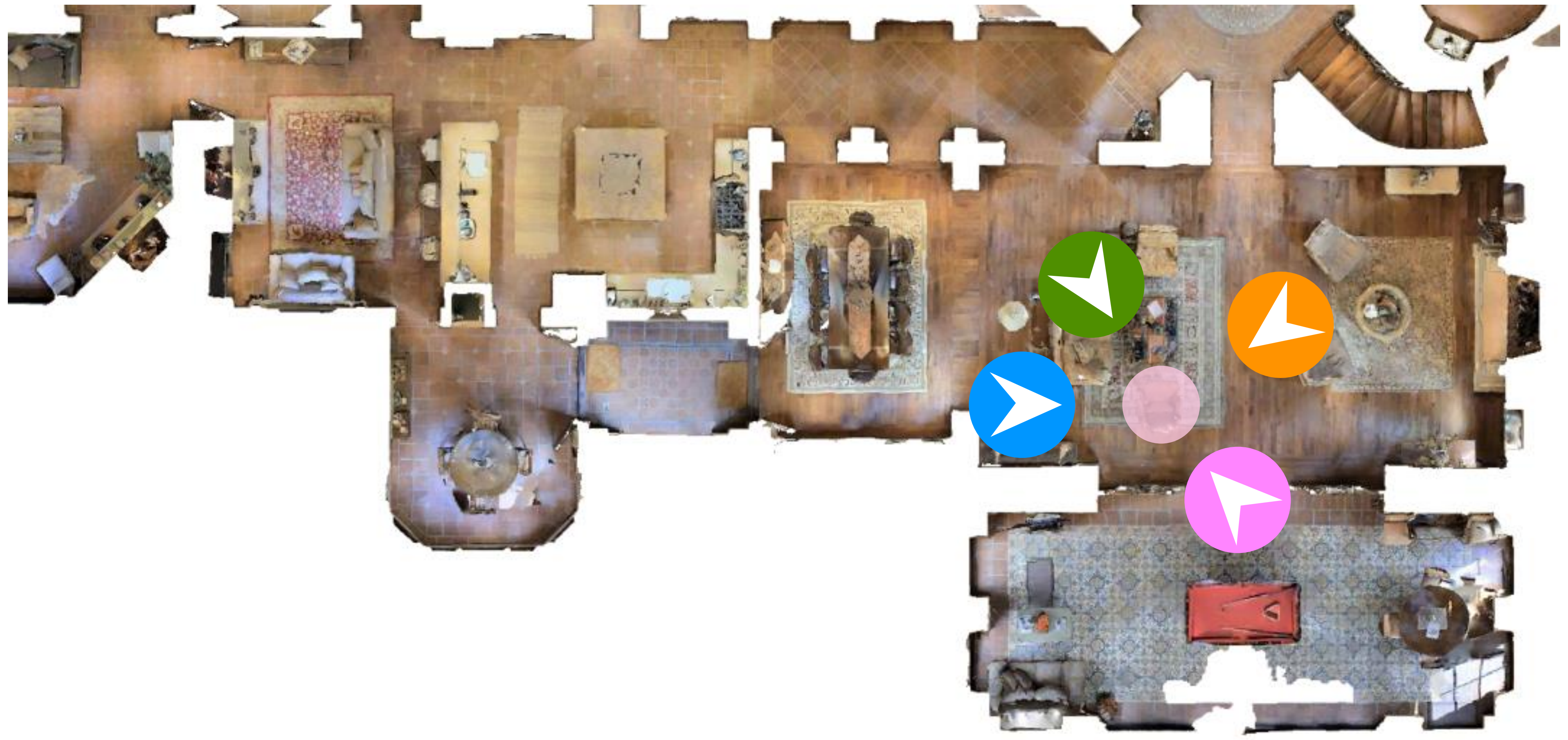


Contrastive Learning



Graph Builder: Object Graph

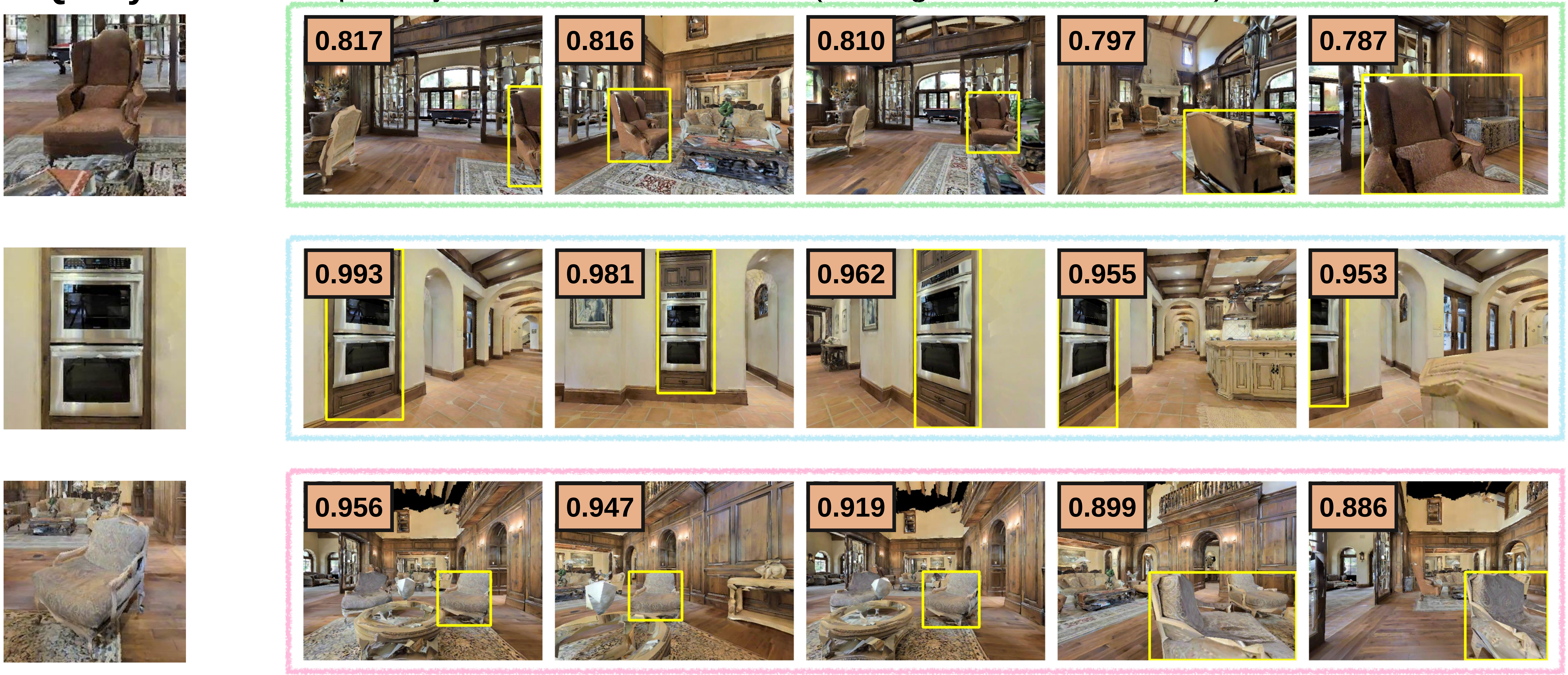
Collect an object from different viewpoints



Graph Builder: Object Graph

RLLAB Robot Learning Laboratory

Query Top 5 objects in the environment (among ~7000 candidates)



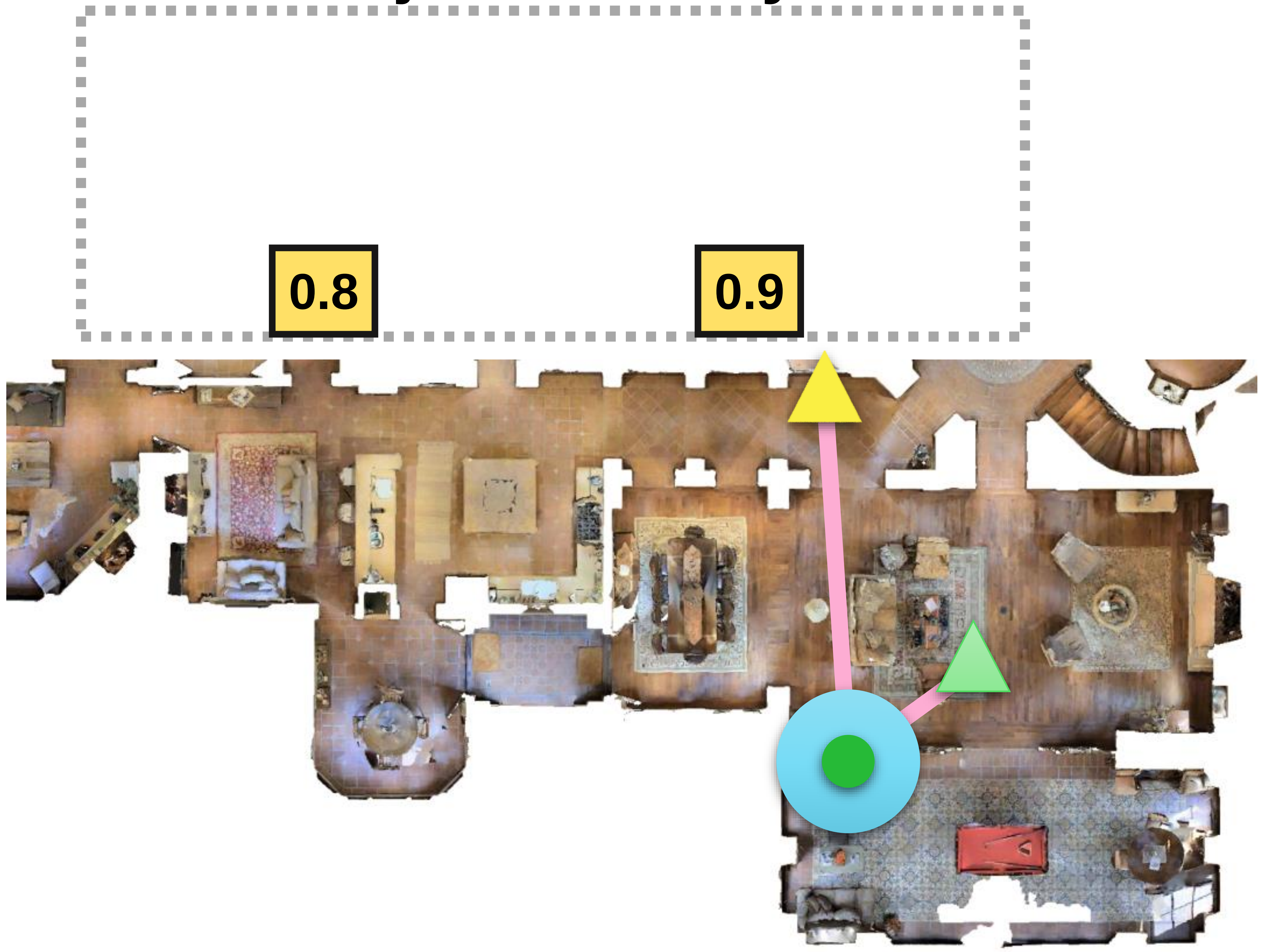
The object encoder finds the query objects in various viewpoints

Graph Builder: Object Graph

RLLAB Robot Learning Laboratory

Observation

Object Memory



Object Nodes: Individual objects

Detected objects are connected to the current node

⁎ Color represents the 3-dim tsne feature of the place

- **Image Nodes**
- **Agent's Current Image Node**
- **Object Nodes**

Graph Builder: Object Graph

RLLAB
Robot Learning Laboratory

*Similarity is **high** and the category is the same.*
*It indicates that the object is **already in the memory.***
*Since **detection score is higher** than the memory node,*
It is used to update the memory node.
The node is connected to the lastly localized image node.

Object Memory

0.7

Observation

Similarity with memory is low.
It is added to a memory as a new node
and connected to the lastly localized image node.

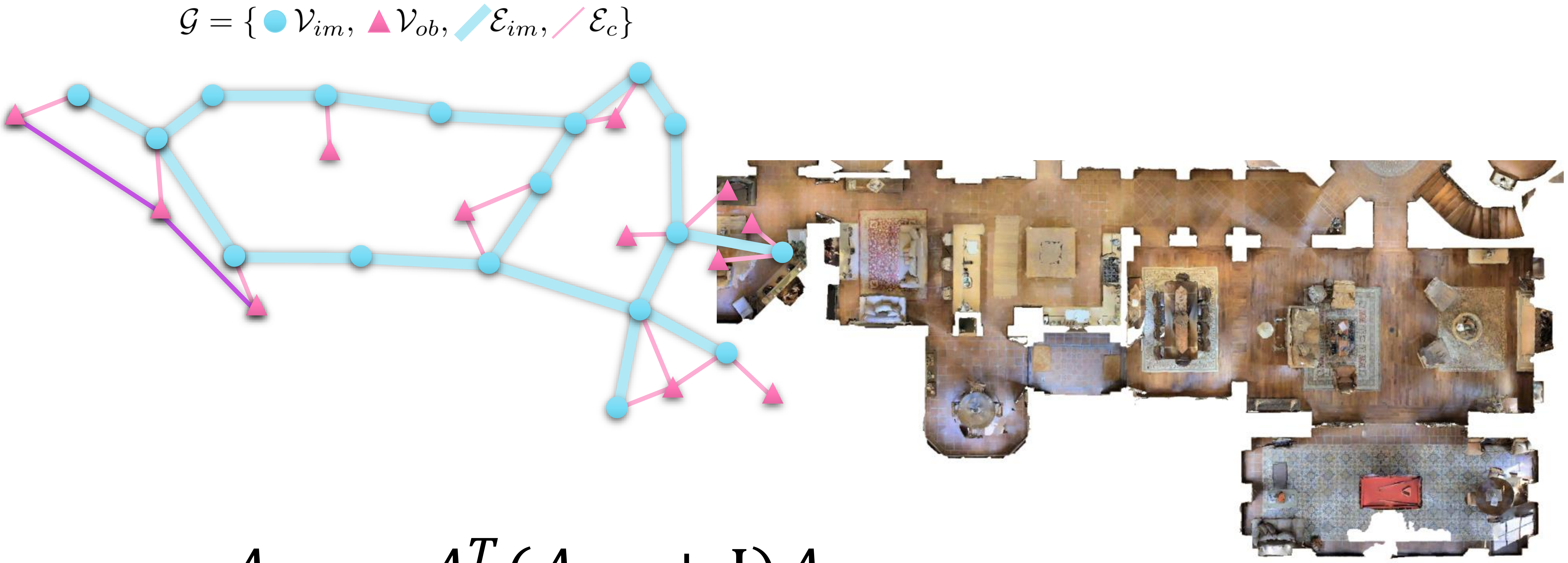
* Color represents the 3-dim tsne feature of the place

Image Nodes

Agent's Current Image Node

Object Nodes

Graph Builder: Object Graph



$$A_{ob} = A_c^T (A_{im} + \mathrm{I}) A_c$$

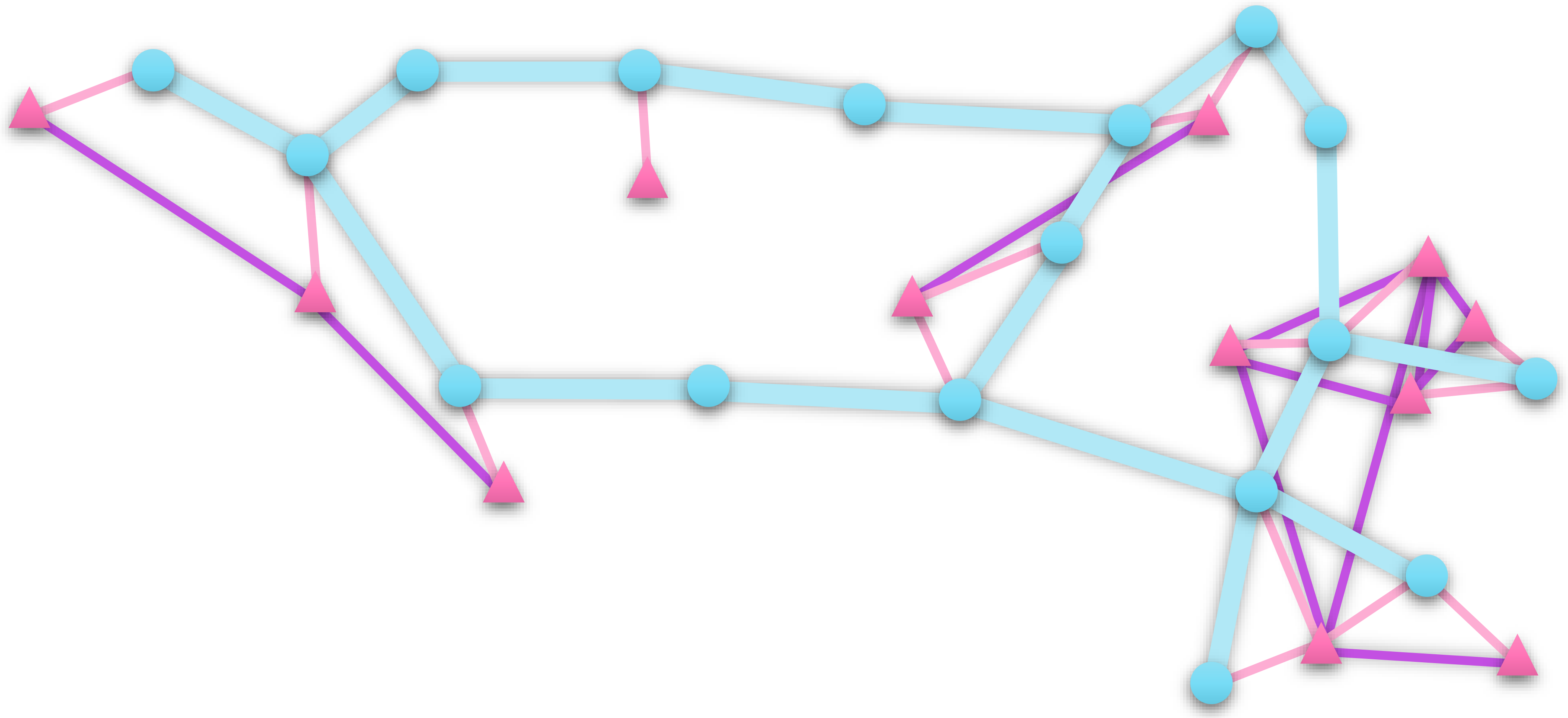
A_{im}: *image affinity matrix*

A_{ob}: *object affinity matrix*

A_c: *image–object affinity matrix*

Graph Builder: Object Graph

$$\mathcal{G} = \{ \mathcal{V}_{im}, \mathcal{V}_{ob}, \mathcal{E}_{im}, \mathcal{E}_{c} \}$$



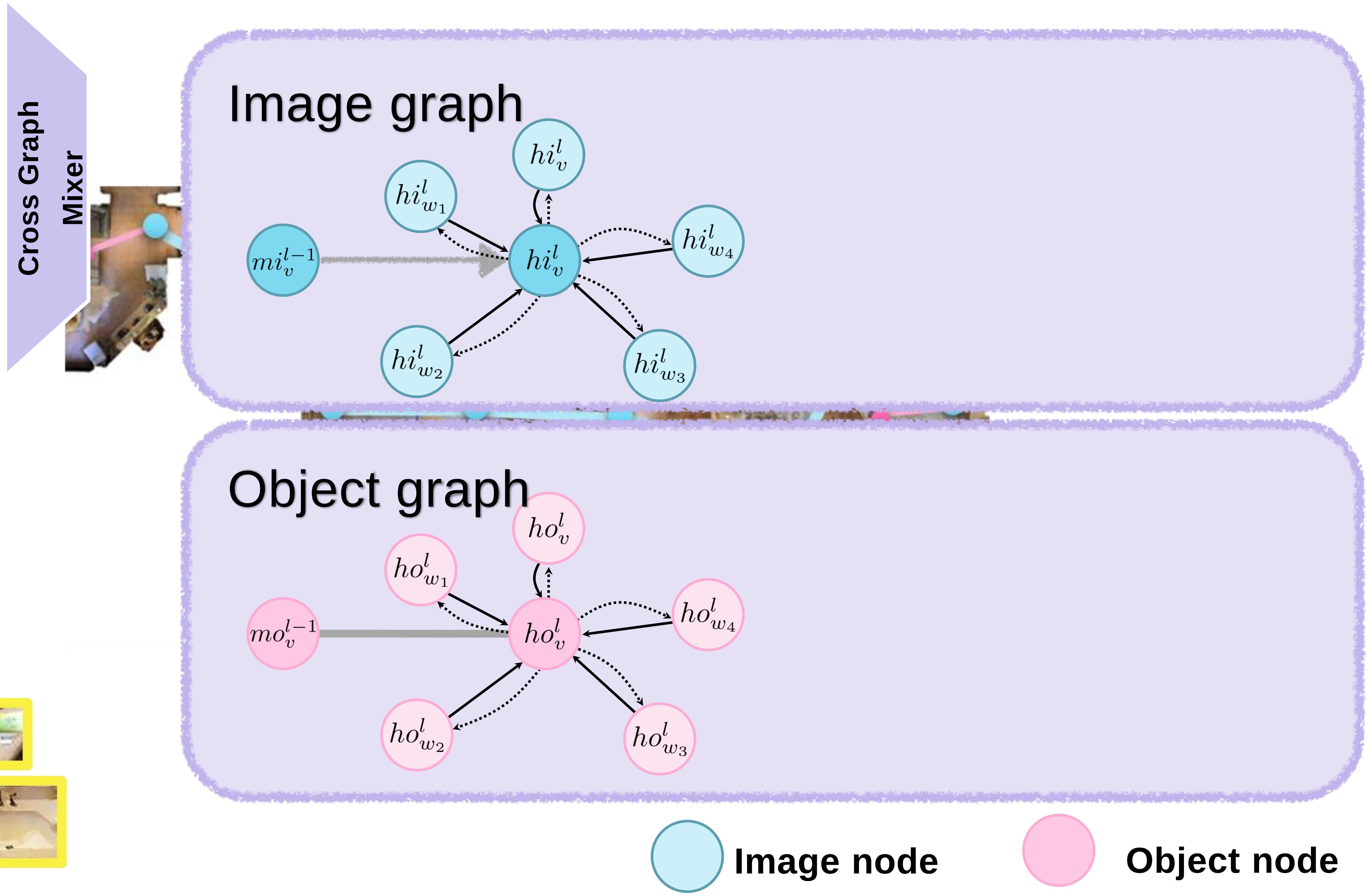
$$A_{ob} = A_c^T (A_{im} + \mathrm{I}) A_c$$

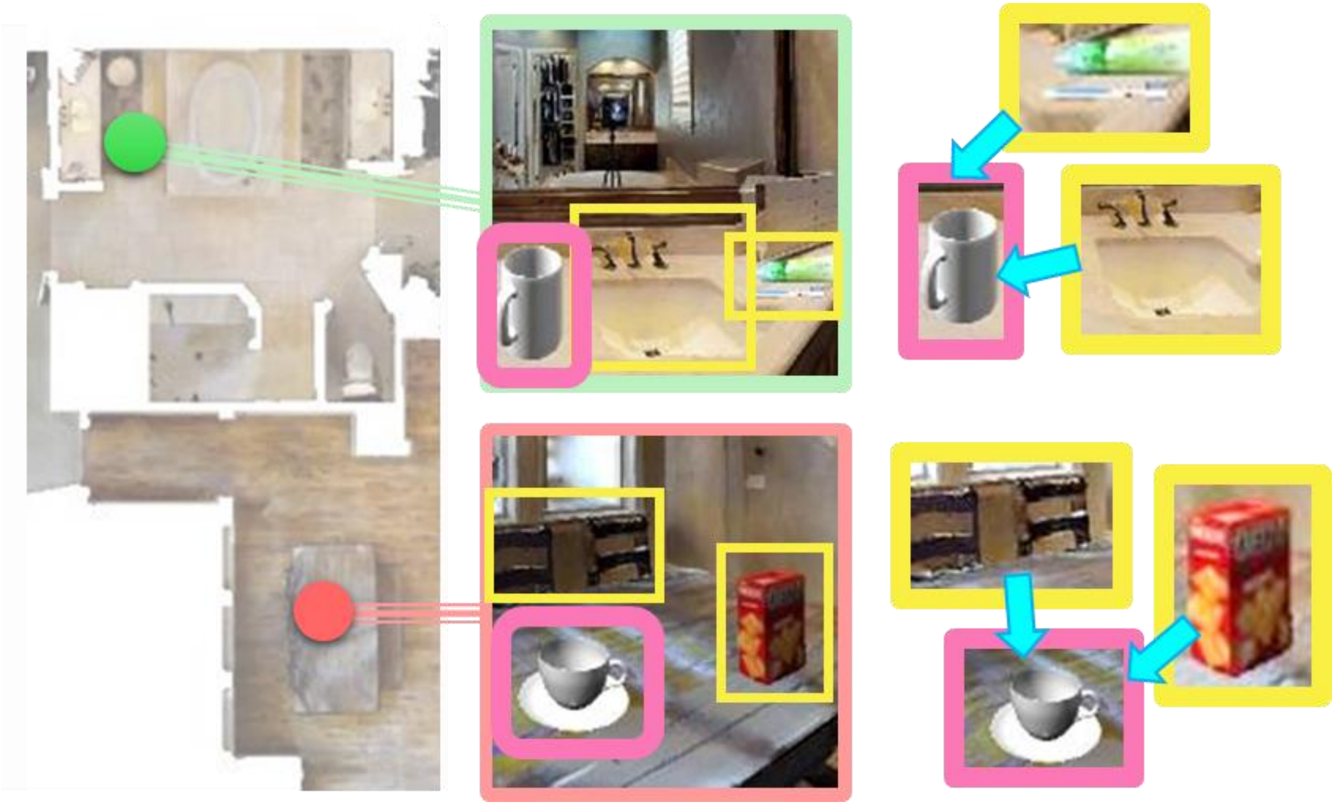
A_{im}: *image affinity matrix*

A_{ob}: *object affinity matrix*

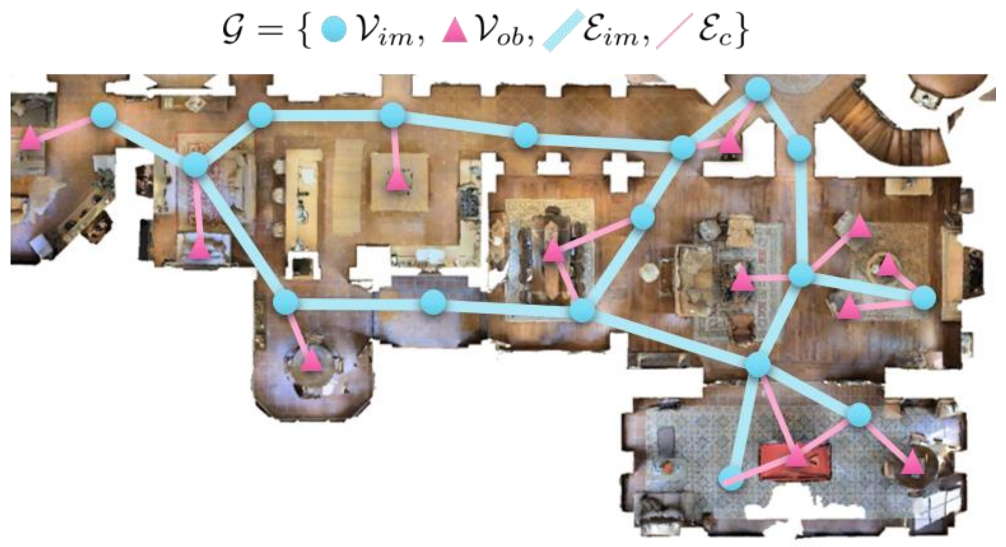
A_c: *image−object affinity matrix*

Cross Graph Mixer

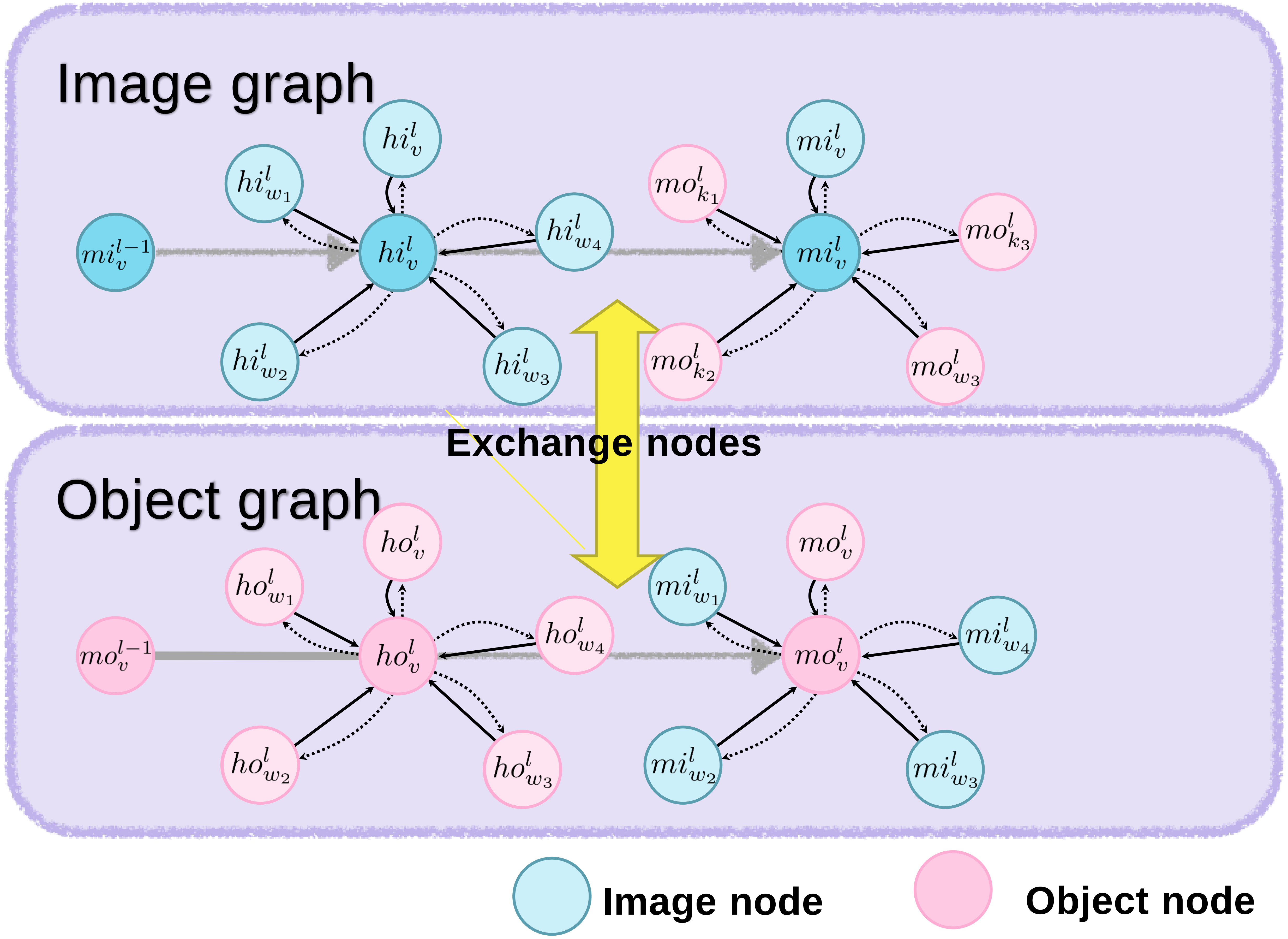


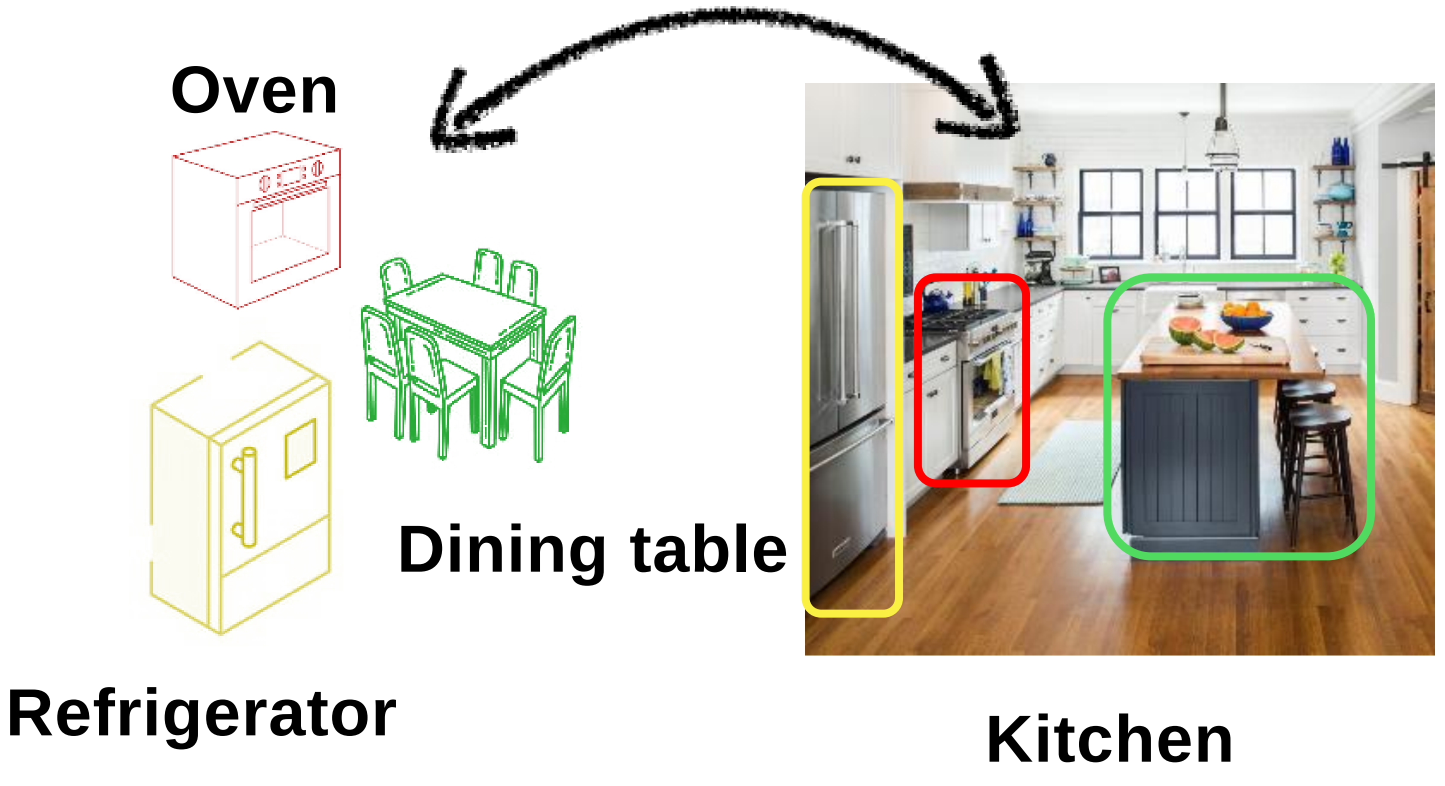


Cross Graph Mixer

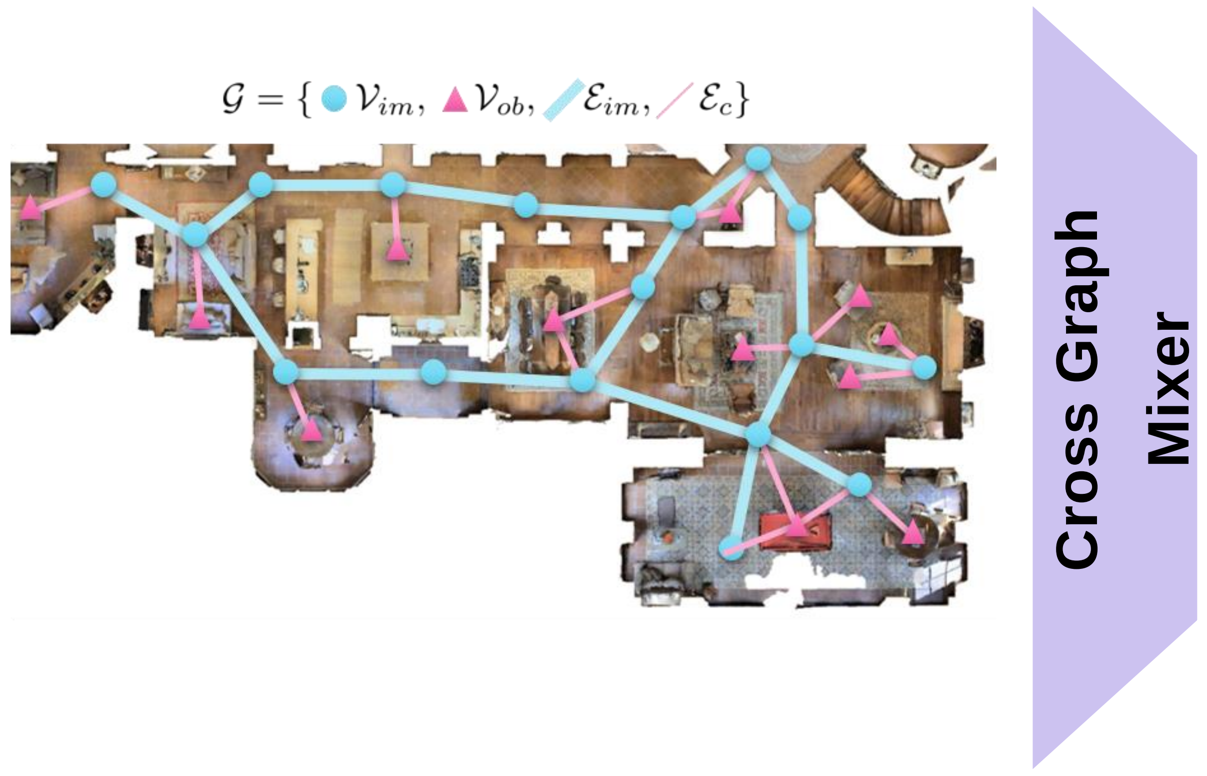


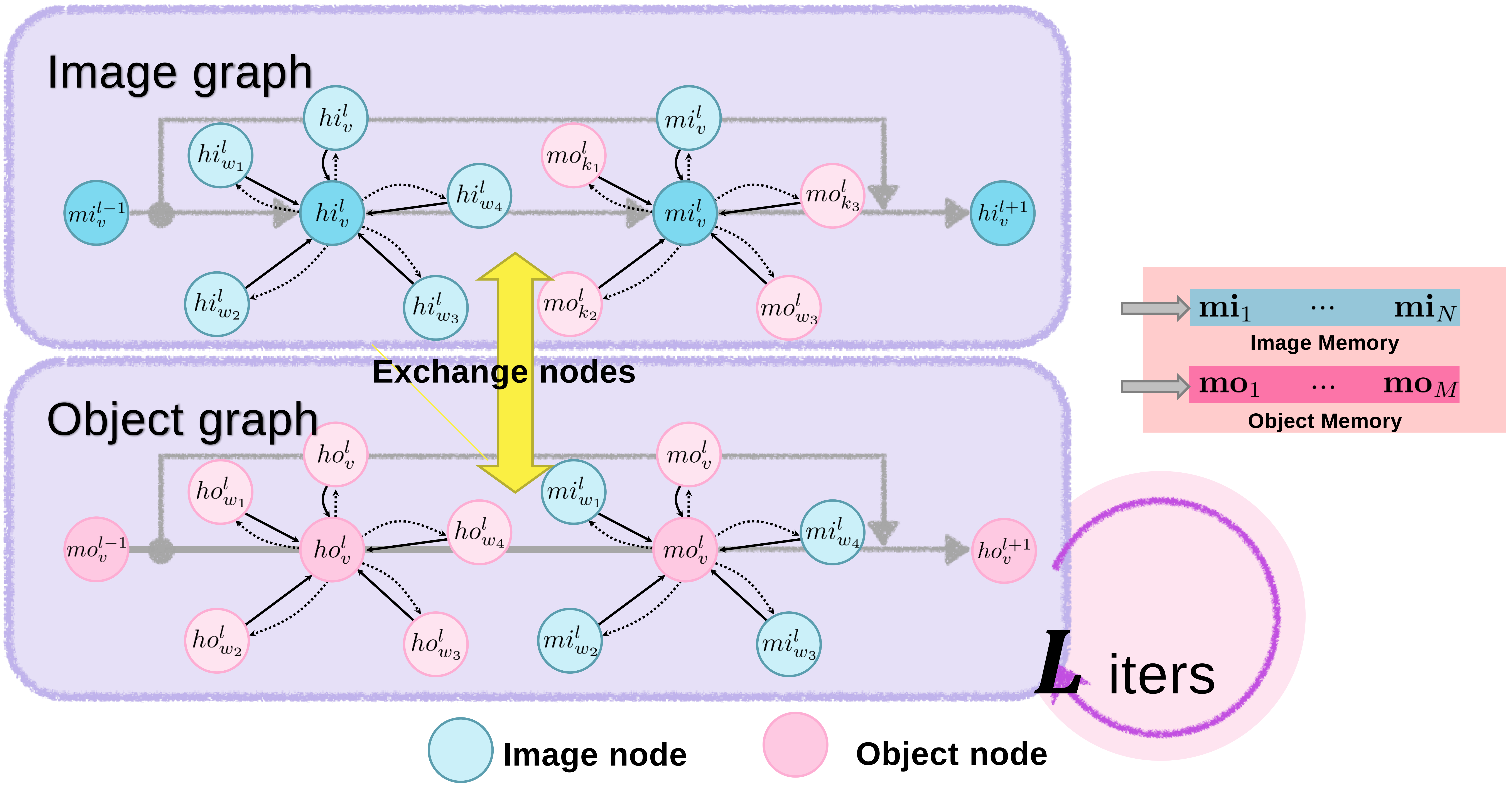
Cross Graph Mixer



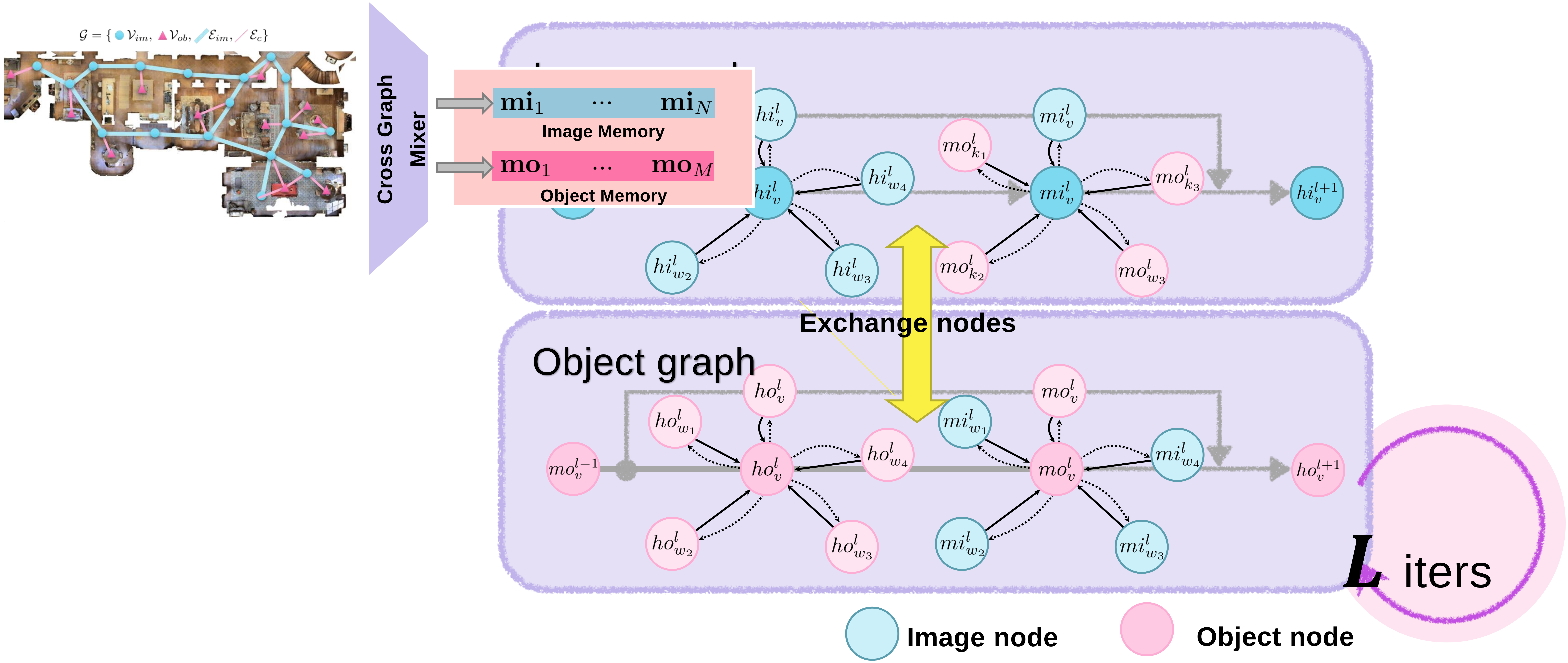


Cross Graph Mixer

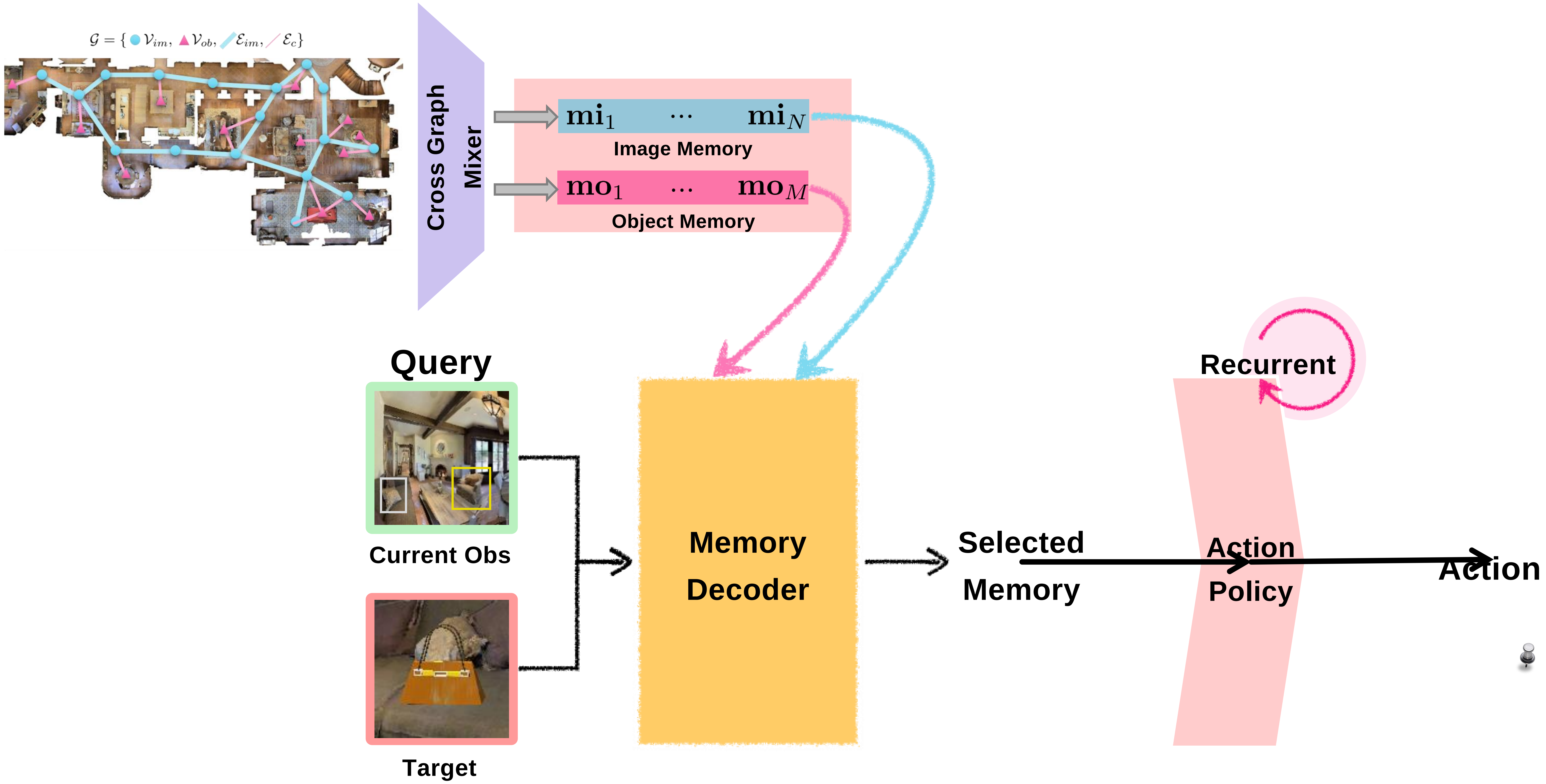




Cross Graph Mixer

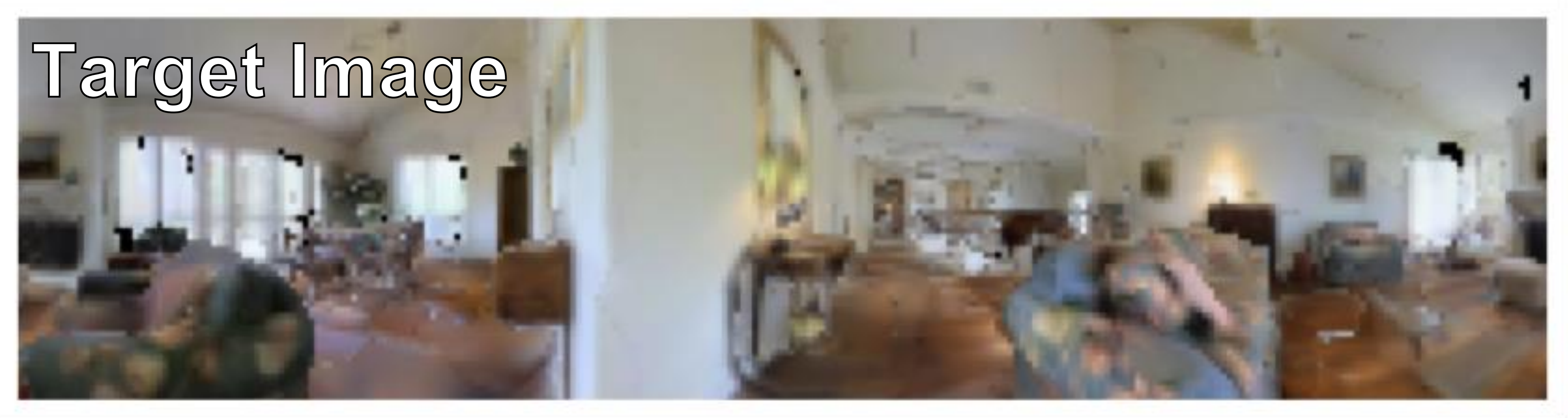


Memory Decoder



- **PPO** algorithm

Demo Video



RLLAB
Robot Learning Laboratory

TSGM (Ours)

VGM [2]

[2] Obin Kwon, et al. "Visual graph memory with unsupervised representation for visual navigation." ICCV 2021.

Real-World Demonstration

Robot specification

Height: 1.2m

Richo Theta 360° Camera (RGB sensor)

Intel Core i7 and GeForce RTX 2080

Jackal UGV from Clearpath

Real-World Demonstration

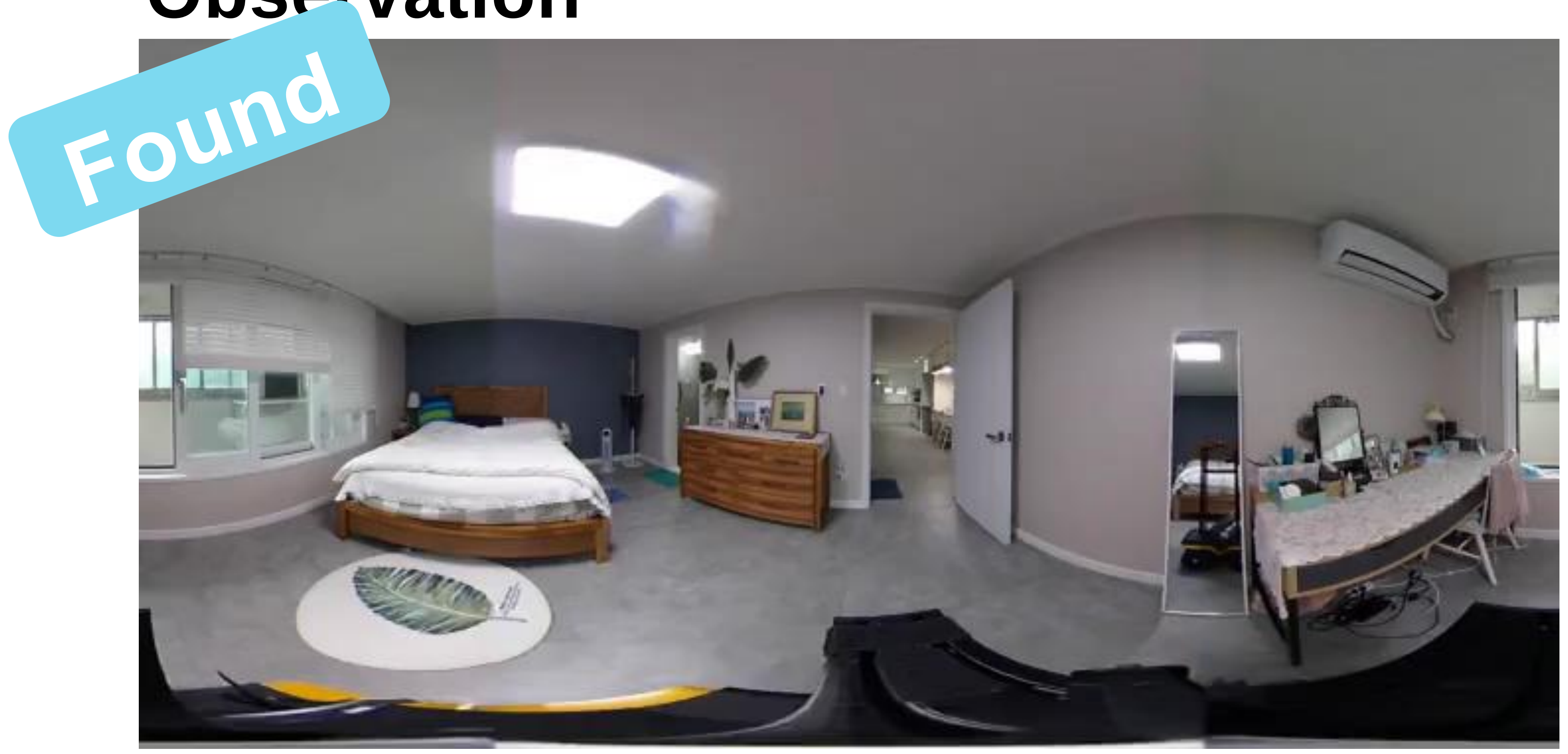
RLLAB Robot Learning Laboratory

Start Position

Goal Position

Goal *7.55m

Observation



* we estimated the robot and object locations to draw graphs on the map

Results

Method	Memory	No Pose	Object	Easy		Medium		Hard		Overall	
				Success	SPL	Success	SPL	Success	SPL	Success	SPL
RGBD + RL [26]	implicit	✗	✗	72.5	69.5	53.1	48.6	22.3	17.7	49.3	45.3
ANS [17]	metric	✗	✗	74.2	20.5	68.4	22.9	29.9	11.0	57.5	18.1
Exp4nav [5]	metric	✗	✗	70.2	61.8	60.6	52.4	46.9	38.5	59.2	50.9
SMT [8]	graph	✗	✗	81.9	77.4	65.6	52.2	55.6	39.7	67.7	56.4
Neural Planner [20]	graph	✗	✗	71.7	41.3	64.7	38.5	42.0	27.0	59.5	35.6
SPTM [9]	graph	✔	✗	66.5	40.6	64.2	38.5	42.1	25.4	57.6	34.8
VGM [18]	graph	✔	✗	86.1	79.6	81.2	**68.2**	60.9	45.6	76.1	64.5
TSGM (Ours)	graph	✔	✔	**91.1**	**83.5**	**82.0**	68.1	**70.3**	**50.0**	**81.1**	**67.2**

Results

RLLAB Robot Learning Laboratory

Path Type	Method	Easy		Medium		Hard		Overall	
		Success	SPL	Success	SPL	Success	SPL	Success	SPL
Straight	NRNS [27]	67.1	57.8	52.4	41.2	32.6	22.4	50.7	40.5
	VGM [18]	81.0	54.4	82.0	69.9	67.3	54.4	76.7	59.6
	TSGM (Ours)	**94.4**	**92.1**	**92.6**	**84.3**	**70.3**	**62.8**	**85.7**	**79.7**
Curved	NRNS [27]	31.7	13.0	29.0	13.6	19.2	10.4	26.6	12.3
	VGM [18]	81.0	45.5	78.8	59.5	62.2	46.9	74.0	50.6
	TSGM (Ours)	**93.6**	**91.0**	**89.7**	**77.8**	**64.2**	**55.0**	**82.5**	**74.1**

SPL: Success weighted by normalized inverse Path Length

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

Ablation Study on Cross Graph Mixer

Update	Success	SPL
No	0.533	0.393
Visual	0.578	0.446
Object	0.613	0.458
Cross	**0.627**	**0.471**

Ablation study on Cross graph mixer updates

Object Goal Navigation

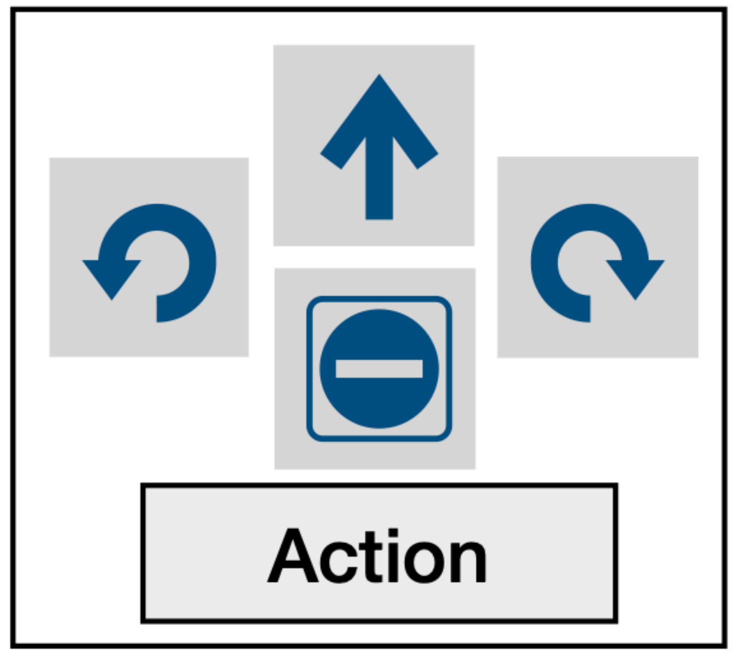
Input

Observation (RGBD)

(x, y, θ)

Pose Sensor

Output



Goal: Chair

Slide from Chaplot, CMU Ph.D. Thesis Defense: Building Intelligent Autonomous Navigation Agents

Results

RLLAB
Robot Learning Laboratory

Method	No Pose	Success Rate (↑)	SPL (↑)	DTS (↓)
BC	✗	12.2	8.3	3.90
DDPPO [1]	✗	15.0	10.7	3.24
FBE [2]	✗	64.3	28.3	1.78
Active Neural SLAM [3]	✗	67.1	34.9	1.66
SemExp [4]	✗	71.7	39.6	1.39
PONI [5]	✗	73.6	**41.0**	**1.25**
TSGM	✔	**75.1**	32.7	1.48

Method	No Pose	Success Rate (↑)	SPL (↑)	DTS (↓)
BC	✗	12.2	8.3	3.90
DDPPO [1]	✗	3.8	2.1	7.5
Red-Rabbit [6]	✗	34.6	1.8	-
THDA [7]	✗	28.4	7.9	5.6
FBE [2]	✗	20.0	7.6	6.5
Active Neural SLAM [3]	✗	21.2	9.4	6.3
PONI [5]	✗	27.8	12.0	5.6
TSGM	✔	**46.1**	**15.3**	**4.7**

*Object Goal Navigation Results on **Gibson** and **MP3D** dataset*

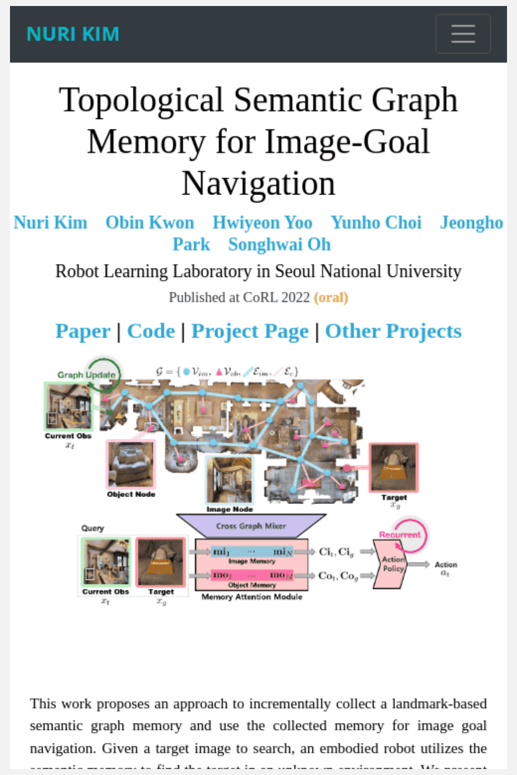
Summary & Dirty Laundry

- ☑ TSGM made robots understand the relationship between an object and a place.
  - A robot can find efficient paths if it knows “the coffee machine” tends to be in the “kitchen,” not in a “bedroom.”
- ☑ TSGM **Integrates semantic information** to topological graph memory, outperforming SOTA methods on image goal navigation.
  - To the best of our knowledge, we firstly constructed object graph on the topological graph.

- ☐ Dirty Laundry
  - Since it does not use any metric map, a robot often bumps into the corner.
  - The image nodes do not exactly represent the semantic places, such as a kitchen.

Topological Semantic Graph Memory

RLLAB Robot Learning Laboratory



Topological Semantic Graph Memory for Image-Goal Navigation

Nuri Kim, Obin Kwon, Hwiyeon Yoo, Yunho Choi, Jeongho Park, and Songhwai Oh

CoRL 2022

Webpage: https://bareblackfoot.github.io/TopologicalSemanticGraphMemory

Nuri Kim

Webpage: http://bareblackfoot.github.io

Email nuri.kim@rllab.snu.ac.kr

Twitter: @nurikimmel