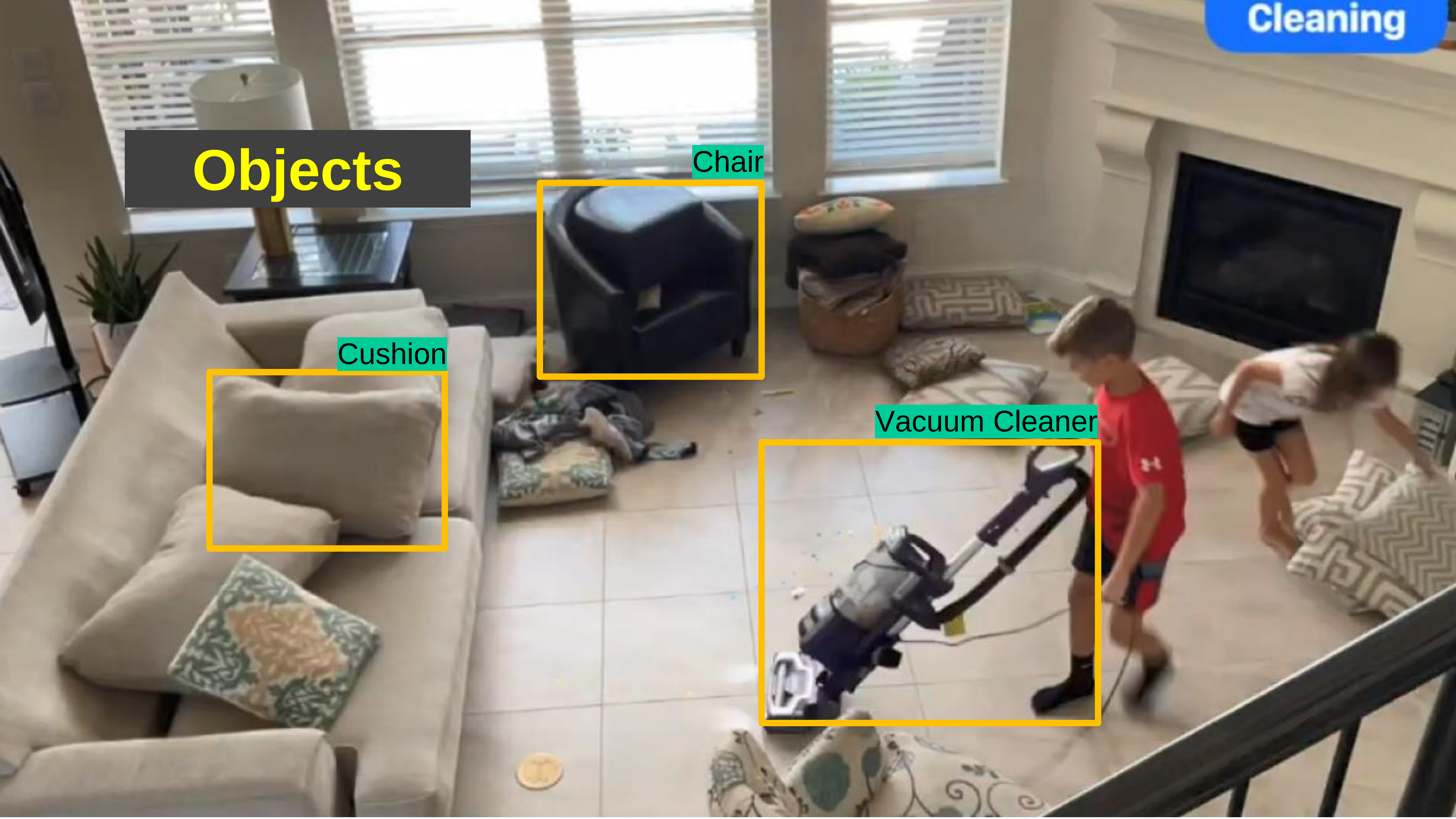
Learning visual representations for robotics

Ivan Laptev

Ivan.Laptev@mbzuai.ac.ae
https://www.di.ens.fr/~laptev
Visiting professor, MBZUAI, United Arab Emirates
External member, Willow Team, Inria, DI ENS, Paris

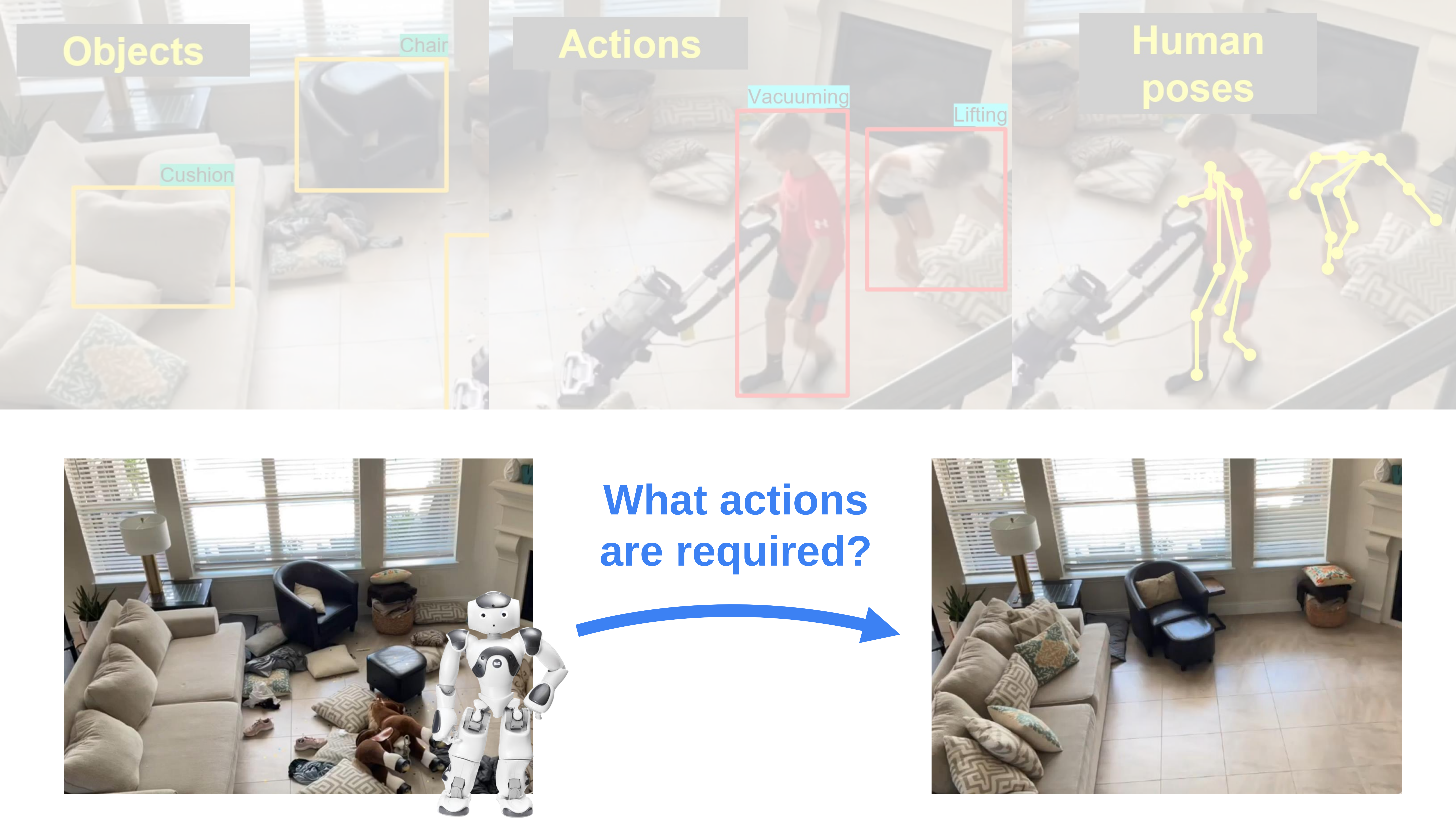

NASA

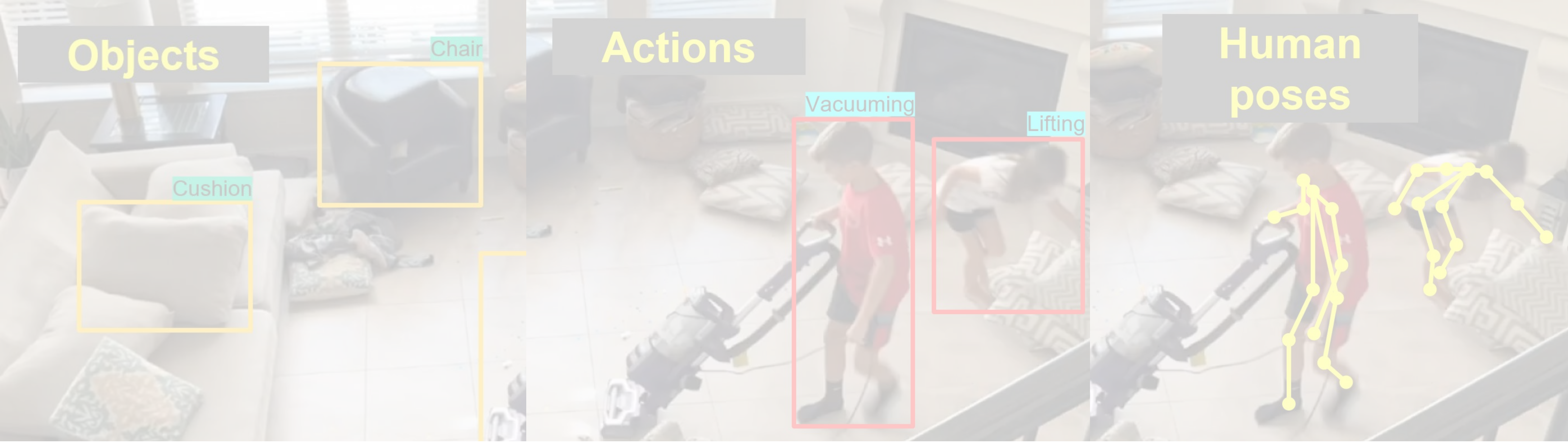

Cleaning
Objects
Chair
Cushion
Vacuum Cleaner

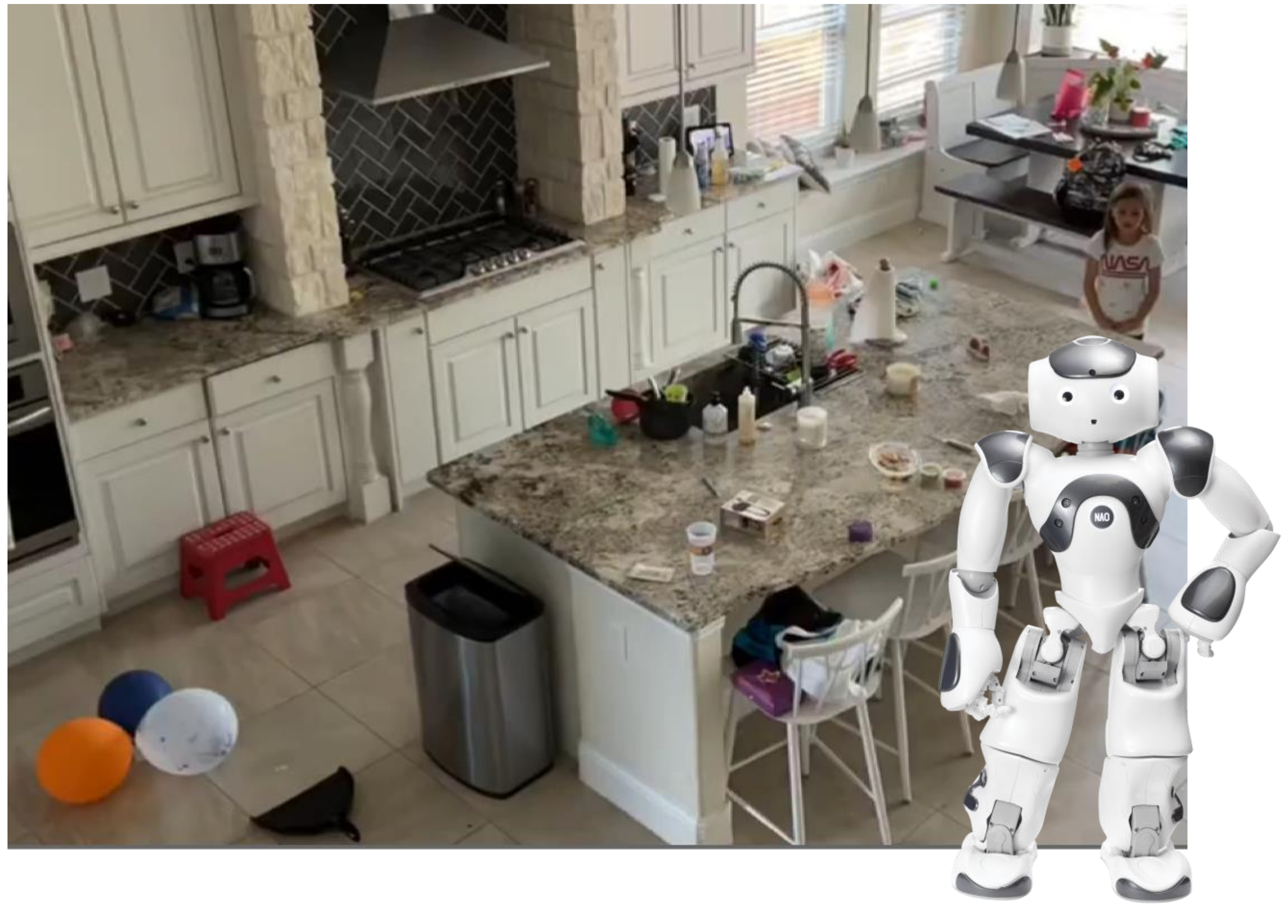

Cleaning
Actions
Vacuuming
Lifting


Cleaning
Human poses

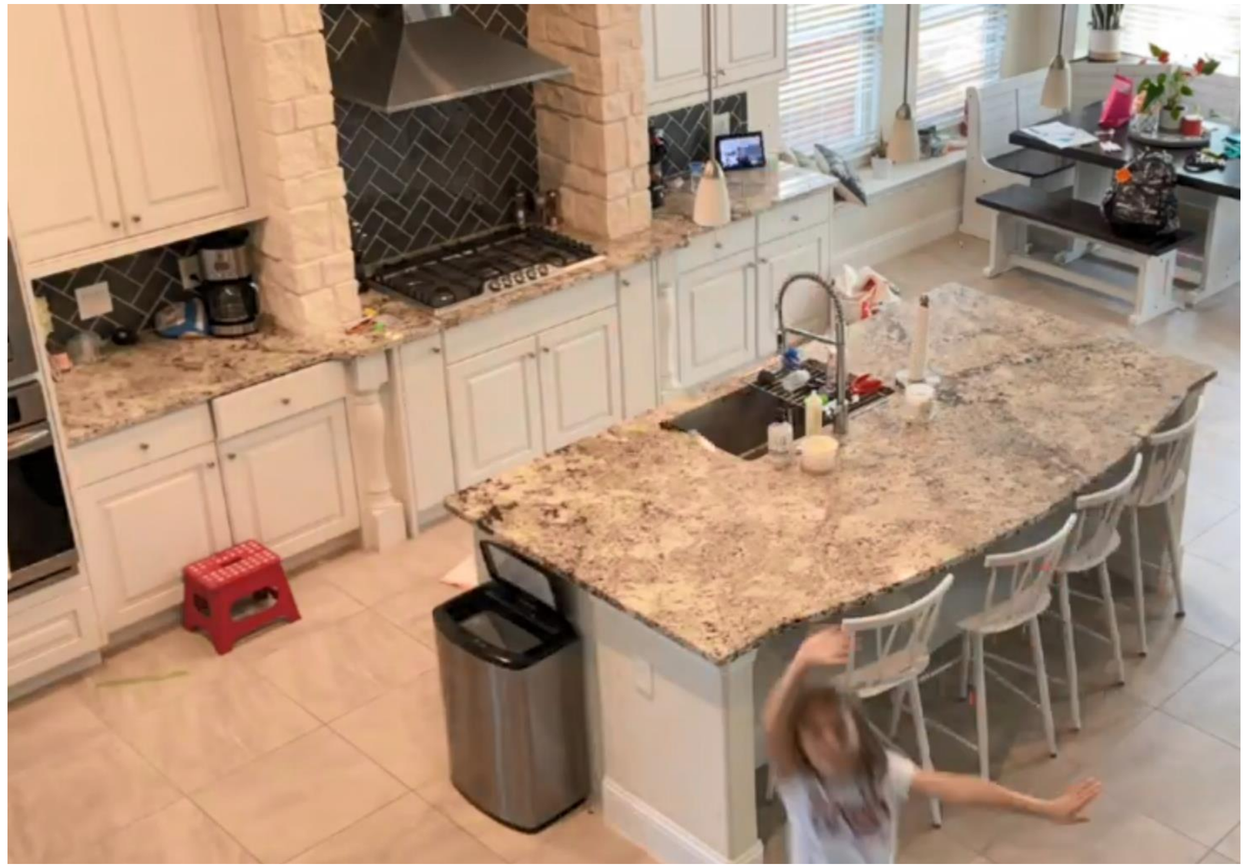

Objects
Chair
Cushion
Actions
Vacuuming
Lifting
Human poses


Objects
Chair
Cushion
Actions
Vacuuming
Lifting
Human poses
What actions are required?



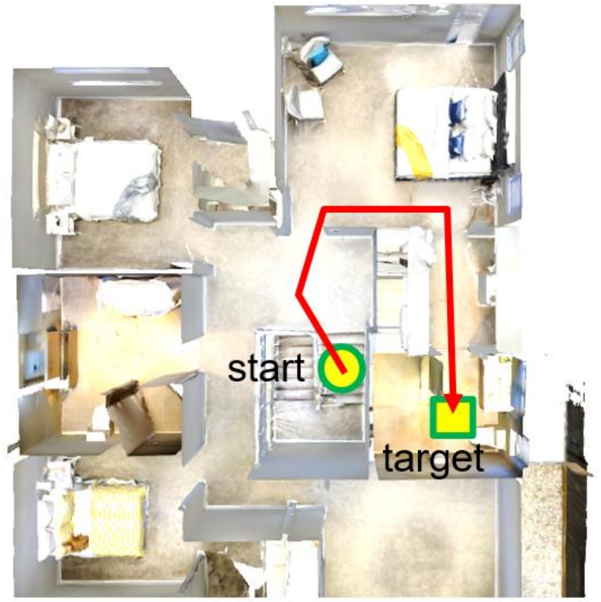


What actions are required?




Navigation
required?
Manipulation

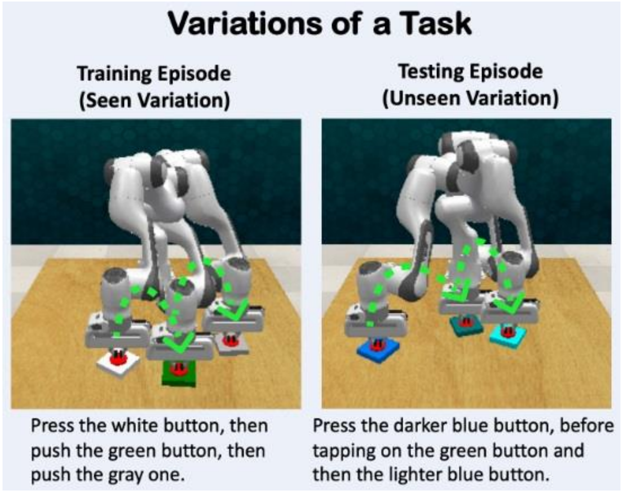
Summary



History Aware Multimodal Transformer for Vision-and-Language Navigation, S. Chen, P.-L. Guhur, C. Schmid and I. Laptev; *in Proc. NeurIPS 2021*

Object Goal Navigation with Recursive Implicit Maps, S. Chen, T. Chabal, I. Laptev and C. Schmid; *In Proc. IROS 2023*

Vision and language **navigation**

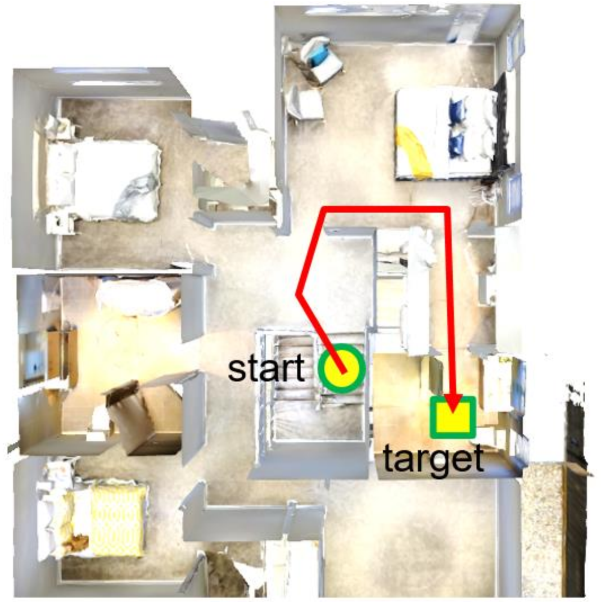


Instruction-driven history-aware policies for robotic manipulations, P.-L. Guhur, S. Chen, R. Garcia, M. Tapaswi, I. Laptev and C. Schmid; *in Proc. CoRL 2022*

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation, S. Chen, R. Garcia, C. Schmid and I. Laptev; *In Proc CoRL 2023*

Vision and language **manipulation**

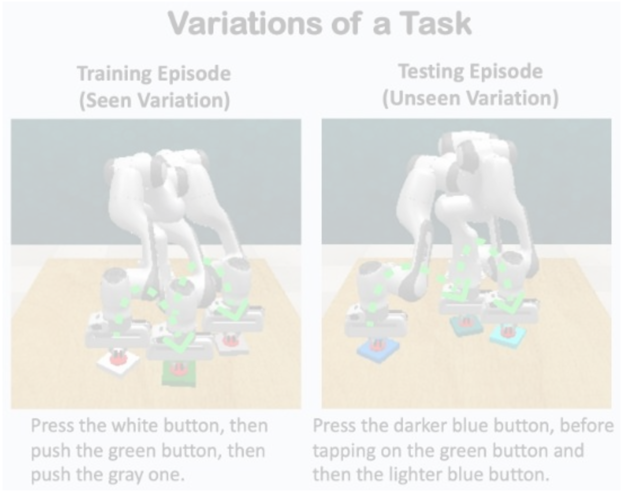
Summary



History Aware Multimodal Transformer for Vision-and-Language Navigation,
S. Chen, P.-L. Guhur, C. Schmid and I. Laptev;
in Proc. NeurIPS 2021

Object Goal Navigation with Recursive Implicit Maps,
S. Chen, T. Chabal, I. Laptev and C. Schmid;
In Proc. IROS 2023

Vision and language **navigation**



Instruction-driven history-aware policies for robotic manipulations, P.-L. Guhur, S. Chen, R. Garcia, M. Tapaswi, I. Laptev and C. Schmid; *in Proc. CoRL 2022*

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation, S. Chen, R. Garcia, C. Schmid and I. Laptev; *In Proc CoRL 2023*

Vision and language **manipulation**

History Aware Multimodal Transformer for Vision-and-Language Navigation




Shizhe Chen Pierre-Louis Guhur Cordelia Schmid Ivan Laptev

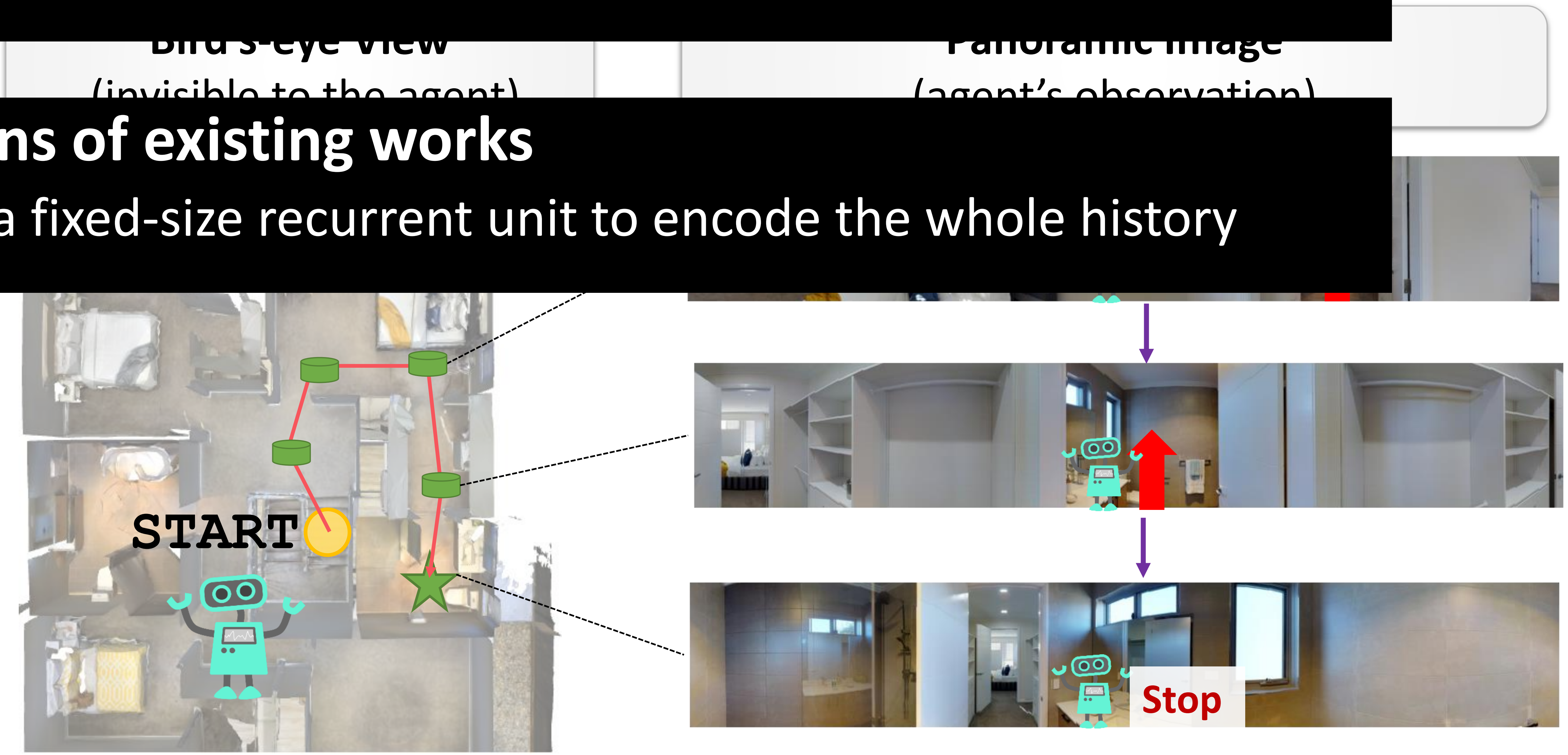
NeurIPS 2021

Webpage: https://cshizhe.github.io/projects/vln_hamt.html

VLN Challenges: Modeling history

- **Keeping track of the navigation state**
  - Environment understanding
  - Instruction grounding

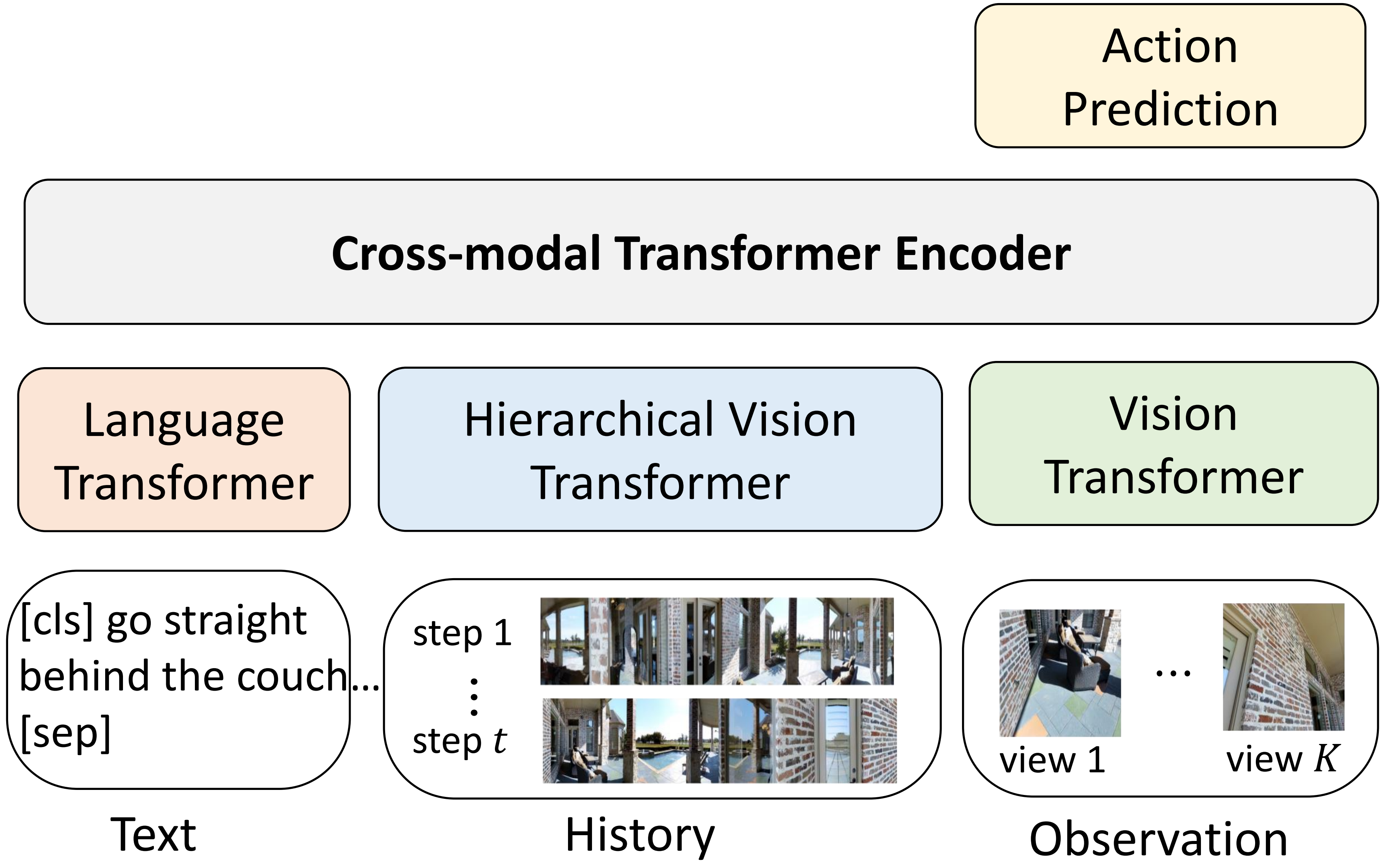
Turn left and continue
up the st
Go straig
the bedr
the right
past the bed.
Turn right again and
go through the closet.
Continue straight, into
the bathroom.
Wait right there, in
front of the mirror.



- **Limitations of existing works**
  - Adopt a fixed-size recurrent unit to encode the whole history

Our Proposed Model: HAMT

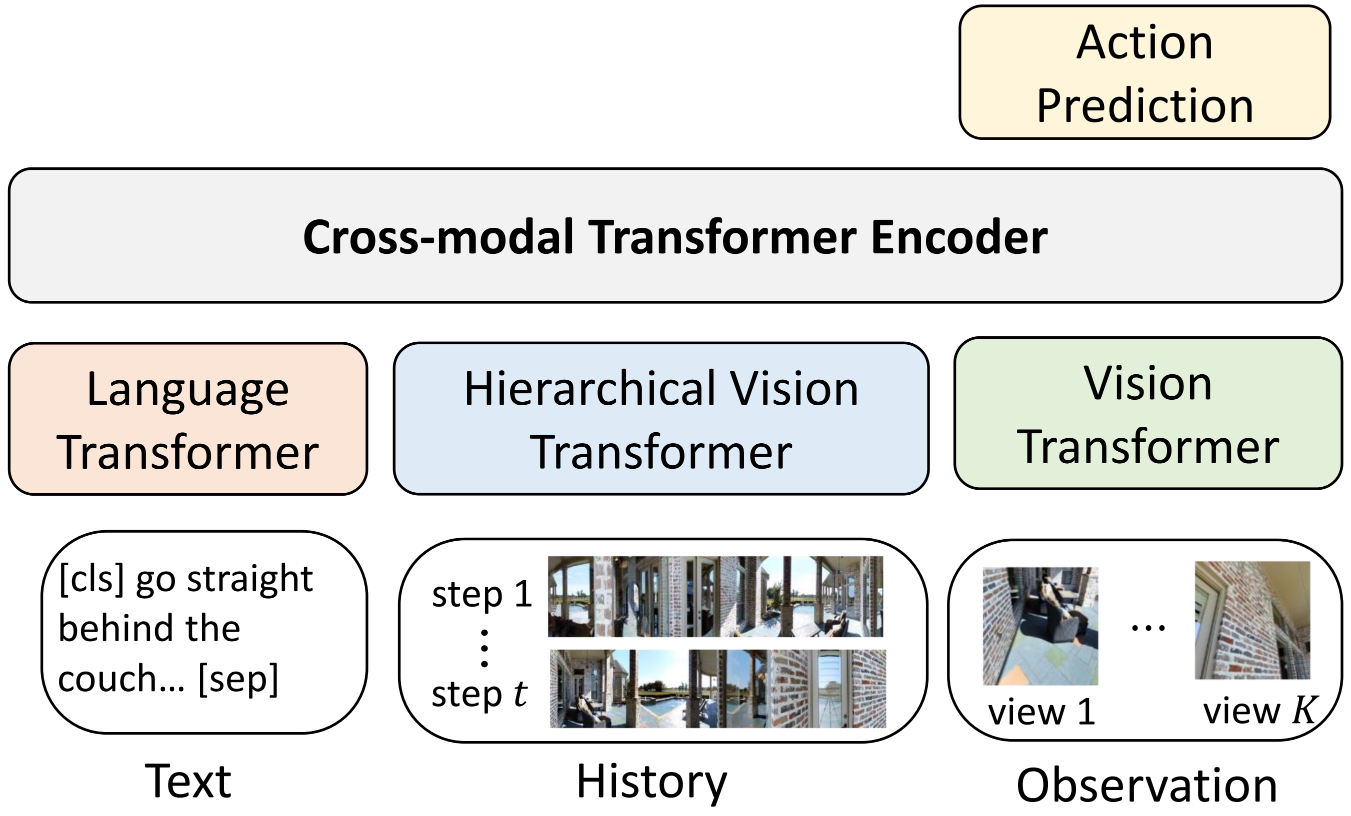
- History Aware Multimodal Transformer (HAMT)



A fully transformer-based architecture for multimodal decision making

Our Proposed Model: HAMT

- Long-horizon history modelling
  - Learn dependency of all panoramic observations and actions in history sequence
- End-to-end optimization for visual representation
  - Fully transformer-based architecture allows efficient training

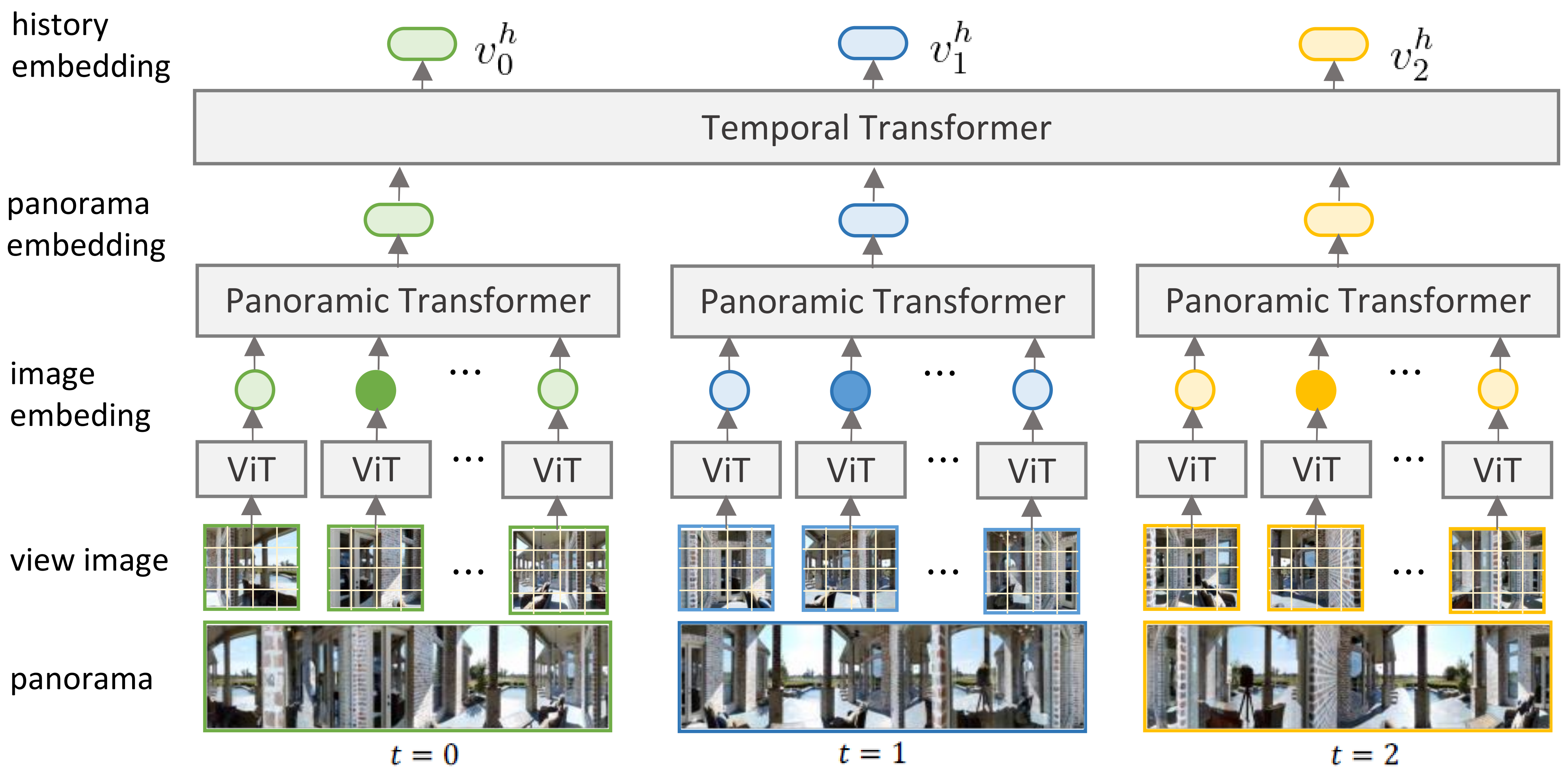


PROBLEMS

- Computationally expensive to encode all panoramas
  - K views, T steps → $O(K^2T^2)$
- The action prediction task alone might be insufficient to learn generalizable models

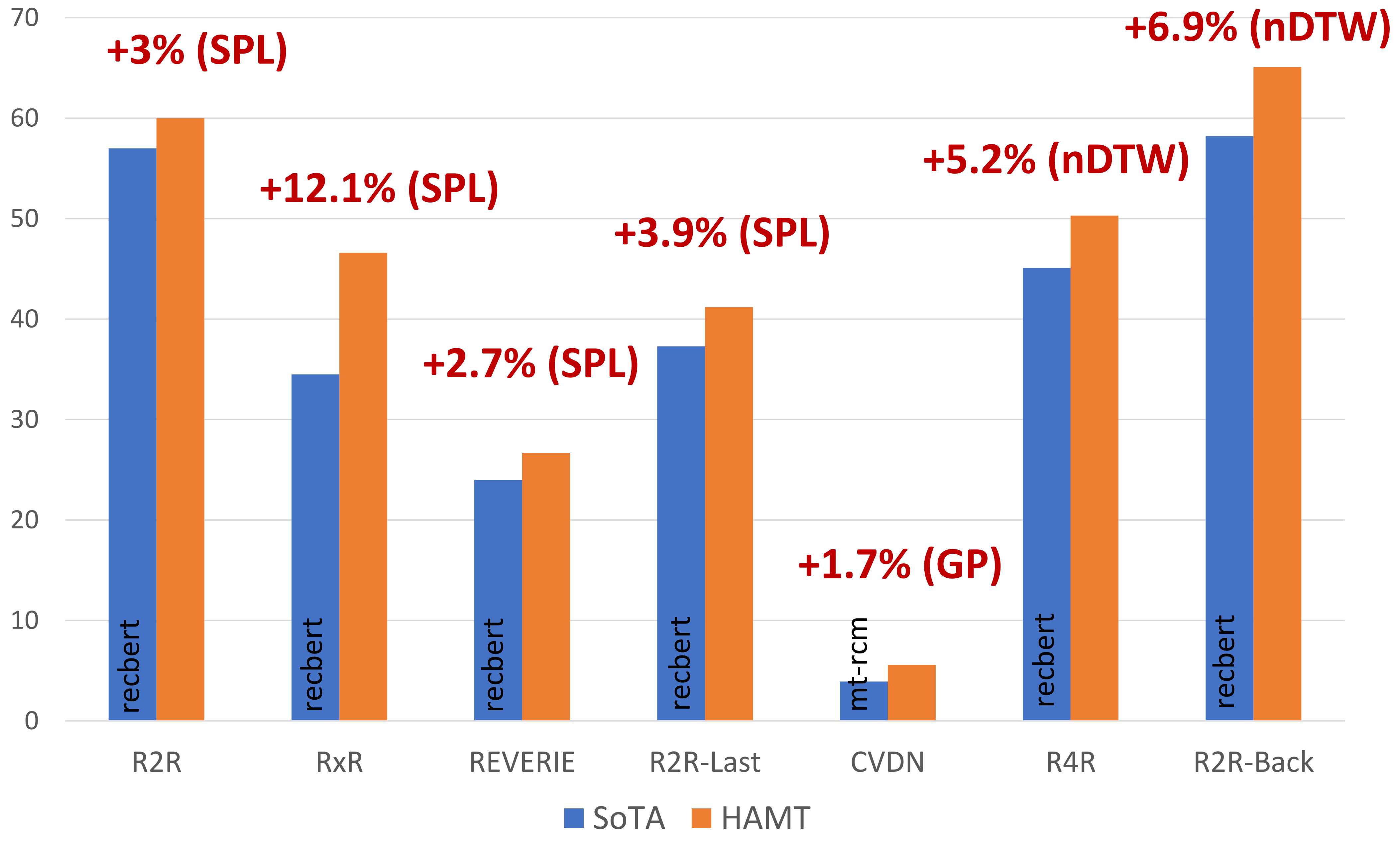
HAMT: Hierarchical History Encoding

- ViT for single view image encoding
- Panoramic Transformer for spatial relation encoding within panorama
- Temporal Transformer for temporal relation encoding across panoramas



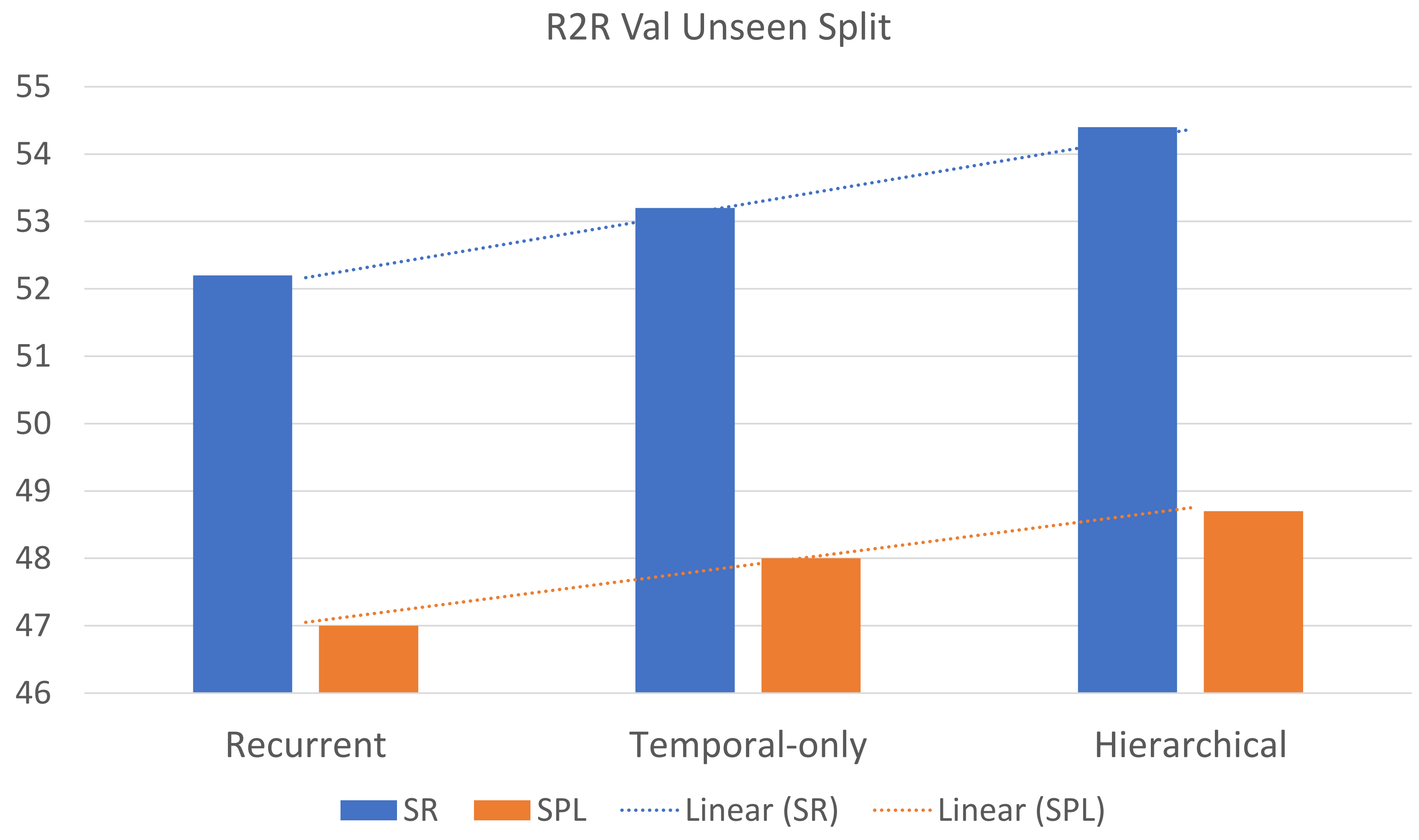
Experiments: Comparison with SoTA

- HAMT outperforms state of the art on all datasets

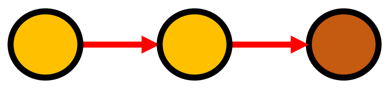
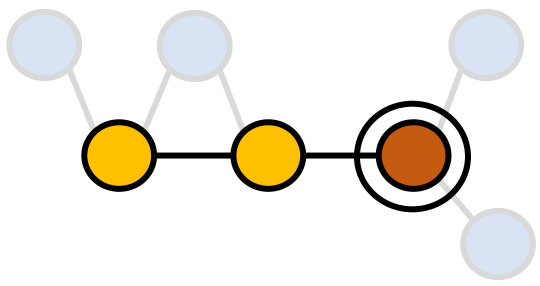
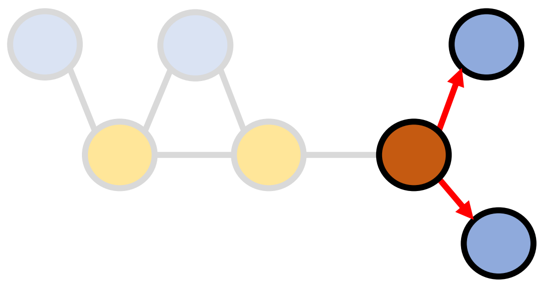


Experiments: Ablation

- How important is the history encoding?
  - Recurrent: a fixed-size vector to encode the whole history
  - Temporal-only: select only one view per panorama to improve efficiency
  - **Hierarchical: hierarchically encode all panoramas**



Limitations of HAMT

	Navigation Memory	Visual Memory	Action Space
HAMT			
		Sequence no structure of the house	**Local actions** hard to backtrack many steps

Current location

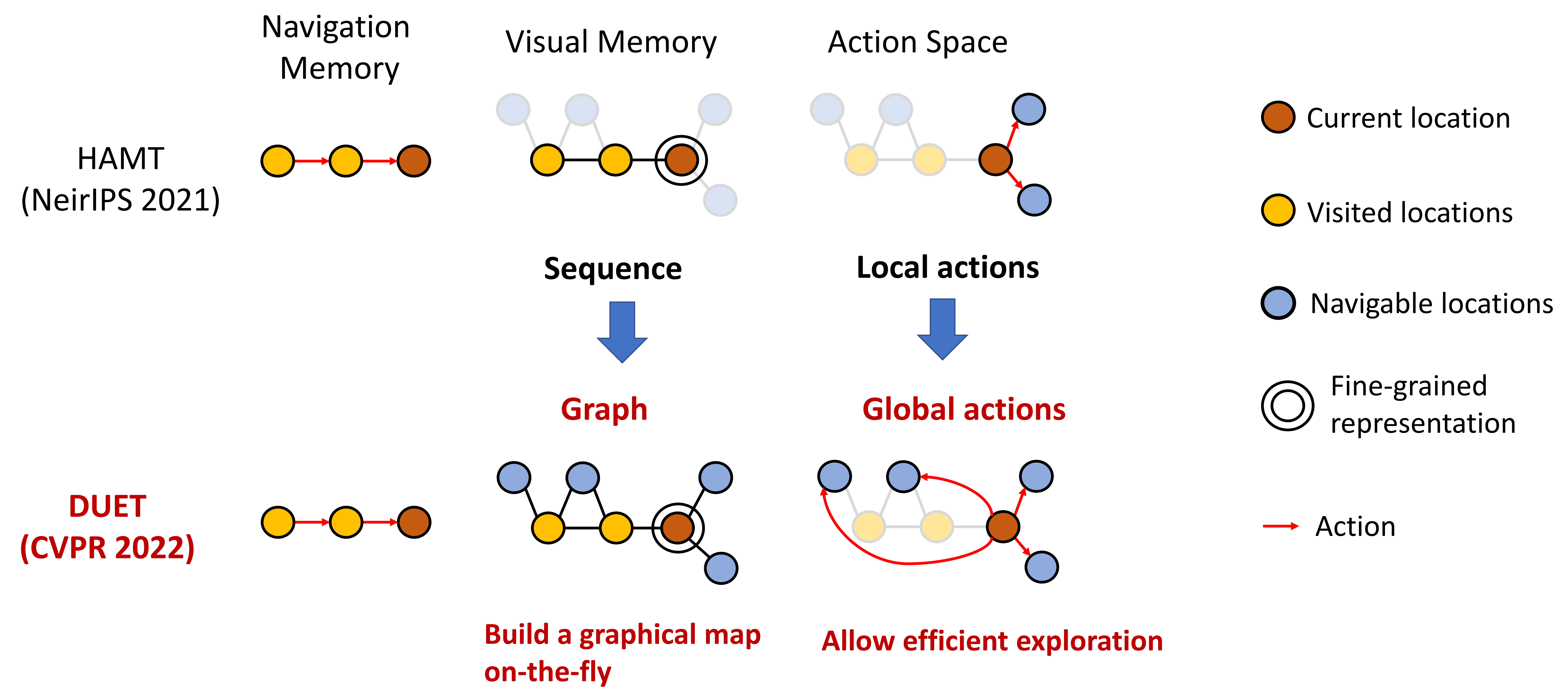
Visited locations

Navigable locations

Fine-grained representation

Action

Improving HAMT with Structured Memory



DUET: Experimental Results

- REVERIE dataset

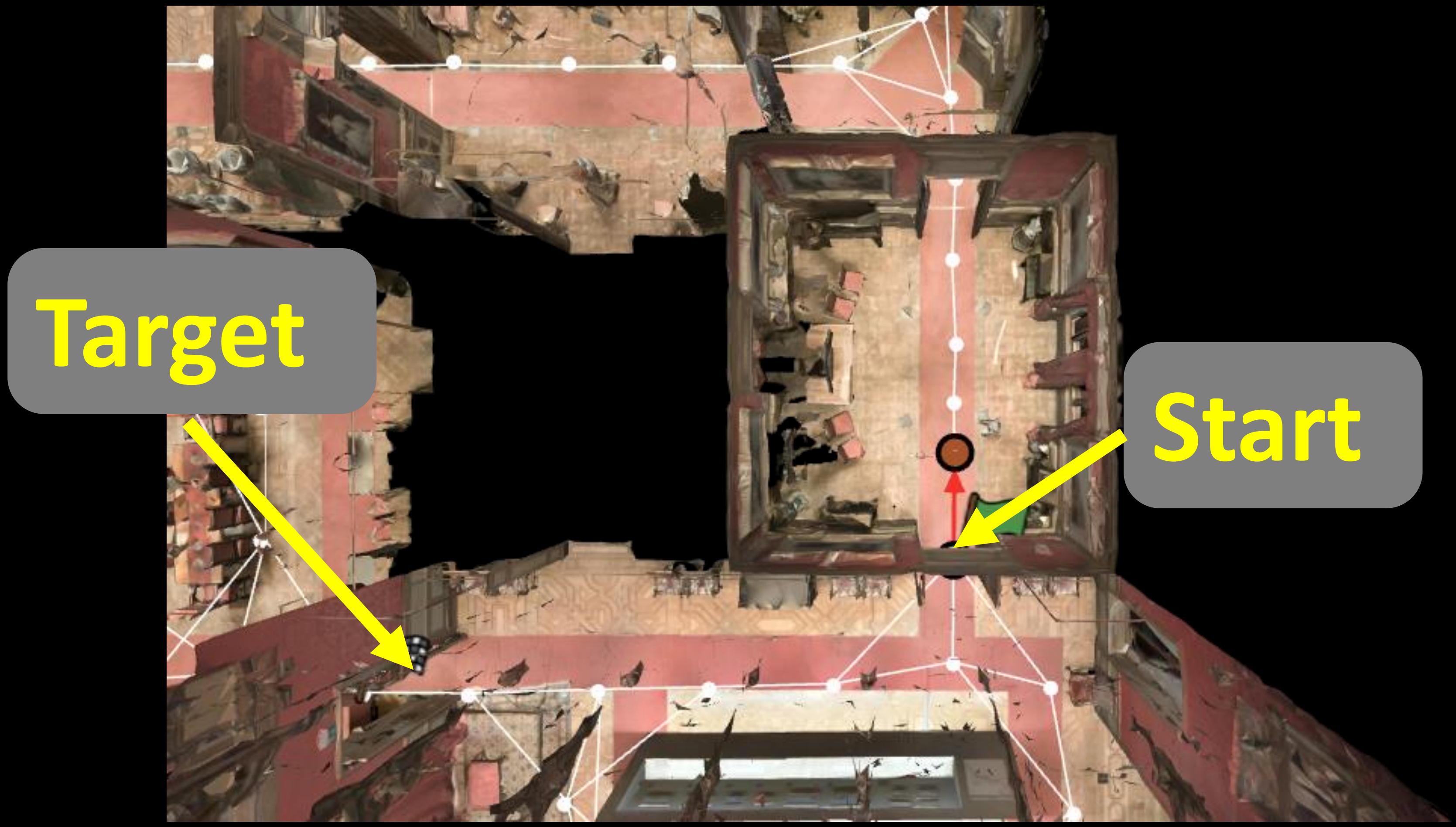
	SR	SPL	RGS	RGSPL
HAMT	30.40	26.67	14.88	13.08
DUET	**52.51**	**36.06**	**31.88**	**22.06**

- SOON dataset

Split	Methods	TL	OSR↑	SR↑	SPL↑	RGSPL↑
Val Unseen	GBE [8]	28.96	28.54	19.52	13.34	1.16
	DUET (Ours)	36.20	**50.91**	**36.28**	**22.58**	**3.75**
Test Unseen	GBE [8]	27.88	21.45	12.90	9.23	0.45
	DUET (Ours)	41.83	**43.00**	**33.44**	**21.42**	**4.17**

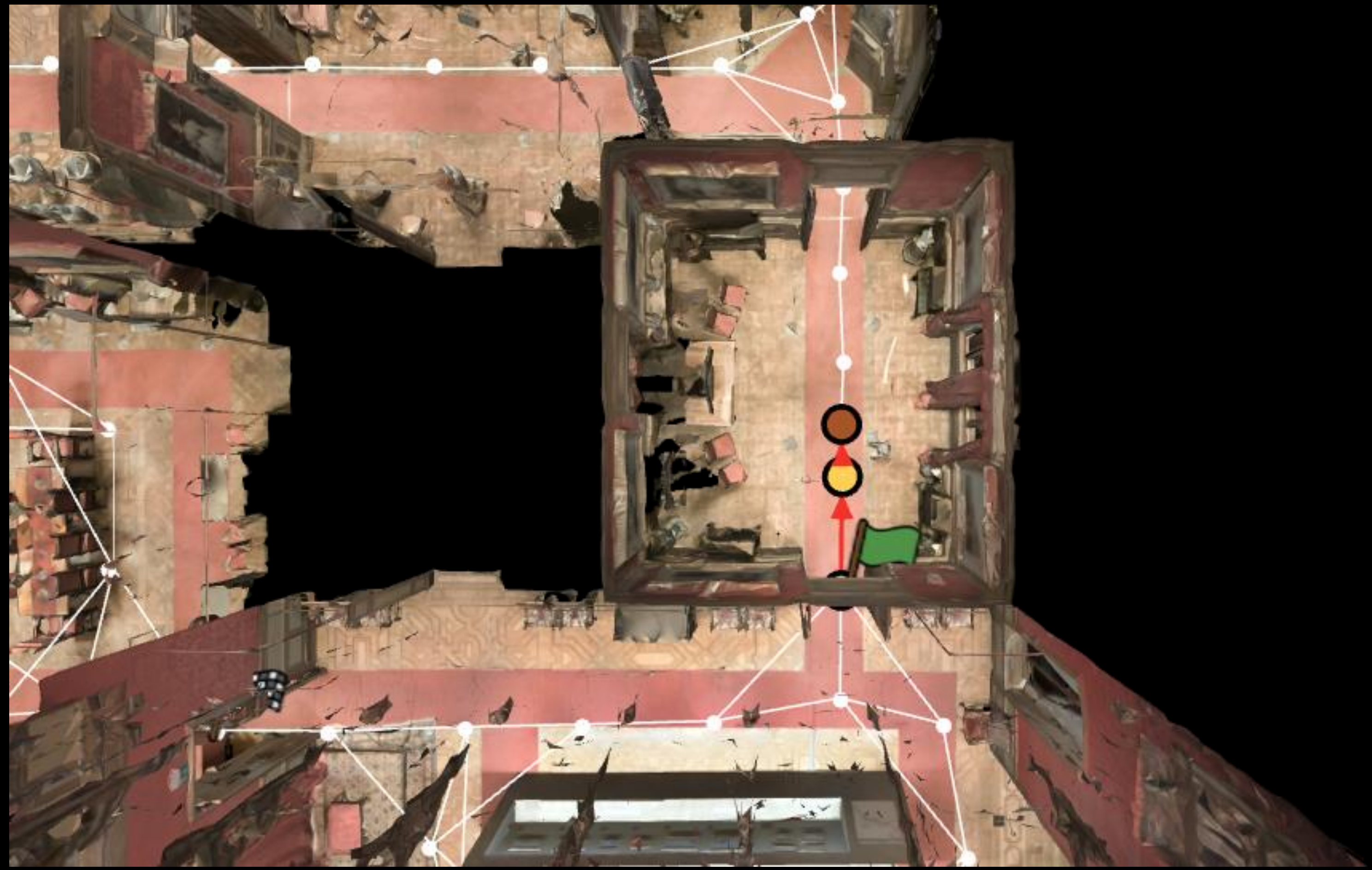
- **Winner of VLN Challenges** hosted in Human Interaction for Robotics Navigation Workshop at ICCV 2021

Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.

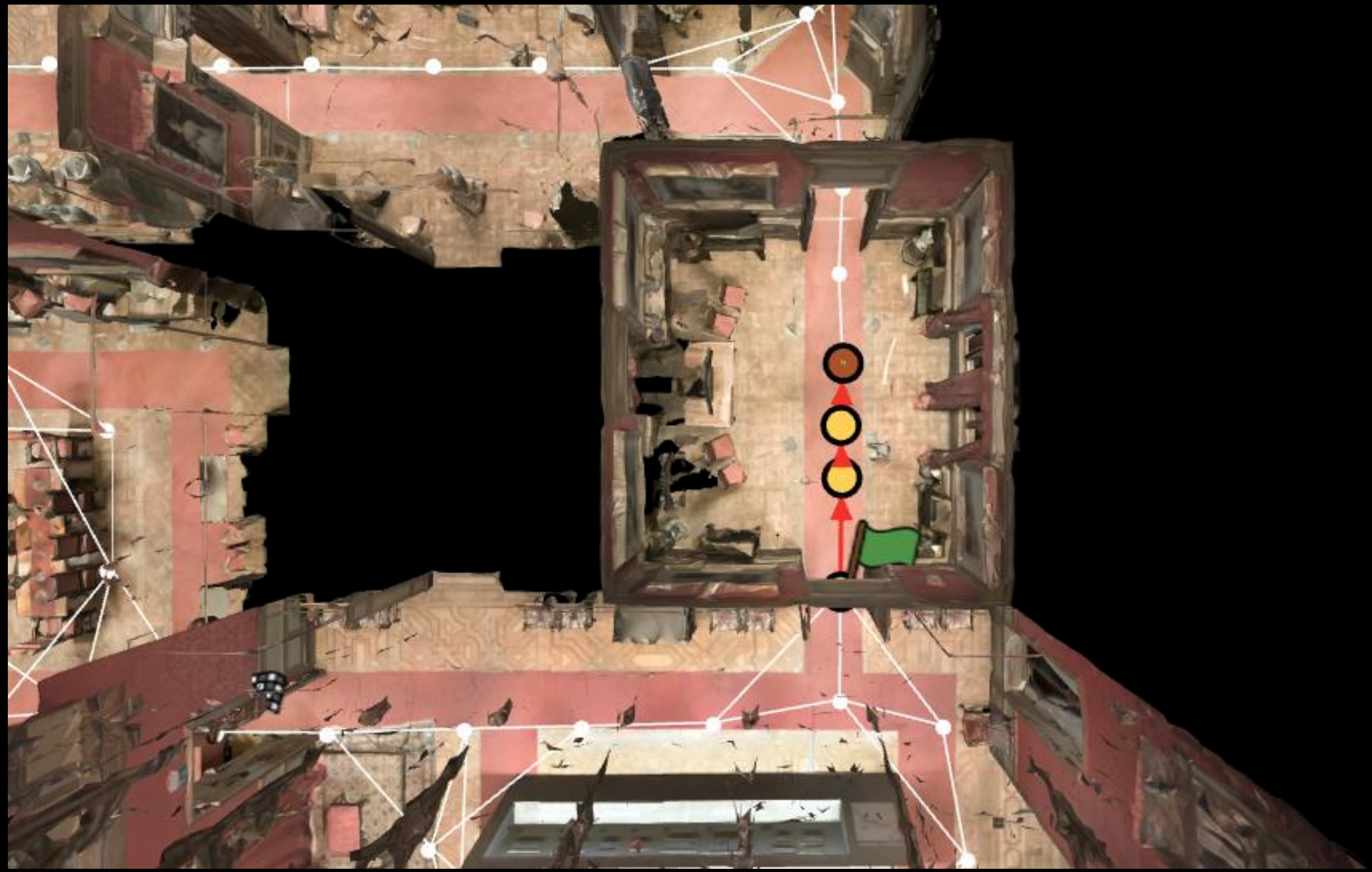


Instruction: **Exit the roped off hall, follow the red carpet**, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.

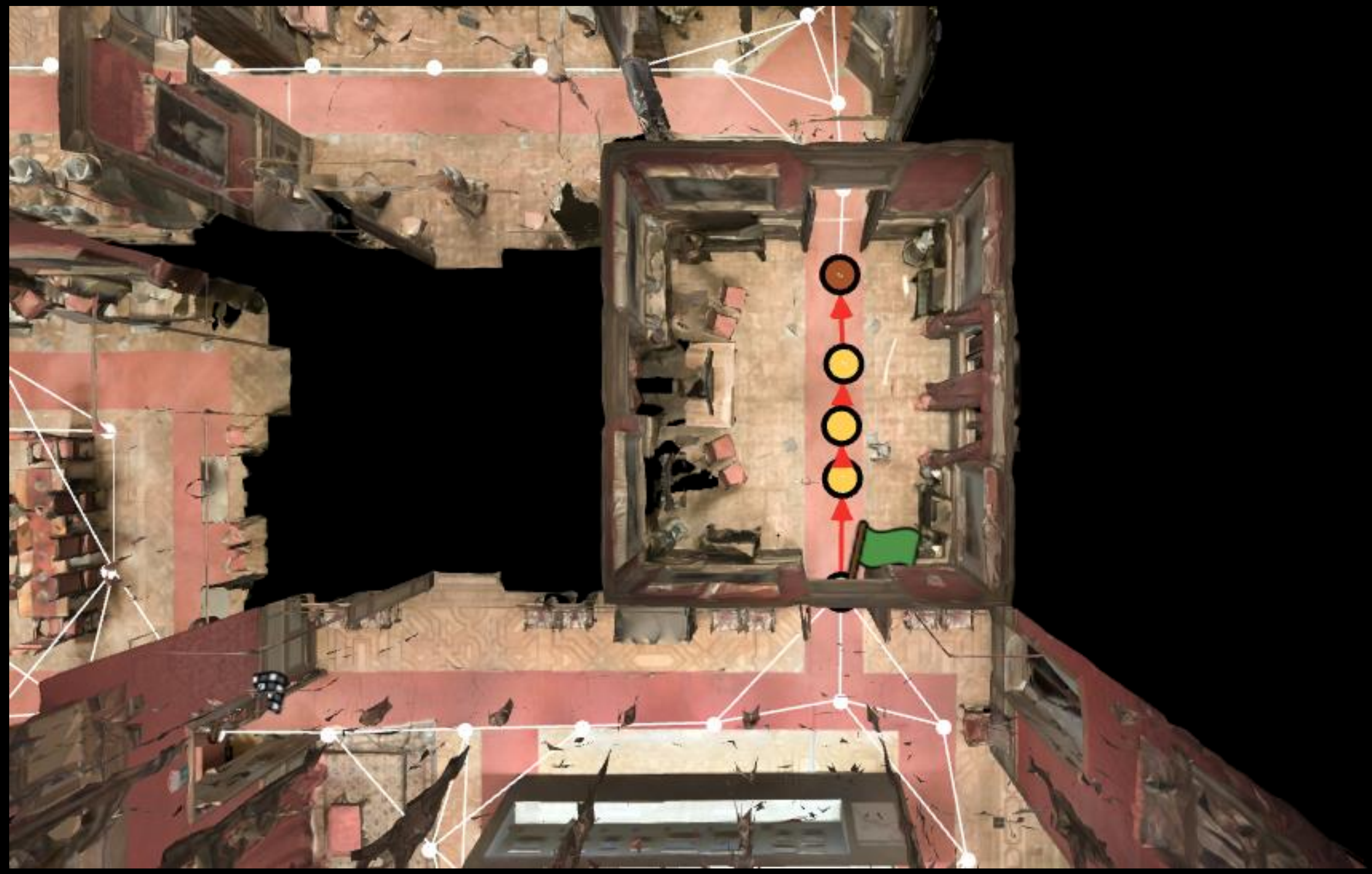
Instruction: **Exit the roped off hall, follow the red carpet**, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.



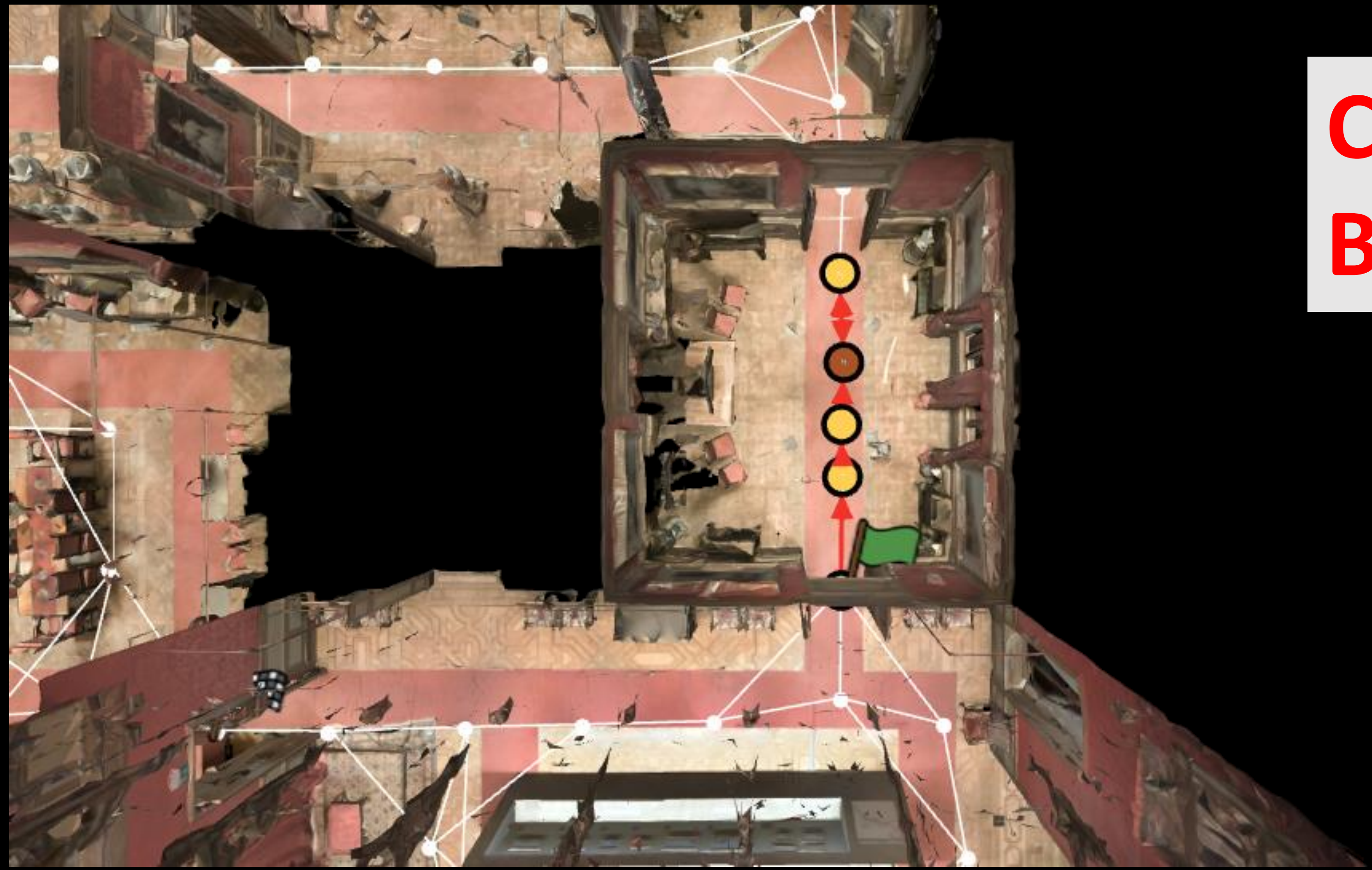
Instruction: **Exit the roped off hall, follow the red carpet**, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.



Instruction: **Exit the roped off hall, follow the red carpet**, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.

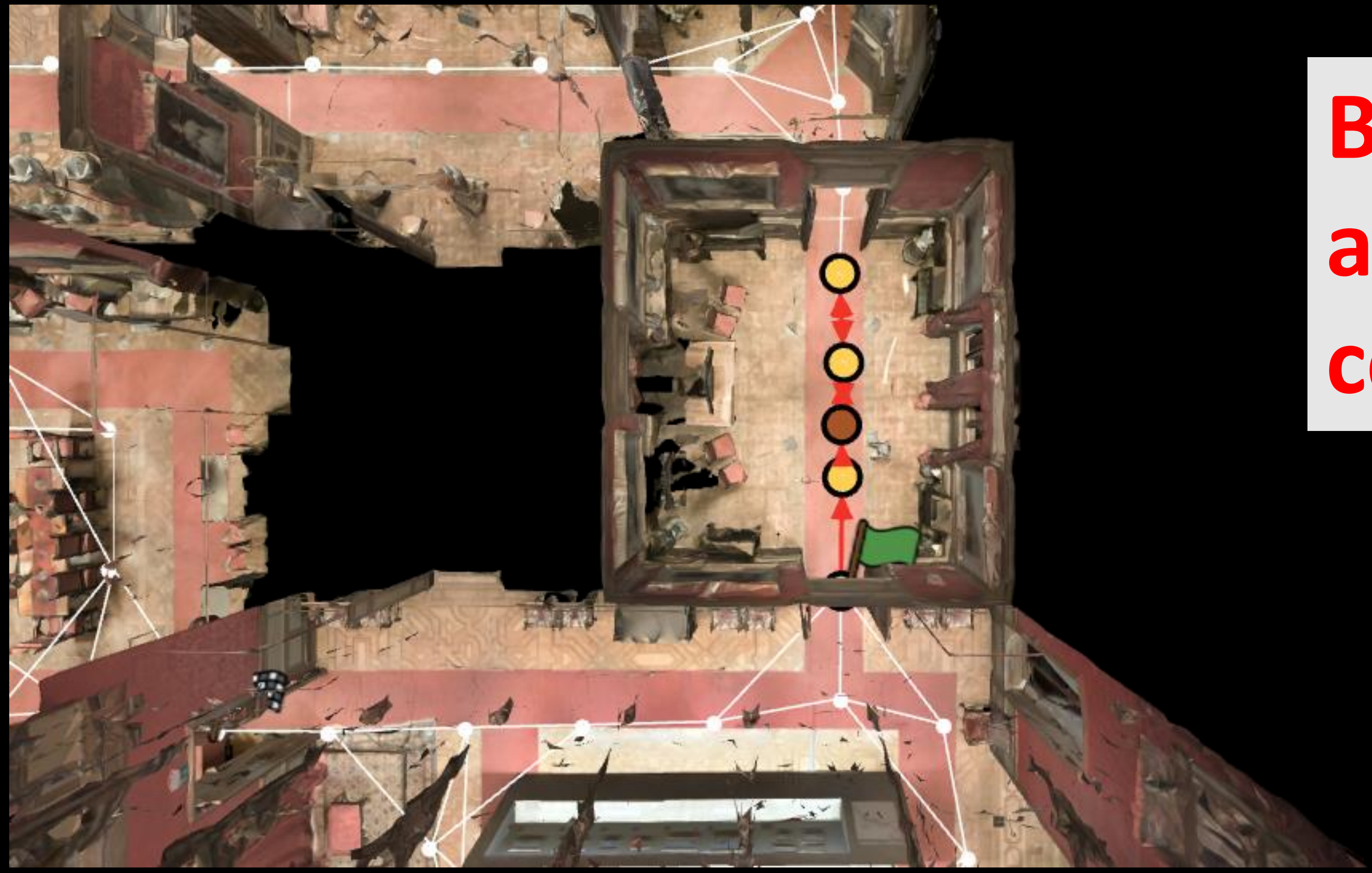


Instruction: Exit the roped off hall, follow the red carpet, **turn right**, continue straight down the red carpet, enter room at the end, stop once inside the room.



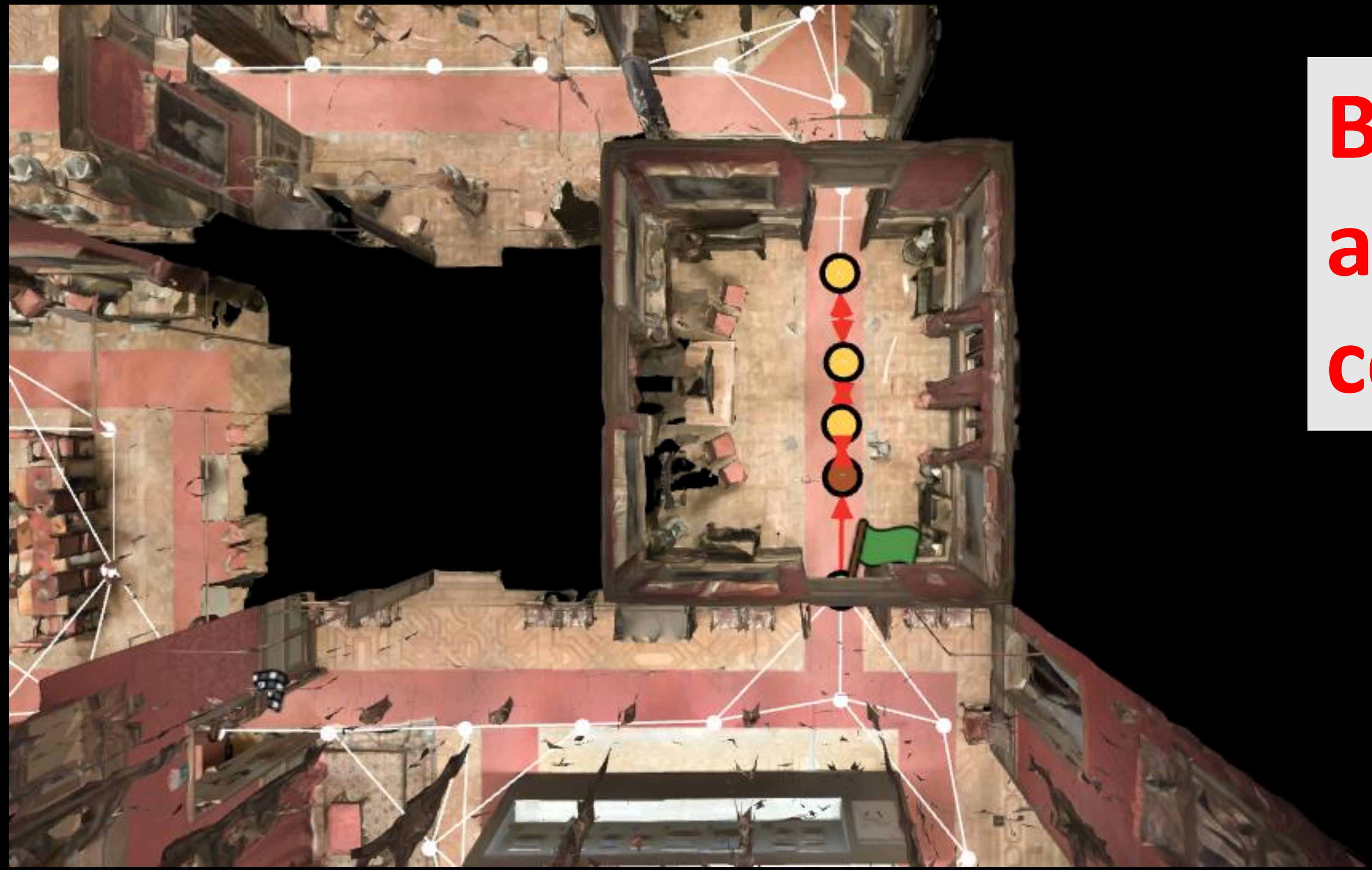
Cannot turn right.
Back Track

Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.



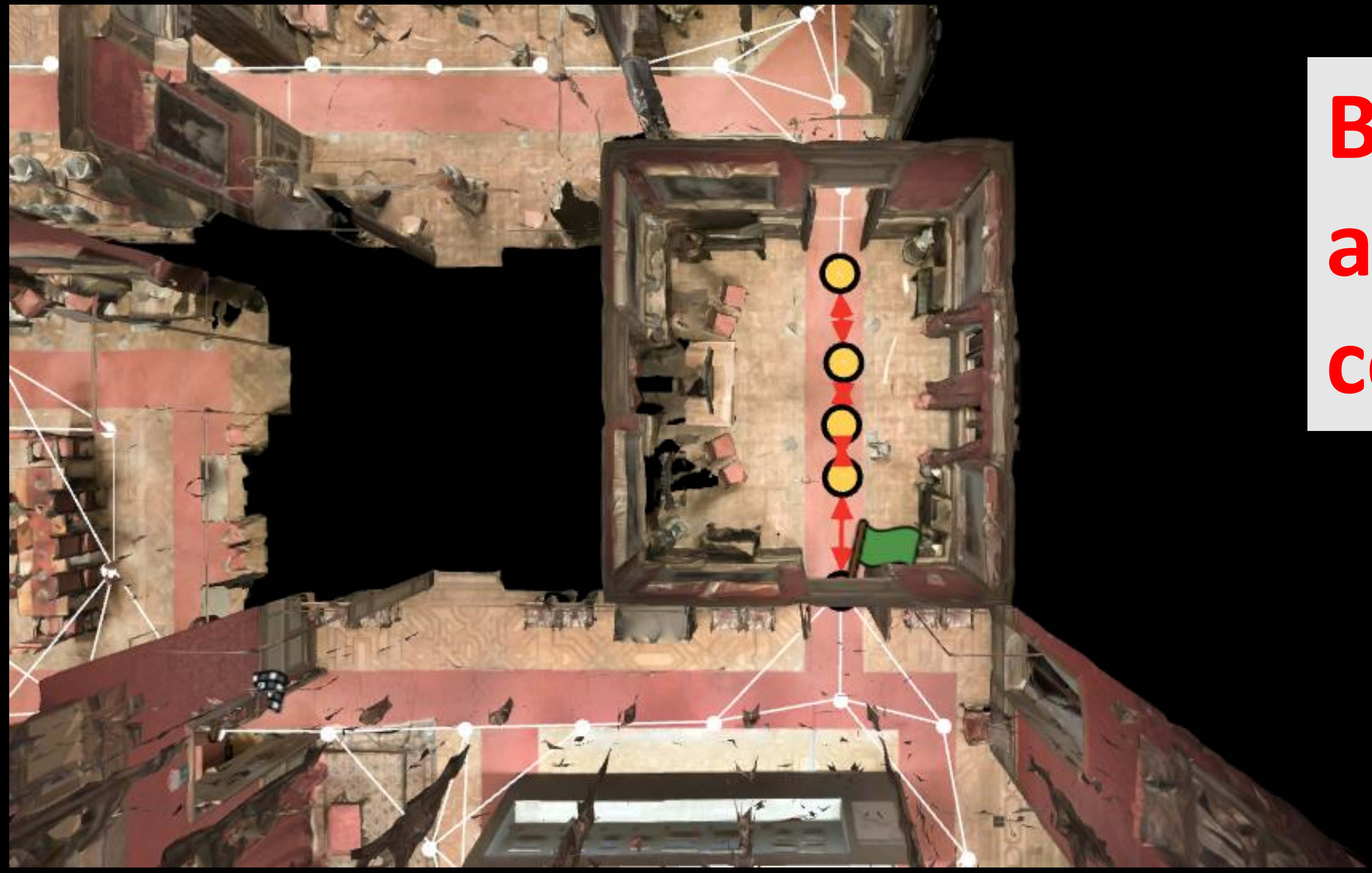
Back tracking according to the constructed map.

Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.



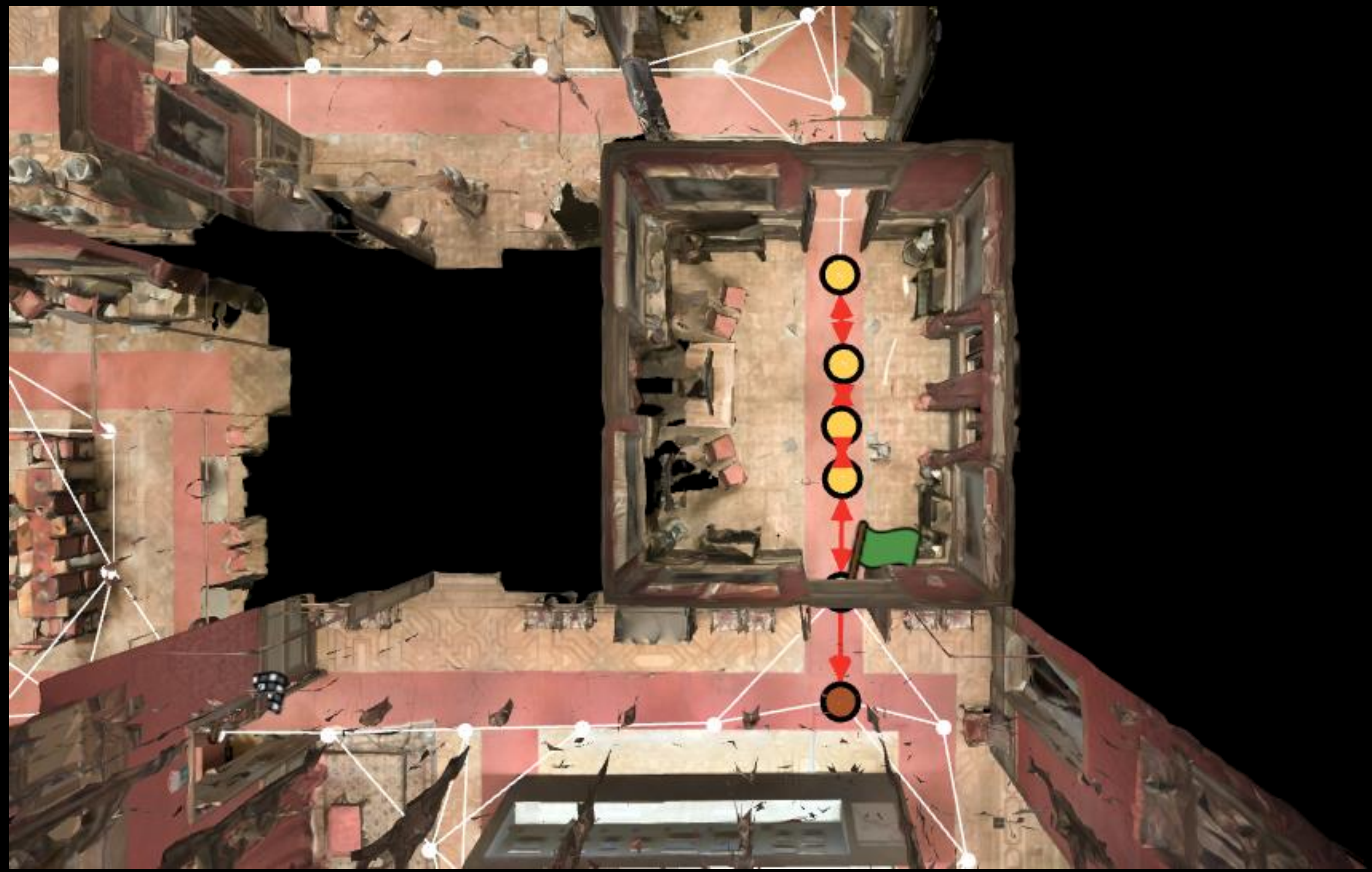
Back tracking according to the constructed map.

Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.

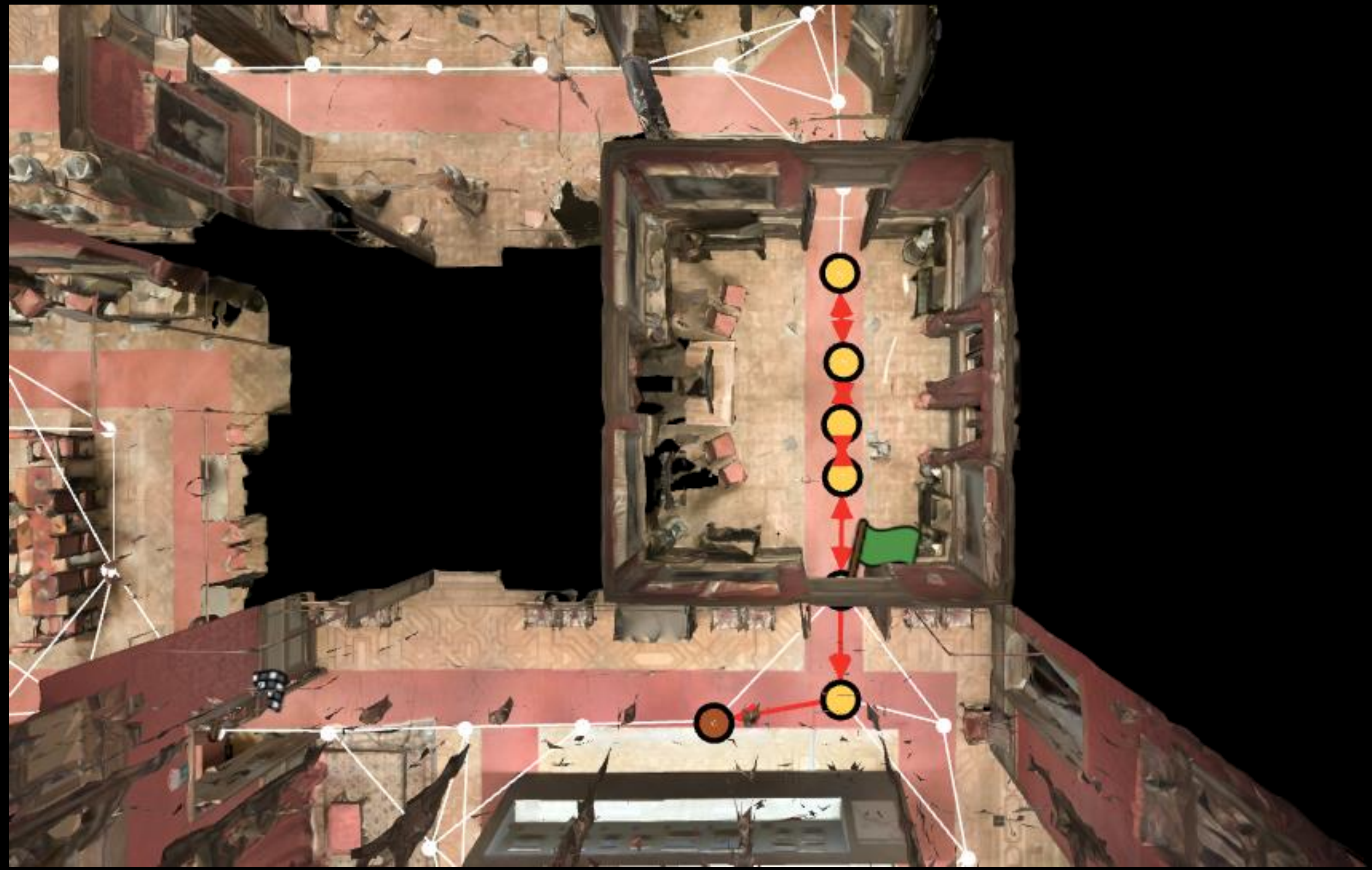


Back tracking according to the constructed map.

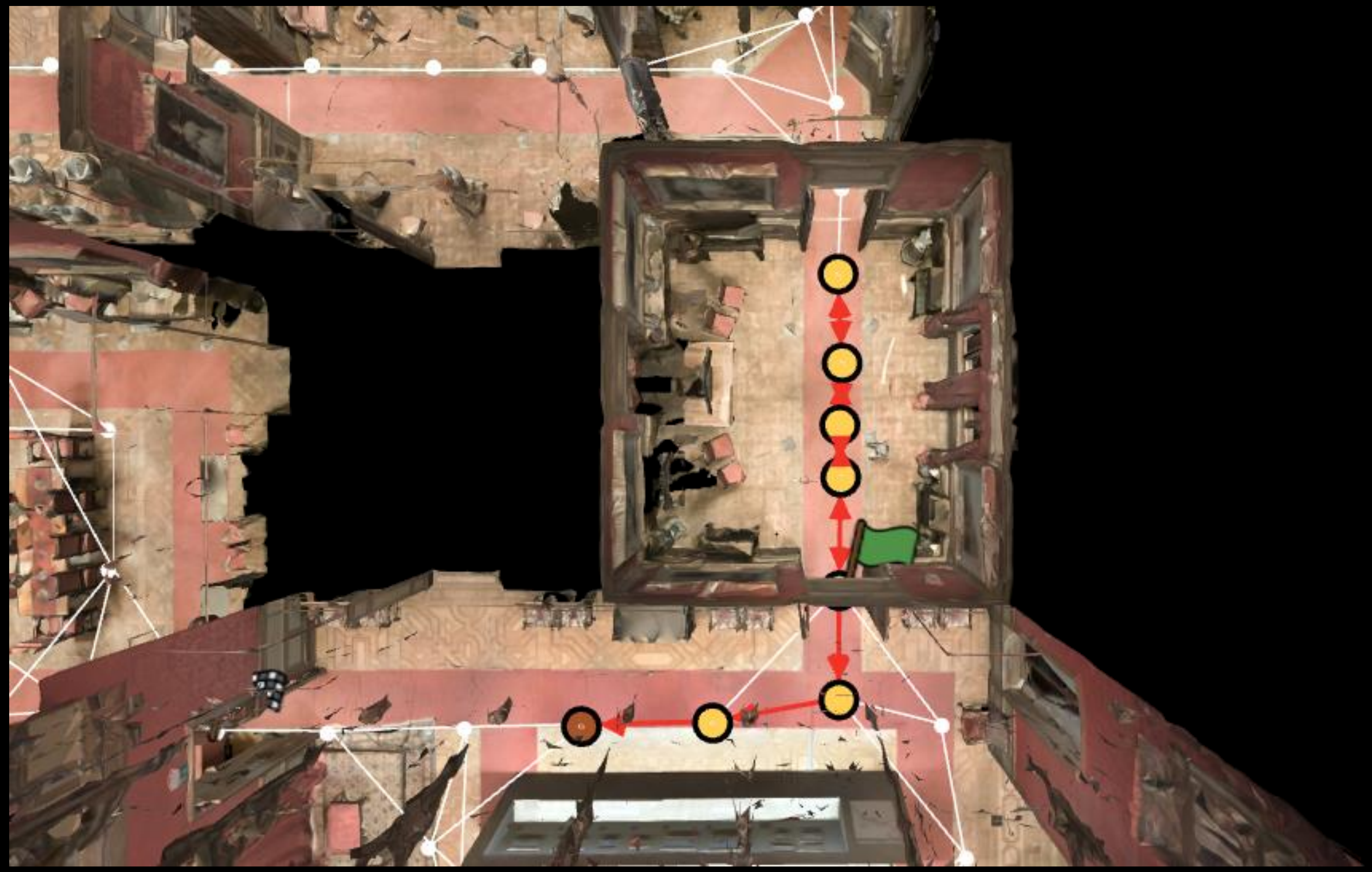
Instruction: **Exit the roped off hall, follow the red carpet**, turn right, continue straight down the red carpet, enter room at the end, stop once inside the room.



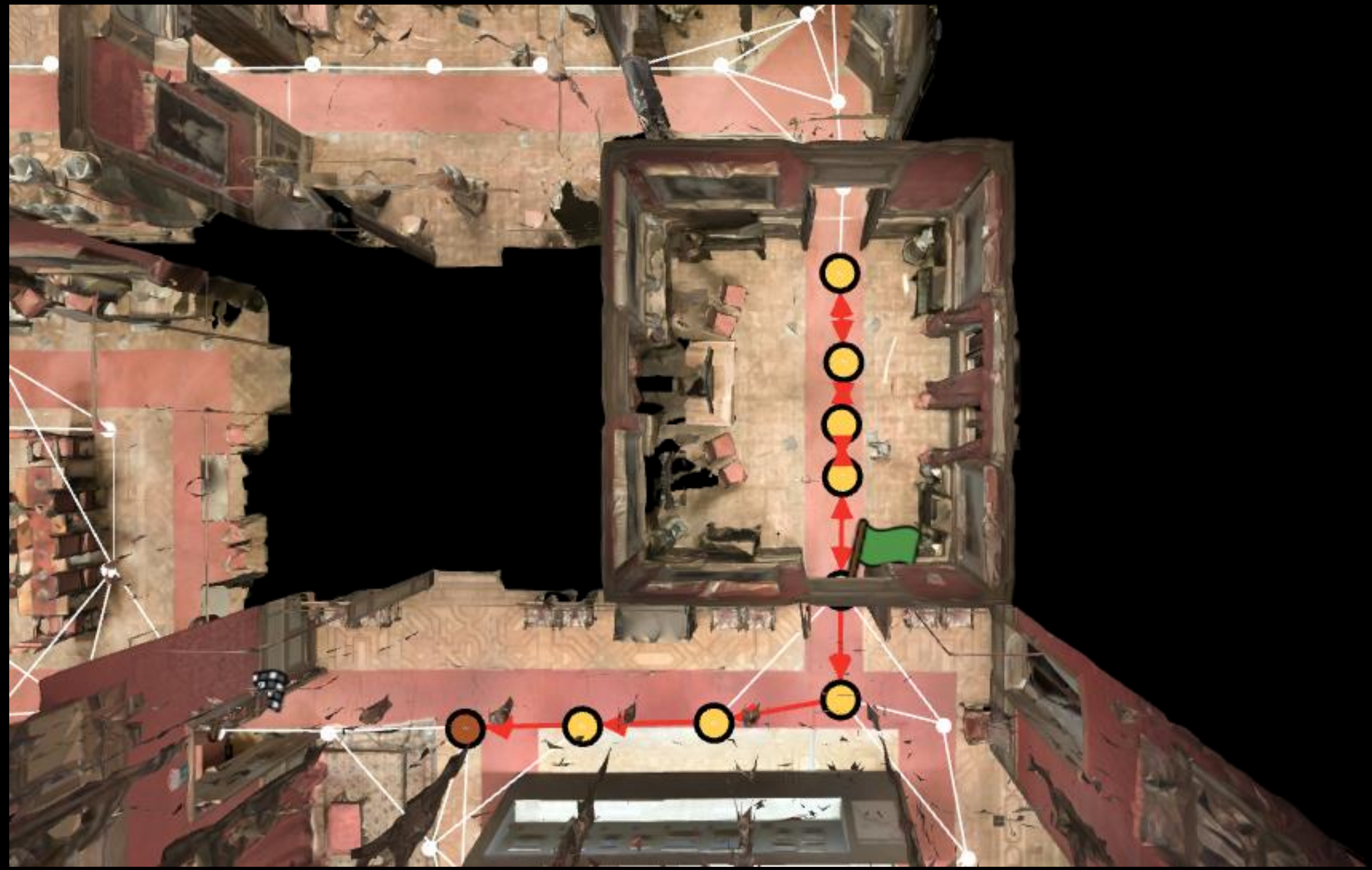
Instruction: Exit the roped off hall, follow the red carpet, **turn right**, continue straight down the red carpet, enter room at the end, stop once inside the room.



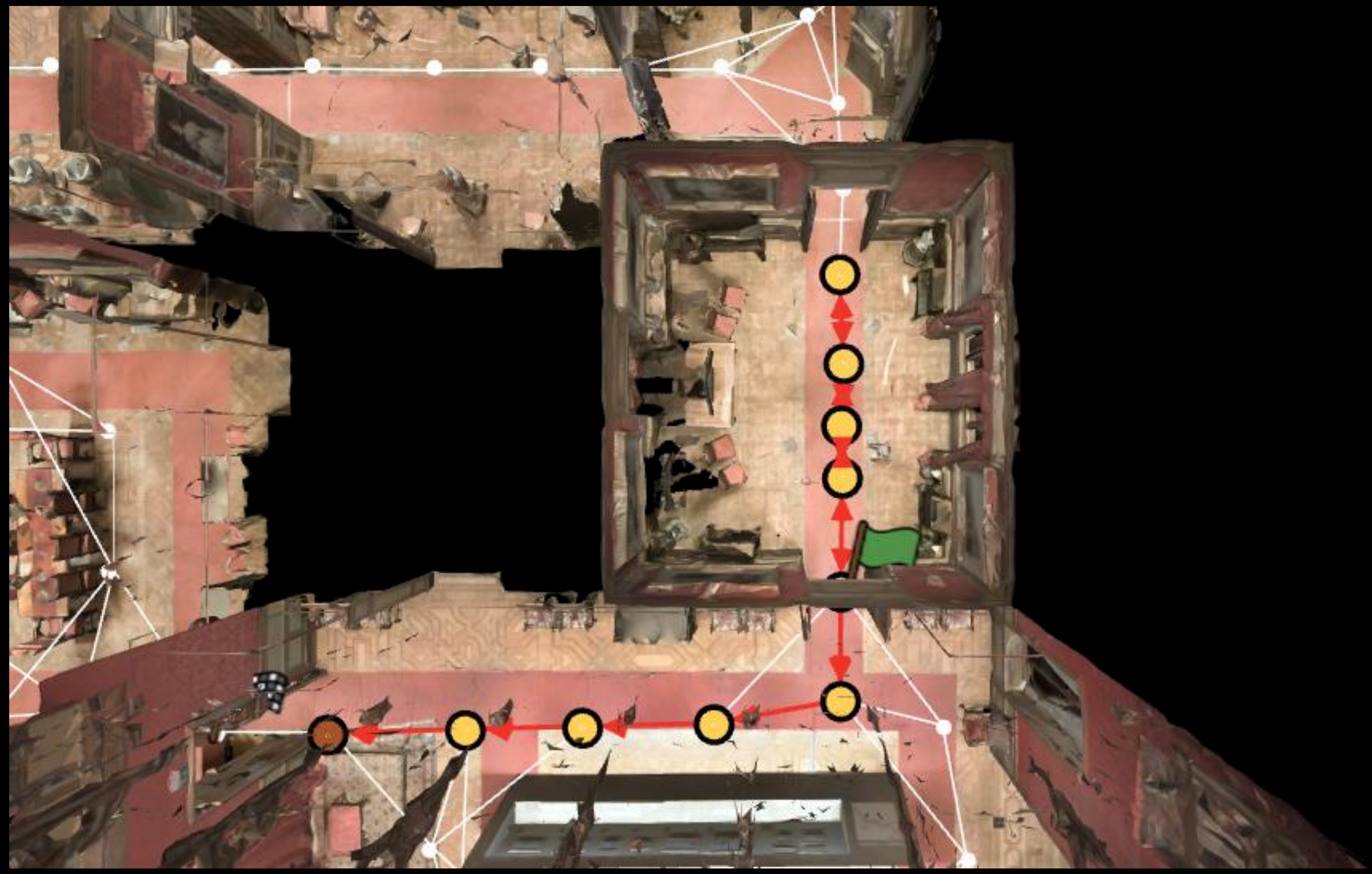
Instruction: Exit the roped off hall, follow the red carpet, turn right, **continue straight down the red carpet,** enter room at the end, stop once inside the room.



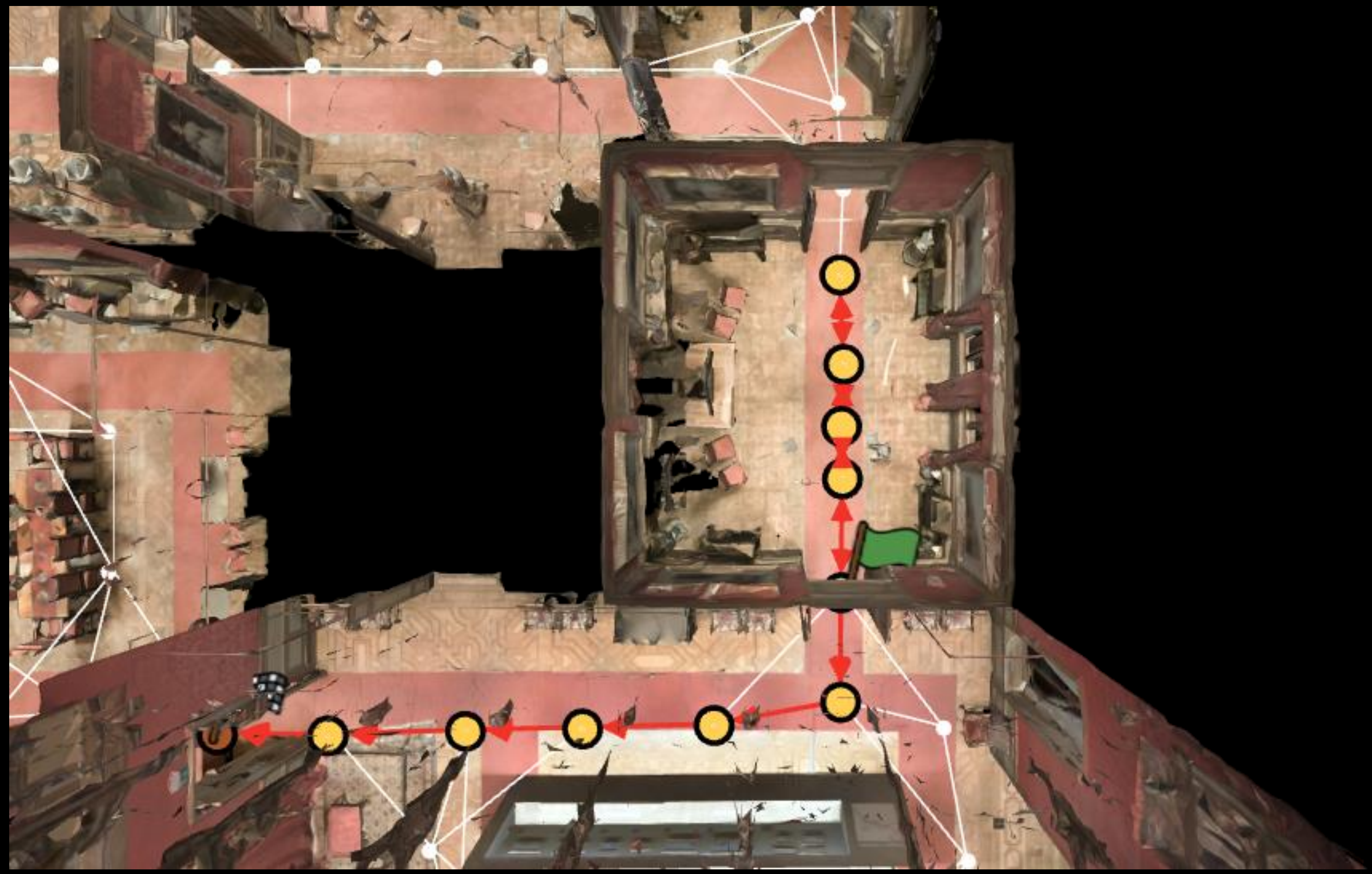
Instruction: Exit the roped off hall, follow the red carpet, turn right, **continue straight down the red carpet,** enter room at the end, stop once inside the room.



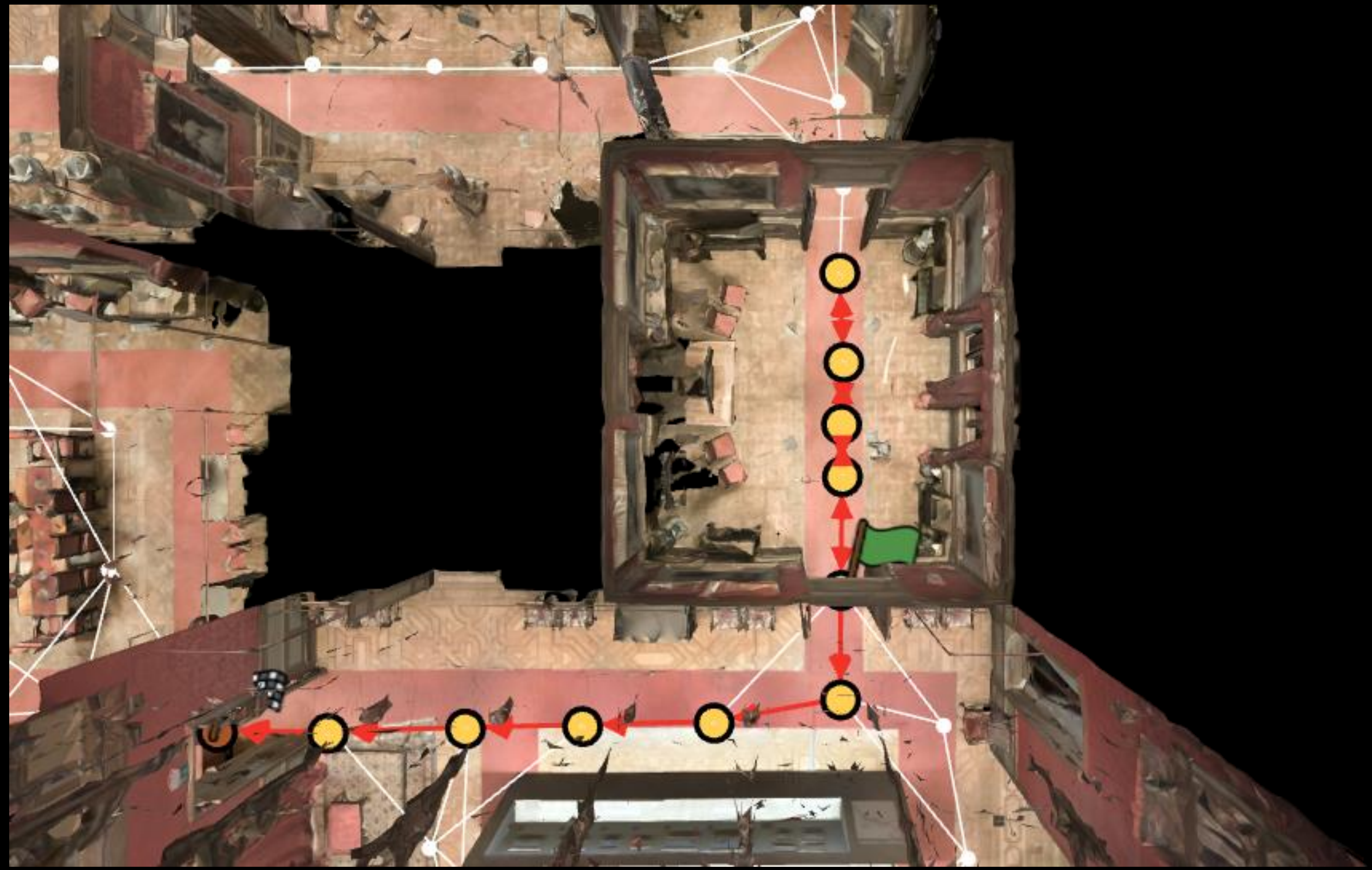
Instruction: Exit the roped off hall, follow the red carpet, turn right, **continue straight down the red carpet,** enter room at the end, stop once inside the room.



Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, **enter room at the end, stop once inside the room.**



Instruction: Exit the roped off hall, follow the red carpet, turn right, continue straight down the red carpet, **enter room at the end, stop once inside the room.**



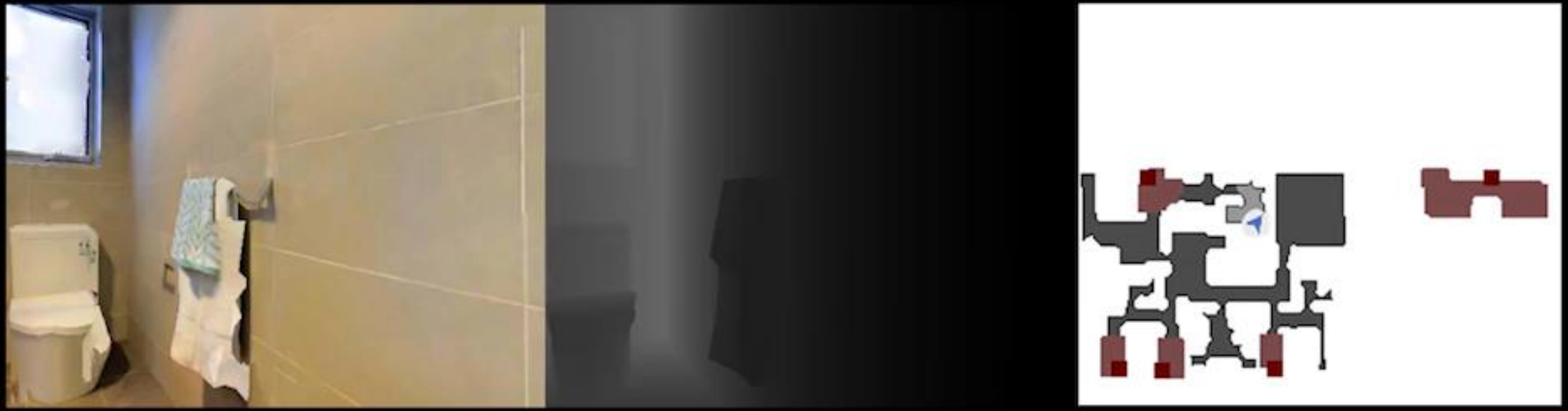
Object Goal Navigation with Recursive Implicit Maps

Shizhe Chen, Thomas Chabal, Ivan Laptev and Cordelia Schmid

Examples in simulation: successful cases

Target: *"cabinet"*

Target: *"chest of drawer"*



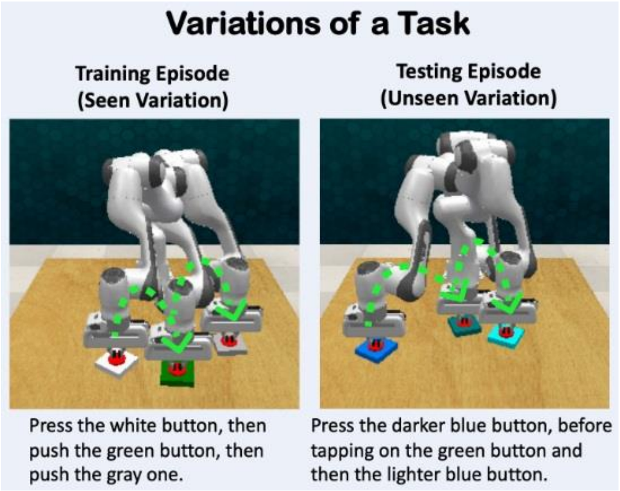
Real world examples

Summary

History Aware Multimodal Transformer for Vision-and-Language Navigation, S. Chen, P.-L. Guhur, C. Schmid and I. Laptev; *in Proc. NeurIPS 2021*

Object Goal Navigation with Recursive Implicit Maps, S. Chen, T. Chabal, I. Laptev and C. Schmid; *In Proc. IROS 2023*

Vision and language **navigation**



Instruction-driven history-aware policies for robotic manipulations, P.-L. Guhur, S. Chen, R. Garcia, M. Tapaswi, I. Laptev and C. Schmid; *in Proc. CoRL 2022*

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation, S. Chen, R. Garcia, C. Schmid and I. Laptev; *In Proc CoRL 2023*

Vision and language **manipulation**

Instruction-driven History-aware Policies for Robotic Manipulation


Pierre-Louis Guhur


Shizhe Chen


Ricardo Garcia Pinel


Makarand Tapaswi


Ivan Laptev

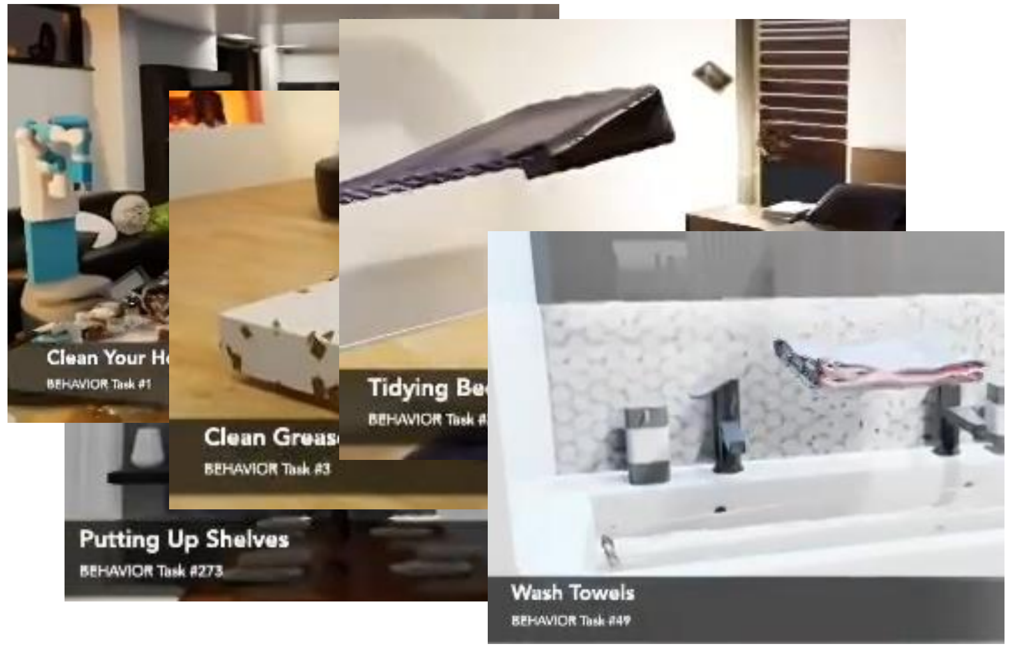

Cordelia Schmid

CoRL 2022

Project page: https://guhur.github.io/hiveformer/

Challenges

1.

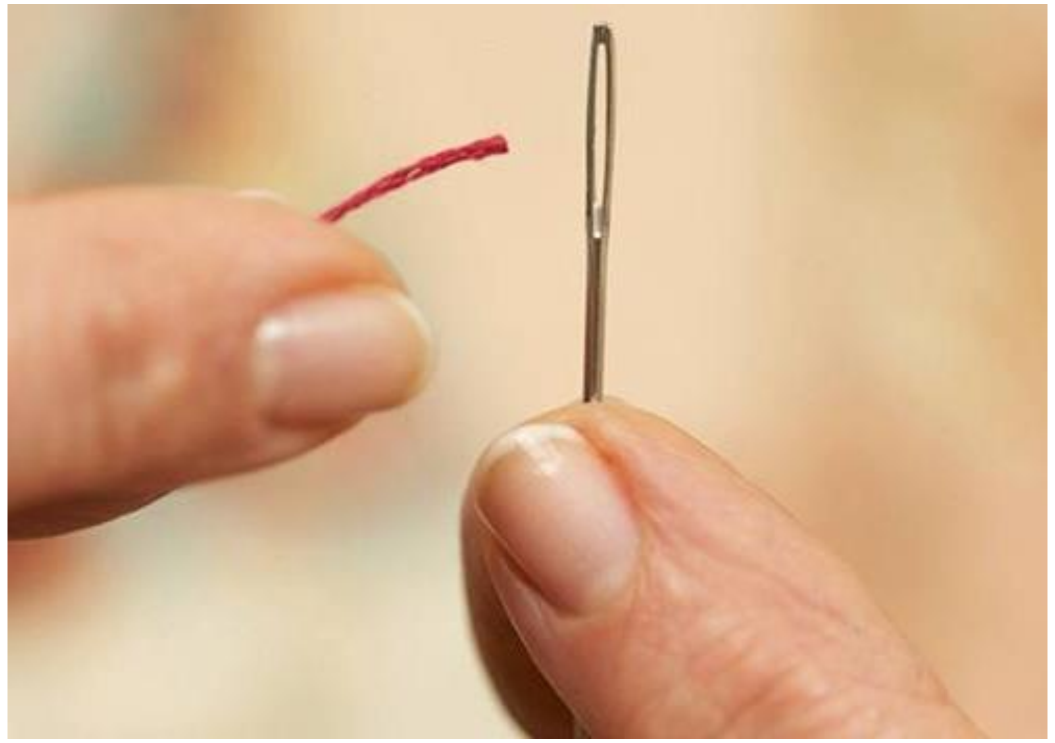


Many tasks and their variations

2.

Current observation is insufficient

3.



Precision can be crucial

4.

Explicit state recovery is too difficult

How to address these challenges?

1.

Define tasks by language, e.g.
Use the broom to brush the dirt into the dustpan

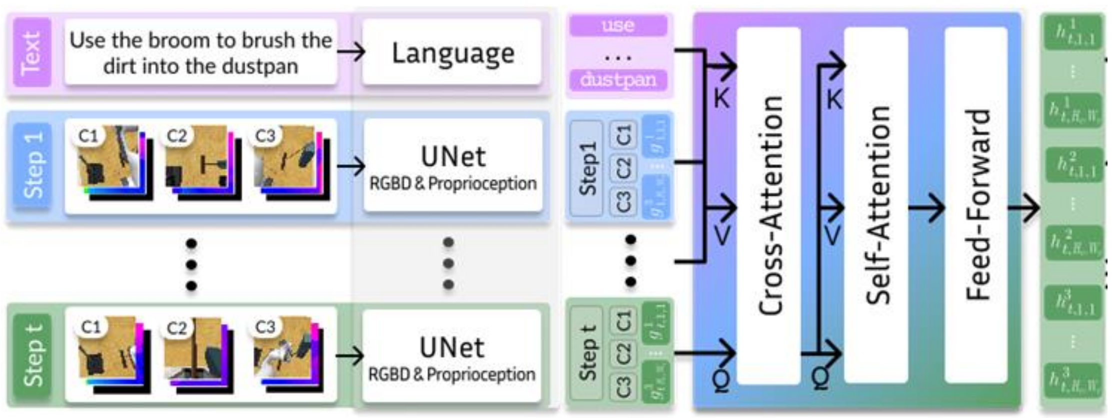
Many tasks and their variations

2.

Encode explicit observation history

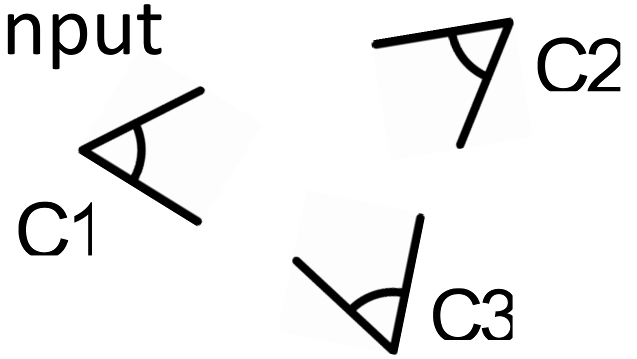
Curre
obser
insufficient

Transformer



3.

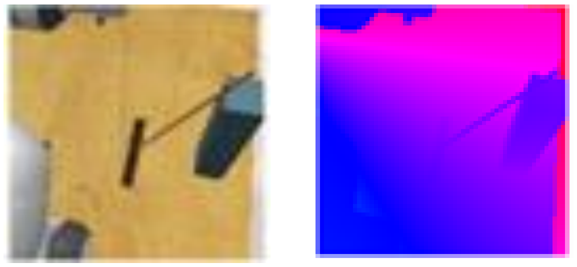
Use multi-camera input



Precision can be crucial

4.

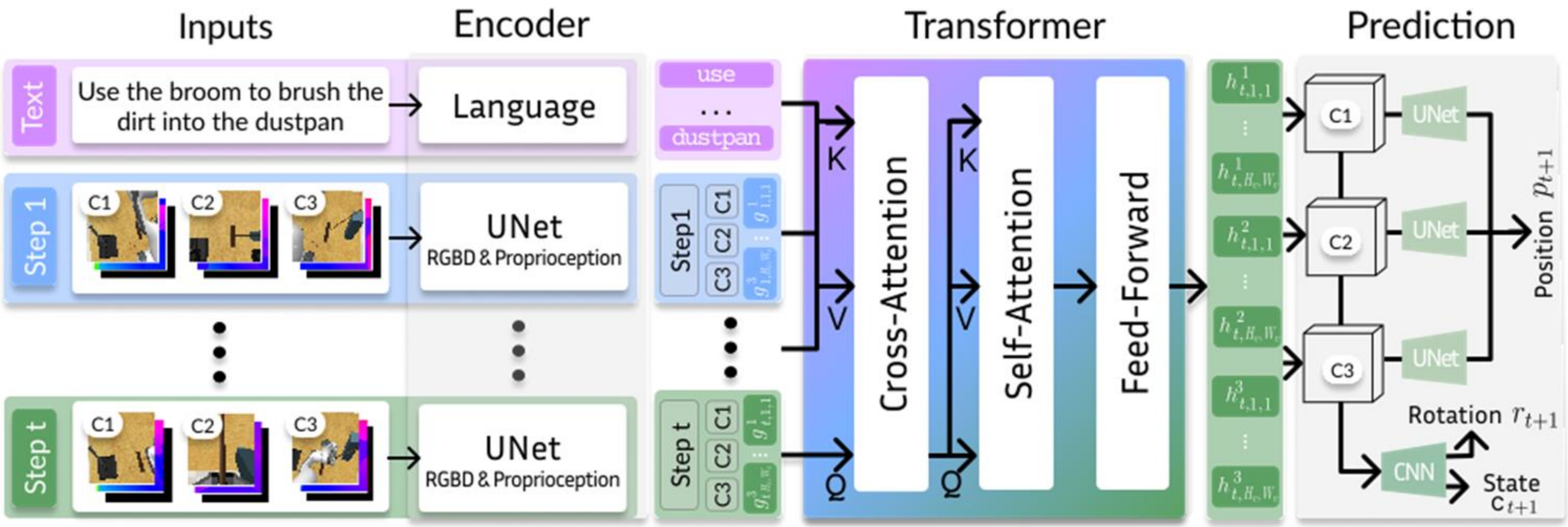
Use raw RGB+D for visuomotor policies



state
recovery is
too difficult

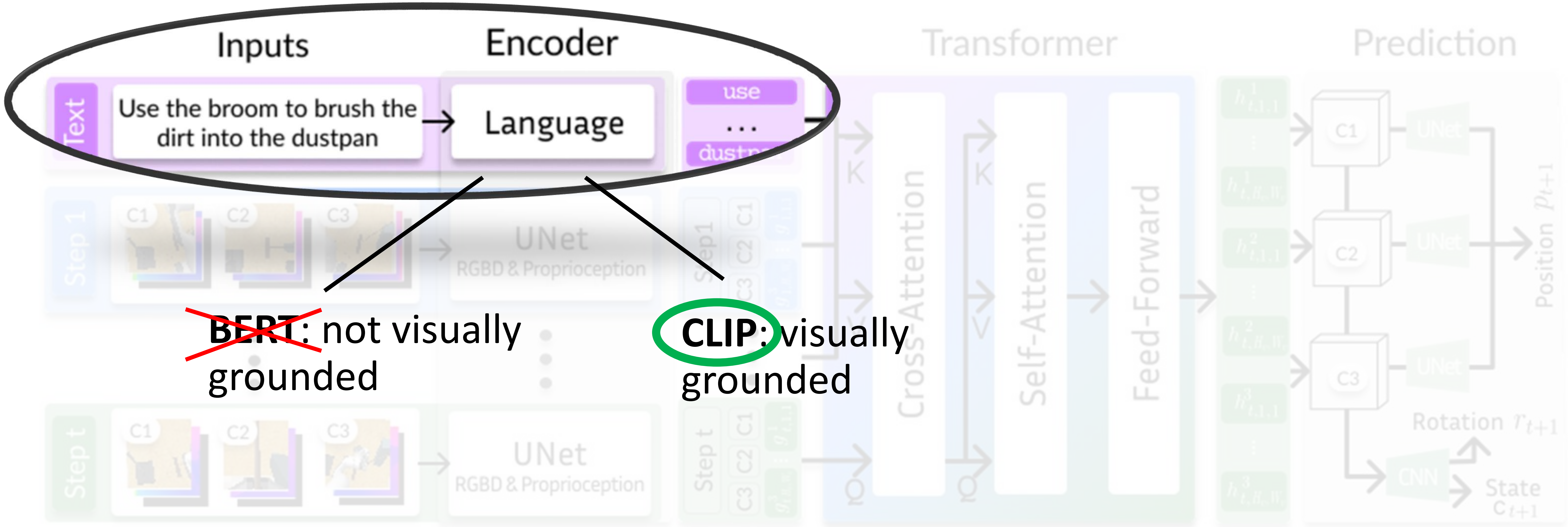
HiveFormer

History-aware instruction-conditioned multi-view transformer



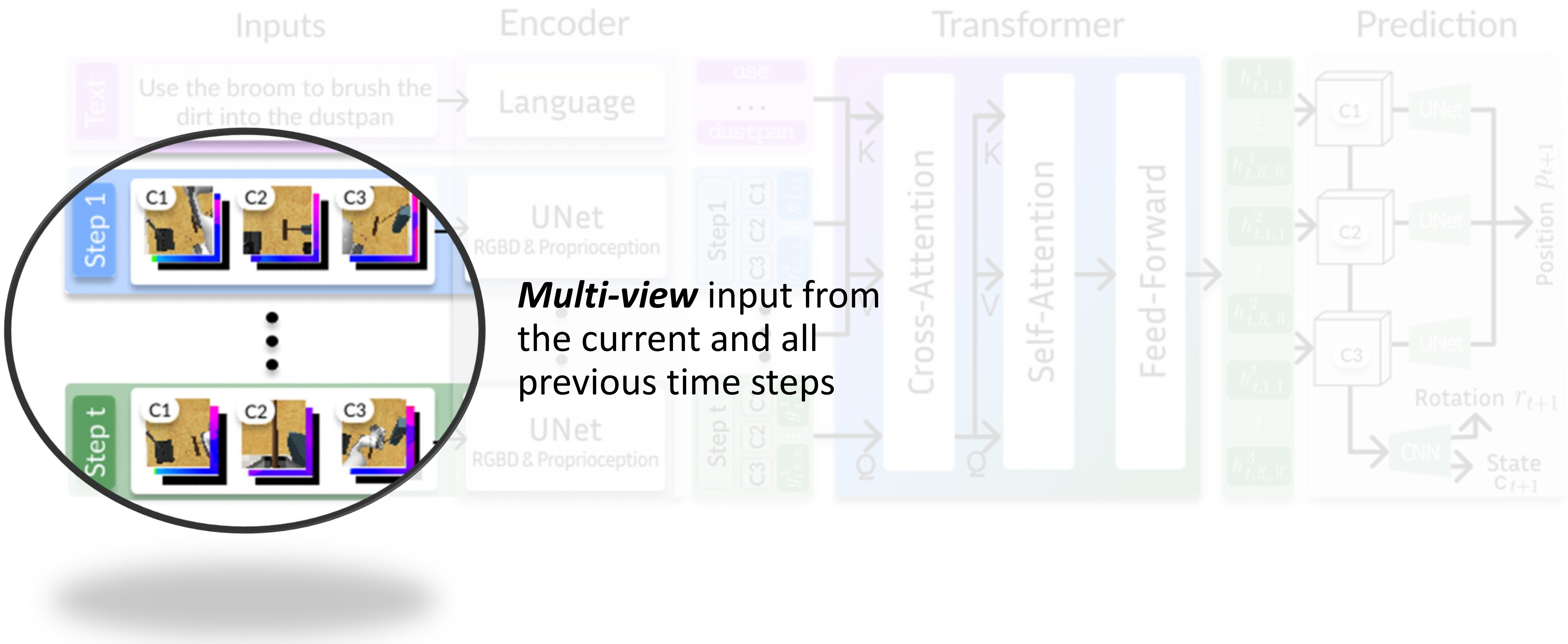
HiveFormer

History-aware instruction-conditioned multi-view transformer



HiveFormer

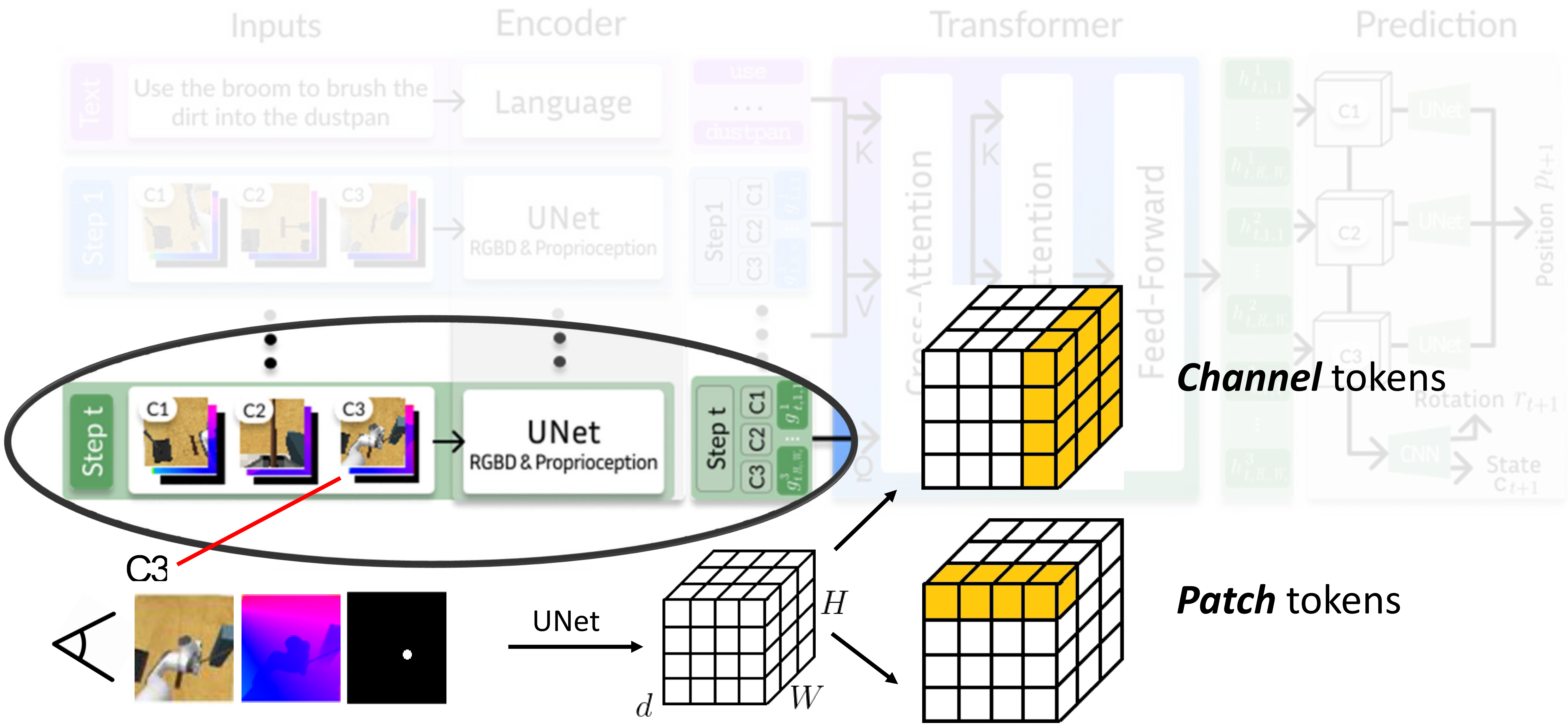
History-aware instruction-conditioned multi-view transformer



Multi-view input from the current and all previous time steps

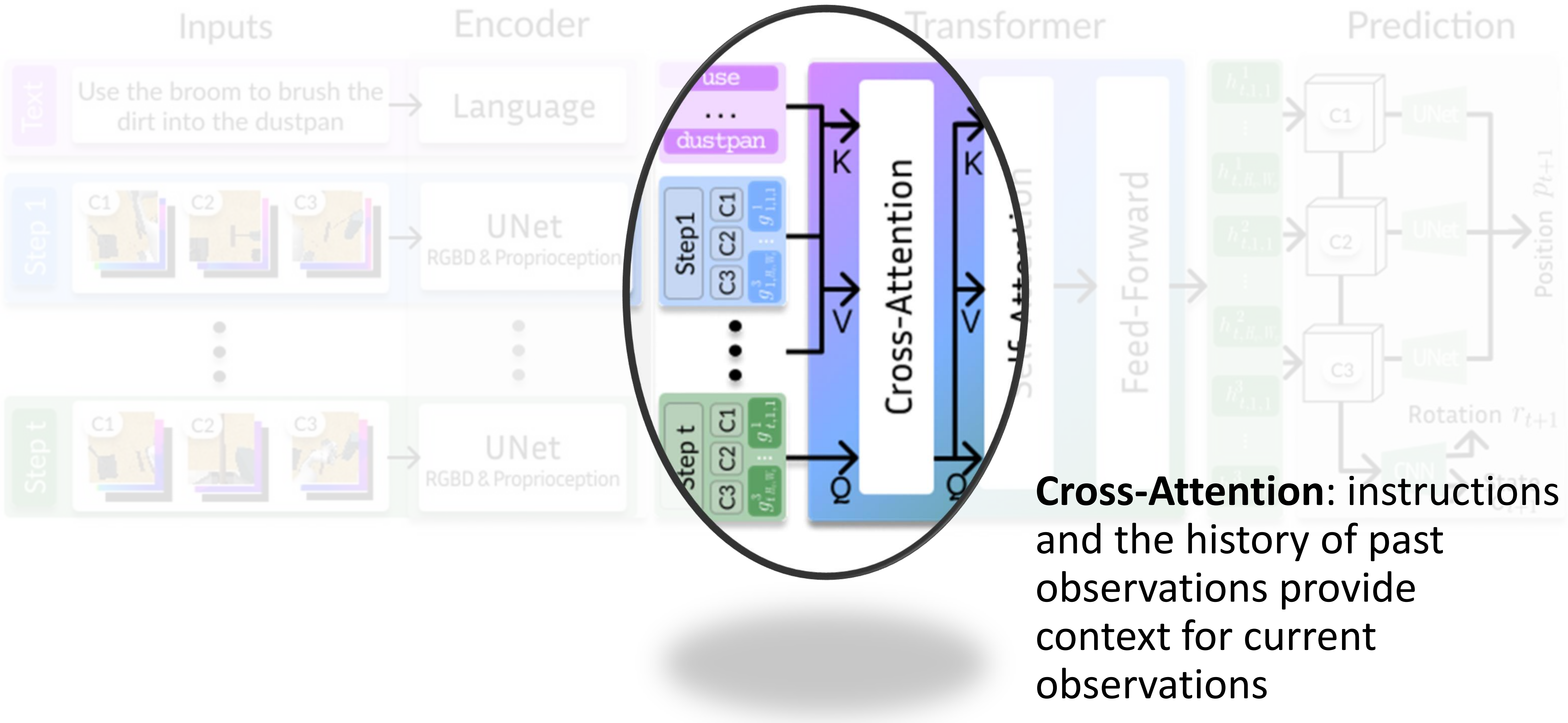
HiveFormer

History-aware instruction-conditioned multi-view transformer



HiveFormer

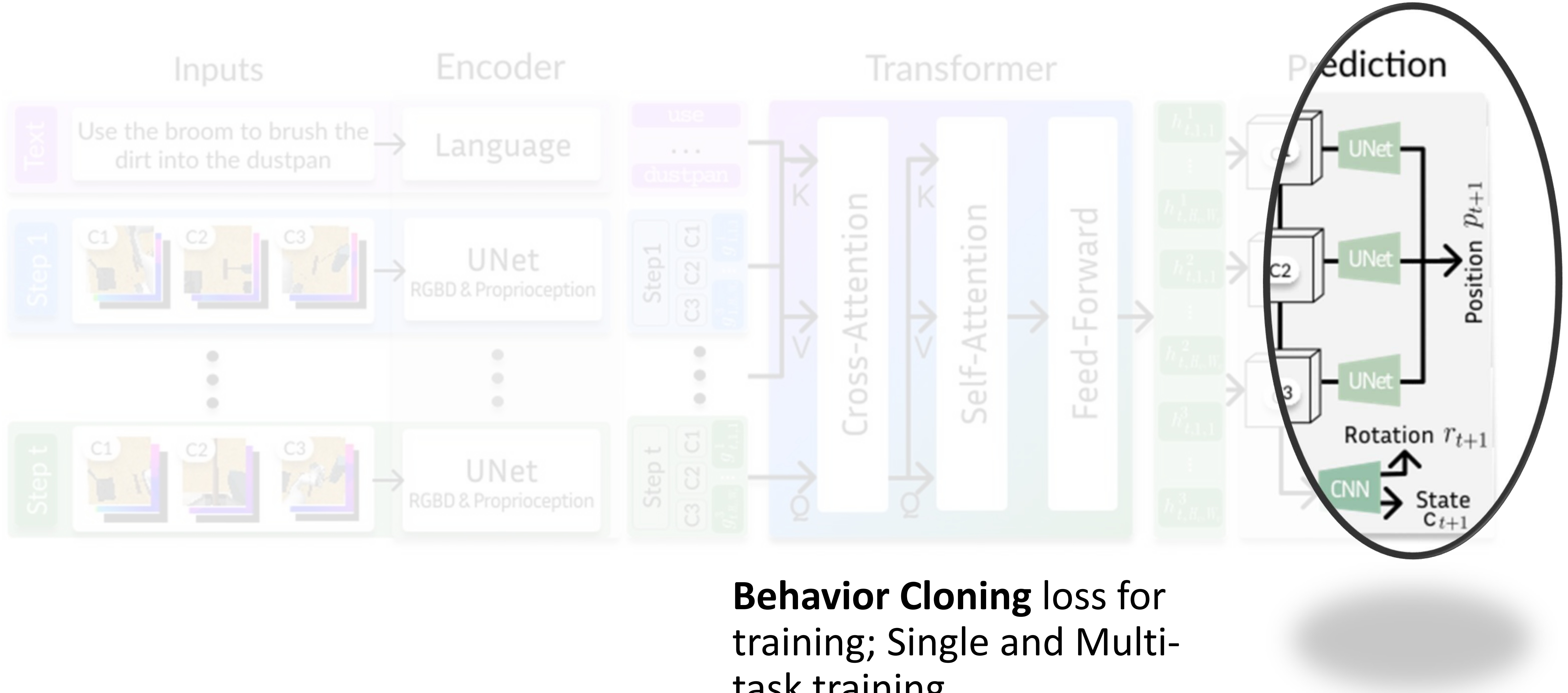
History-aware instruction-conditioned multi-view transformer



Cross-Attention: instructions and the history of past observations provide context for current observations

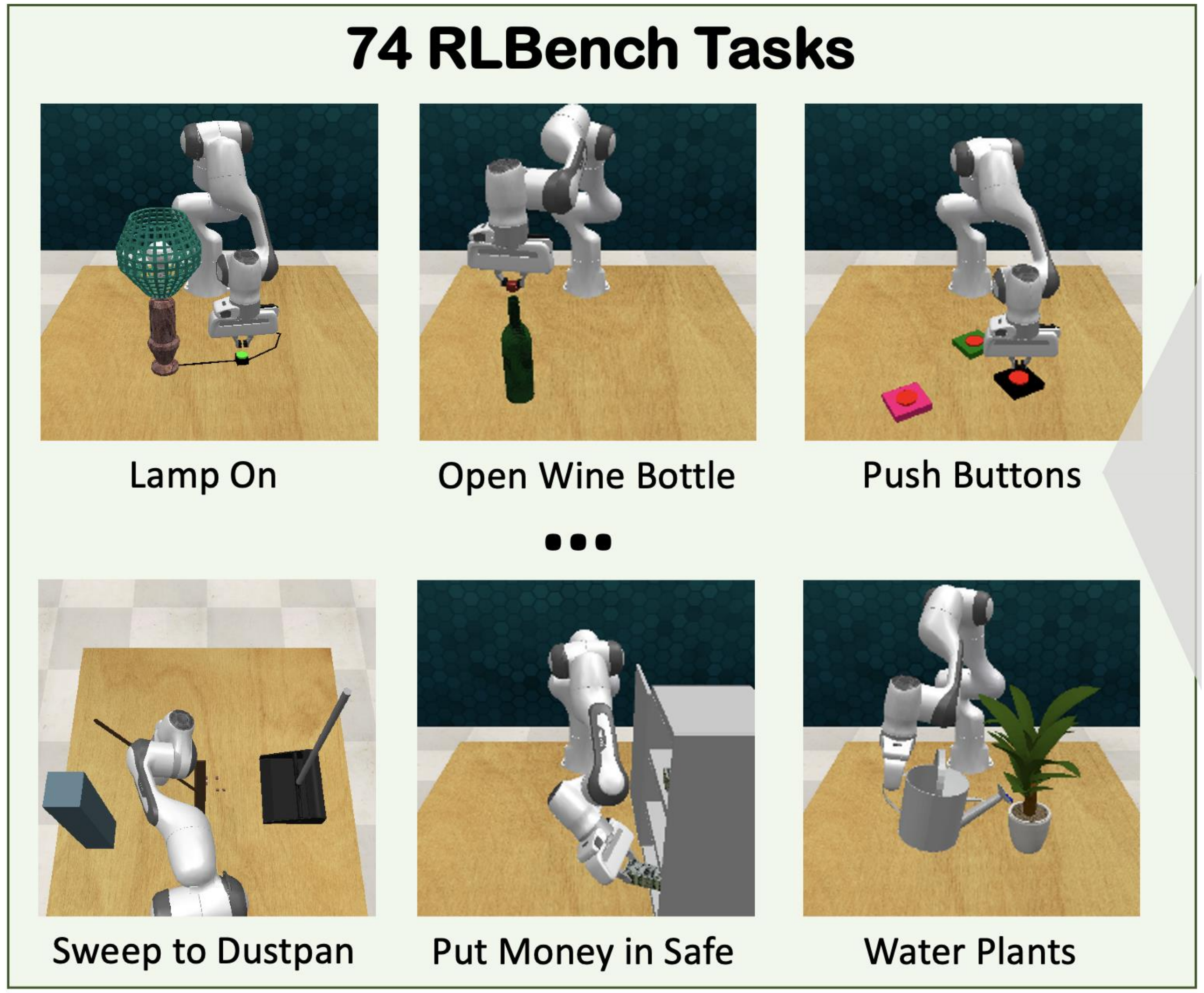
HiveFormer

History-aware instruction-conditioned multi-view transformer



Behavior Cloning loss for training; Single and Multi-task training

Evaluation: RLBench tasks



- 100 hand-designed tasks
- Multi-view RGB-D images
  Franka Emika Panda 7 DoF arm
- Text description for each task

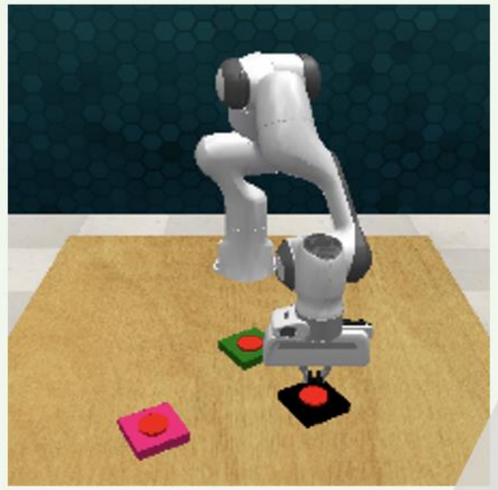
⇨ Select 74 tasks we could simulate

⇨ Evaluate in single and multi-task settings

(Task text descriptions are not needed)

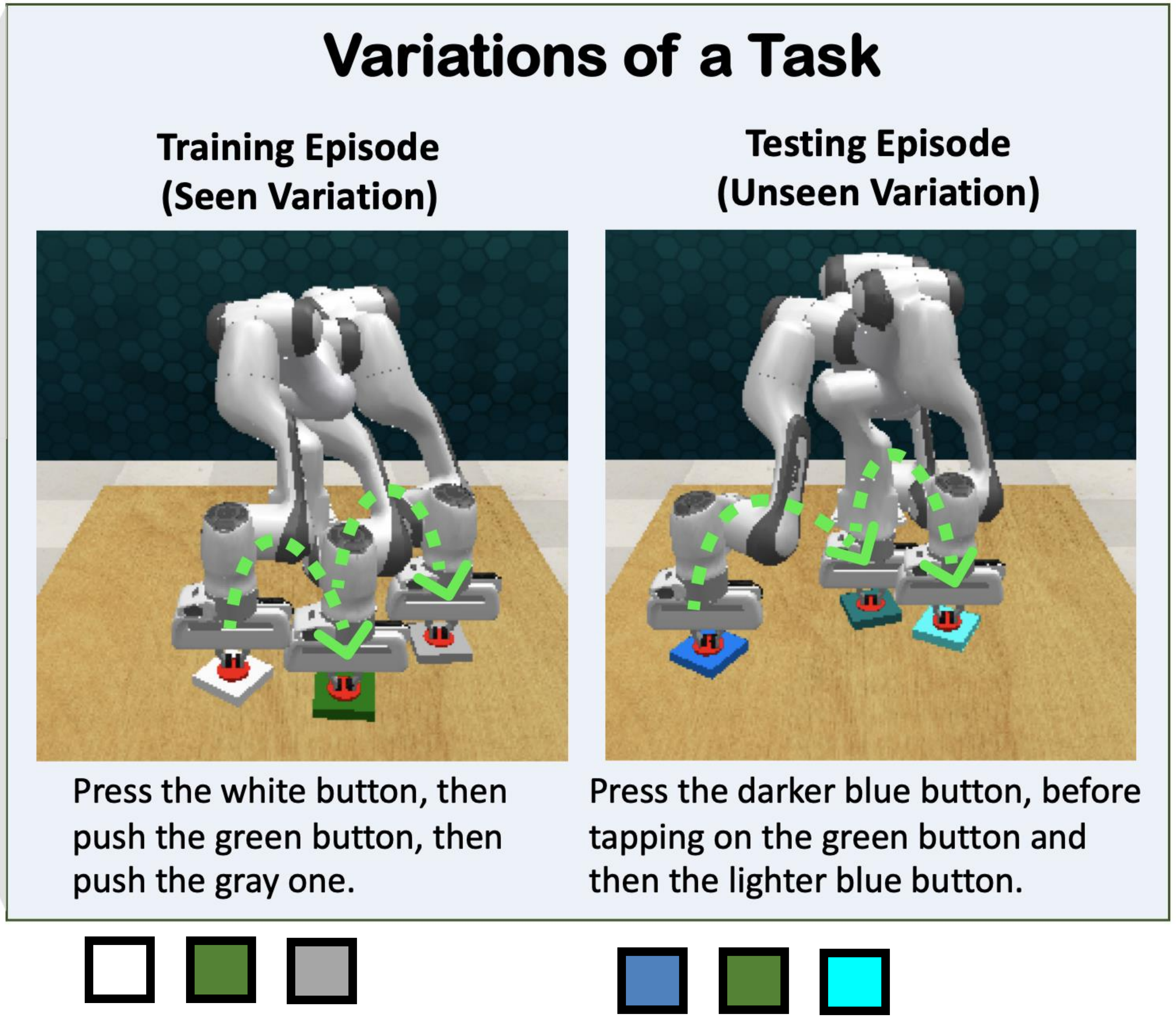
James, S., Ma, Z., Arrojo, D. R., & Davison, A. J. (2020). RLBench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, *5*(2), 3019-3026.

Evaluation: RLBench task variations



Push Buttons

Water Plants



Unseen sequence of colors during training

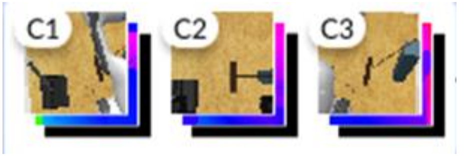
- Evaluate on *unseen task variations*
- Task text descriptions become crucial

Results: 10 tasks • Single-task setting

	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

Results: 10 tasks • Single-task setting

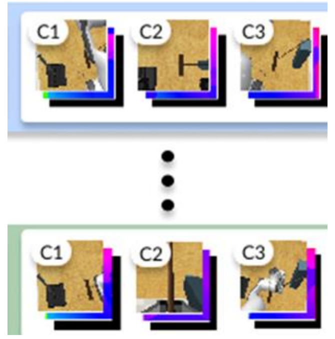


	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

+5.2%

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

Results: 10 tasks • Single-task setting

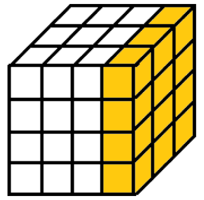


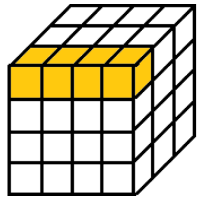
	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

+3.7%

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

Results: 10 tasks • Single-task setting



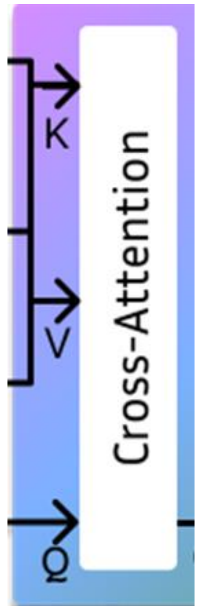


	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

+2.1%

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

Results: 10 tasks • Single-task setting

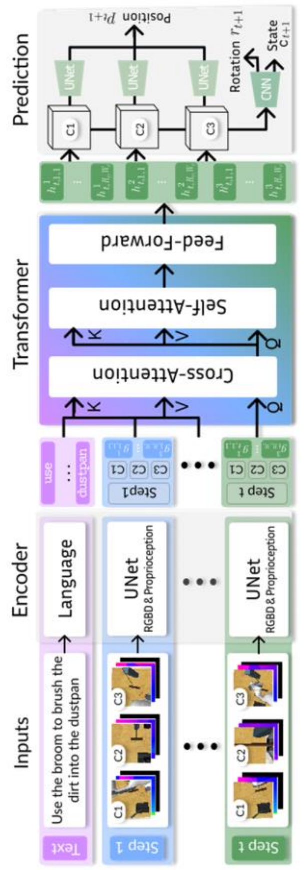


	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

+4%

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

Results: 10 tasks • Single-task setting

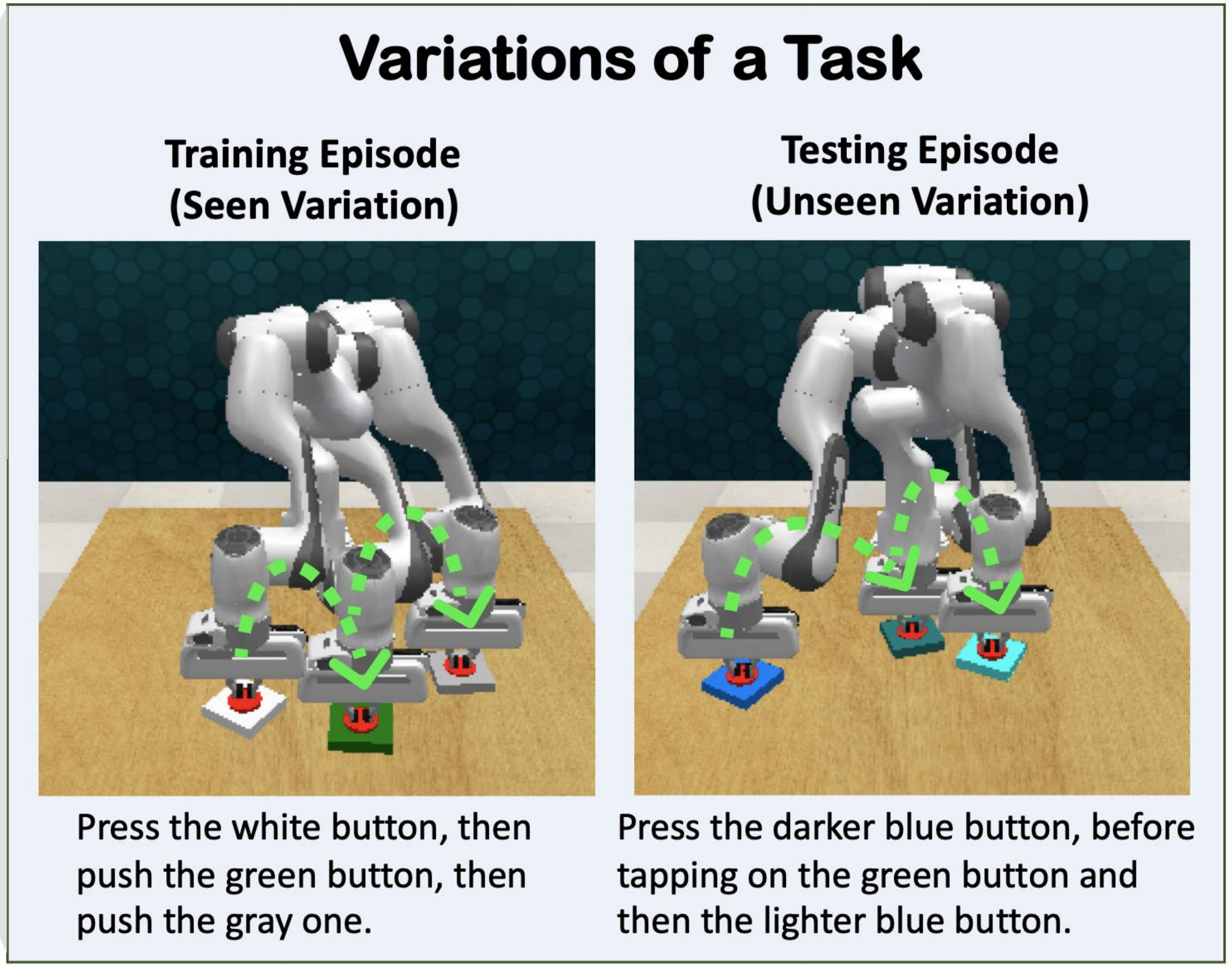


	Visual Tokens	Point Clouds	Gripper Position	Multi-View	History	Attn	Mask Obs	SR
R1	×	×	×	×	×	×	×	72.9 ± 4.1
R2	Channel	×	×	✓	×	Self	×	73.1 ± 4.5
R3	Channel	✓	×	✓	×	Self	×	77.1 ± 5.8
R4	Channel	✓	✓	✓	×	Self	×	78.1 ± 5.8
R5	Channel	✓	✓	✓	✓	Self	×	81.8 ± 5.2
R6	Channel	✓	✓	✓	✓	Self	✓	82.3 ± 5.3
R7	Patch	✓	✓	✓	✓	Self	✓	84.4 ± 6.4
R8	Patch	✓	✓	✓	✓	Cross	✓	88.4 ± 4.9

+15.5%

- Transformer with multi-view, depth and gripper: +5.2%
- w/ vs. w/o history: +3.7%
- Patch vs. channel tokens: +2.1%
- Cross- vs. Self-Attention: +4%
- Overall: +15.5%

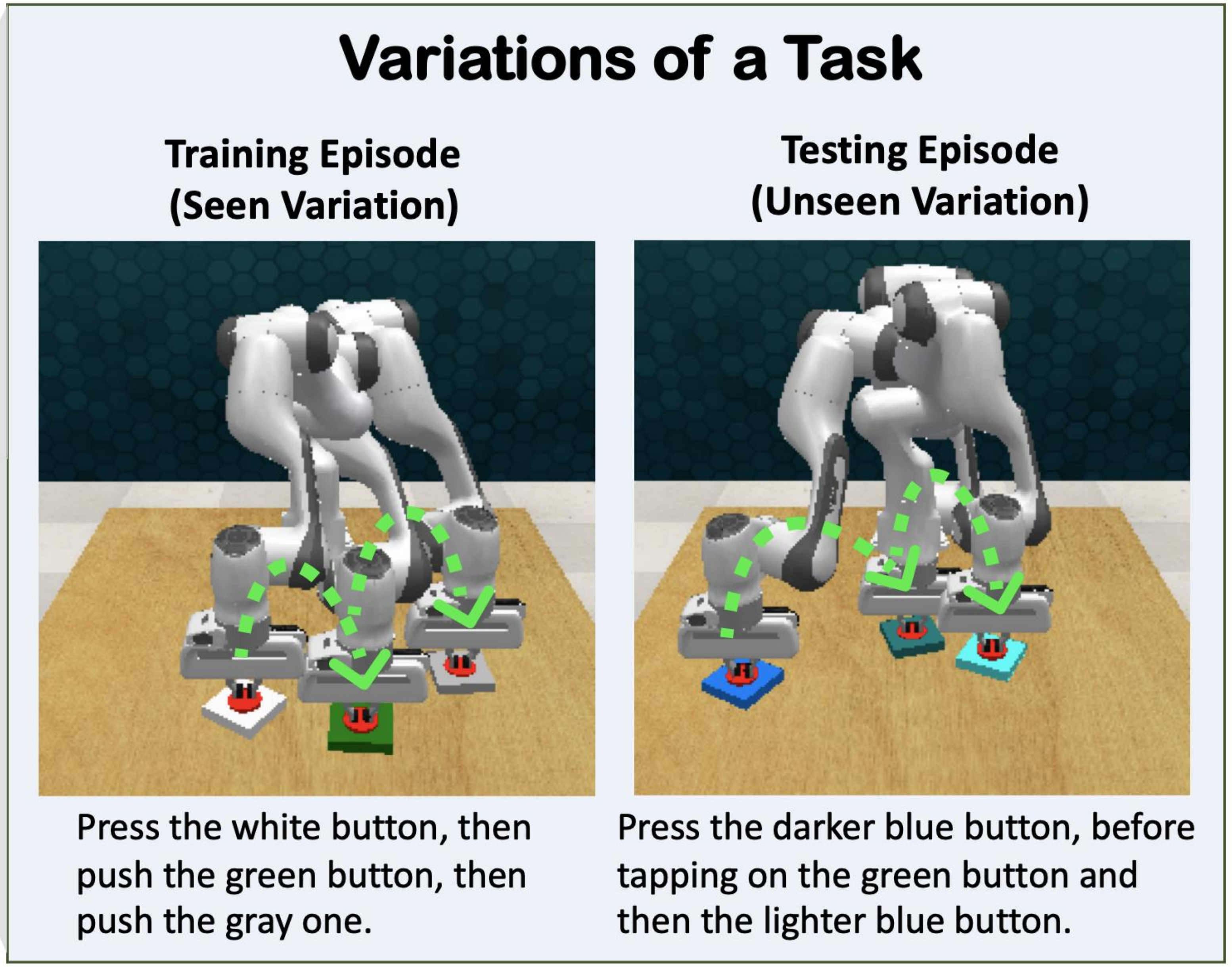
Results: Task variations



# Demos Per Variation	Instr.	Push Buttons Seen Synt.	Push Buttons Unseen Synt.	Push Buttons Unseen Real	Tower Seen Synt.	Tower Unseen Synt.	Tower Unseen Real
10	Seq.	96.4	71.1	65.7	71.6	49.8	19.4
50	Seq.	99.4	83.1	70.9	74.3	52.1	20.6
100	Seq.	100	86.3	74.2	77.4	56.2	24.1

- Generalization to unseen variations
- Generalization to natural language extractions

Results: Task variations



# Demos Per Variation	Instr.	Push Buttons Seen Synt.	Push Buttons Unseen Synt.	Push Buttons Unseen Real	Tower Seen Synt.	Tower Unseen Synt.	Tower Unseen Real
10	Seq.	96.4	71.1	65.7	71.6	49.8	19.4
50	Seq.	99.4	83.1	70.9	74.3	52.1	20.6
100	Seq.	100	86.3	74.2	77.4	56.2	24.1

- Generalization to unseen variations
- Generalization to natural language expressions

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation


Ricardo Garcia Pinel


Shizhe Chen

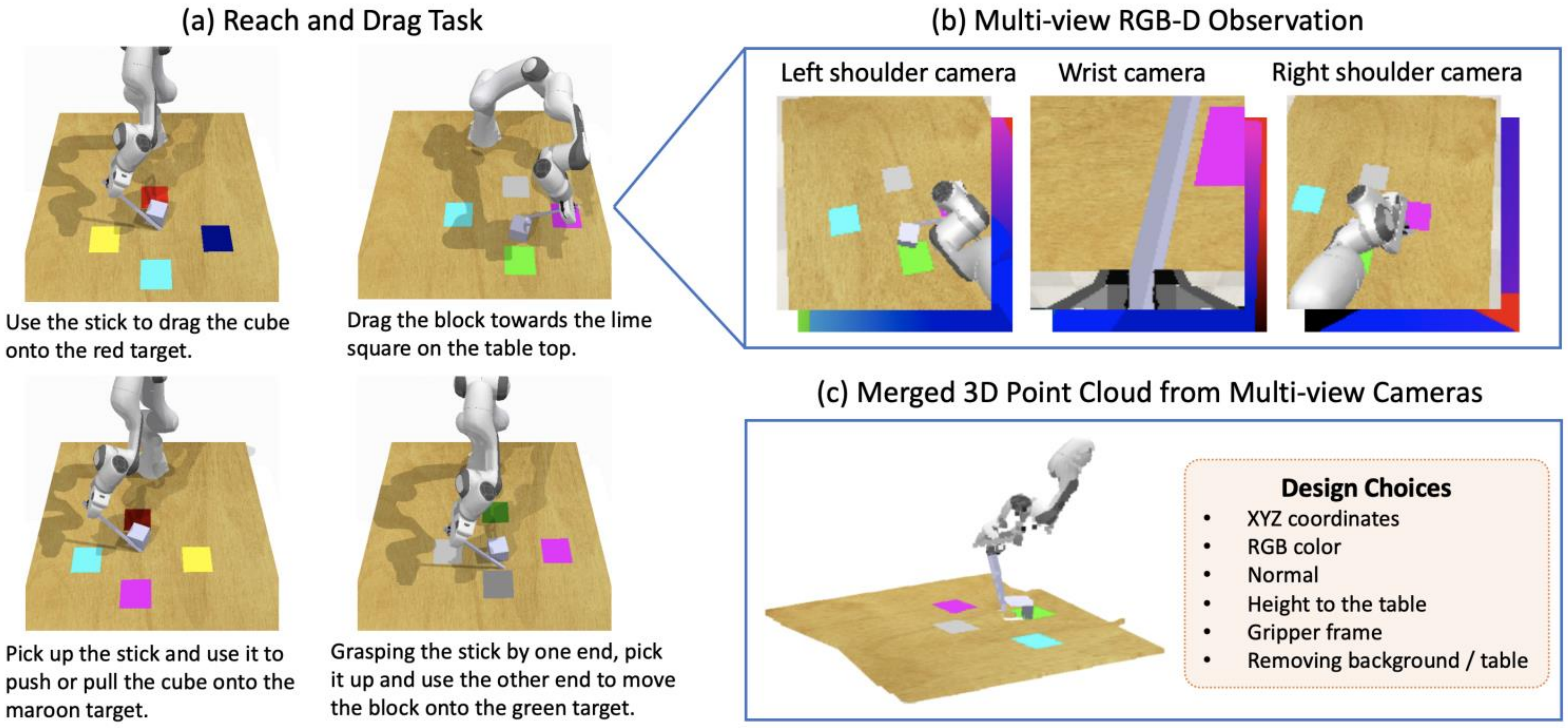

Cordelia Schmid


Ivan Laptev

CoRL 2023

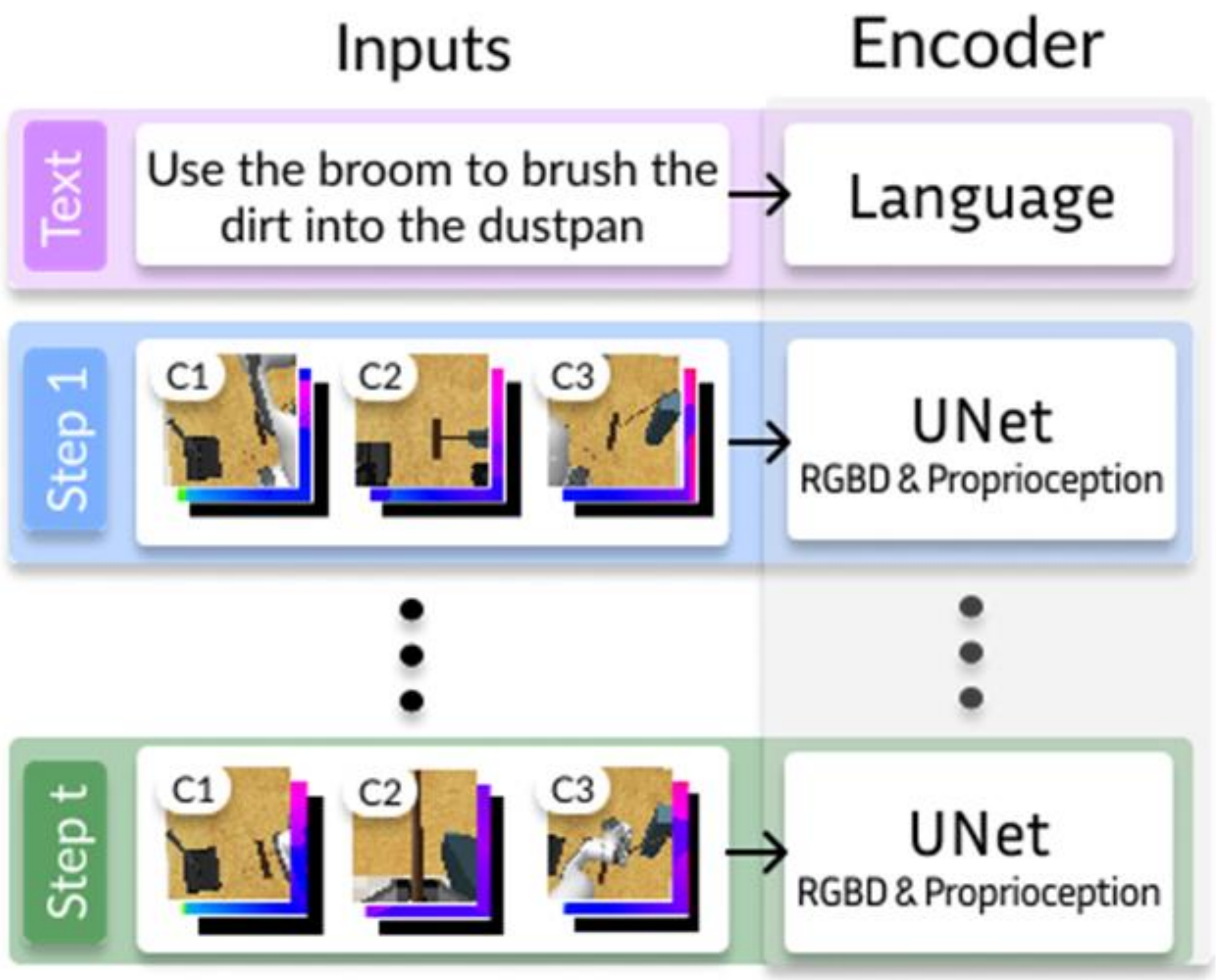
Project page: https://www.di.ens.fr/willow/research/polarnet

Example task: Reach and Drag



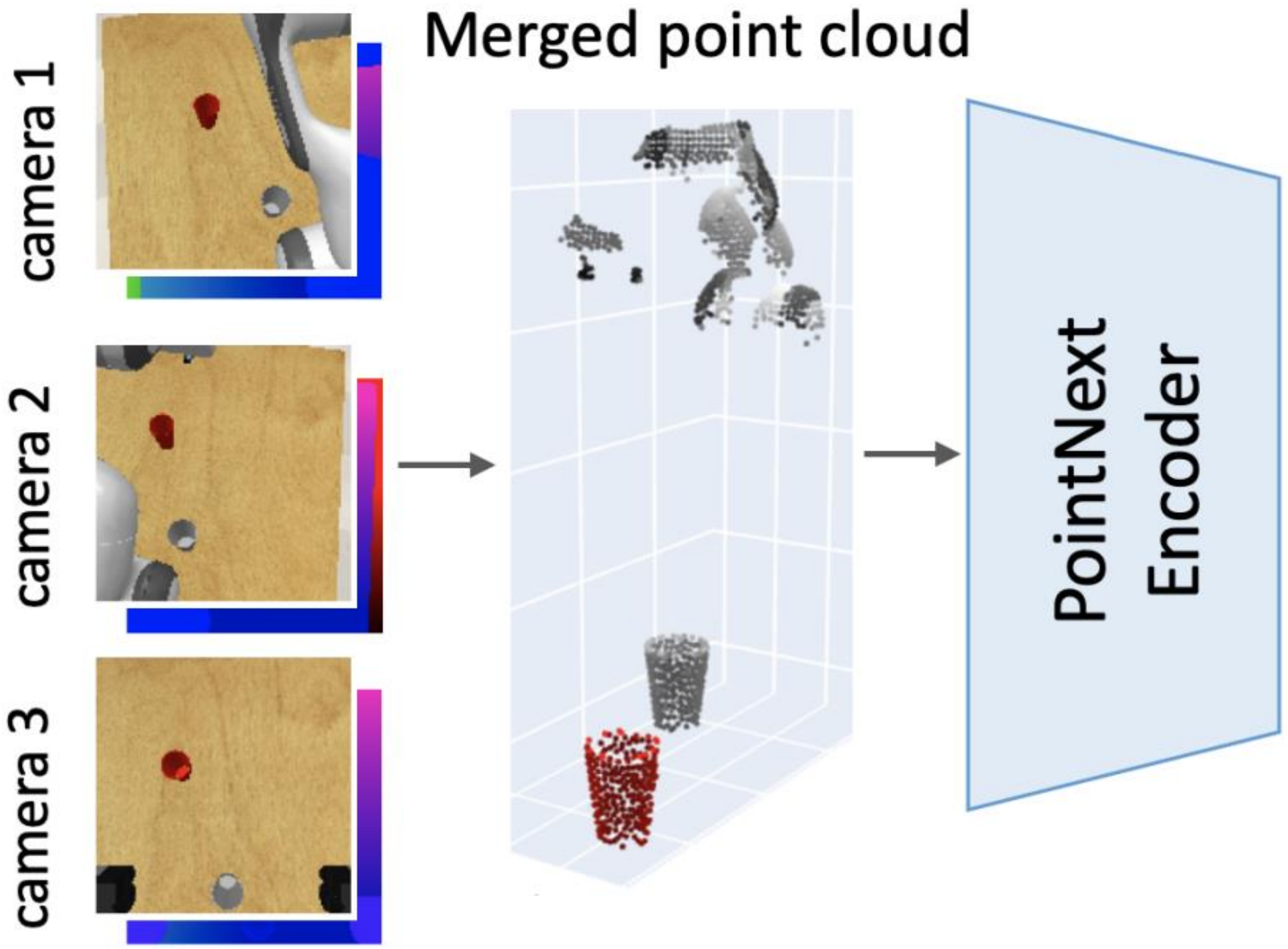
XYZ-RGB Point Cloud Representation

HiveFormer



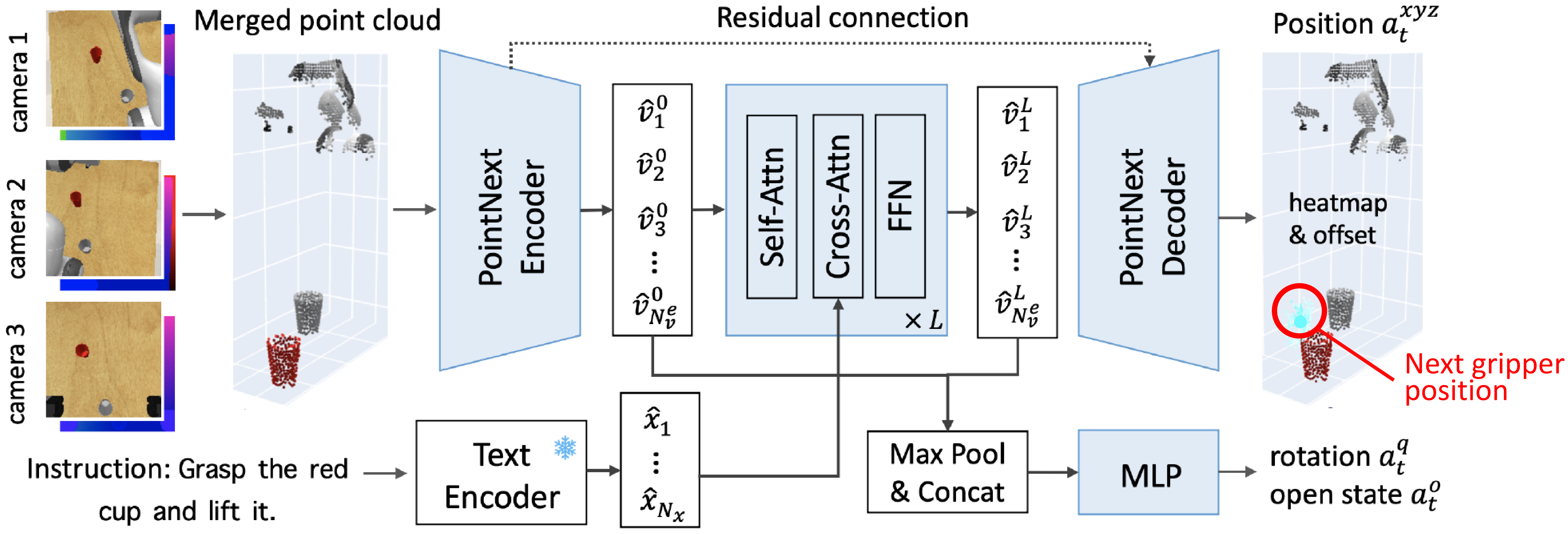
- 2.5D image representation
- Multiple views encoded independently
- History encoding

PolarNet (new)



- 3D color point cloud
- Multiple views are merged before encoding
- No history encoding

PolarNet: Point cloud-based language-guided robotic manipulation network



Training with Behavioral Cloning

PolarNet: Ablation

RGB	Normal	Height	Pick & Lift	Pick-Up Cup	Put Knife	Put Money	Push Button	Reach Target	Slide Block	Stack Wine	Take Money	Take Umbrella	Avg.
×	×	×	26.2	44.0	81.1	95.9	99.6	27.8	89.3	91.0	70.3	95.3	72.1 ±4.4
✓	×	×	97.9	94.7	79.5	95.8	100.0	100.0	91.0	91.1	65.9	97.3	91.3 ±1.6
✓	✓	×	94.9	94.2	77.1	95.9	90.3	100.0	93.1	94.1	69.4	94.4	90.3 ±3.1
✓	×	✓	96.2	94.3	82.6	95.3	100.0	99.9	91.5	90.3	67.5	97.7	91.5 ±1.4
✓	✓	✓	96.7	91.9	82.5	96.1	99.9	99.9	93.5	94.1	68.6	97.5	**92.1** ±0.4

+19%

By adding color to 3D point cloud

Left	Right	Wrist	Avg.
✓	×	×	37.6 ±4.8
×	✓	×	48.0 ±4.5
×	×	✓	35.0 ±5.5
✓	✓	×	67.0 ±4.7
✓	×	✓	80.2 ±3.0
×	✓	✓	76.6 ±5.6
✓	✓	✓	**92.1** ±0.4

1 camera

2 cameras

3 cameras

Remove Table	Remove Background	Avg.
×	×	81.6 ±3.2
×	✓	89.9 ±2.8
✓	✓	**92.1** ±0.4

+11%

By removing background

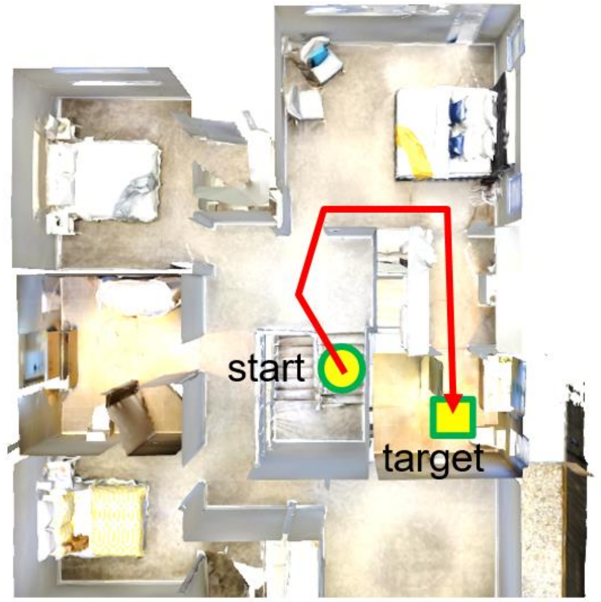
PolarNet: State-of-the-art comparison

	10 Tasks		74 Tasks	
	Single-task	Multi-task	Single-task	Multi-task
Auto-λ [2]	55.0	69.3	-	-
PerAct [14]	-	-	-	-
Hiveformer [3]	88.4	83.3	66.1	49.2
PolarNet (Ours)	**92.1**	**89.8**	**69.0**	**60.3**

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation

CoRL Submission #247

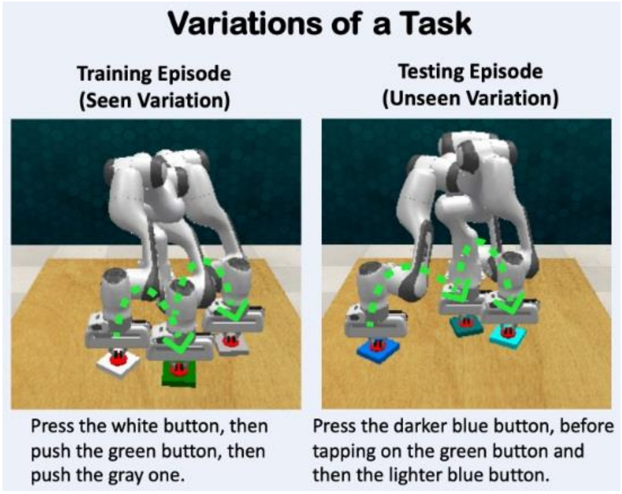
Summary



History Aware Multimodal Transformer for Vision-and-Language Navigation, S. Chen, P.-L. Guhur, C. Schmid and I. Laptev; *in Proc. NeurIPS 2021*

Object Goal Navigation with Recursive Implicit Maps, S. Chen, T. Chabal, I. Laptev and C. Schmid; *In Proc. IROS 2023*

Vision and language **navigation**



Instruction-driven history-aware policies for robotic manipulations, P.-L. Guhur, S. Chen, R. Garcia, M. Tapaswi, I. Laptev and C. Schmid; *in Proc. CoRL 2022*

PolarNet: 3D Point Clouds for Language-Guided Robotic Manipulation, S. Chen, R. Garcia, C. Schmid and I. Laptev; *In Proc CoRL 2023*

Vision and language **manipulation**

Vision, language and robotics

Goal: learn Embodied Vision-Language Models

Open challenges

What happens if…?

…shaking an apple tree

…pulling tablecloth

Open challenges

describe this image

The image features a man in a black jacket standing in a room, holding a white napkin. He is positioned near a dining table with a vase of roses on it. The vase is filled with several roses, creating a beautiful and elegant display.

In addition to the roses, there is a bottle of wine placed on the table, and a wine glass is also present. The man appears to be in the process of setting the table for a special occasion or a romantic dinner.

GPT-4V

what is the man trying to do?

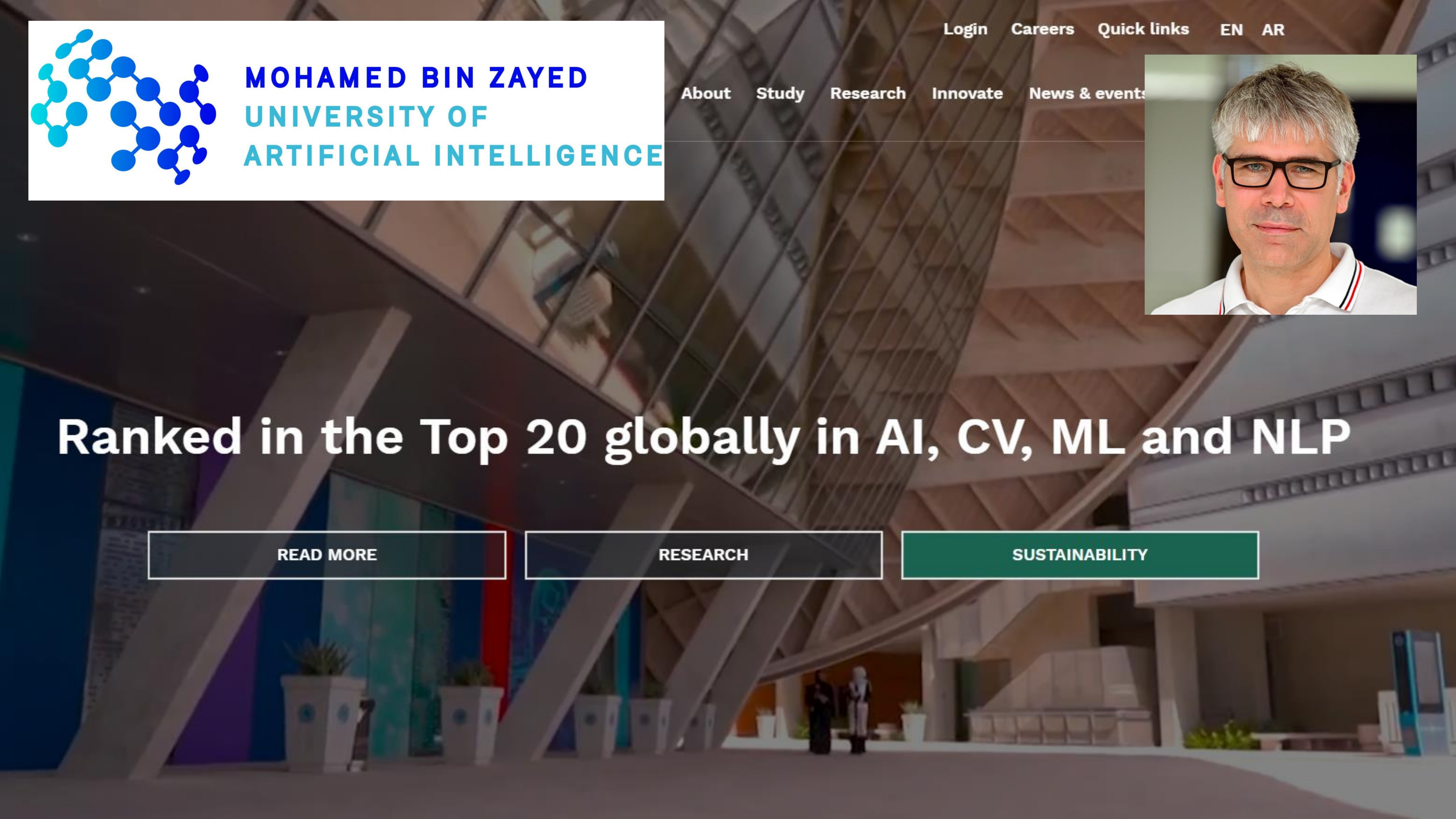
The man is trying to set the table for a special occasion or a romantic dinner, as he is holding a napkin and arranging items on the table, such as the vase of roses and the wine bottle.

Open challenges

What will happen if people in this image will turn and walk left?

GPT-4V

If the people in the image turn and walk left, they will be heading towards the bridge that crosses the river. This suggests that they might be planning to cross the river using the bridge, possibly to explore the other side or reach a specific destination. The presence of backpacks among the group indicates that they might be tourists or travelers, so crossing the river could be part of their


MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE
Login
Careers
Quick links
EN
AR
About
Study
Research
Innovate
News & events
Ranked in the Top 20 globally in AI, CV, ML and NLP
READ MORE
RESEARCH
SUSTAINABILITY

MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE

Building a new lab for Embodied and Language-Aware Visual Models

- Internships are available
- PhD application is open
- Competitive Internship and PhD salaries
- Departments of CV, NLP, ML, Robotics

Contact: Ivan.Laptev@mbzuai.ac.ae