PALMER: Perception-Action Loop with Memory for Long-Horizon Planning

Onur Beker, Mohammad Mohammadi, Amir Zamir

Swiss Federal Institute of Technology

[palmer.epfl.ch]

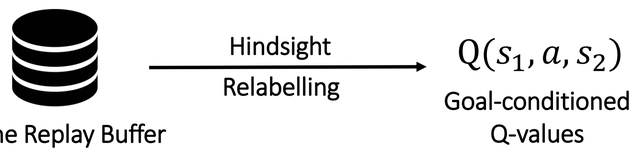
NEURAL INFORMATION PROCESSING SYSTEMS

Overview

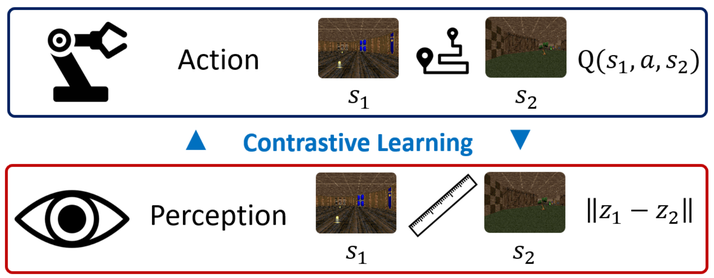
- Goal: Agents that can **plan far into the future** solely using **their own sensory experience** (e.g., images from an onboard camera)
- Key Idea: **Retrieve past trajectory segments** from a replay buffer and **restitch them into new paths**
- How to retrieve trajectory segments?: by using **reinforcement learning for representation learning**
- How to restitch trajectory segments?: by combining **learning-based representations** with **classical sampling-based planning**

Representation Learning by RL

1) Estimate $Q(s_1, a, s_2)$

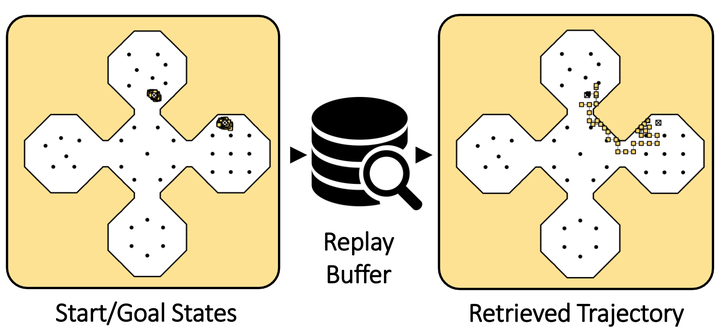


2) Use $Q(s_1, a, s_2)$ for contrastive representation learning



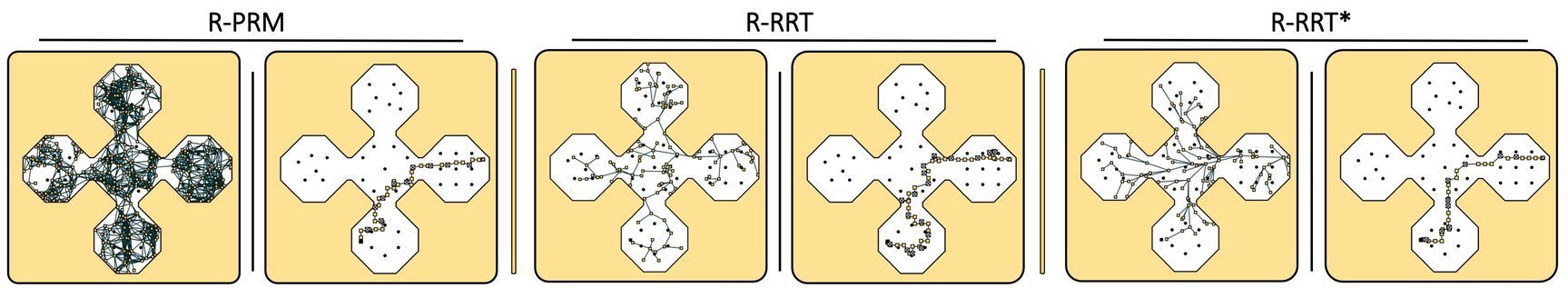
- $\|z_1 - z_2\|$ captures how many **timesteps it takes for the optimal policy to go from one state to another**

Retrieving Trajectories from the Replay-Buffer



- Given a pair of start/goal states as a query, **retrieve a trajectory segment** from the replay-buffer whose **endpoints lie close to the start/goal pair** in representation space

Planning by Stitching Trajectory Segments



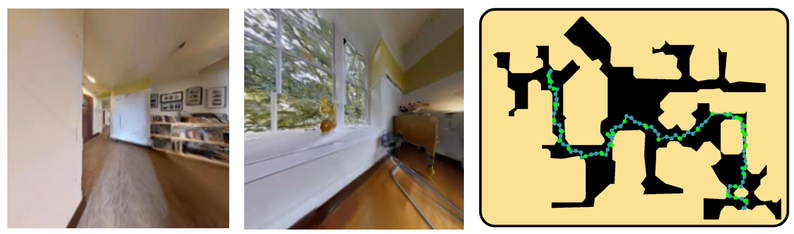
- Repurpose **RRT / PRM** so that **whenever an edge is created a trajectory is retrieved** and stored in that edge, to build a planning graph
- **Compute shortest paths** over this graph to **restitch the trajectories stored in edges**

Results and Summary

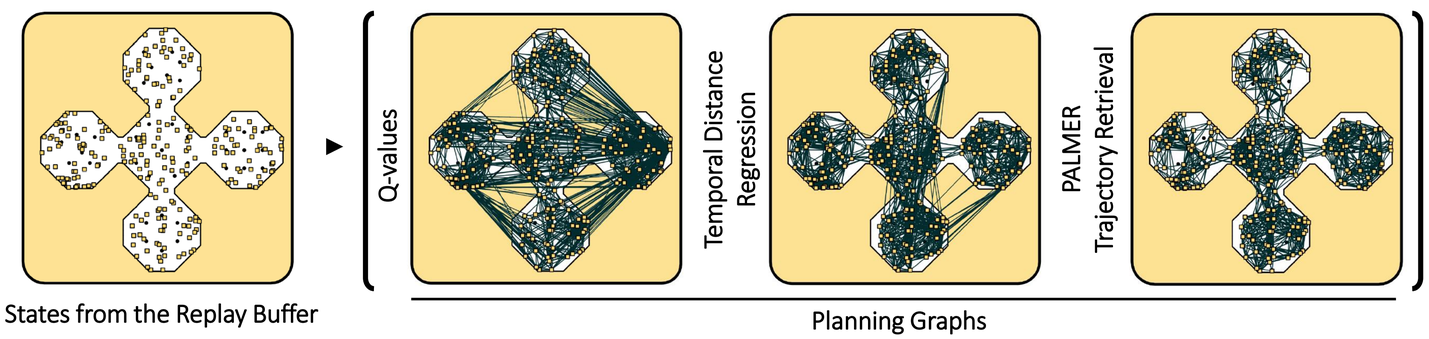
- Visual representations that are ***shaped by available actions*** (i.e., physical capabilities of an agent) and ***their consequences*** (i.e., state transitions)
- A learning-based planning method that is ***scalable*** *and* ***sample-efficient***:
  - Can navigate between any two points in reconstructions of ***large-scale real-world apartments***
  - This requires only ***150k environment steps of training data collected from a random-walk***
  - Does not assume a geometric model or a map of the environment, instead ***plans directly over past sensory observations***

Goal Distance ($n \times \Delta$)	**8**	**16**	**24**	**32**	**36**	**44**
SPTM Success Ratio (%)	0.28	0.01	0.0	0.0	0.0	0.0
SoRB Success Ratio (%)	0.42	0.0	0.0	0.0	0.0	0.0
PALMER Success Ratio (%)	0.99	0.97	0.91	0.93	0.93	0.94

Success Ratios for Visual Navigation in Habitat



- Resulting planning graphs are ***significantly more robust*** to hallucinated shortcuts compared to baselines, as creating an edge between two states ***requires there to be an actual trajectory segment*** in the replay-buffer that connects them



Comparing Distance Metrics