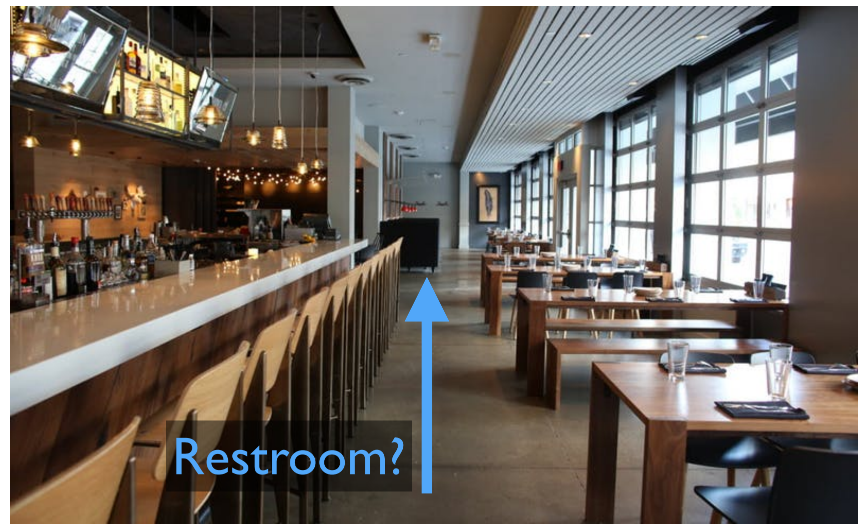
ILLINOIS

Semantic Visual Navigation by Watching YouTube Videos

Matthew Chang   Arjun Gupta   Saurabh Gupta

Code, data, video, models available: https://matthewchang.github.io/value-learning-from-videos/

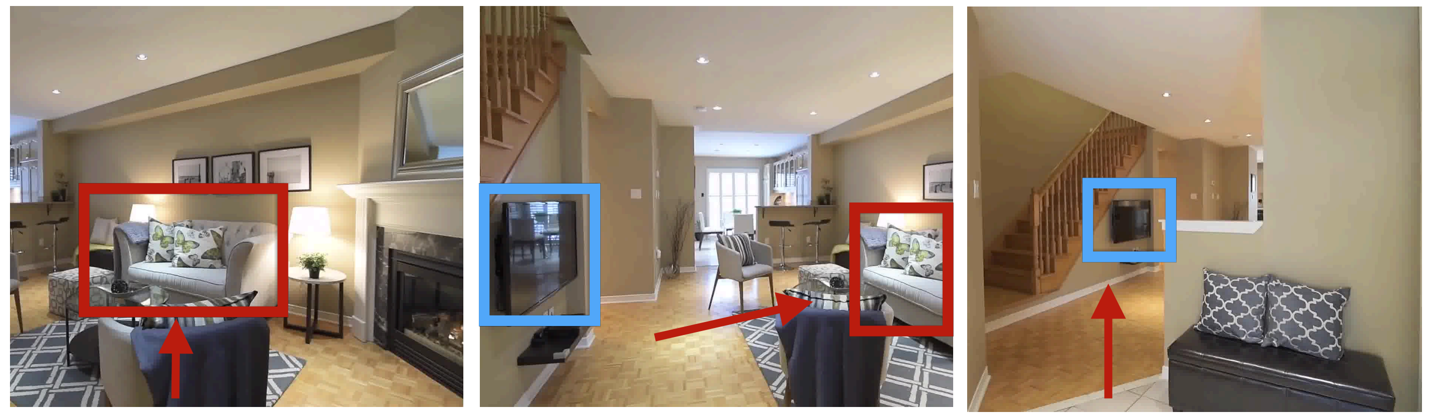
Motivation



Semantic cues and statistical regularities allow humans to efficiently navigate in novel environments.

This paper seeks to learn such cues from videos.

Instead of training with reinforcement learning or direct interaction, we learn semantic cues from object co-occurrence in videos.

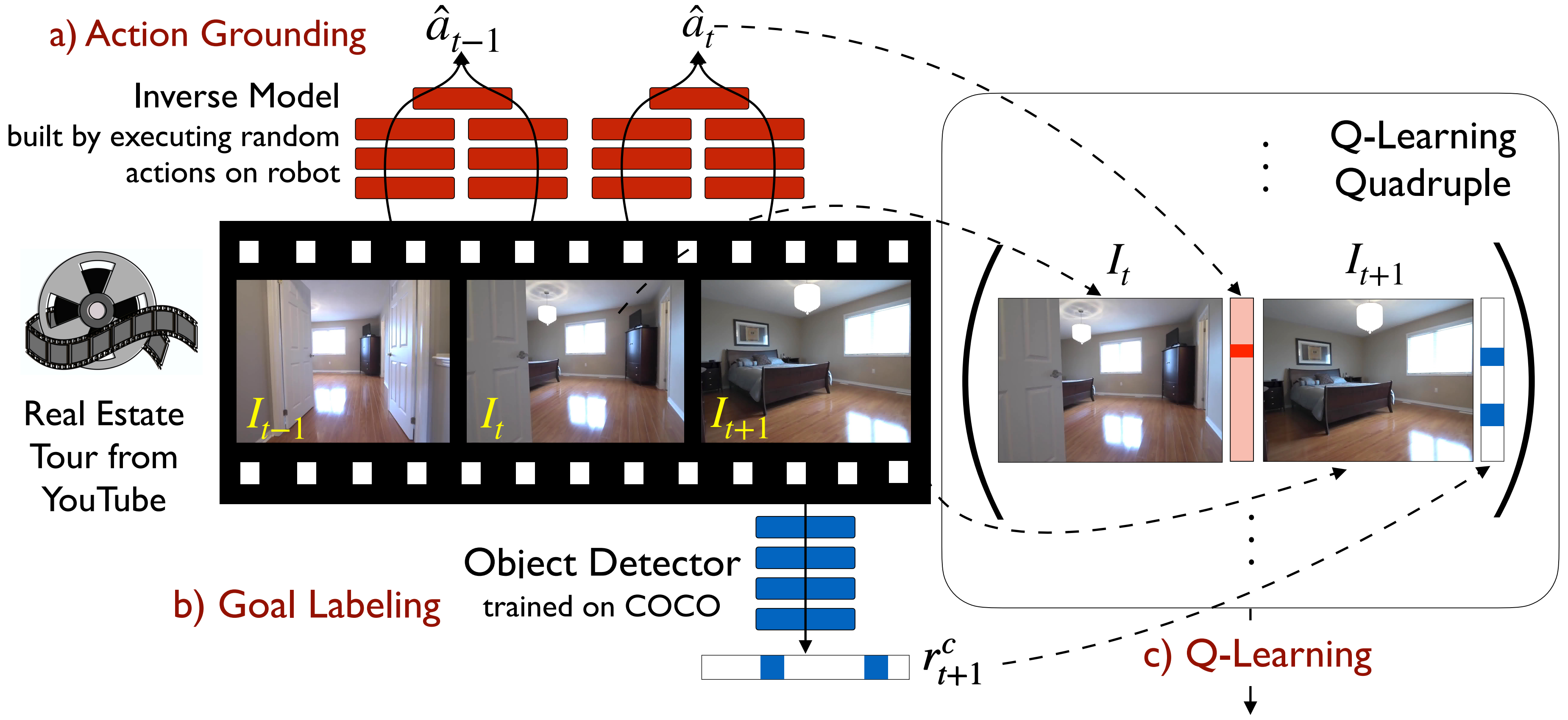


Challenges

- Videos don't come with action labels
  - ⟹ Action Grounding via an Inverse Model [1]
- Goals and intents are not known
  - ⟹ Use off-the-shelf Object Detectors to label frames with desired objects
- Depicted trajectories may not be optimal
  - ⟹ Use Q-learning to learn optimal behavior from sub-optimal data [2]
- Dataset not in existing literature
  - ⟹ Collected YouTube House Tours Dataset (1387 videos, 119 Hours)

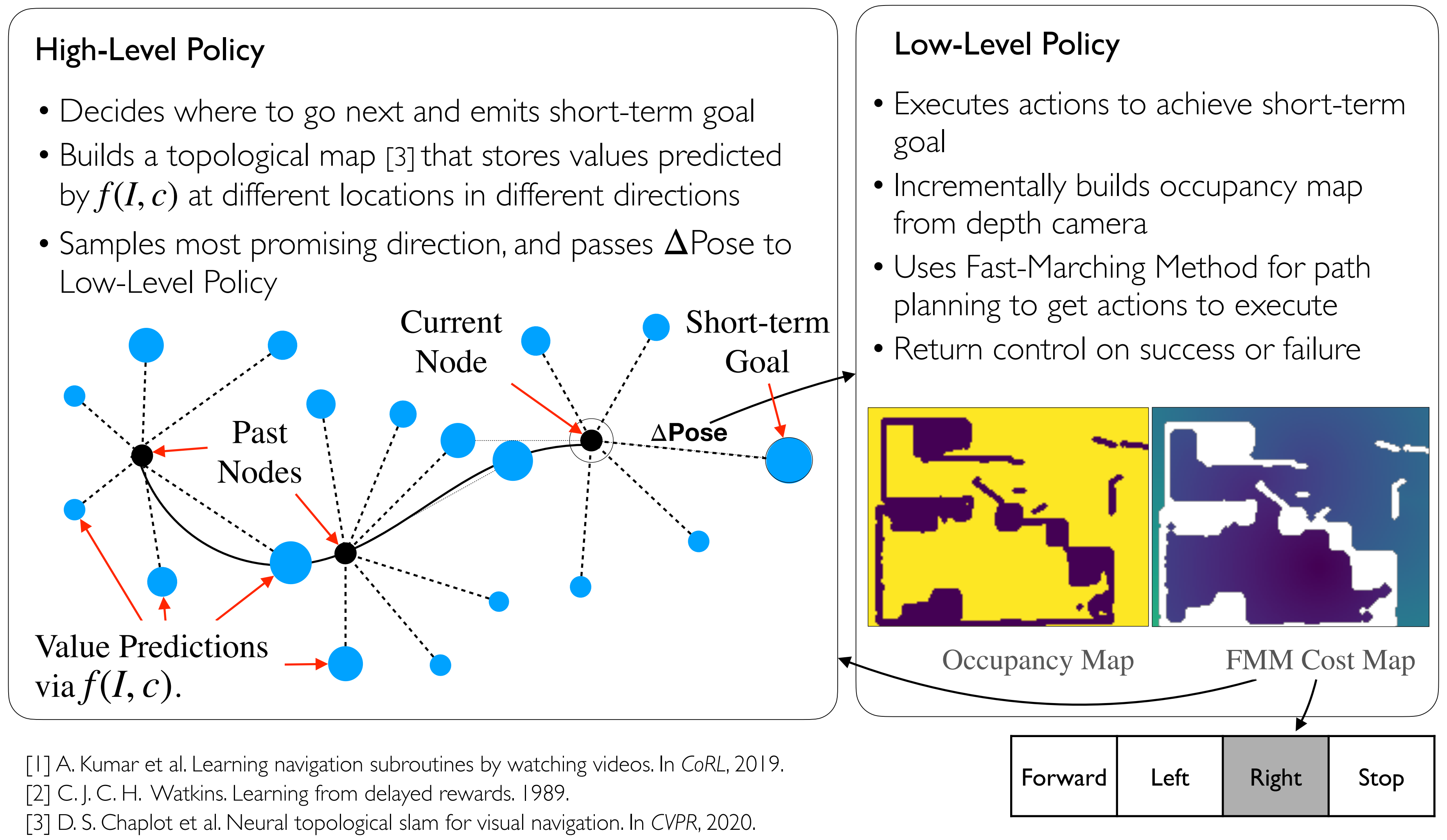
Approach

Value Learning from Videos



Value function that predicts nearness to goal: $f(I, c) = \max_a Q^*(I, a, c)$

Hierarchical Policy



High-Level Policy

- Decides where to go next and emits short-term goal
- Builds a topological map [3] that stores values predicted by $f(I, c)$ at different locations in different directions
- Samples most promising direction, and passes ΔPose to Low-Level Policy

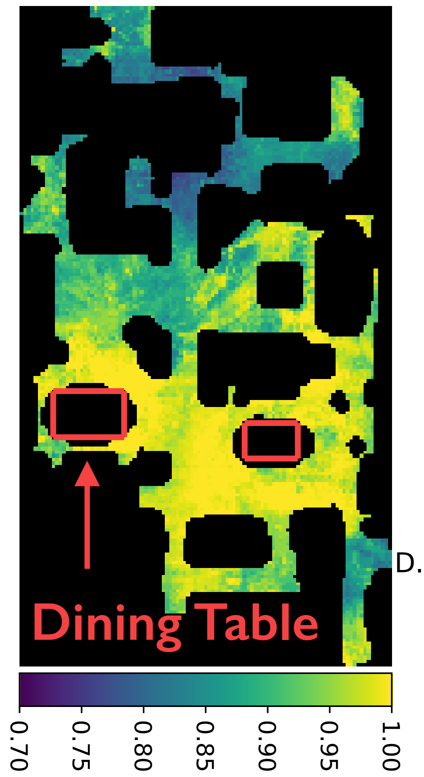
Low-Level Policy

- Executes actions to achieve short-term goal
- Incrementally builds occupancy map from depth camera
- Uses Fast-Marching Method for path planning to get actions to execute
- Return control on success or failure

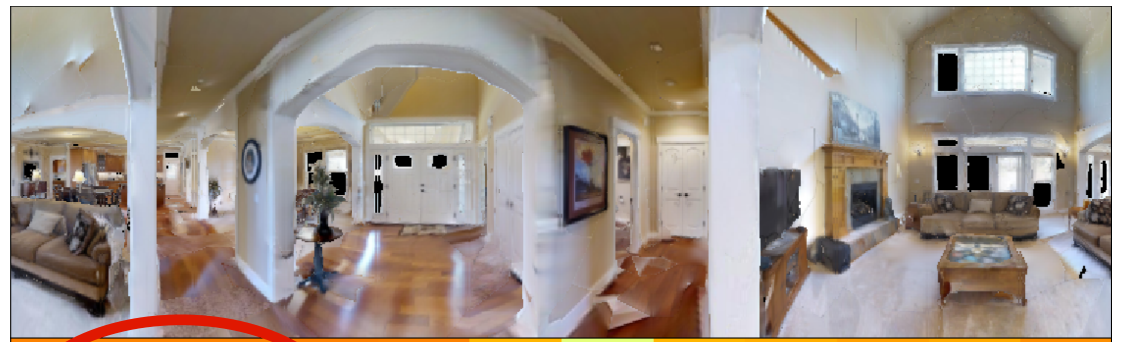
[1] A. Kumar et al. Learning navigation subroutines by watching videos. In *CoRL*, 2019.
[2] C. J. C. H. Watkins. Learning from delayed rewards. 1989.
[3] D. S. Chaplot et al. Neural topological slam for visual navigation. In *CVPR*, 2020.

Acknowledgement: We thank Sanjeev Venkatesan for help with data collection. We also thank Rishabh Goyal, Ashish Kumar, and Tanmay Gupta for feedback on the paper. This material is based upon work supported by NSF under Grant No. IIS-2007035, and DARPA Machine Common Sense. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or DARPA.

Visualizations

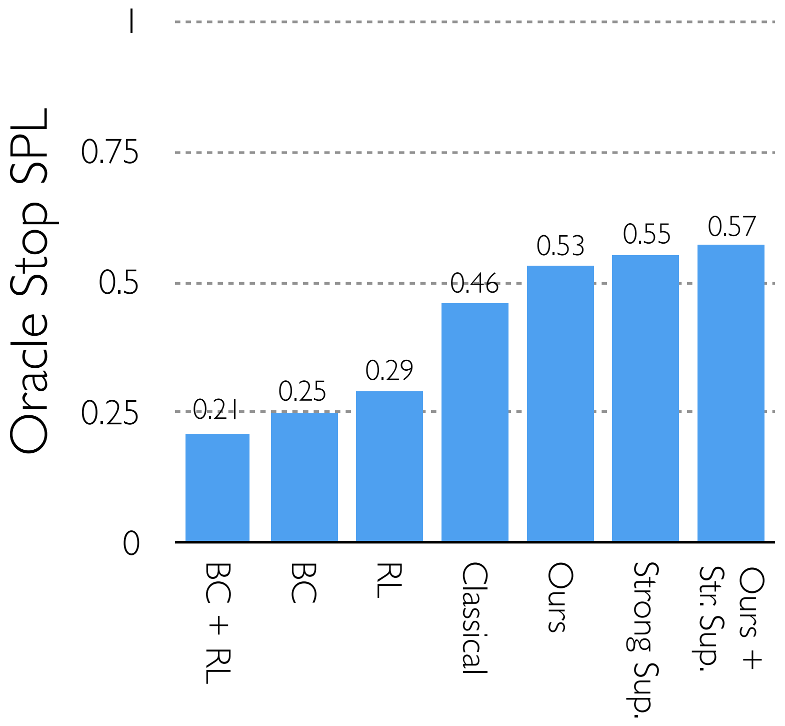


Top down map and panorama predictions for nearness to goal



Chair	0.99	0.99	0.99	1.00	0.99	0.96	0.92	0.96	0.97	0.98	0.97	0.99
Couch	0.99	0.95	0.84	0.80	0.82	0.82	0.80	0.84	0.87	0.90	0.94	0.99
D. Table	0.87	0.97	0.99	1.01	0.92	0.88	0.82	0.84	0.85	0.85	0.84	0.83
Bed	0.78	0.78	0.80	0.80	0.83	0.83	0.84	0.84	0.84	0.83	0.80	0.78
Toilet	0.62	0.63	0.65	0.63	0.71	0.68	0.71	0.71	0.71	0.66	0.63	0.62

Evaluation

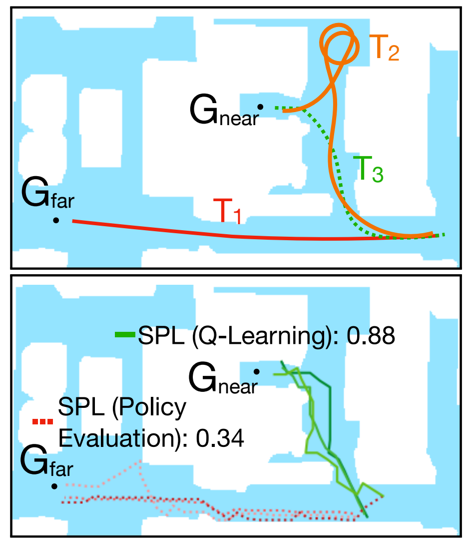


- Stronger than behavior cloning on videos and BC + RL
- Stronger than even RL methods trained with dense rewards with 250x more interaction samples and 6x more environments with direct interaction access
- Better than strong exploration baselines
- Improves performance when combined with strongly supervised model

Ablations

Oracle Stop SPL

Method	Easy	Medium	Hard	Overall
Base Setting	0.62	0.42	0.23	0.40
True Actions	0.61	0.45	0.25	0.41
True Detections	0.62	0.45	0.22	0.40
True Rewards	0.64	0.46	0.21	0.41
Optimal Trajectories	0.65	0.46	0.25	0.43
Detector Score	0.73	0.48	0.26	0.46
Train on 360° Videos	0.66	0.51	0.32	0.47
No Hierarchy	0.38	0.10	0.02	0.15



- Inverse model and detector do not hurt performance significantly
- Detector at test time helps for close objects, panorama helps for far objects
- Q-Learning outperforms simple policy evaluation for challenging environments
- Hierarchical policy is a major factor in strong performance