NRI: Self-Supervised Object Detection and Visual Navigation

Award # IIS 1925231

Jana Kosecka, George Mason University

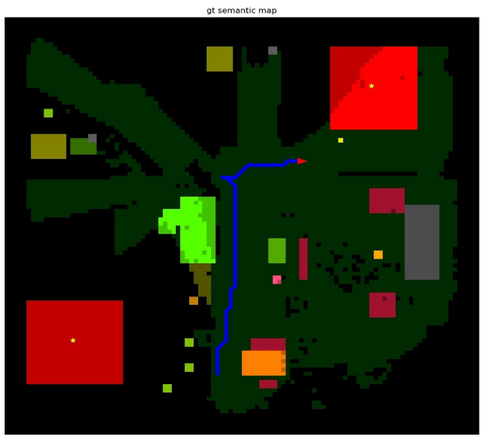
Challenge

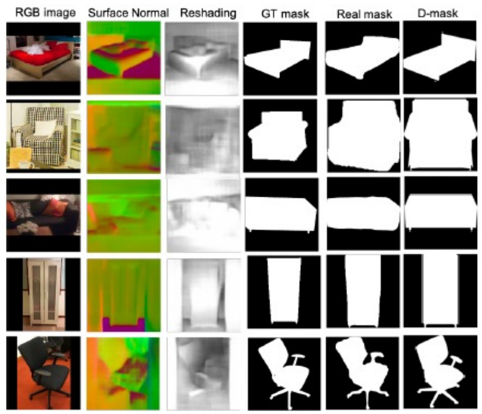
- Task 1: Target driven visual navigation in indoor environmemts
- Task 2: Improvements in semantic mapping using object pose estimation
- Task 3: Self-supervised fine-tuning of semantic segmentation using temporal consistency

Students: Yimeng Li (GMU)

Negar Nejatishahidin

Sulabh Shresta





Broader Impact

- Improving robustness and functionality for fetch and delivery tasks for service robotics
- Pose estimation benchmark dataset
- Education and Outreach
- *Improvement in the state of the art in target driven navigation*

Task 1: Object Goal Navigation in a Novel Environment

Target Objects:

couch, potted plant, fridge, oven, tv, chair, toilet, cup, sink

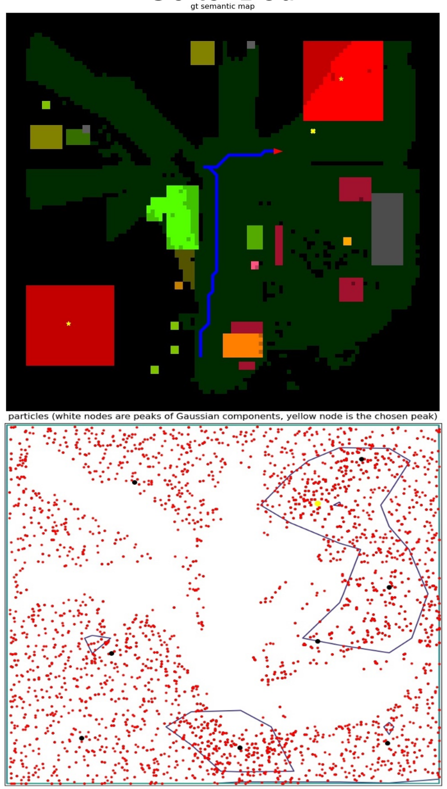
Belief maintenance

- Particle filters updated based on observations from semantic ego-centric maps and semantic priors on the object/object co-occurrence and room/object co-occurrence

Navigation:

- Predict subgoal as peak in the object belief space, short-range navigation to reach the subgoal

Go to 'Bed'



'Bed location' belief space

Evaluation

- Evaluate feasibility of method similar to [1] Matterport3D scenes, object goal navigation task
- Performance degrades in large environments
- Comparable performance to the SOTA Approach (PONI)

	Success	SPL
PONI (STOA)	0.87	0.52
Ours	**0.97**	**0.64**

[1] Semantic Linking Maps for Active Visual Object Search with Interaction-free Learning
Z. Zeng, A. Röfer, O. Chadwicke Jenkins
[2] PONI: Potential Functions for ObjectGoal Navigation.
S. Ramakrishnan, D. Chaplot. Z. Al-Halah et. al

Task 2: Object Pose Estimation using Mid-level Visual Representations

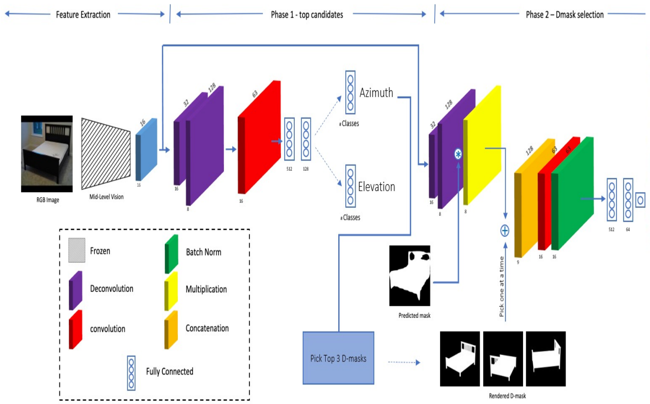
Challenge

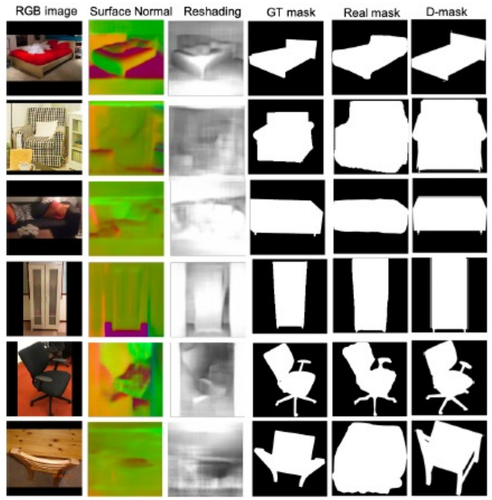
- Pose estimation for highly occluded and truncated objects in real-world indoor environments.
- Need for large amount of data and costly labeling.

Solution

- Novel object pose estimation model built on top of generic mid-level representation features. Pre-trained feature maps of surface normal and re-shading
- Competitive performance in low training data regime, transfer to real-world unseen objects

[1] Object Pose Estimation using Mid-level Visual Representations
N. Nejatishahidin, P. Fayyazsanavi, J. Kosecka, arXiv:2203.01449





Scientific Impact

- New benchmark for challenging real-word object pose estimation, 6337 objects' pose of furniture categories and 3D bounding box labels.
- The model significantly outperforms other models in low training data regime

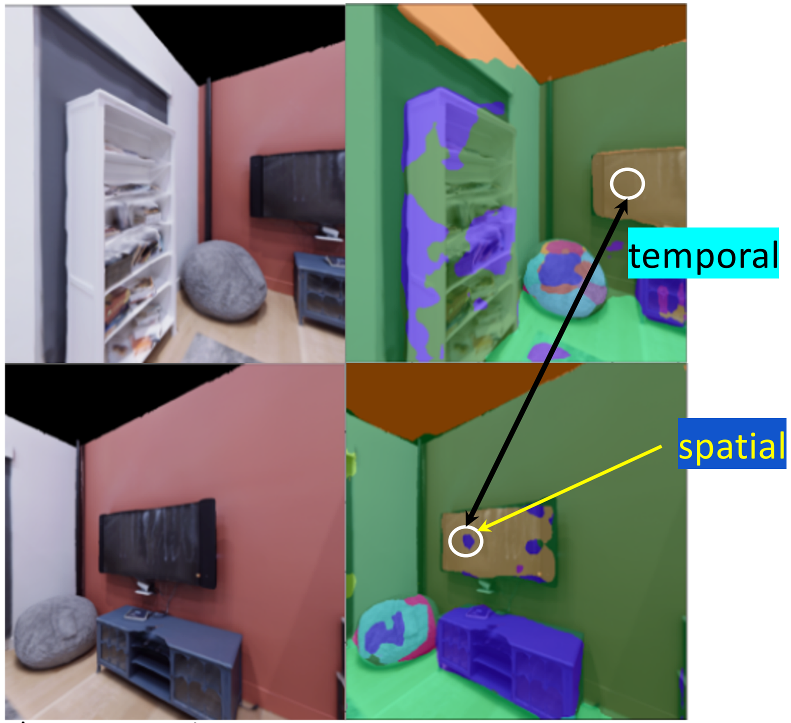
Task 3: Self-supervised fine-tuning of semantic segmentation

Challenge

- Perception models in absence/scarcity of data in an indoor environment

Solution

- Take high confidence or ground truth predictions
- Associate pixels across views and perform self-supervised learning



Solution

- Principle foundation: temporal and spatial consistency
- Take high confidence or ground truth predictions
- Associate pixels across views, across local regions and perform self-supervised learning
- Demonstrated the effectiveness of contrastive learning approach for this tasks.

[1] Object Pose Estimation using Mid-level Visual Representations
S. Shresta, J. Kosecka (in preparation)