Object Goal Navigation using Goal-oriented Semantic Exploration

Team Arnold (SemExp)

Devendra Singh Chaplot
Dhiraj Gandhi
Abhinav Gupta
Ruslan Salakhutdinov
Object Goal Navigation
Object Goal Navigation

Object Goal: dining table
Object Goal Navigation

Object Goal: dining table

Geometric Scene Understanding
- Understanding navigable space

Semantic Scene Understanding
- Object detection and segmentation

Passive
Object Goal Navigation

Object Goal: dining table

Passive

Geometric Scene Understanding
Understanding navigable space

Semantic Scene Understanding
Object detection and segmentation

Active

Learning Semantic Priors
Where is ‘dining table’ more likely to be found?

Episodic Memory
Keeping track of explored and unexplored areas
Active Neural SLAM

Sensor Pose Reading ($x_t'$)

Observation ($s_t$)

Action ($a_t$)

Neural SLAM ($f_{SLAM}$)

Local Policy ($\pi_L$)

Neural SLAM ($f_{SLAM}$)

Local Policy ($\pi_L$)

Pose Estimate ($\hat{x}_t$)

Global Policy ($\pi_G$)

Map ($m_t$)

Short-term goal ($g^t$)

Long-term goal ($g^l_t$)

Map ($m_t$)

[fPlan]

[Chaplot et al. ICLR-20]
Incorporating Semantics

Obstacle Map Representation
(Active Neural SLAM)

Obstacle Map \((2 \times M \times M)\)

- Obstacles
- Explored Area
Incorporating Semantics

Obstacle Map Representation (Active Neural SLAM)

Obstacle Map \((2 \times M \times M)\)

Semantic Map Representation (SemExp)

Semantic Map \((K \times M \times M)\)

\(K = C + 2\)
Semantic Mapping

RGB ($I_j$)

Depth ($D_j$)
Semantic Mapping

RGB ($I_i$) → Mask RCNN → First-person Semantic Predictions

Depth ($D_i$)
Semantic Mapping

RGB ($I_t$)

Depth ($D_t$)

First-person Semantic Predictions

X Y Z

Point Cloud

Mask RCNN
Semantic Mapping

RGB ($I_t$) $\rightarrow$ Mask RCNN $\rightarrow$ First-person Semantic Predictions

Depth ($D_t$) $\rightarrow$ Point Cloud $\rightarrow$ Semantic Labels
Semantic Mapping

RGB ($I_t$) → Mask RCNN → First-person Semantic Predictions

Depth ($D_t$) → Voxel $(C + 1) \times H \times M \times M$
Semantic Mapping

RGB ($I_t$) → Mask RCNN → First-person Semantic Predictions

Depth ($D_t$) → Point Cloud → Semantic Labels

Voxel $(C + 1) \times H \times M \times M$

Projection Map $(C + 2) \times M \times M$

sum across height

All obstacles

All cells

Category-wise

Obstacles

Explored Area

Semantic categories ($C$)
Semantic Mapping

RGB ($I_t$)

Depth ($D_t$)

Mask RCNN

First-person Semantic Predictions

Point Cloud

Semantic Labels

Voxel

Projection Map

Semantic Map Prediction

Denoising Network

RGB ($I_t$)

Depth ($D_t$)

Mask RCNN

First-person Semantic Predictions

Point Cloud

Semantic Labels

Voxel

Projection Map

Semantic Map Prediction

Denoising Network
SemExp Model Overview

Sensor Pose Reading ($x_i$)

Observation ($s_i$) (RGBD)

Object Goal ($G = \text{“chair”}$)
SemExp Model Overview

Sensor Pose Reading ($x_t$)

Observation ($s_t$) (RGBD)

Object Goal ($G = \text{“chair”}$)

Semantic Mapping

Semantic Map ($m_t$)
SemExp Model Overview

1. Sensor Pose Reading ($x_t$)
2. Observation ($s_t$) (RGBD)
3. Object Goal ($G = \text{“chair”}$)

- Semantic Mapping
  - Semantic Map ($m_t$)
- Long-term goal ($g_t$)
- Goal-Oriented Semantic Policy

Observed data flows through the model to generate a semantic map, which is then used to guide the agent towards its long-term goal.
SemExp Model Overview

Sensor Pose Reading ($x_t$)

Observation ($s_t$) (RGBD)

Object Goal ($G = \text{“chair”}$)

Semantic Mapping

Semantic Map ($m_t$)

Long-term goal ($g_t$)

Goal-Oriented Semantic Policy

Deterministic Local Policy ($\pi_L$)

Action ($a_t$)
Demo Video

Observation  Object Goal: Chair  Predicted Semantic Map  Ground Truth
ObjectGoal Navigation Results
ObjectGoal Navigation Results

Success Rate

<table>
<thead>
<tr>
<th>Method</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.004</td>
</tr>
<tr>
<td>RGBD + RL [1]</td>
<td>0.082</td>
</tr>
<tr>
<td>RGBD + Semantics + RL [2]</td>
<td>0.159</td>
</tr>
<tr>
<td>Classical Map + FBE</td>
<td>0.403</td>
</tr>
<tr>
<td>Active Neural SLAM [3]</td>
<td>0.446</td>
</tr>
<tr>
<td>SemExp</td>
<td>0.544</td>
</tr>
</tbody>
</table>

SPL

<table>
<thead>
<tr>
<th>Method</th>
<th>SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.004</td>
</tr>
<tr>
<td>RGBD + RL [1]</td>
<td>0.027</td>
</tr>
<tr>
<td>RGBD + Semantics + RL [2]</td>
<td>0.049</td>
</tr>
<tr>
<td>Classical Map + FBE</td>
<td>0.124</td>
</tr>
<tr>
<td>Active Neural SLAM [3]</td>
<td>0.145</td>
</tr>
<tr>
<td>SemExp</td>
<td>0.199</td>
</tr>
</tbody>
</table>

ObjectGoal Navigation Results

Success Rate

<table>
<thead>
<tr>
<th>Method</th>
<th>MP3D</th>
<th>SPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>RGBD + RL [1]</td>
<td>0.037</td>
<td>0.027</td>
</tr>
<tr>
<td>RGBD + Semantics + RL [2]</td>
<td>0.031</td>
<td>0.049</td>
</tr>
<tr>
<td>Classical Map + FBE</td>
<td>0.311</td>
<td>0.124</td>
</tr>
<tr>
<td>Active Neural SLAM [3]</td>
<td>0.321</td>
<td>0.145</td>
</tr>
<tr>
<td>SemExp</td>
<td>0.36</td>
<td>0.199</td>
</tr>
</tbody>
</table>

# Habitat Challenge Leaderboard

<table>
<thead>
<tr>
<th>Method</th>
<th>Test-standard</th>
<th>Minival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPL</td>
<td>Success</td>
</tr>
<tr>
<td>Arnold (SemExp)</td>
<td>0.071</td>
<td>0.179</td>
</tr>
<tr>
<td>Active Exploration</td>
<td>0.041</td>
<td>0.089</td>
</tr>
<tr>
<td>DD-PPO</td>
<td>0.021</td>
<td>0.062</td>
</tr>
<tr>
<td>Blue Ox</td>
<td>0.017</td>
<td>0.060</td>
</tr>
<tr>
<td>SRCB-robot-sudoer</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td>PPO RGBD</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Random</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
# Habitat Challenge Leaderboard

<table>
<thead>
<tr>
<th>Method</th>
<th>Test-standard</th>
<th>Minival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPL</td>
<td>Success</td>
</tr>
<tr>
<td>Arnold (SemExp)</td>
<td>0.071</td>
<td>0.179</td>
</tr>
<tr>
<td>Active Exploration</td>
<td>0.041</td>
<td>0.089</td>
</tr>
<tr>
<td>DD-PPO</td>
<td>0.021</td>
<td>0.062</td>
</tr>
<tr>
<td>Blue Ox</td>
<td>0.017</td>
<td>0.060</td>
</tr>
<tr>
<td>SRCB-robot-sudoer</td>
<td>0.002</td>
<td>0.004</td>
</tr>
<tr>
<td>PPO RGBD</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Random</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Real-world Transfer

Observation

Object Goal: Oven

Predicted Semantic Map

Object Goal: Oven
Thank you

Devendra Singh Chaplot

Webpage:  http://devendrachaplot.github.io/
Email:     chaplot@cs.cmu.edu
Twitter:   @dchaplot