Action Conditioned Segmentation for ALFRED

Kush Bhagat | GLAMOR lab [http://glamor.rocks/]

USC Viterbi School of Engineering

Introduction

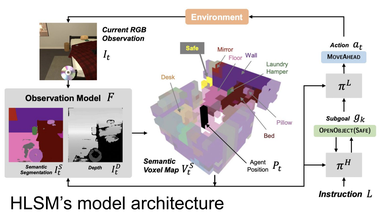
- *A robot that can carry out a natural-language instruction has been a dream since before the Jetsons cartoon series imagined a life of leisure mediated by a fleet of attentive robot helpers*
- Increasing segmentation accuracy reflects a 11% performance increase in Hierarchical Language-Conditioned Spatial Model (HLSM)

ALFRED demonstrations within the AI2Thor environment
[ALFRED, Shridhar et al., CVPR 2020]

Methods

- Segmentation is trained over 300x300 RGB images alongside class masks from “rollouts” of tasks in the AI2Thor environment
- An action-conditioned segmentation model was trained by isolating object by affordances that match a HLSM subgoal

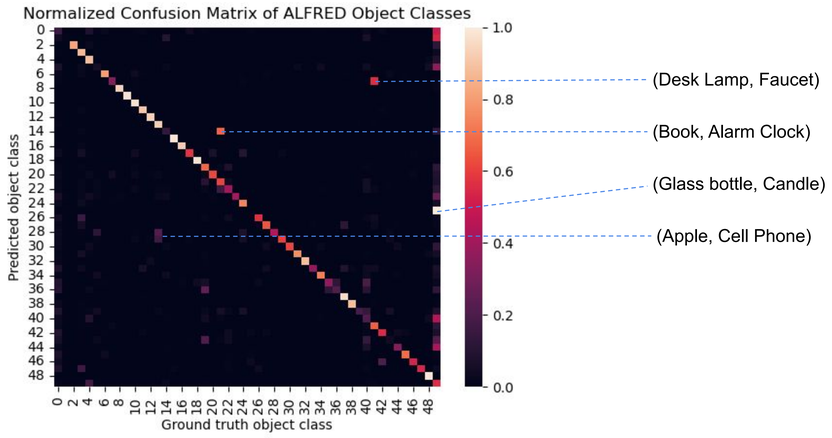
Objects that are not “pickupable” are removed from the ground truth class mask during training the action conditioned model

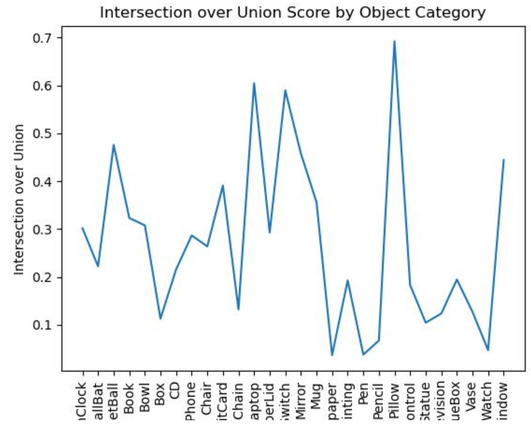


HLSM’s model architecture
[A Persistent Spatial Semantic Representation for High-level Natural Language Instruction Execution, Blukis et al. CoRL 2021]

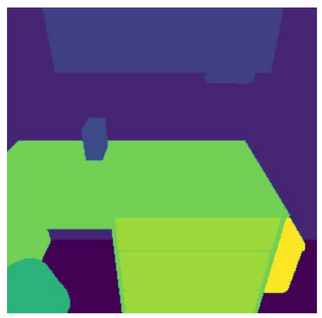
- Classification errors were collected over validation runs to measure performance

Results

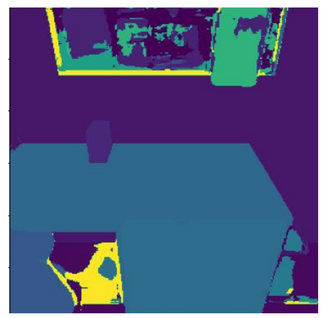




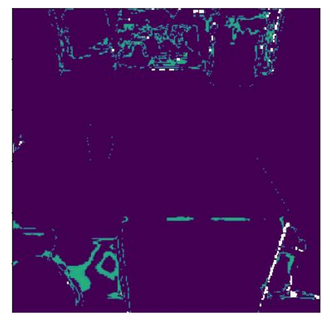
300x300 RGB snapshot from ALFRED demonstration



Ground truth class mask



Predicted class mask by HLSM’s base segmentation



Entropy regions of model predictions

Conclusions

- Action conditioning the segmentation model allows for better detection of objects associated with a subgoal
- More accurate interaction masks improves the performance of HLSM’s low-level controller and reduce API errors
- Physical deployment would require greater interaction accuracy
- Recognizing small distant objects aids in HLSM’s navigation phases

Future Directions

- Create conditioned models for each skill type and incorporate them into the HLSM framework
- Utilize HLSM’s map in cross training segmentation and depth perception
- Explore methods of navigation to include ego motion to aid depth perception and segmentation of distant low confidence objects
- Sample semantic curiosity to aid in informing HLSM’s map during early exploration phases