Irdi Balla

1 Related Work

1.1 Grounded Language + Abstract Markov Decision Process (GLAMDP)

GLAMDP uses an Object-Oriented Markov Decision Process to model the environment and actions of the robot. This provides a policy that maximizes the total expected discounted reward. GLAMDP is configured for a mobile manipulator object and it does not generalize well for other environments. The Cleanup World[CITE] is partitioned into rooms and the robot need to move objects around depending on the command it gets. GLAMDP requires for the action to be a movement action and this is the main drawback for this approach. On the other hand the model works for varying granularities of the command. In that sense it can differentiate between commands like "go to the green room"(high level) and "move up one tile" (low level), but still it expects all the command to be of the same granularity.

1.2 Gate-Attention Architecture for Task-Oriented Language Grounding(DeepRL-Grounding)

DRLG proposes an Architecture that assumes no prior linguistic knowledge and can be trained end-to-end. It comprises of a State Processing Module that takes the current state and outputs a joint representation of the image and the instruction. This Module includes a CNN for the image, a Gated Recurrent Unit for the instruction and a Multimodal fusion unit. DRLG is tested in an environment where the agent is asked to move to a certain object and here it is limited to having the "go" command in the instruction and that is the main drawback.

1.3 Textual Grounding using Image Concepts

This approach first extracts some word priors from the most common words in the training set. Then from the image they extract the semantic segmentation and the detection maps and use all these 3 'image concepts' to train the model. Instead of using a limited set of proposal bounding boxes, they go through a larger number of bounding boxes as an energy minimization approach. This approach is limited to grounding the image and does not take commands as

input but rather only a description and grounds the latter's elements in the image.