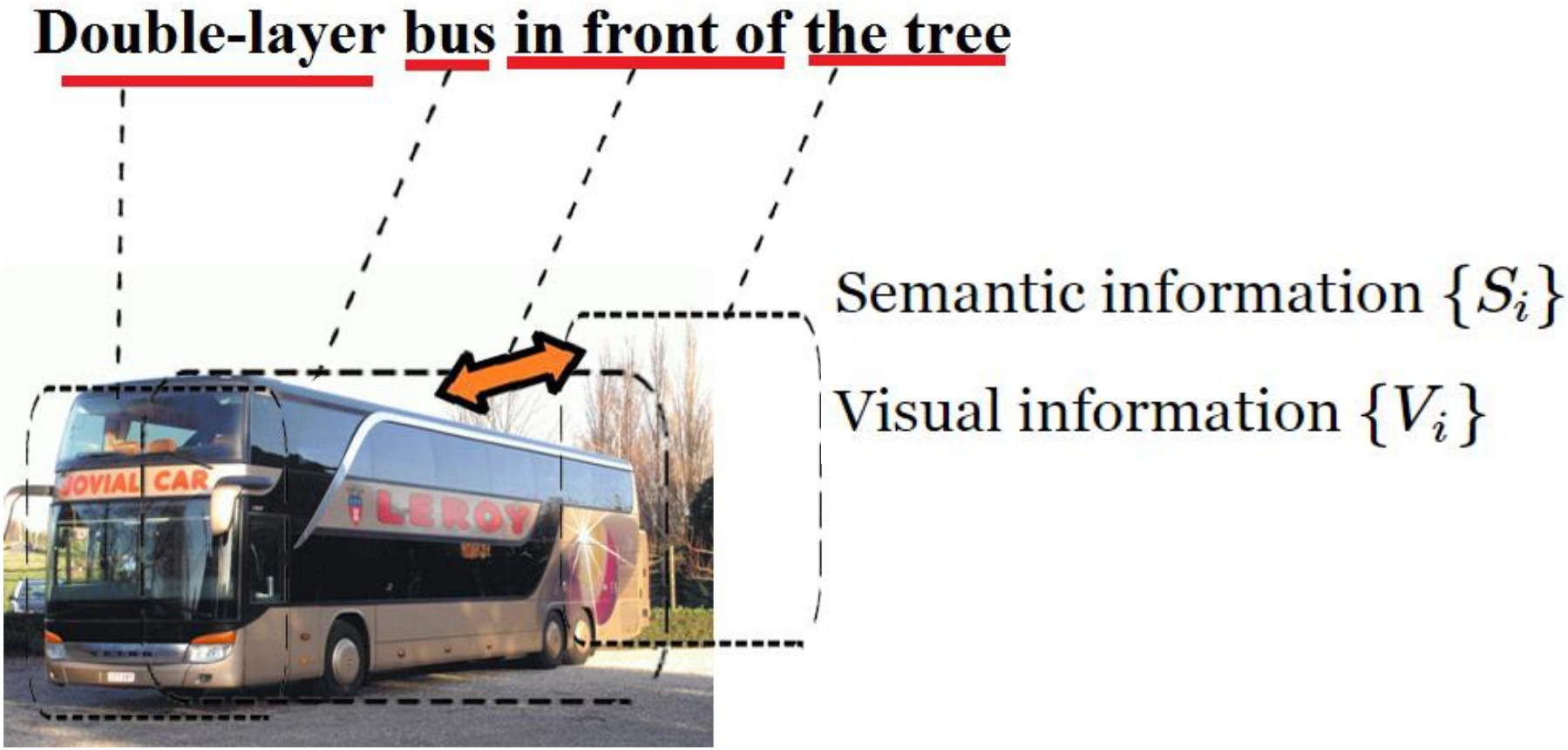
Reasoning Based on Visual + Semantic

October 25, 2018

401所 信息中心
羅敏中

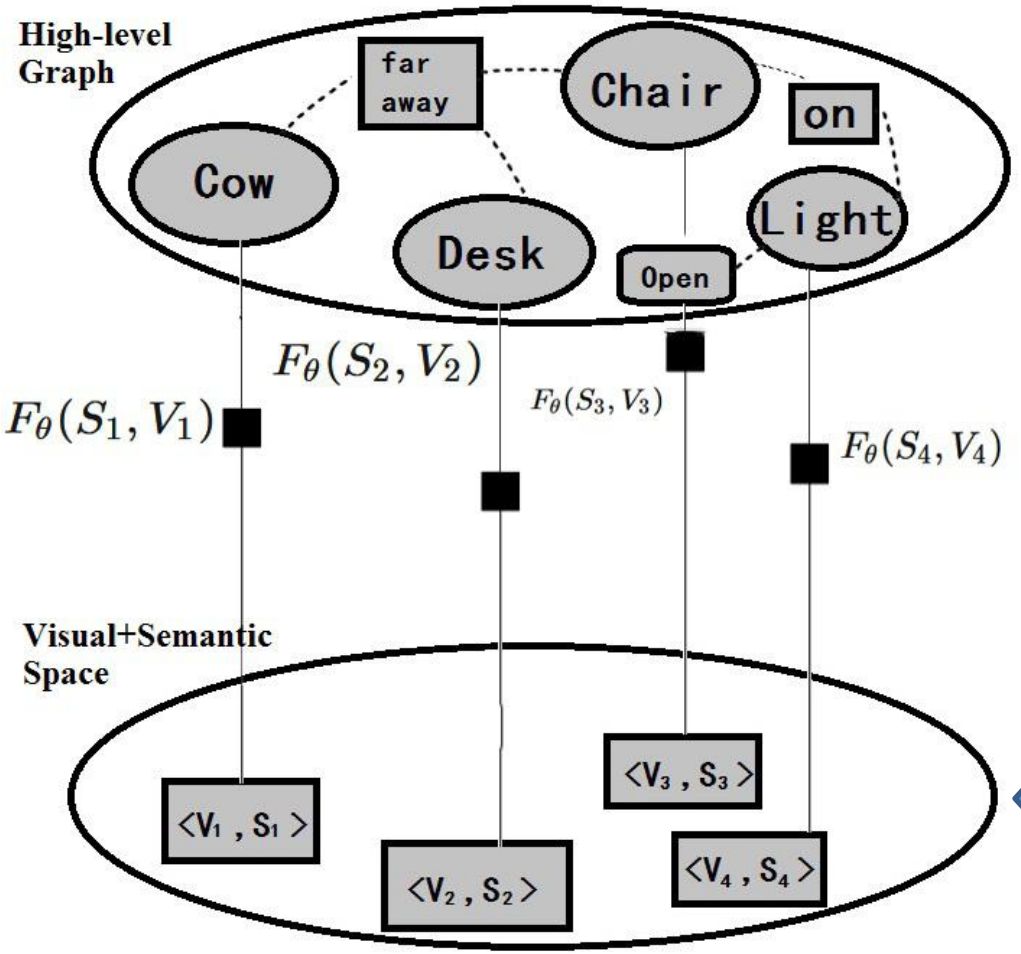
The Theory of visual semantic based reasoning

- One-to-one correspondence between semantic information and visual information



The Theory of visual semantic based reasoning

- Learning the high-level logic graph over the Visual and Semantic

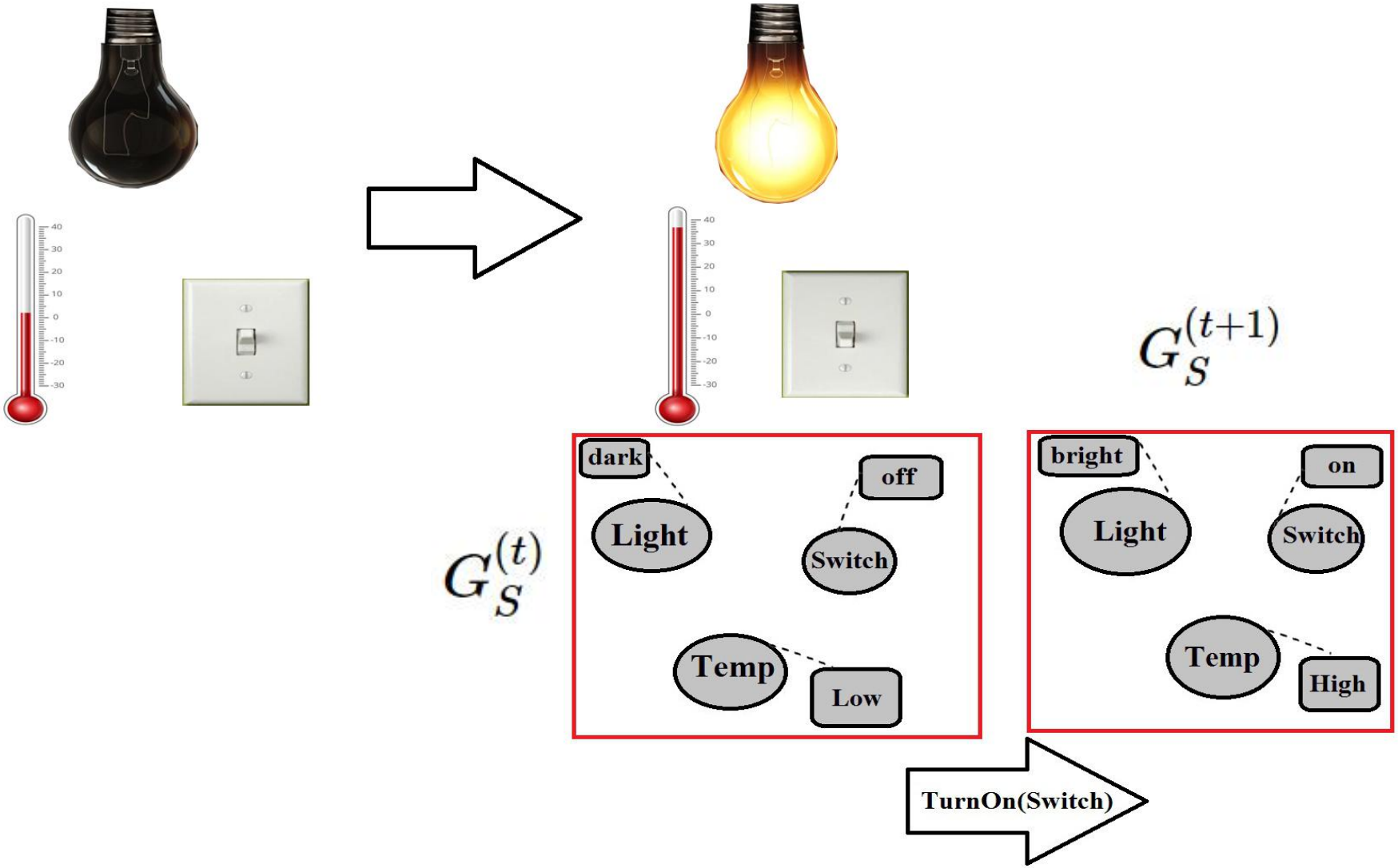


Used for logical reasoning and tasks completing

Image recognition and semantic matching

The Theory of visual semantic based reasoning

- The logic reasoning based on observation

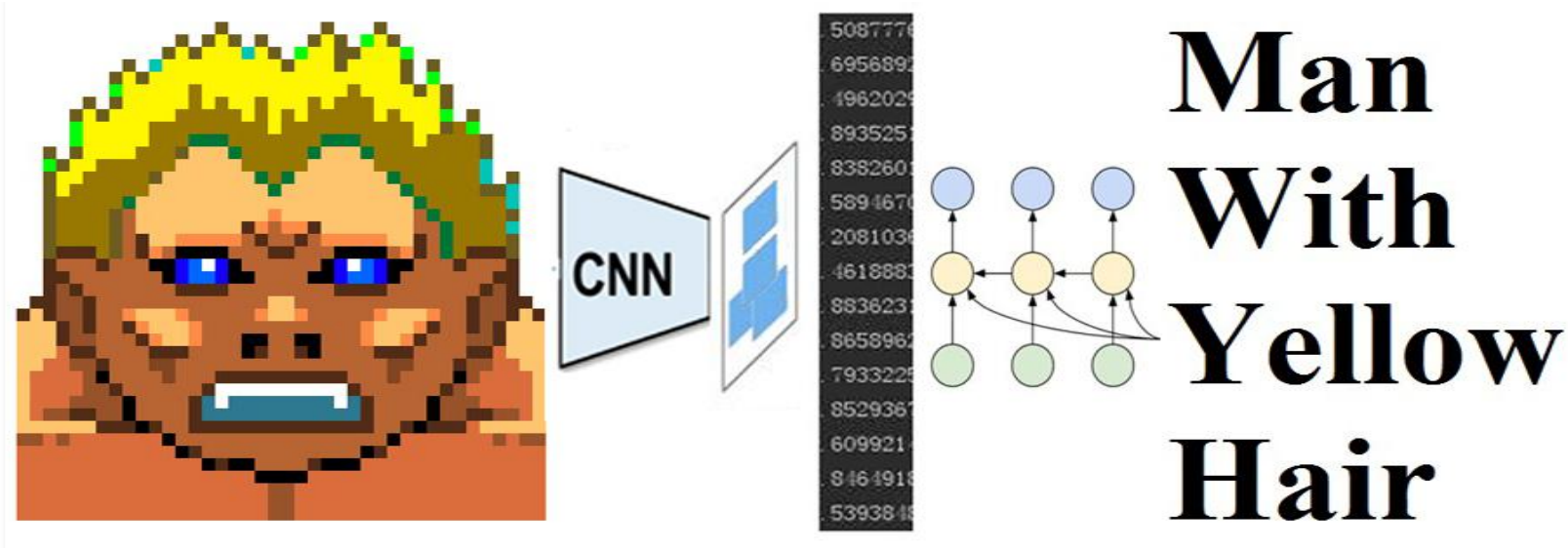


The Theory of visual semantic based reasoning

- The challenge of integrating visual and semantic information

- Language is a set of symbols with advanced coding.

- Visual images is a set of low-order noisy pixels.



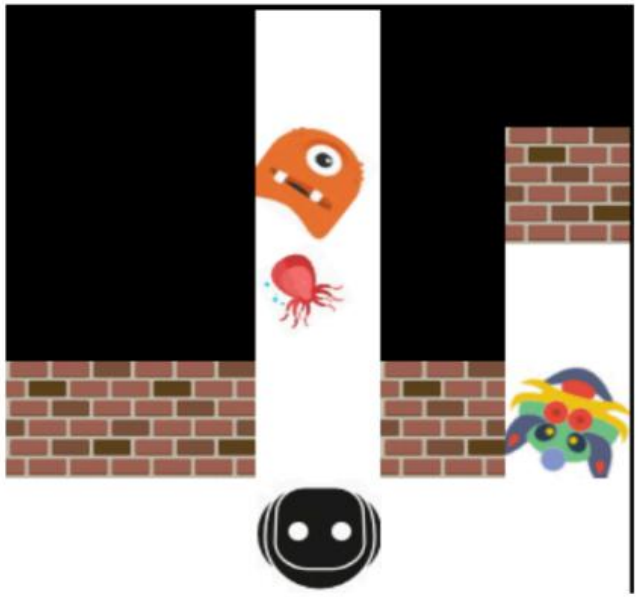
The XWorld

- A 3D simulation tool for RL

Inputs:

Raw pixel inputs + unstructured commands + sparse rewards

XWorld2D



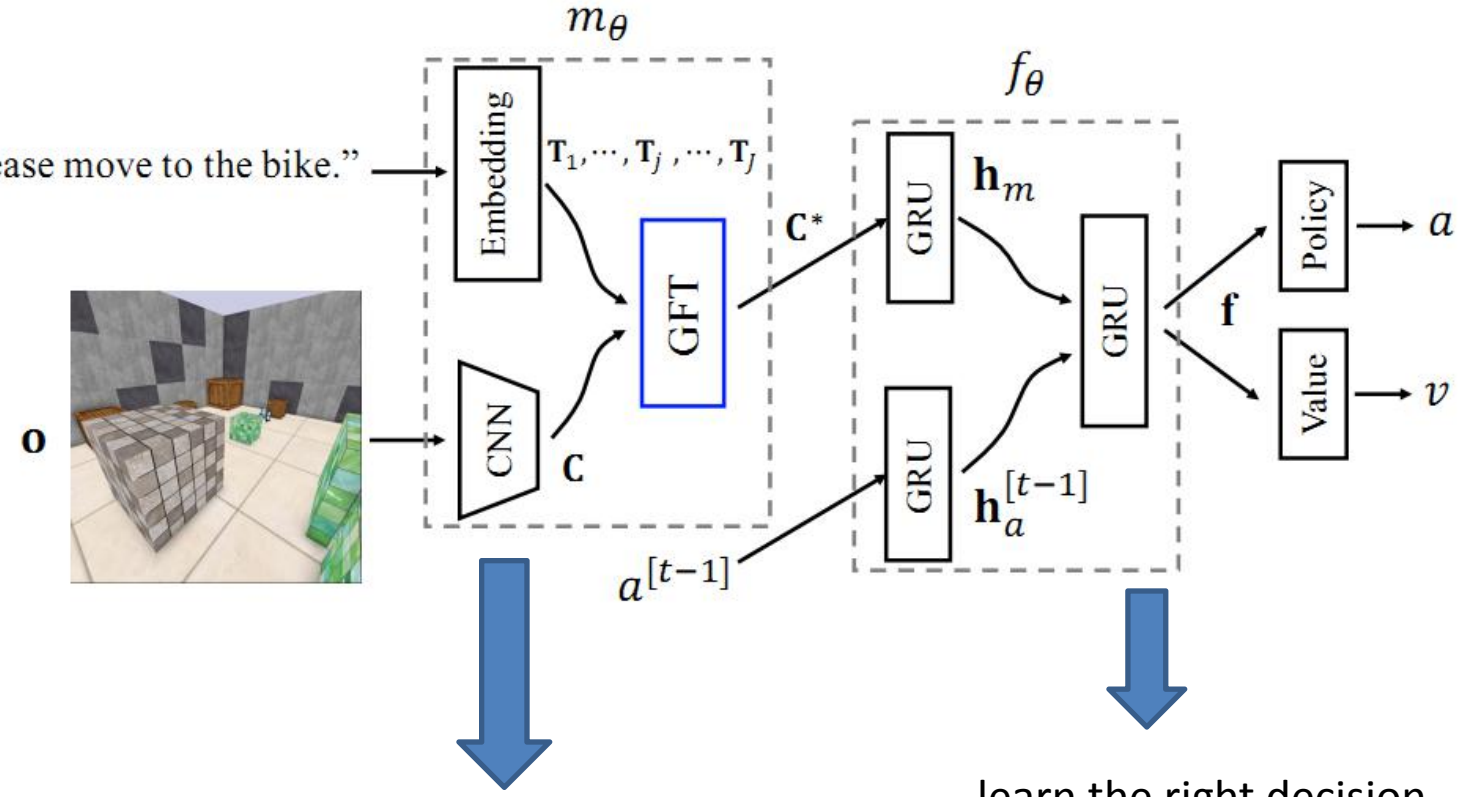
"Navigate to the object in front of the monster."

XWorld3D

"Can you please go to the dog?"

The XWorld

- Learning to navigate under a language command

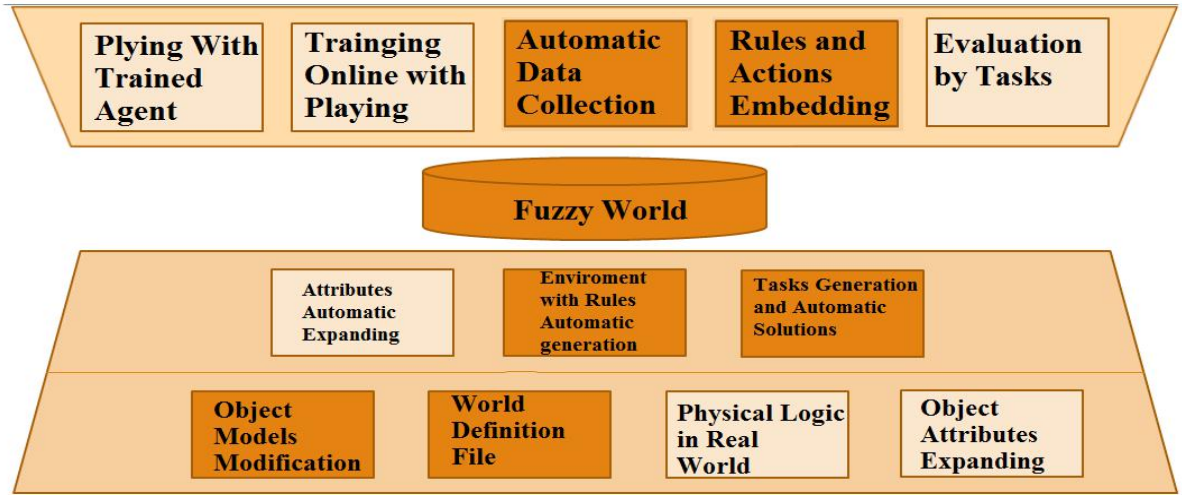


learn to fuse language and images to produce a representation

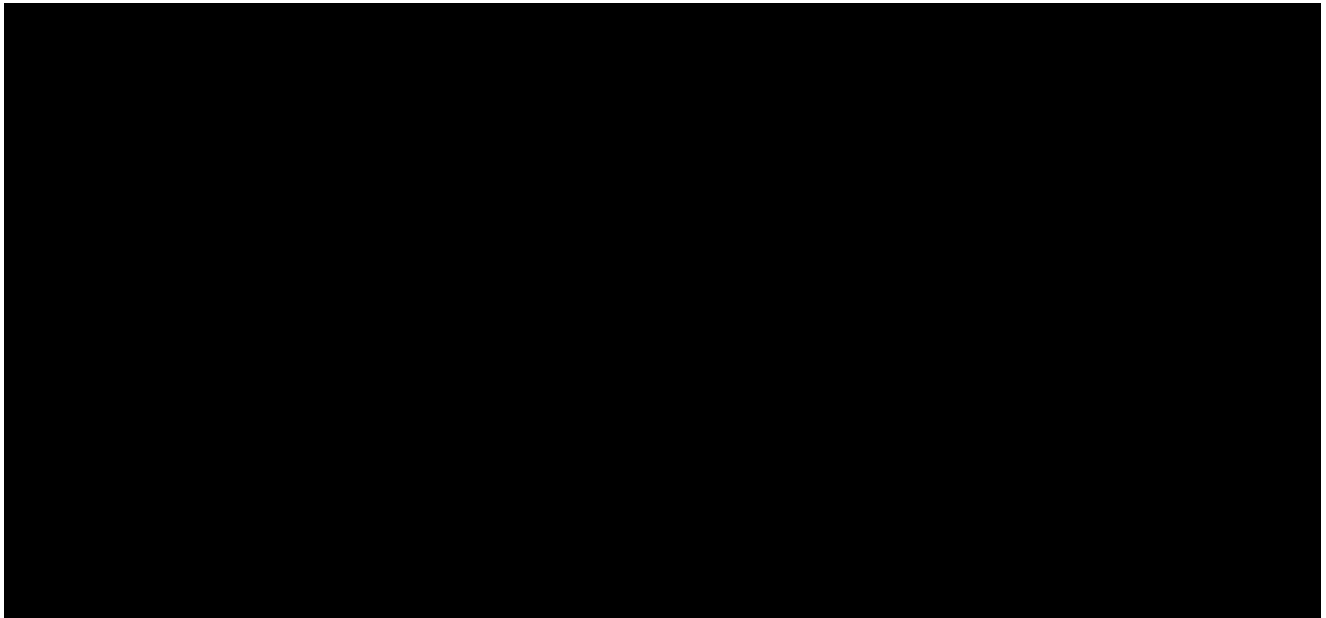
learn the right decision and evaluation

The Fuzzy World: A little tool

- Make worlds with first-order logic tasks

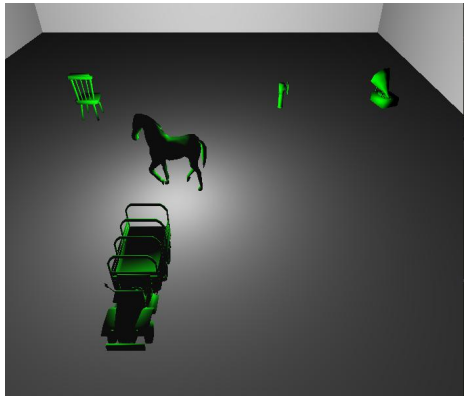


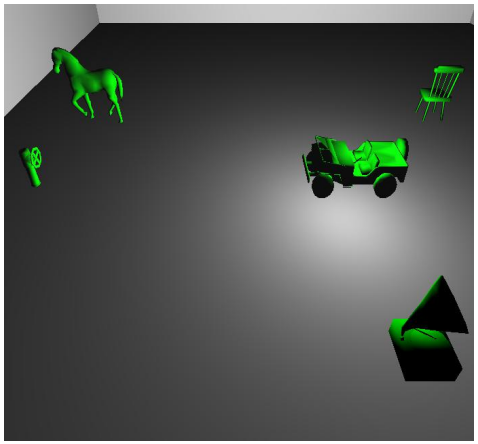
Move(desk)
=>
Move(cow)

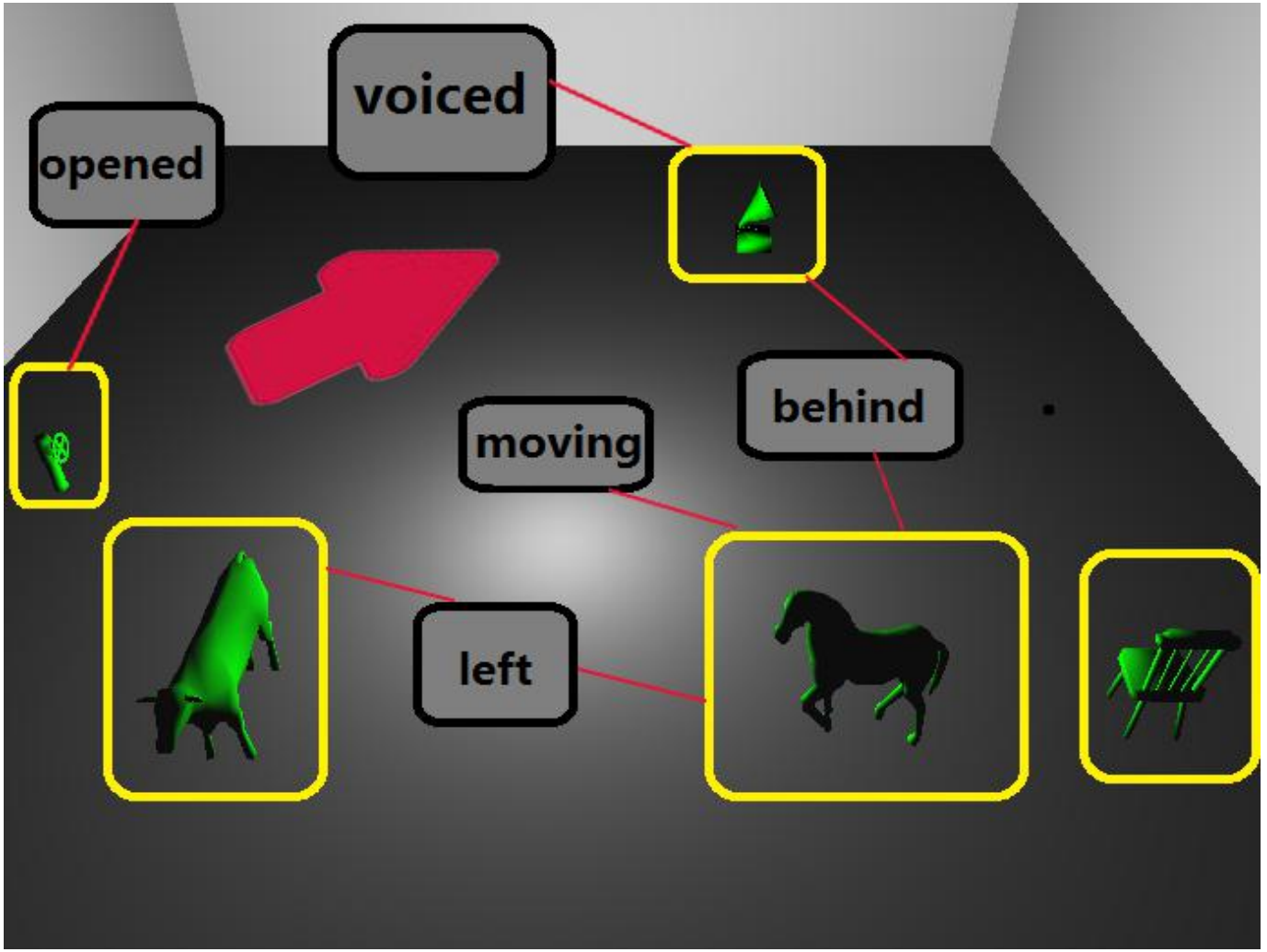


The Fuzzy World: A little tool

- Make worlds with first-order logic tasks

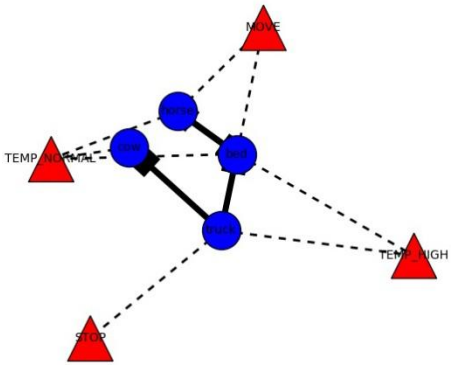


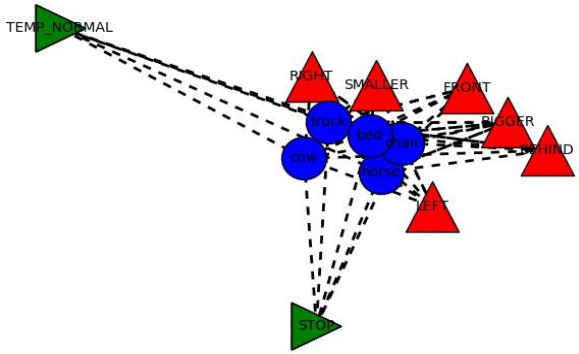




The Semantic Graph

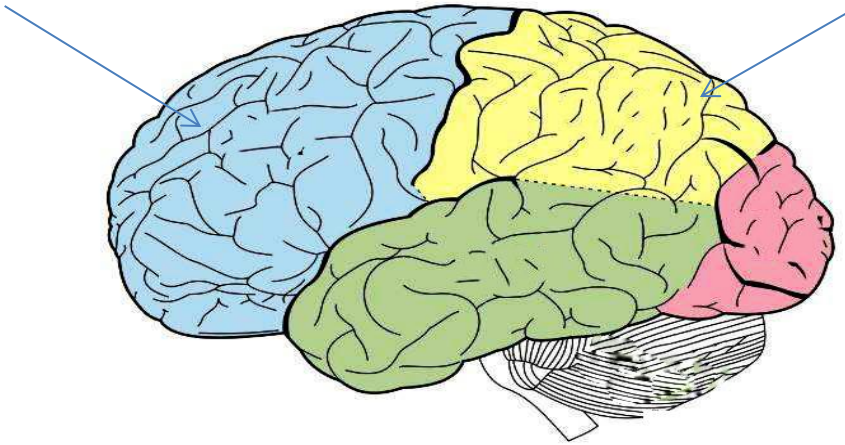
- Understand The World and Reasoning





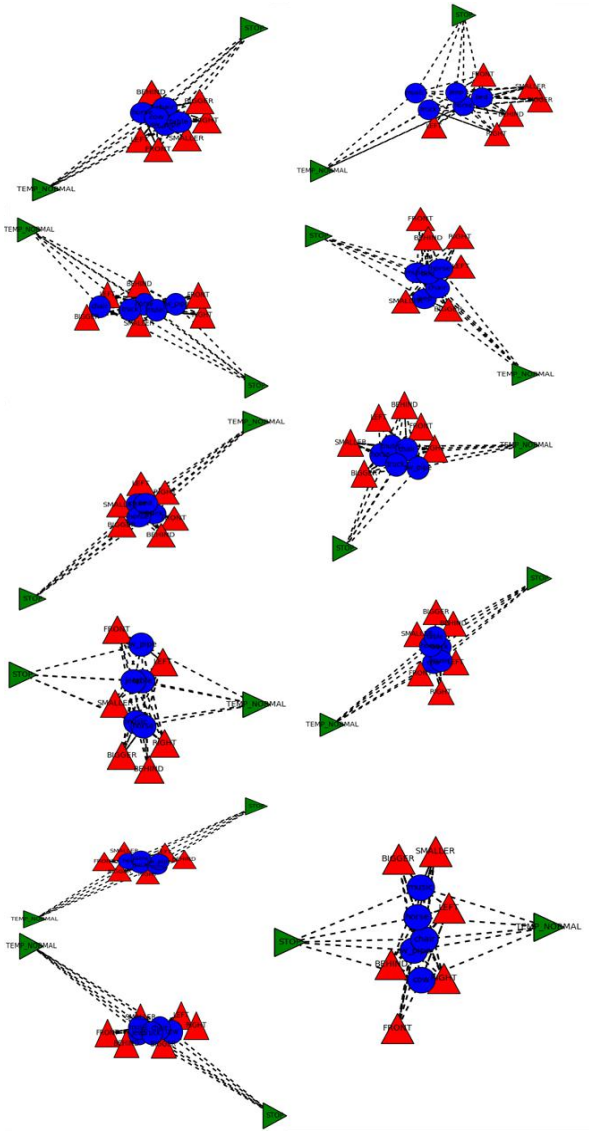
Reasoning $P(G_I|G_S^{(t-1)}, G_S^{(t)})$

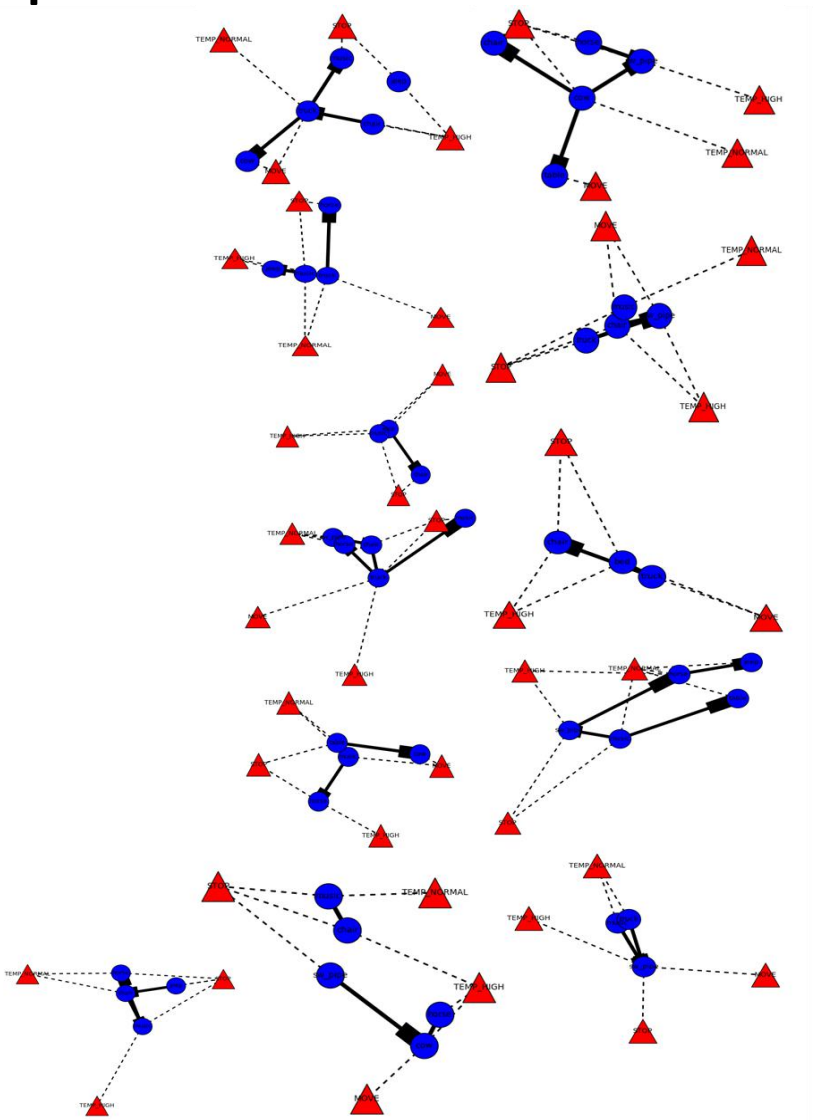
Cognition $P(G_S|V^{(t)}, S^{(t)})$



The Semantic Graph

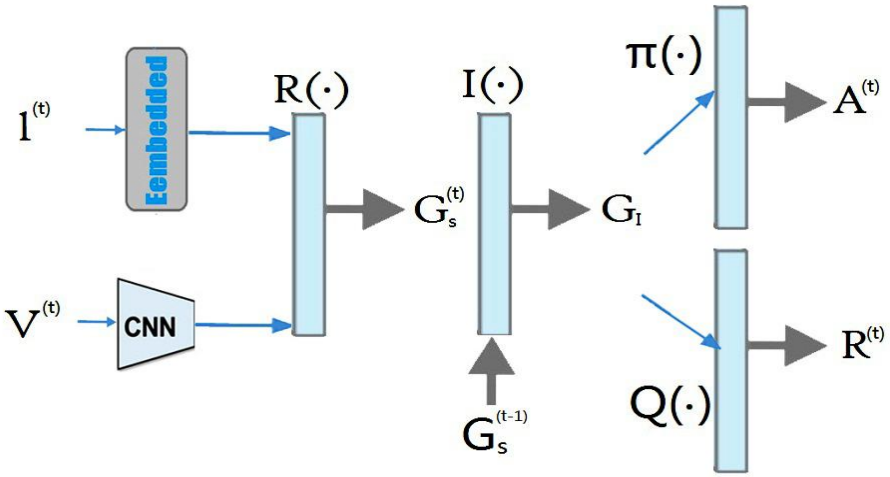
- The Visualization of Graphs from Worlds





The Agent Network with Semantic Graph

- Learn Cognition, Reasoning, Decision, Evaluation.



$$G_S^{(t)} = R_\theta(V^{(t)}, S^{(t)})$$

$$G_I = I_\theta(G_S^{(t-1)}, G_S^{(t)})$$

$$A^{(t)} = \Pi(A^{(t)}|V^{(t)}, S^{(t)}, G_S^{(t)}, G_I)$$

Objective Function

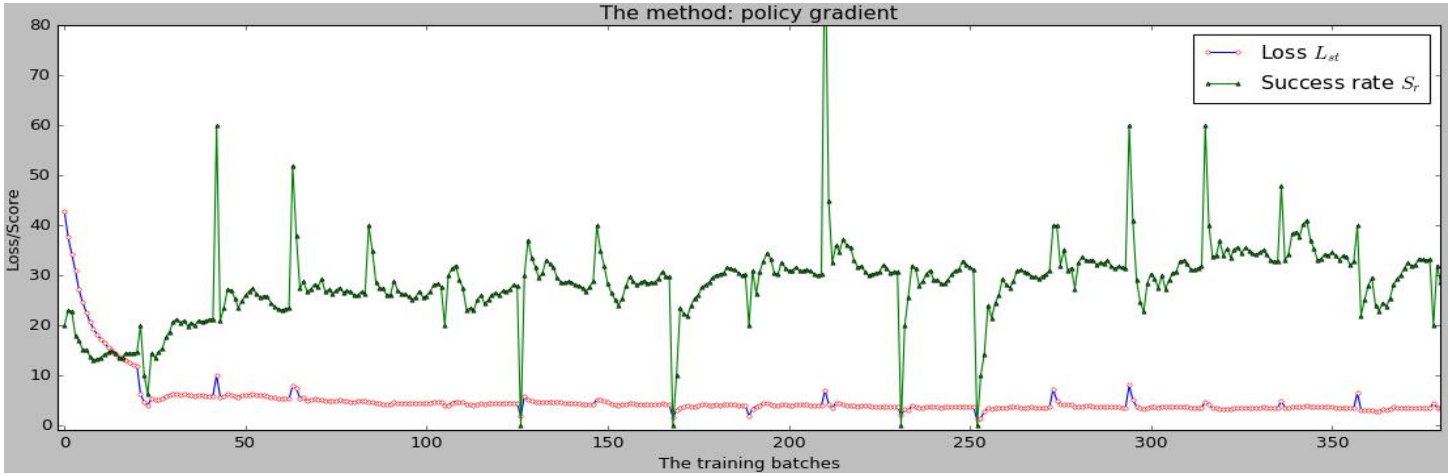
$$\mathbb{E}_{V^{[t]},S^{[t]}\sim Env}[\underbrace{\log \pi_\theta(a^{[t]}|V^{(t)}, S^{(t)}, G_S^{(t)}, G_I)}_{Policy} + \underbrace{\lambda v_\theta(G_S^{(t)}, G_I, S^{(t)})}_{Value\ Evaluation}$$

$$+ \underbrace{\gamma(R_\theta(V^{(t)}, S^{(t)}) + I_\theta(G_S^{(t-1)}, G_S^{(t)}))}_{Cognition\ and\ Reasoning} + \underbrace{\kappa||\theta||}_{norm})]$$

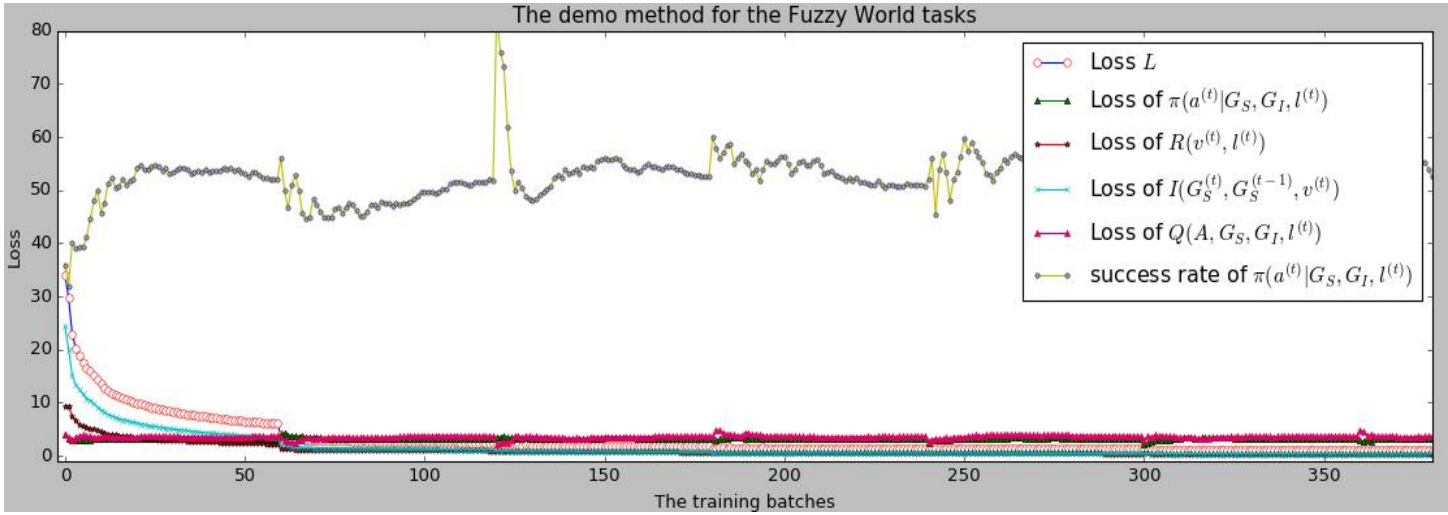
The Agent Network with Semantic Graph

Results

Baidu



Demo Agent



Discussion