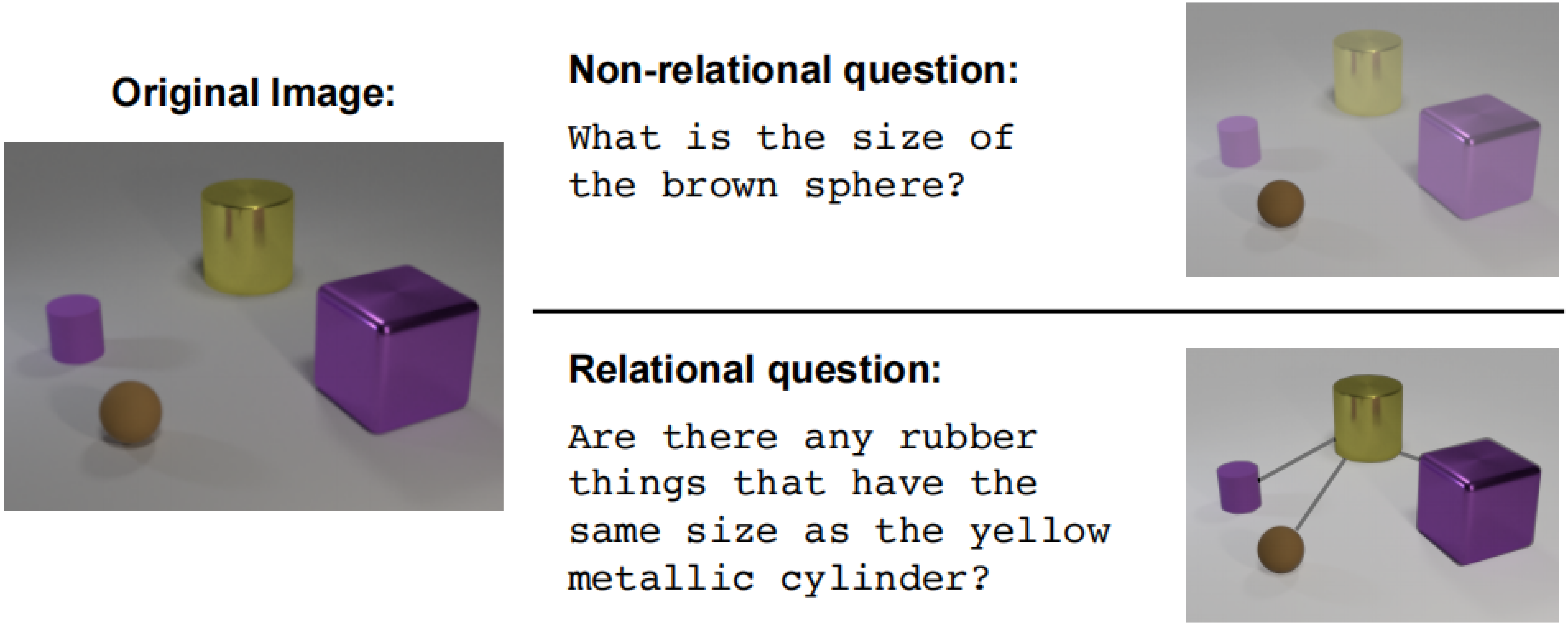
Reasoning over graphs

https://neuralreasoning.github.io/

Presented by Vuong Le

Reasoning on Graphs

- Relational questions: requiring explicit reasoning about the relations between multiple objects



Original Image:

Non-relational question:

What is the size of the brown sphere?

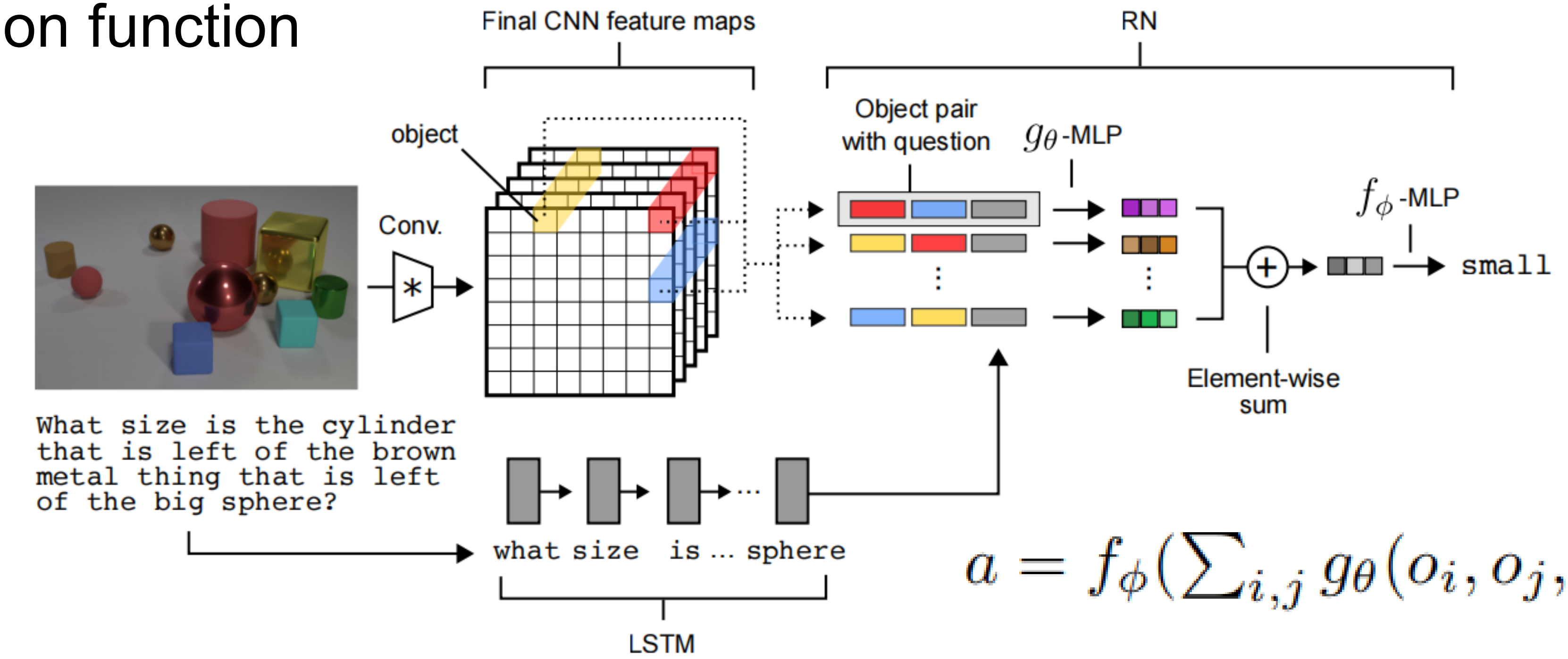
Relational question:

Are there any rubber things that have the same size as the yellow metallic cylinder?

Figure credit: Santoro et al 2017

Relation networks (Santoro et al 2017)

- Relation networks $\mathrm{RN}(O) = f_\phi\left(\sum_{i,j} g_\theta(o_i, o_j)\right)$
- f_ϕ and g_θ are neural functions
- g_θ generate “relation” between the two objects
- f_ϕ is the aggregation function



What size is the cylinder that is left of the brown metal thing that is left of the big sphere?

$$a = f_\phi\left(\sum_{i,j} g_\theta(o_i, o_j, q)\right)$$

→ The relations here are implicit, over-complete, pair-wise

→ inefficient, and lack expressiveness

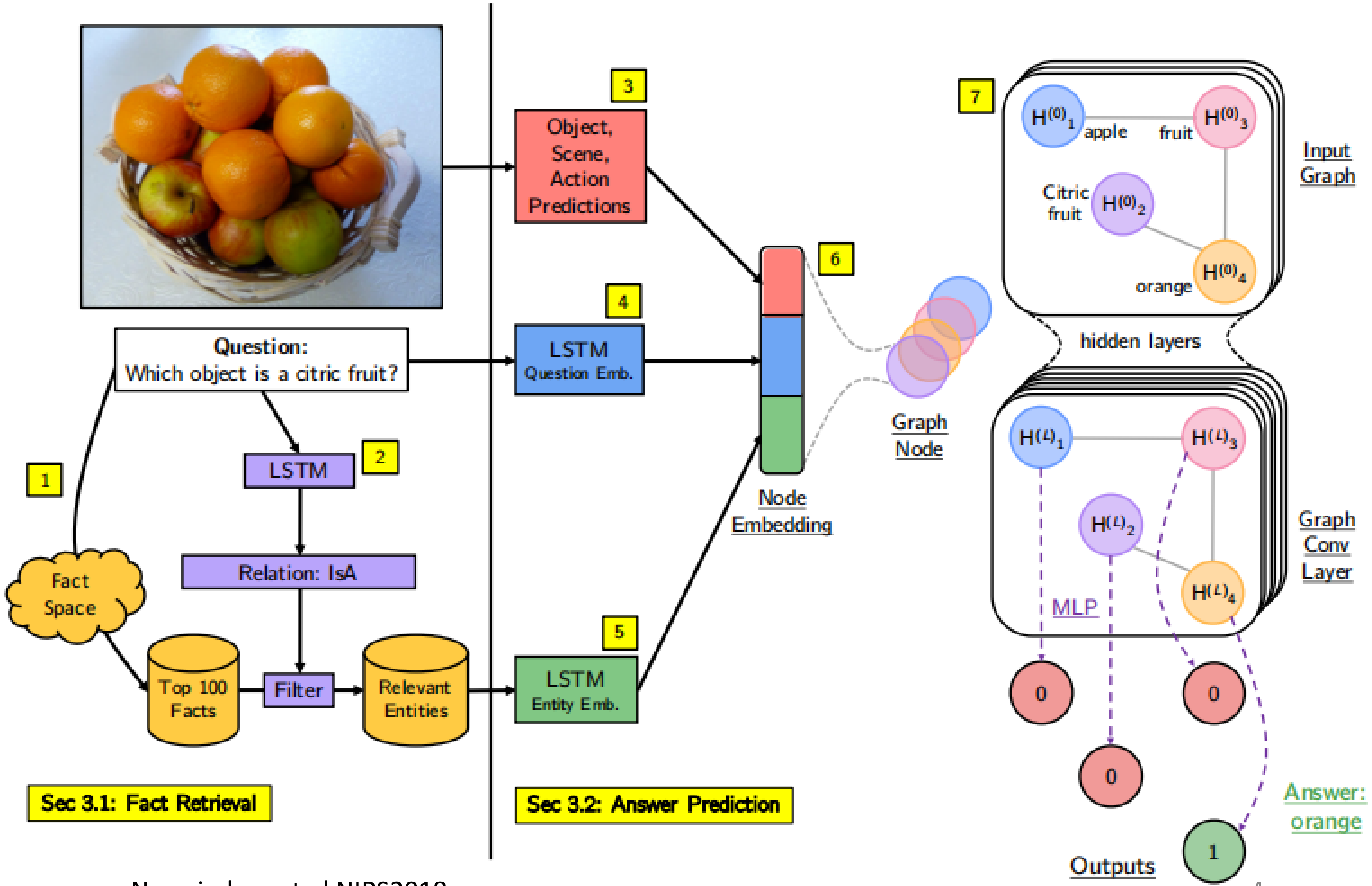
Reasoning with Graph convolution networks

- Input graph is built from image entities and question
- GCN is used to gather facts and produce answer

→ The relations are now explicit and pruned

→ But the graph building is very stiff:

- Unrecoverable from mistakes
- Information during reasoning are not used to build graphs



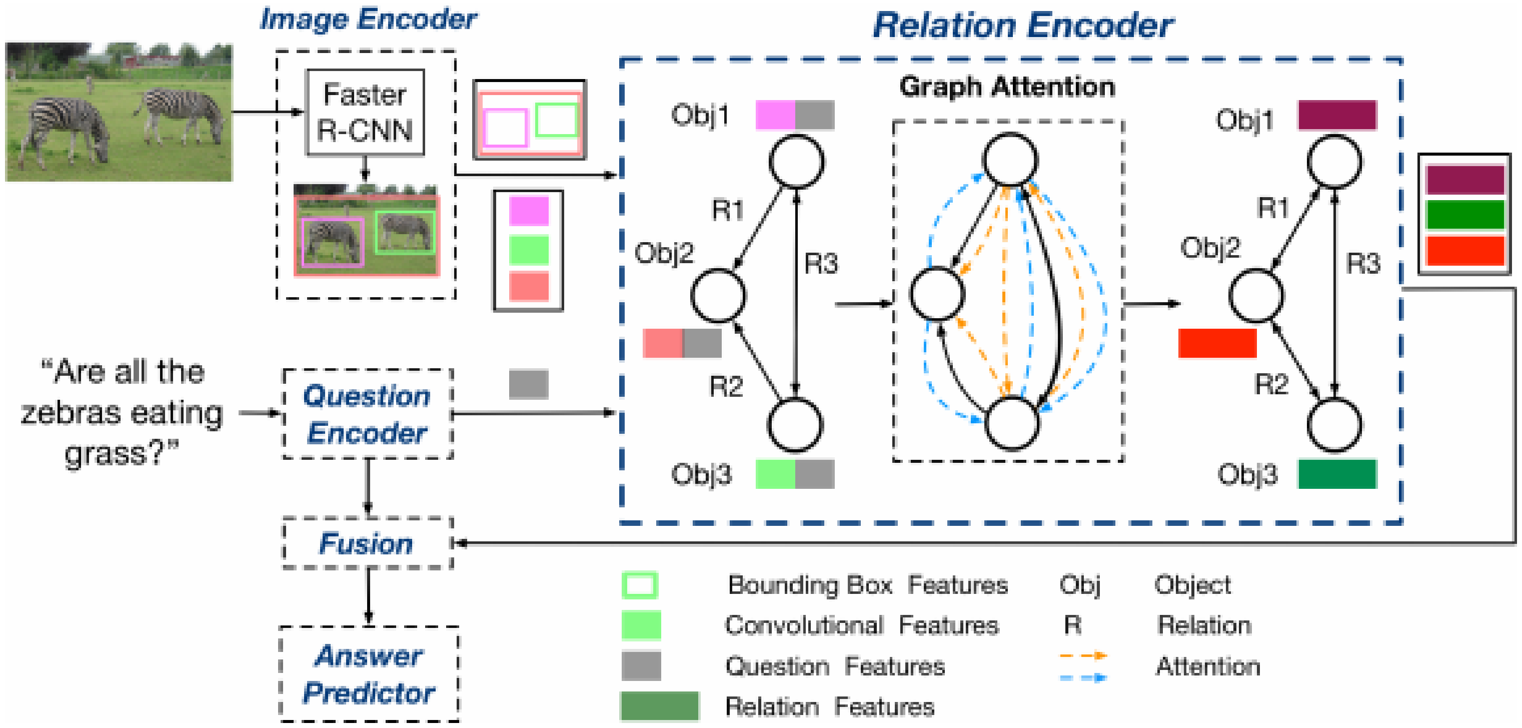
Narasimhan et.al NIPS2018

Reasoning with Graph attention networks

- The graph is determined during reasoning process with attention mechanism

→The relations are now adaptive and integrated with reasoning

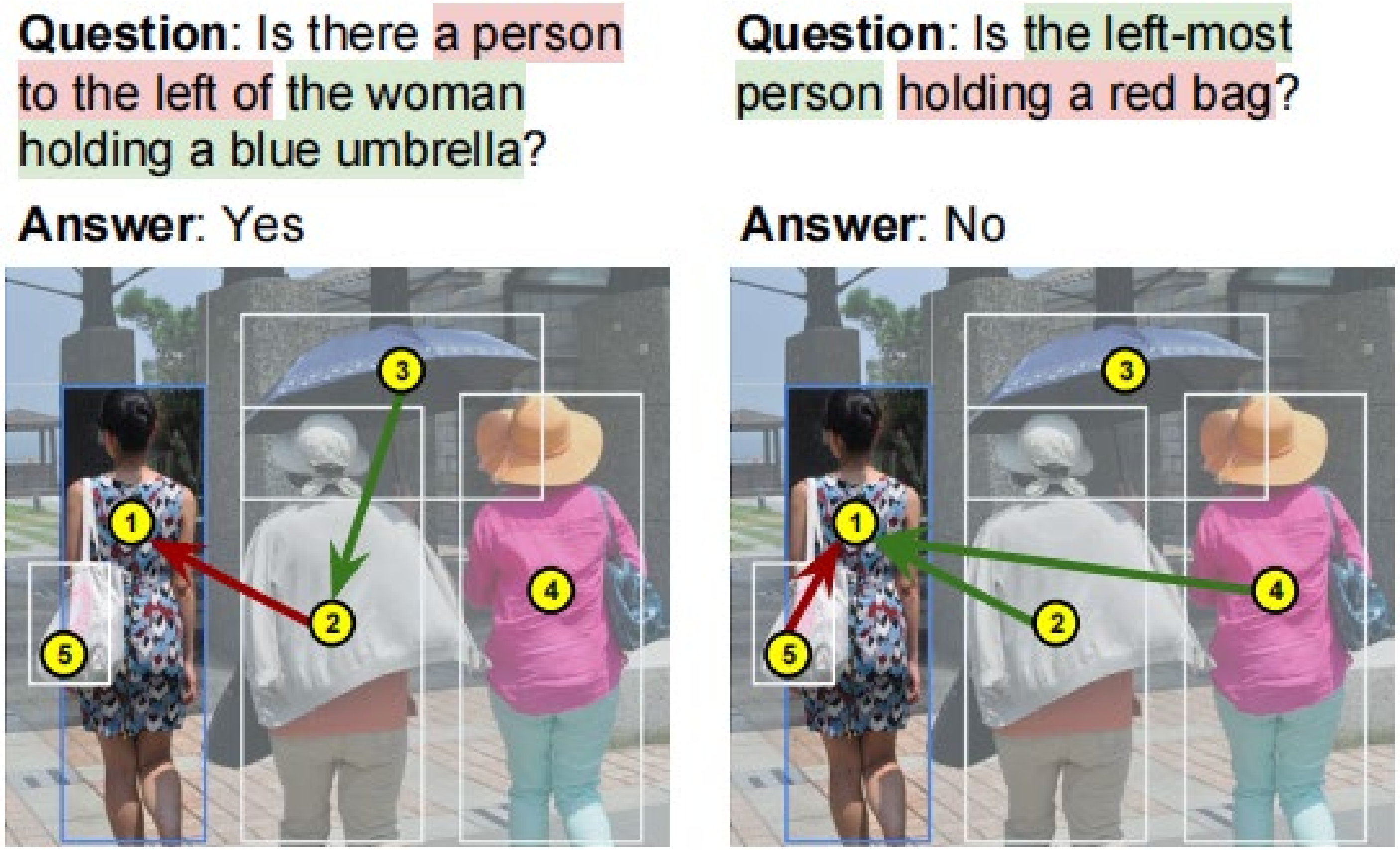
→ Are the relations singular and static?



ReGAT model, Li et.al. ICCV19

Dynamic reasoning graphs

- On complex questions, multiple sets of relations are needed
- We need not only multi-step but also multi-form structures
- Let’s do multiple dynamically–built graphs!



Question: Is there a person to the left of the woman holding a blue umbrella?

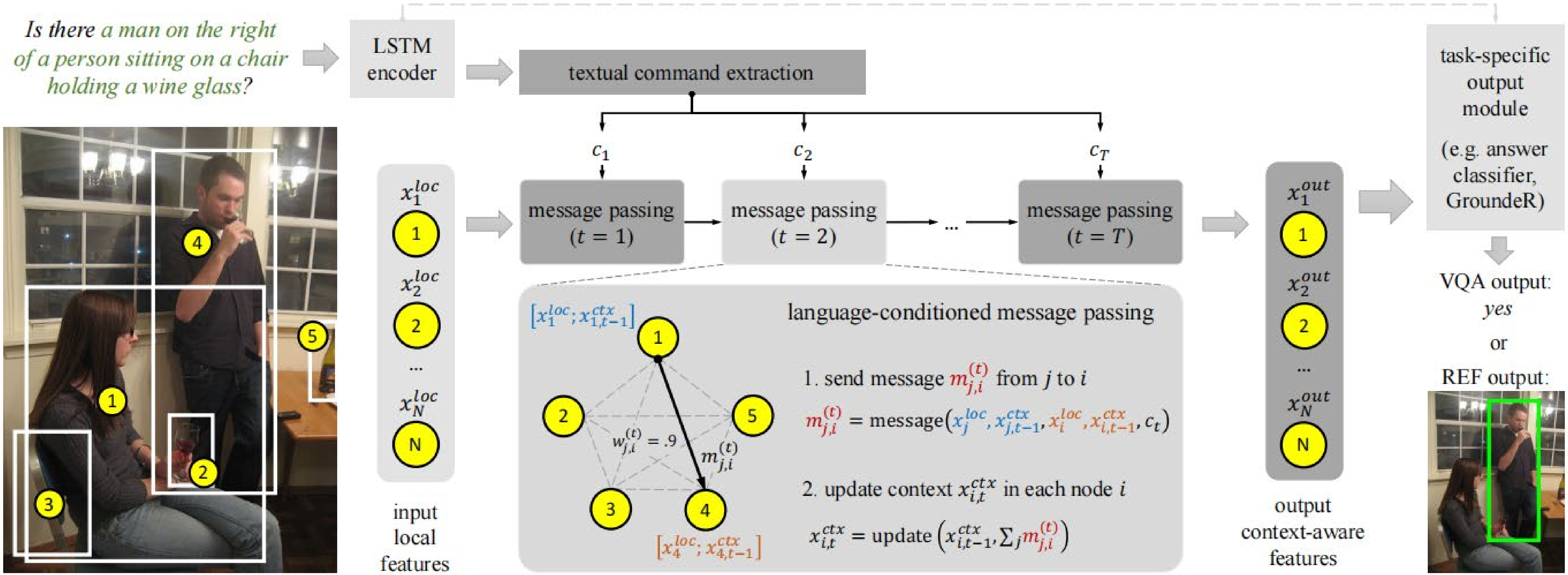
Answer: Yes

Question: Is the left-most person holding a red bag?

Answer: No

LCGN, Hu et.al. ICCV19

Dynamic reasoning graphs


Is there a man on the right of a person sitting on a chair holding a wine glass?
LSTM encoder
textual command extraction
c_1
c_2
c_T
x_1^{loc}
x_2^{loc}
...
x_N^{loc}
1
2
N
3
4
5
input local features
message passing ($t = 1$)
message passing ($t = 2$)
...
message passing ($t = T$)
language-conditioned message passing
$[x_1^{loc}; x_{1,t-1}^{ctx}]$
$w_{j,i}^{(t)} = .9$
$m_{j,i}^{(t)}$
$[x_4^{loc}; x_{4,t-1}^{ctx}]$
1. send message $m_{j,i}^{(t)}$ from j to i
$m_{j,i}^{(t)} = \text{message}(x_j^{loc}, x_{j,t-1}^{ctx}, x_i^{loc}, x_{i,t-1}^{ctx}, c_t)$
2. update context $x_{i,t}^{ctx}$ in each node i
$x_{i,t}^{ctx} = \text{update}\left(x_{i,t-1}^{ctx}, \Sigma_j m_{j,i}^{(t)}\right)$
x_1^{out}
x_2^{out}
x_N^{out}
output context-aware features
task-specific output module
(e.g. answer classifier, GroundeR)
VQA output: yes
or
REF output:

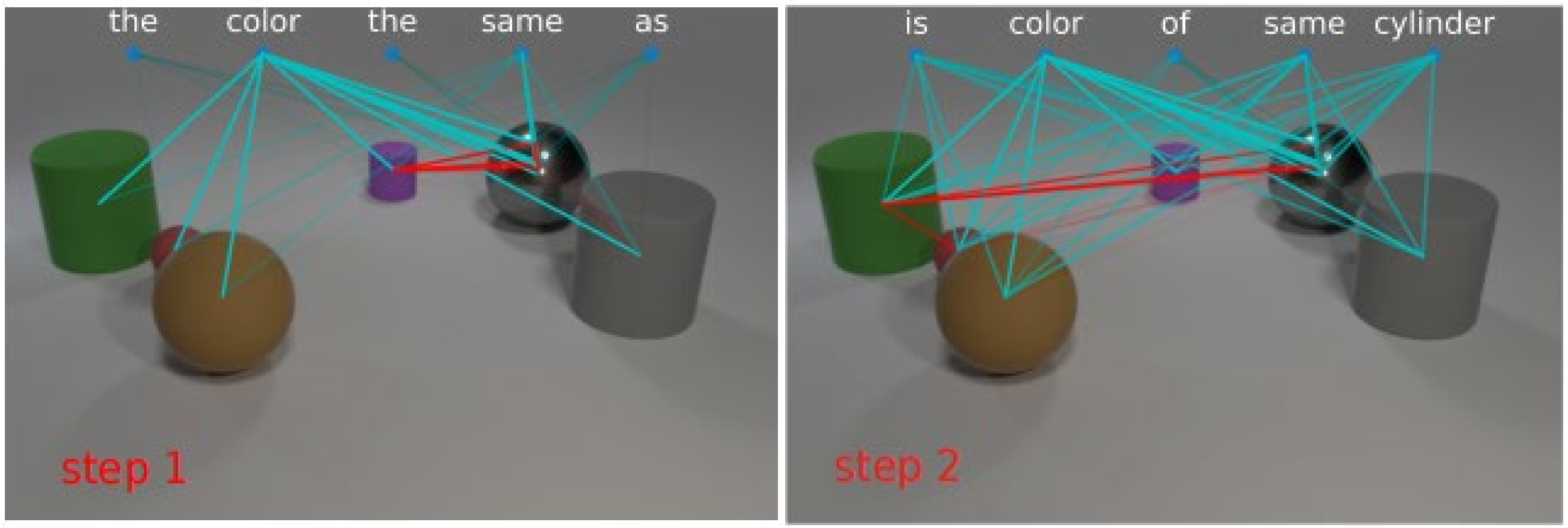
→ The questions so far act as an unstructured command in the process

→ *Aren't their structures and relations important too?*

LCGN, Hu et.al. ICCV19

Reasoning on cross-modality graphs

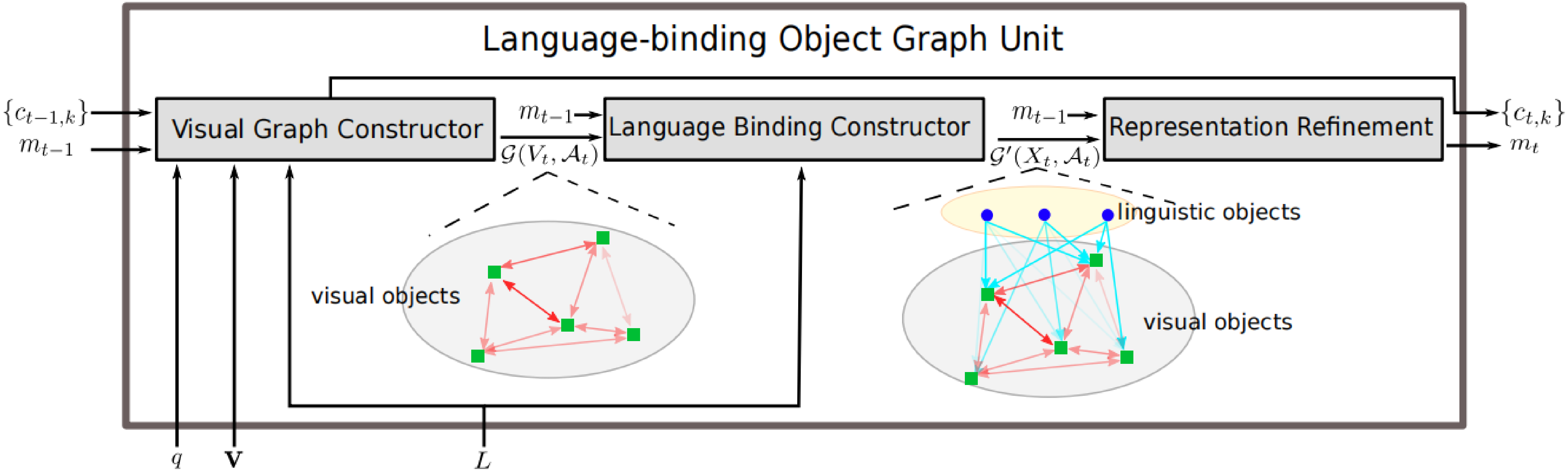
- Two types of nodes: Linguistic entities and visual objects
- Two types of edges:
  - Visual relations
  - Linguistic-visual binding *(as a fuzzy grounding)*
- Adaptively updated during reasoning



LOGNet, T.M Le et.al. IJCAI 2020

Language-binding Object Graph (LOG) Unit

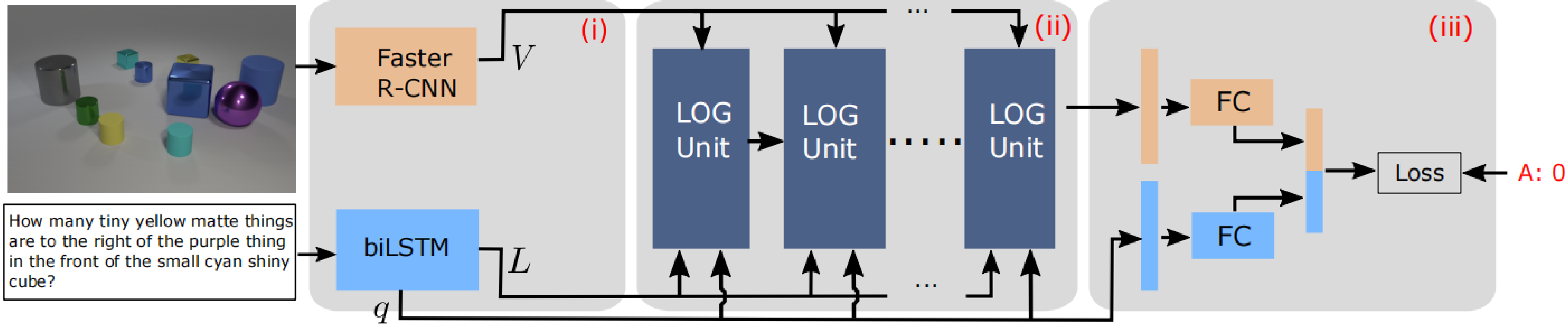
- Graph constructor: build the dynamic vision graph
- Language binding constructor: find the dynamic L-V relations


Language-binding Object Graph Unit
$\{c_{t-1,k}\}$
m_{t-1}
Visual Graph Constructor
m_{t-1}
$\mathcal{G}(V_t, \mathcal{A}_t)$
Language Binding Constructor
m_{t-1}
$\mathcal{G}'(X_t, \mathcal{A}_t)$
Representation Refinement
$\{c_{t,k}\}$
m_t
visual objects
linguistic objects
visual objects
q
$\mathbf{V}$
L

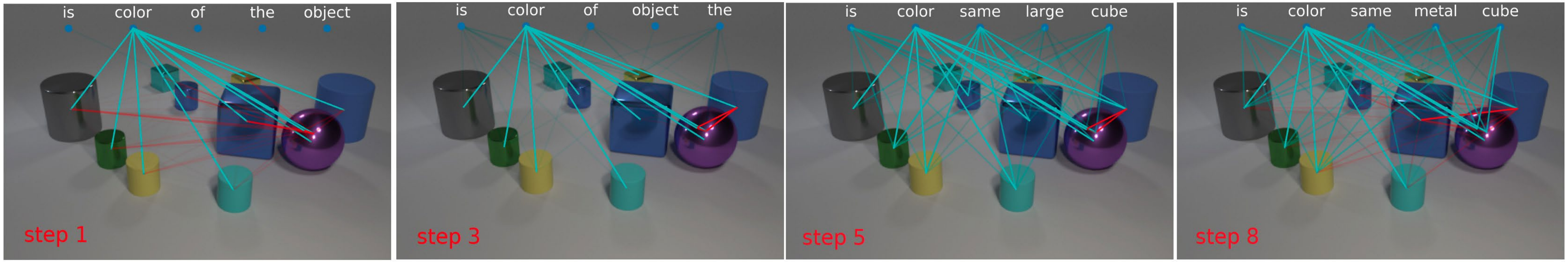
LOGNet, T.M Le et.al. IJCAI2020

LOGNet: multi-step visual-linguistic binding

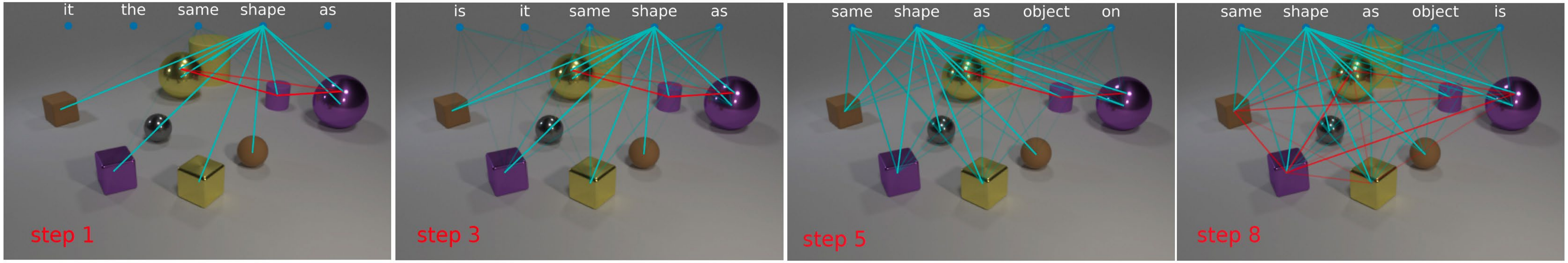
- Object-centric representation ✓
- Multi-step/multi-structure compositional reasoning ✓
- Linguistic-vision detail interaction ✓


How many tiny yellow matte things are to the right of the purple thing in the front of the small cyan shiny cube?

Dynamic language-vision graphs in actions



Question: Is the color of the big matte object the same as the large metal cube?
Prediction: yes **Answer**: yes

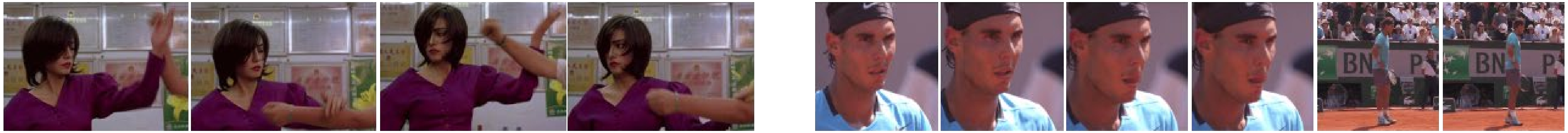


Question: There is a tiny purple rubber thing; does it have the same shape as the brown object that is on the left side of the rubber sphere?
Prediction: no **Answer**: no

We got sets and graphs, how about sequences?

- Videos pose another challenge for visual reasoning: the dynamics through time.
- Sets and graphs now becomes sequences of such.
- Temporal relations are the key factors
- The size of context is a core issue

→Lecture 8 will address these



(a) Question: What does the girl do 9 times?

Ground truth: blocks a person's punch

(b) Question: What does the man do before turning body to left?

Ground truth: breath