

Learning to Compose Neural Networks for Question Answering

발표자 : 정민지

Dynamic Neural Module Network

기존의 NMN(Neural Module Network)를 개선 & 확장

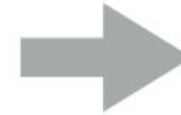
1) Learn

- network structure predictor jointly with module parameters themselves

2) Extend

- Visual Question Answering
 - Knowledge base와 같은 structured world representation까지 확장
- input : (World, Question, Answer)

*What color is
the necktie?*



yellow

*What rivers
are in South
Carolina?*

name	type	coastal
<i>Columbia</i>	city	no
<i>Cooper</i>	river	yes
<i>Charleston</i>	city	yes



Cooper

Neural Module Network의 한계

1. Syntactic parser의 사용
 - eg) dog & puppy
2. Fixed modules
 - heuristic을 사용하여 모듈의 순서와 구조를 제한

Neural Module Network의 한계

1. Syntactic parser의 사용

- eg) dog & puppy

→ Semantic parser 사용

2. Fixed modules

- heuristic을 사용하여 모듈의 순서와 구조를 제한

→ End-to-End 방식을 통해 network 스스로 layout을 구성 & layouts들을 비교하여 선택할 수 있도록 함(dynamic)

Model - variables

- Variables

w a world representation

x a question

y an answer

z a network layout

θ a collection of model parameters

Models - distributions

- Layout model : chooses a layout

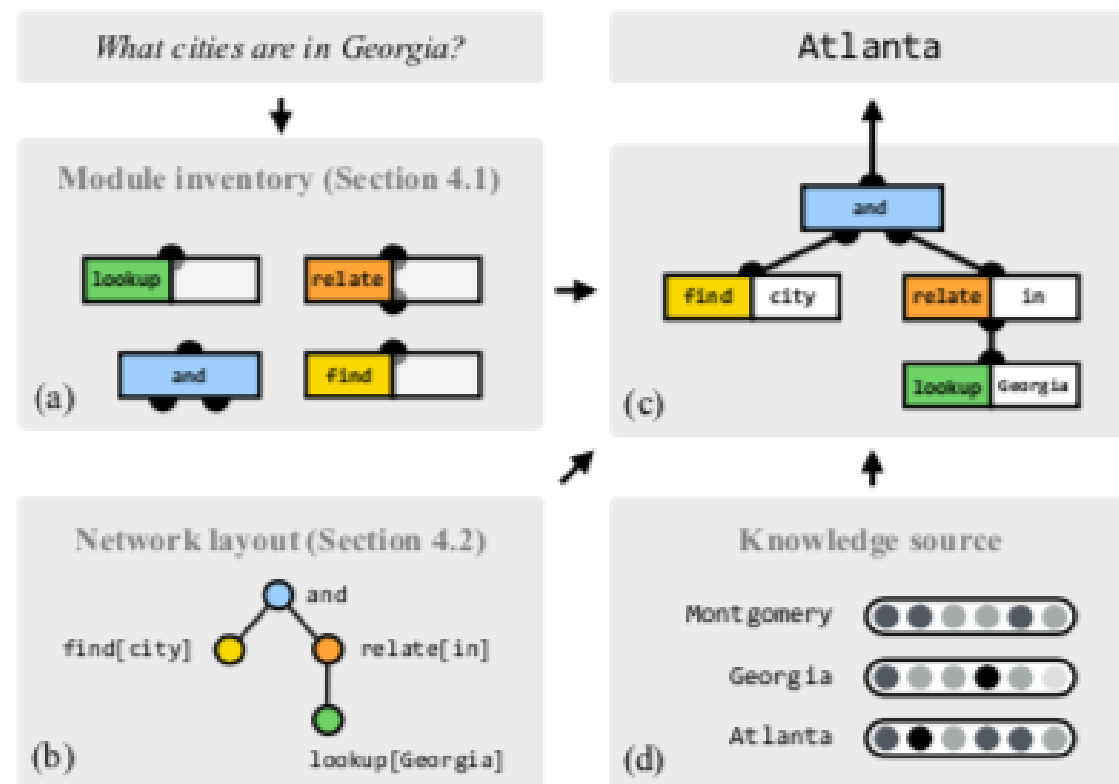
$$p(z|x; \theta_\ell)$$

- Execution model : applies the network specified by z to w

$$p_z(y|w; \theta_e)$$

Evaluating Modules

- Assume a layout z is given,
 1. assemble the corresponding modules into a full neural network
 2. apply the network to the knowledge representation



Module Inventory

- Type Constraints
 - operating directly on the input representation
eg) find
 - depending on input from specific earlier modules
eg) describe
- Attention : a distribution over pixels or entities
- Labels : a distribution over answers

Module Inventory

- Parameter arguments [args]
 - provided by the layout
 - used to specialize module behavior for particular lexical items
- Ordinary inputs (inputs)
 - the result of computation lower in the network

Module Inventory

1. Lookup

lookup[i] (\rightarrow Attention)

- producing an attention

$$\llbracket \text{lookup}[i] \rrbracket = e_{f(i)}$$

Module Inventory

2. Find

find[i] (\rightarrow Attention)

- computing a distribution over indices

$$[\text{find}[i]] = \text{softmax}(a \odot \sigma(Bv^i \oplus CW \oplus d))$$

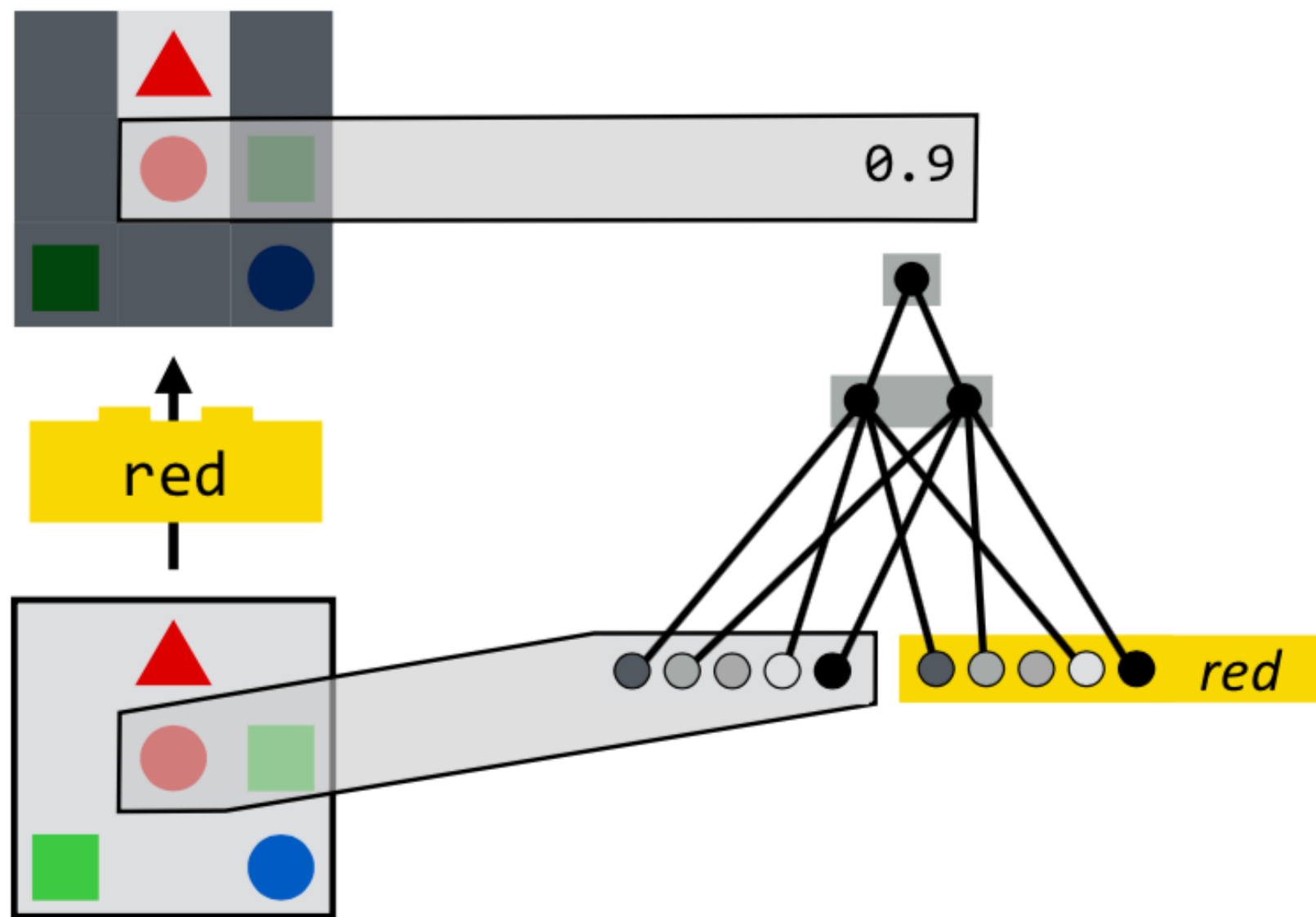


Is there a **red** shape above a **circle**?

What color is the **triangle**?

Who is **running** in the **grass**?

What **cities** are south of **San Diego**?



Module Inventory

3.Relate

relate (Attention \rightarrow Attention)

- move focus from one to another

$$\llbracket \text{relate}[i](h) \rrbracket = \text{softmax}(a \odot \sigma(Bv^i \oplus CW \oplus D\bar{w}(h) \oplus e))$$



*Is there a red shape **above** a circle?*

What color is the triangle?

*Who is running **in** the grass?*

*What cities are **south of** San Diego?*

Module Inventory

4. And

and $(\text{Attention}^* \rightarrow \text{Attention})$

- setting intersection for attentions

$$\llbracket \text{and}(h^1, h^2, \dots) \rrbracket = h^1 \odot h^2 \odot \dots$$



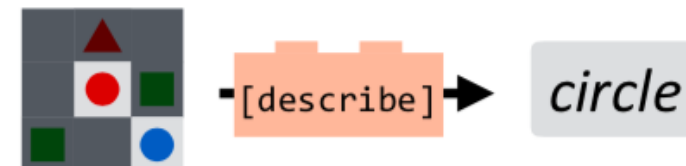
Module Inventory

5. Describe

describe[i] (Attention → Labels)

- computing a weighted average of w under the input attention
- used to predict an answer representati

$$[\text{describe}[i](h)] = \text{softmax}(A\sigma(B\bar{w}(h) + v^i))$$

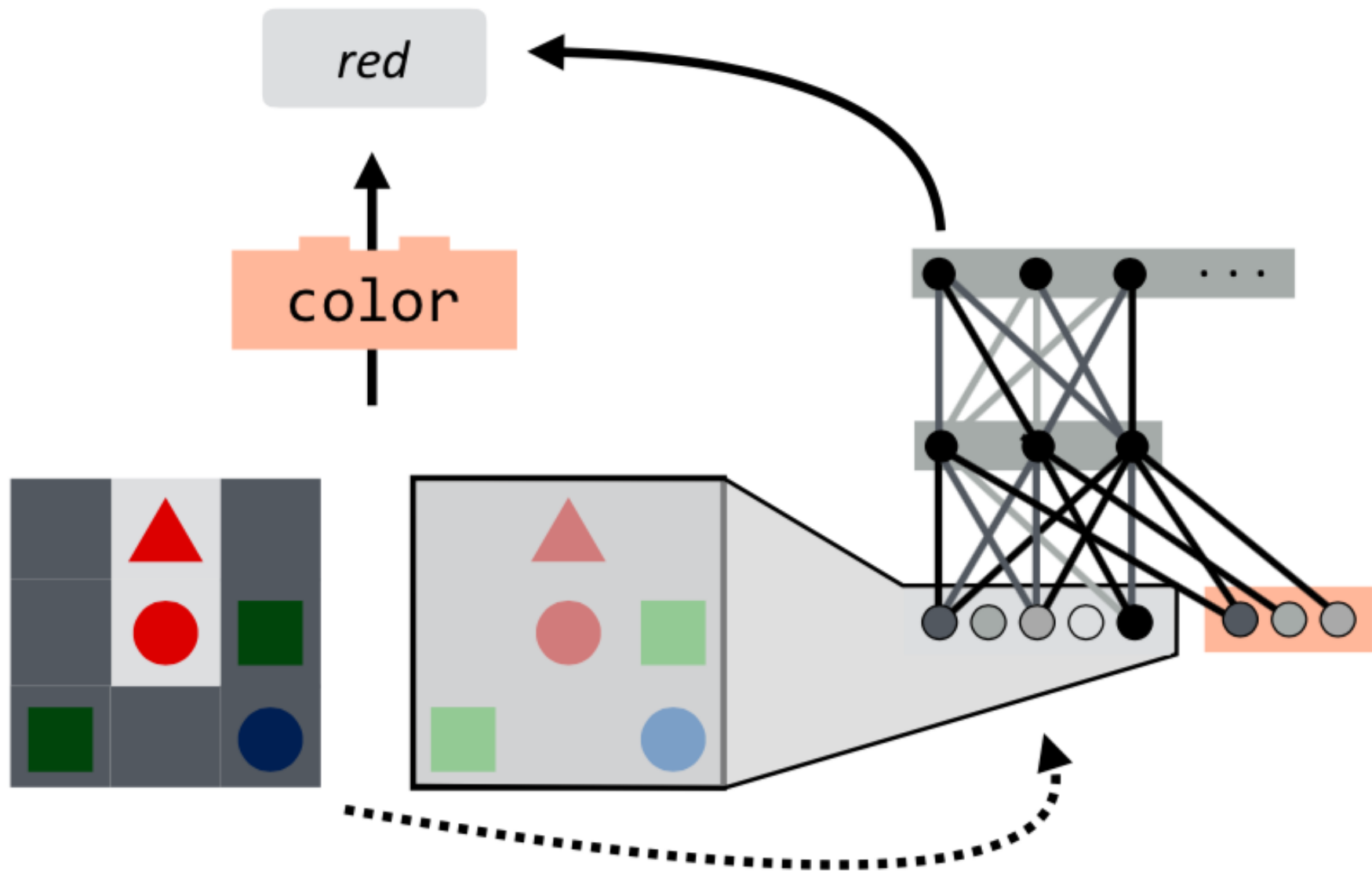


Is there a red shape above a circle?

What color is the triangle?

Who is running in the grass?

What cities are south of San Diego?



Module Inventory

6. Exists

exists (Attention → Labels)

- existential quantifier
- inspects the incoming attention directly to produce a label, rather than an intermediate feature vector

$$[\text{exists}](h) = \text{softmax}\left(\left(\max_k h_k\right)a + b\right)$$



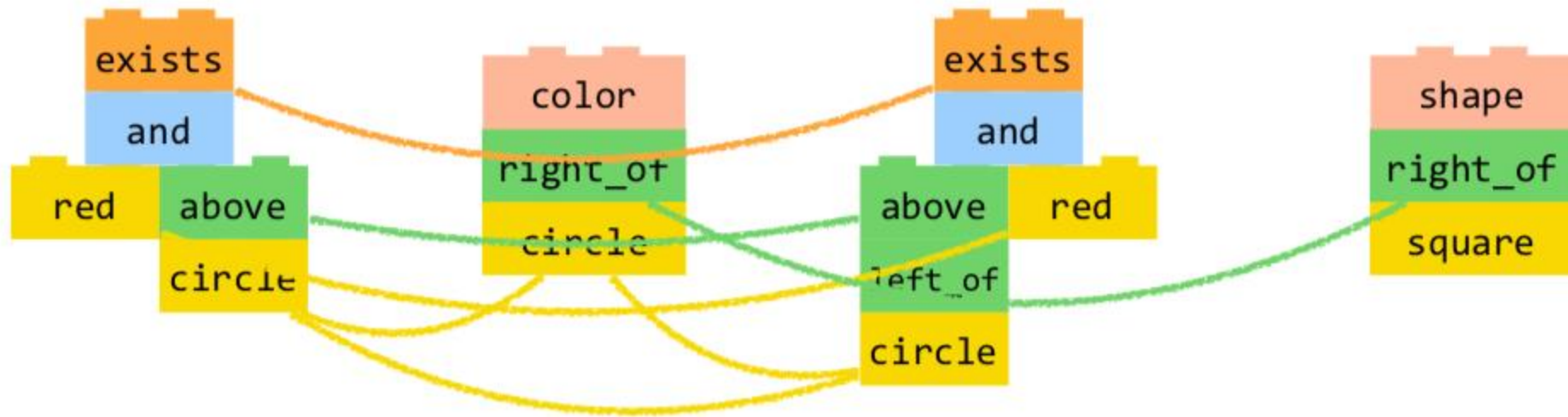
Is there a red shape above a circle?

What color is the triangle?

Who is running in the grass?

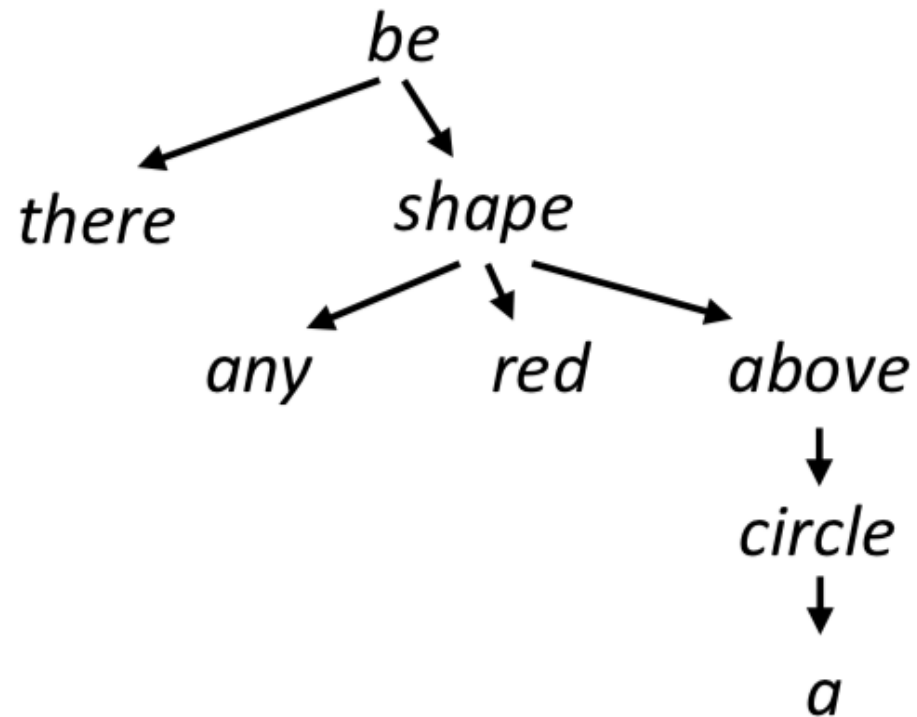
What cities are south of San Diego?

Module Inventory



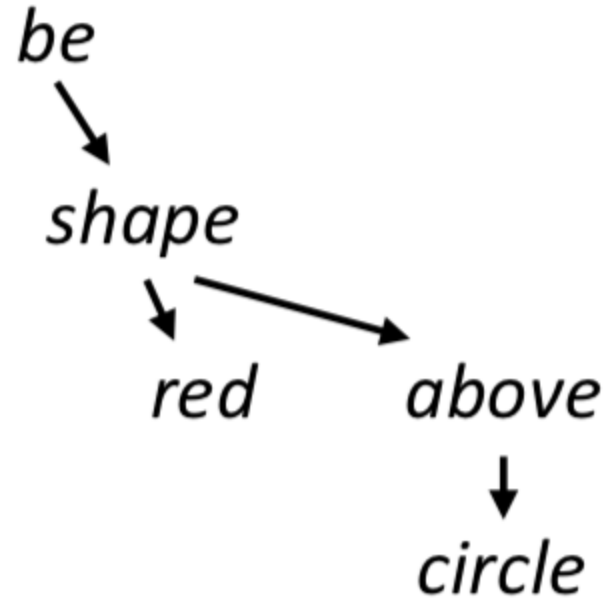
Assembling Networks

Is there a red shape above a circle?



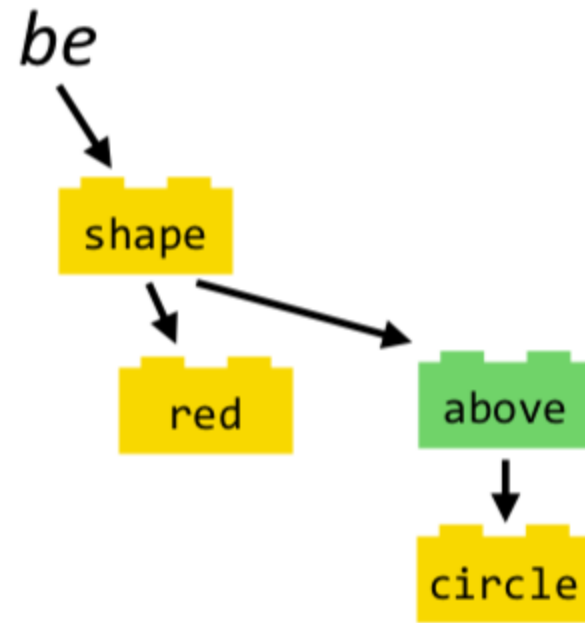
Assembling Networks

Is there a red shape above a circle?



Assembling Networks

Is there a red shape above a circle?



Assembling Networks

Is there a red shape above a circle?

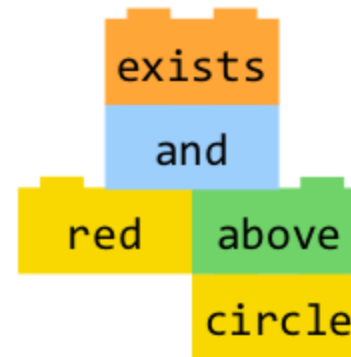
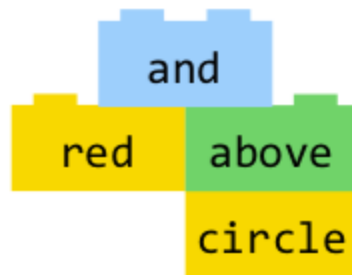
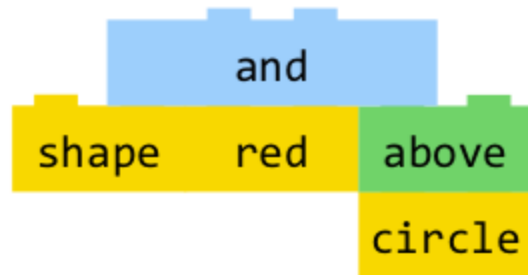
shape

red

above
circle

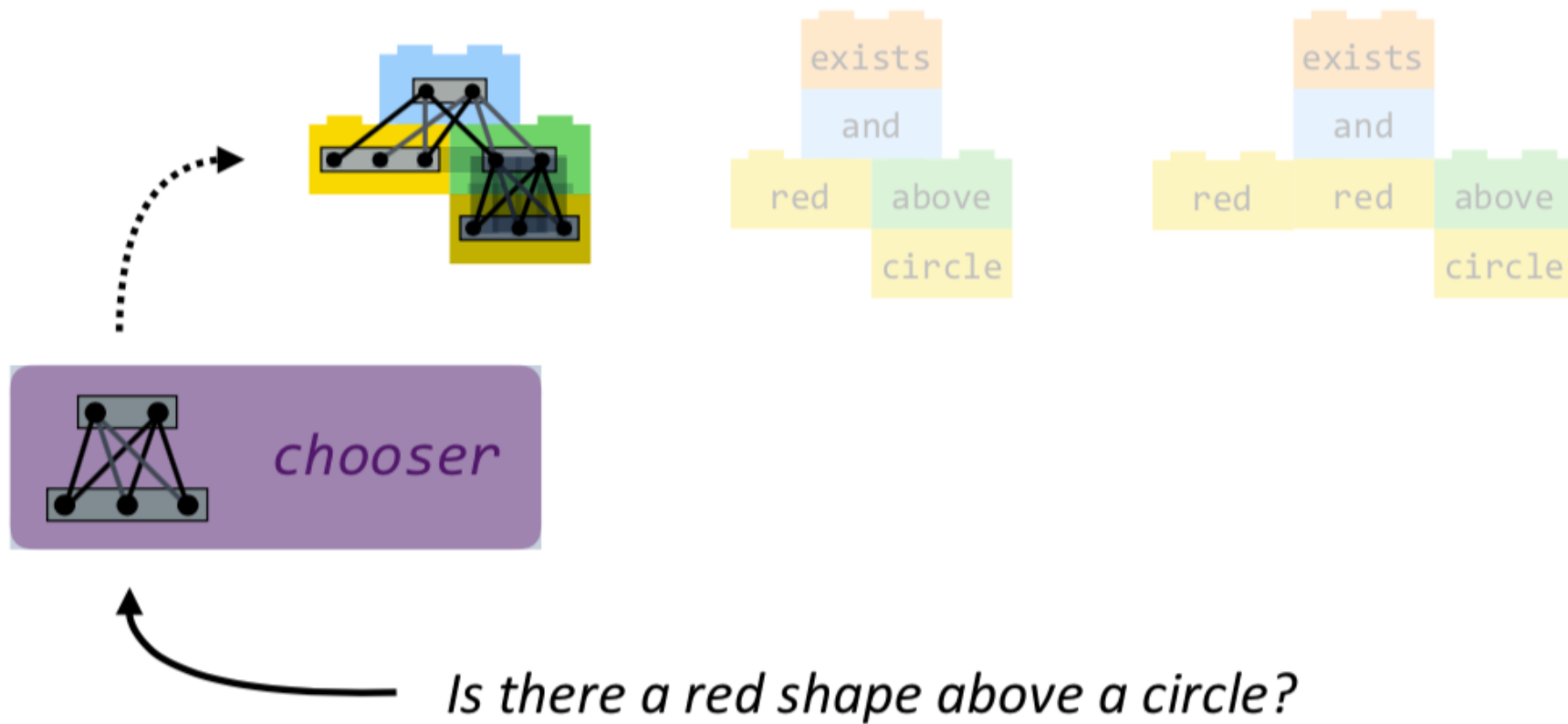
Assembling Networks

Is there a red shape above a circle?



...

Assembling Networks



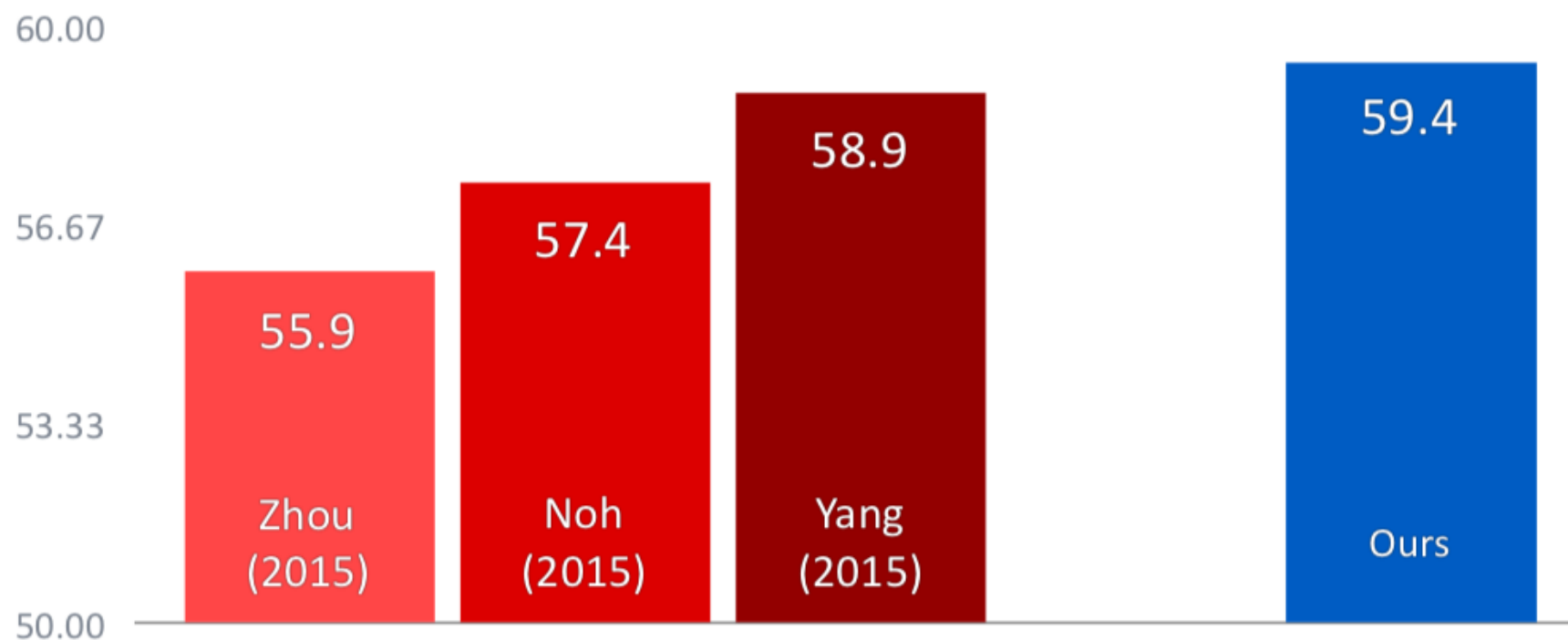
Experiments

1. Question about images

- VQA datasets
 - 200,000 images
 - input images are scaled to 448 x 448
- weakness
 - the candidate layouts were relatively simple
 - prior knowledge
 - Eg) most bears are brown

Experiments

1. Question about images



Experiments

2. Questions about geography

- GeoQA Datasets

- 263 examples
- entities & relations
- GeoQA+Q

Eg) What cities are in Texas?

Are there any cities in Texas?

Experiments

2. Question about geography

*What are some
beaches in Florida?*

name	type	coastal
Columbia	city	no
Cooper	river	yes
Charleston	city	yes



Daytona Beach

*Is Key Largo
an island?*

name	type	coastal
Miami	city	no
Daytona Beach	city	yes
Everglades	park	no



Yes

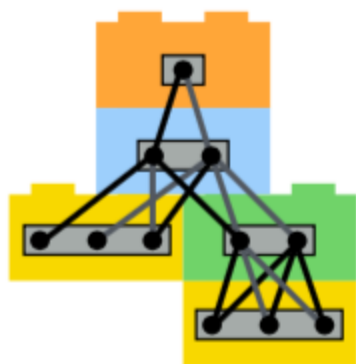
Experiments

2. Question about geography



Conclusion

Linguistic structure dynamically generates model structure



Combines advantages of:

- Representation learning (like a neural net)
- Compositionality (like a semantic parser)