



A Simple Neural Network For Relational Reasoning

AILAB 엄희송

A Simple Neural Network For Relational Reasoning

DeepMind에서 2017년에 발표한 논문

Simple Neural Network이라고 해서 골랐는데... 16장...

RN 자체에 대한 자세한 설명은 Discovering objects and their relations from entangled scene representations 에서...

이 논문에서는 구현한 RN 모델의 결과 위주로 focus

Object 사이의 relation, 즉 관계를 추론하는 것을 학습하는 것이 목표



Object 사이의 relation, 즉 관계를 추론하는 것을 학습한다.



Day 001

재석, 명수, 준하, 하하, 광희 다섯 사람은 무한상사의 직원입니다. 얼마 전부터 무한상사에서는 지각이 잦은 직원들에게 패널티를 주기 위해 출근 순서에 따라 상금을 주기로 했는데요. 다섯 사람이 모두 사실만을 말할 때, 네 번째로 출근한 사람은 누구일까요?

재석 : 나는 광희, 명수보다 늦게 출근했다.

명수 : 나는 광희보다 늦게, 준하보다 빨리 왔다.

준하 : 나는 재석보다 빨리 출근했다.

하하 : 내 앞으로 명수가 들어오고 재석이 뒤따라 출근하는 걸 봤다.

광희 : 내가 오자마자 준하가 들어왔다.

- ① 재석
- ② 명수
- ③ 준하
- ④ 하하
- ⑤ 광희

Symbolic approach

GOFAI

Symbolic grounding problem

Statistical learning approach

deep learning

data-poor problem

Symbolic approach

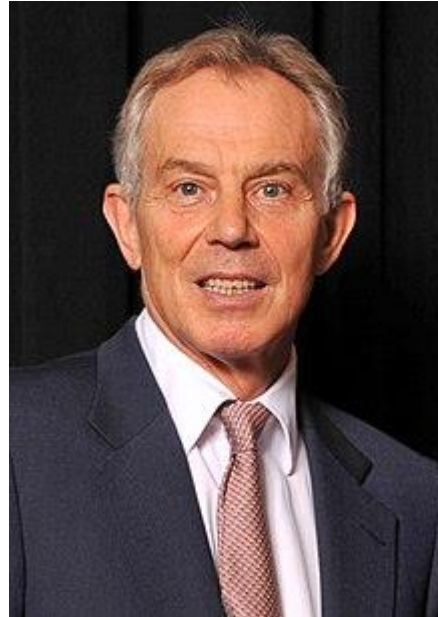
GOFAI

Symbolic grounding problem

Statistical learning approach

deep learning

data-poor problem



Tony Blair

The UK's former prime minister

Cheri Blair's husband

Symbolic approach

GOFAI

Symbolic grounding

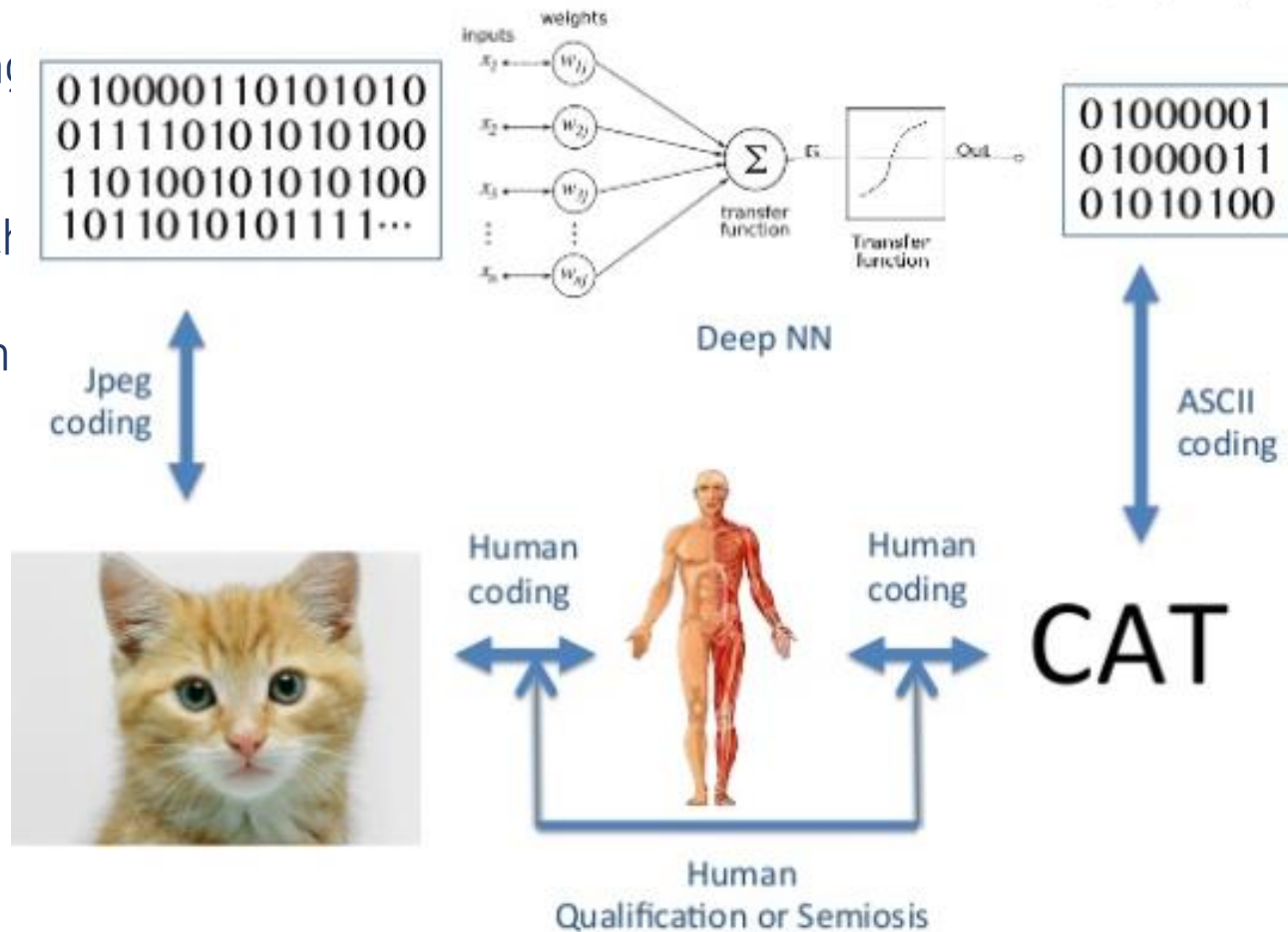
Statistical learning approach

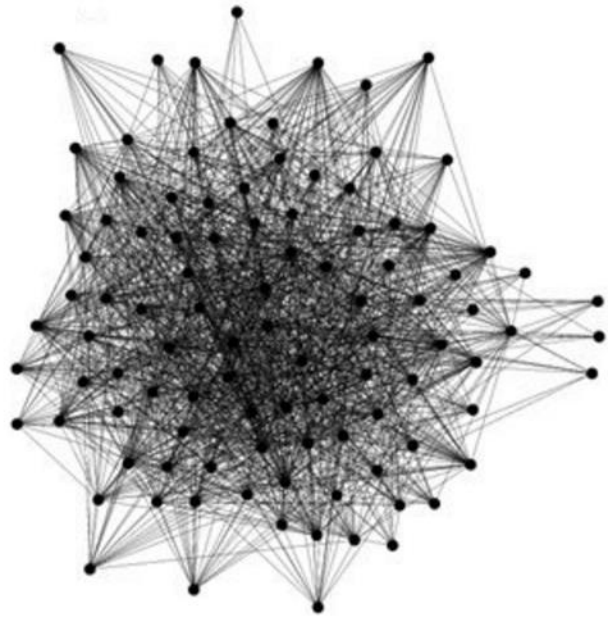
deep learning

data-poor problem

The Symbol Grounding Problem

Harnad, S. (1990)





Dense Graph



Sparse Graph

Past solutions

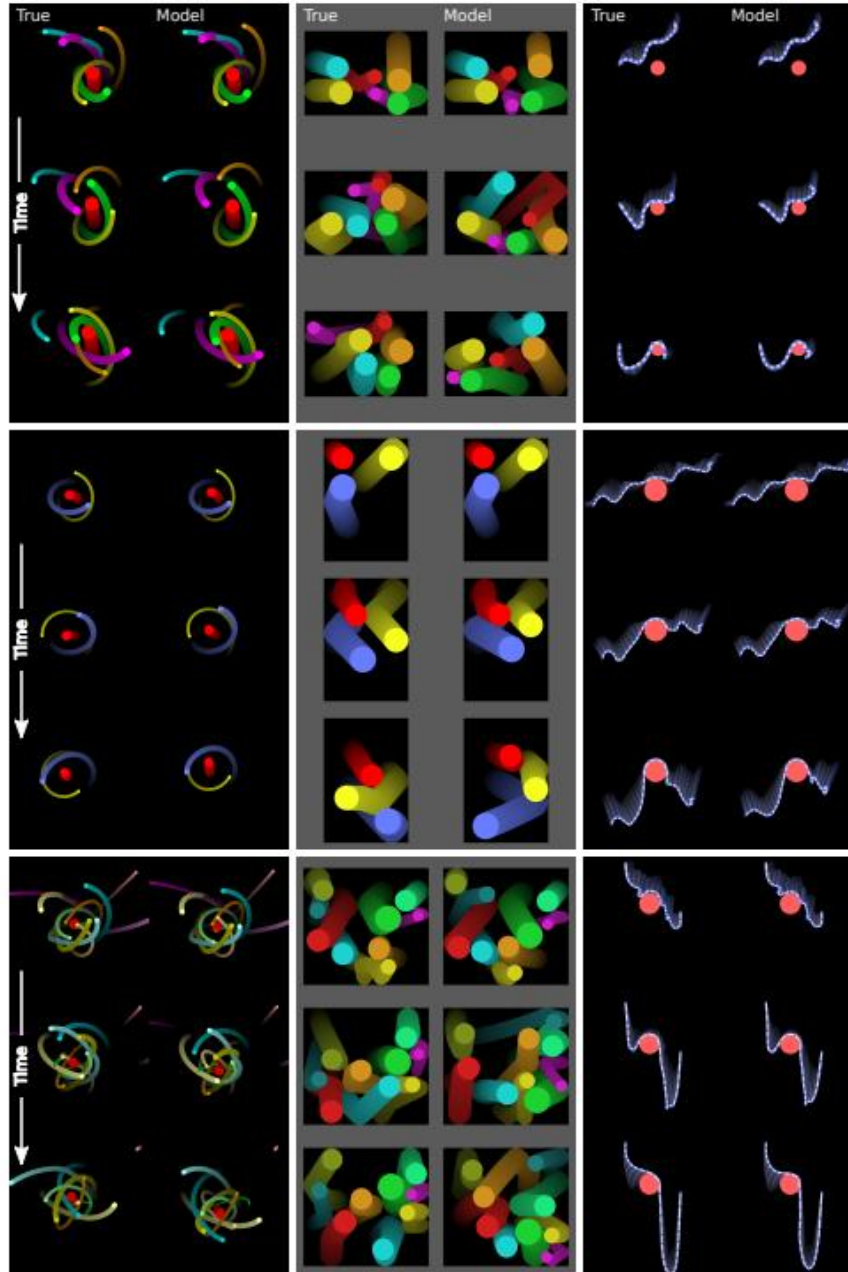
Graph Neural Networks, Gated Graph Sequence Neural Networks, Interaction Networks...

너무 구조가 복잡함... 제약조건...

Past solutions

Graph Neural Network

너무 구조가 복잡함...



orks, Interaction Networks...

RN은 단순하고 plug-and-play 방식이며, reasoning에 집중함.

g_θ 가 i 번째 object와 j 번째 object 사이의 relation을 찾아내는 함수
 f_ϕ 가 relation reasoning(prediction) 하는 함수

$$\text{RN}(O) = f_\phi \left(\sum_{i,j} g_\theta(o_i, o_j) \right)$$

RN의 강점

RN은 relation 자체를 추론하는 것을 학습함.

실제로 object 사이에 relation이 존재하는 지 몰라도 됨.

관계의 의미를 몰라도 됨.

Object의 semantic이 무엇인지 몰라도 됨.

기본적으로 all-to-all이지만, relation이 존재할 object에 대한 정보가 확실하면 some object pair

RN은 data efficient함.

D.E = expected benefits from applying I.T. to a given task / cost of application of I.T.

object pair에 대해 weight를 공유하는 하나의 function으로 계산함.

RN은 set of objects에 대해 동작함.

RN의 Equation이 order invariant를 보장함.

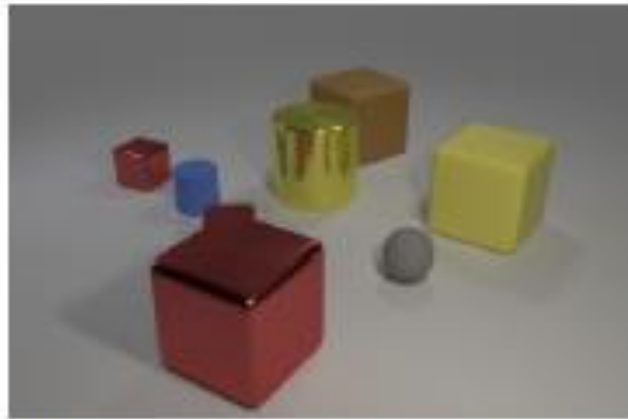
RN이 내놓은 output이 object set 안에 있는 relation에 대한 정보임을 보장함.

CLEVR – visual QA model.

- 1) pixel version.
- 2) state description version.

Matrix의 각 row에 single object의 state description이 표현됨.

(3D coordinates x, y, z; material rubber, metal, etc; size small, large, etc;)



What shape is the small object that is in front of the yellow matte thing and behind the gray sphere?



What number of things are either tiny green rubber objects or shiny things that are behind the big metal block?

Sort-of-CLEVR – 6개의 2D object로 구성된 image의 set.

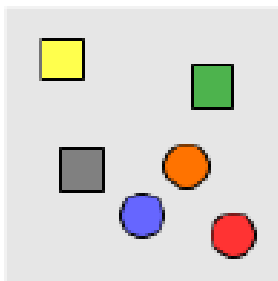
각각의 object는 random shape and color.

(square or circle):(red, blue, green, orange, yellow, gray)

각각의 image에 대해 10개의 non-relational,

10개의 relational question으로 구성.

Question은 fixed-length binary string으로 hard-coded



Non-relational question

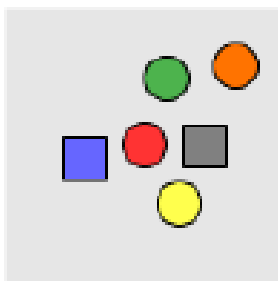
Q: Is the green object on the left or on the right?

A: right

Relational question

Q: How many objects have the shape of the orange object?

A: 3



Non-relational question

Q: Is the yellow object on the top or on the bottom?

A: bottom

Relational question

Q: What is the color of the object that is closest to the blue object?

A: red

bAbI – text-based QA dataset.

Deduction, induction, counting 등 reasoning의 종류를 담당하는
20개의 task로 나누어짐. 각 task는 10k의 예제를 가짐.
각각의 질문은 set of supporting facts와 연관됨.

Supporting facts : "Sandra picked up the football", "Sandra went to the office"

Question : "Where is the football?"

Answer : "office"

Dynamic physical systems

MuJoCo physics engine으로 움직이는 물체들을 만든 dataset.

10개의 각각 다른 색의 공.

어떤 공은 독립적으로, 어떤 공들은 서로 묶여 있음 (rigid 또는 spring)

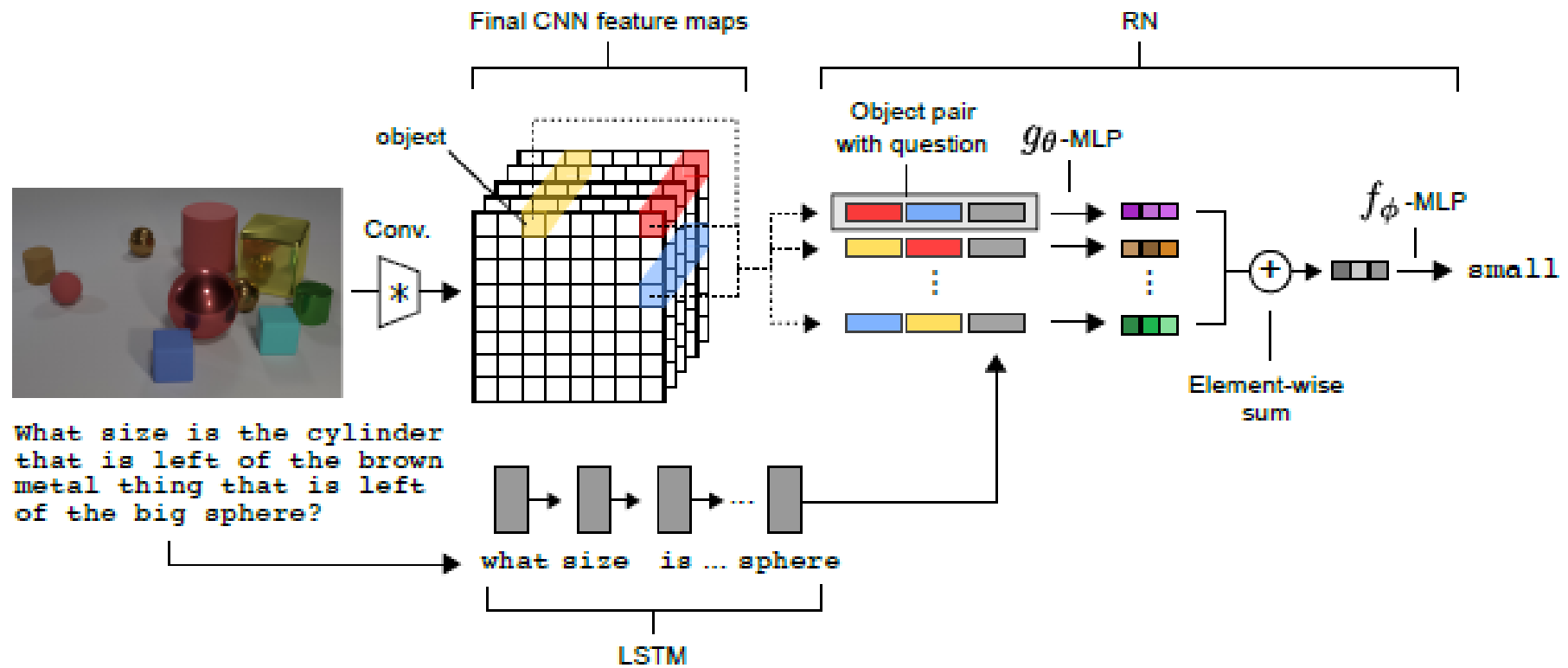
input : 각 공의 state description(RGB color, coordinates x and y)

16개의 time step으로 matrix의 row에 입력됨.

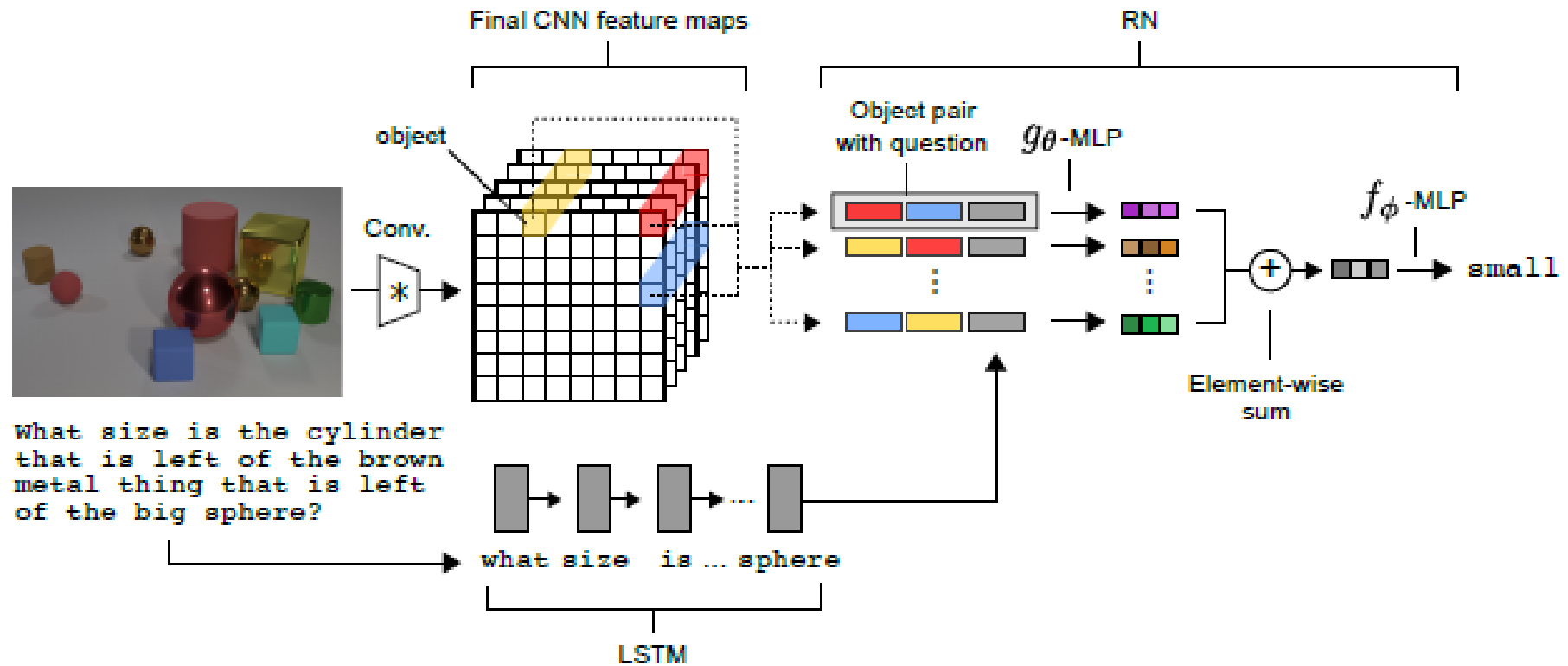
공 사이에 connection이 있는지를 추측하는 task와

Connection system이 몇 개인지 세는 task.





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

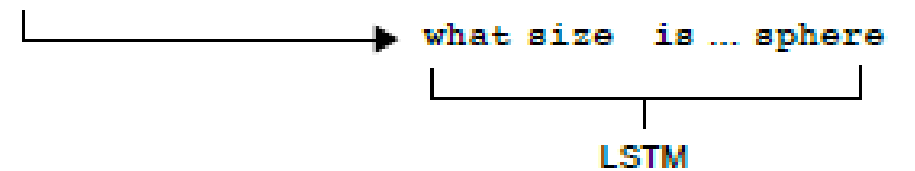


Dealing with pixels

128x128 크기의 image를 4 convolutional layer,
 마지막 layer의 kernel이 k개인 CNN에 통과시킴. k개의 $d \times d$ feature map이 생성됨.
 각각의 d^2 개의 k-dimensional cell은 object처럼 취급함.
 이 "object"는 배경, 특정한 물리적 객체, texture, 또는 물리적 객체의
 겹친 부분 등등을 표현할 수 있음.

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

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



Conditioning RNs with question embeddings

질문에 따라 object 사이의 relation이 결정됨.

예측에 조건을 주기 위해 RN의 입력으로 q 를 추가해 condition을 줌.

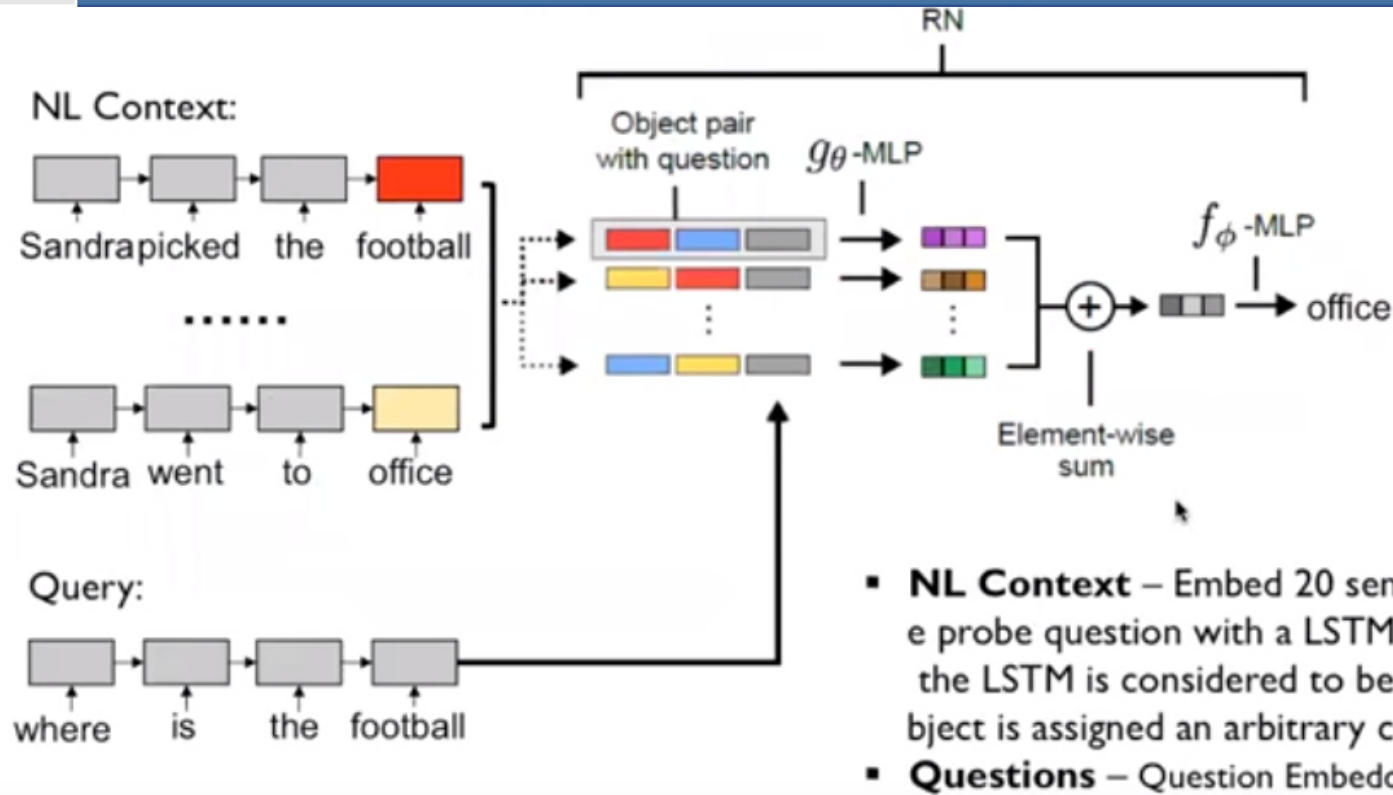
Question을 한 단어 단위로 LSTM에 통과시키고 마지막 단어가

통과한 final state를 q 로 사용.

Dealing with state descriptions

state description 자체가 pre-factored object representation이기 때문에

직접 RN의 input으로 사용하면 됨.



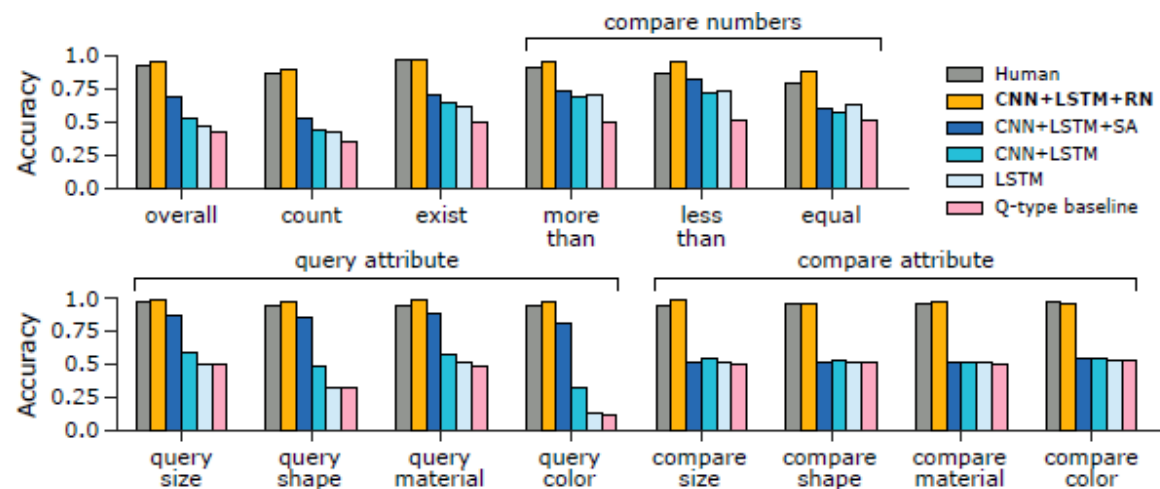
Dealing with natural language

bAbI에서는 natural language input을 set of object로 변환해야 함.
 Support set에서 질문과 연관이 있는 문장을 최대 20개까지 뽑아낸 후,
 support set에서의 relative position에 따라 labeling함.
 그 후 word-by-word로 LSTM에 통과 후 final state가 그 문장의 object가 됨.

Results

Model	Overall	Count	Exist	Compare Numbers	Query Attribute	Compare Attribute
Human	92.6	86.7	96.6	86.5	95.0	96.0
Q-type baseline	41.8	34.6	50.2	51.0	36.0	51.3
LSTM	46.8	41.7	61.1	69.8	36.8	51.8
CNN+LSTM	52.3	43.7	65.2	67.1	49.3	53.0
CNN+LSTM+SA	68.5	52.2	71.1	73.5	85.3	52.3
CNN+LSTM+SA*	76.6	64.4	82.7	77.4	82.6	75.4
CNN+LSTM+RN	95.5	90.1	97.8	93.6	97.9	97.1

* Our implementation, with optimized hyperparameters and trained fully end-to-end.



Overall accuracy가 사람보다 높게 나온 것이 제일 놀라운 부분.
state descriptions 형태의 input에 대해서는 96.4%
Object set이 실제 object를 잘 반영한다면, 매우 잘 동작.
그렇지 못한 경우("object-like")의 set에 대해서도, 성능을 보장함.

➔ RN이 visual problems에 국한된 모델이 아님. Relational reasoning 관련된 여러 문제에 적용 가능.

Results



Image

Non-relational question

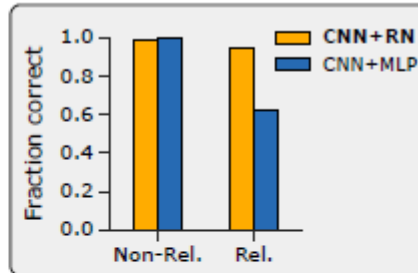
Q: What is the shape of the gray object?

A: circle

Relational question

Q: What is the shape of the object that is furthest from the gray object?

A: square



Non-relational question

Q: Is the green object on the left or on the right?

A: right

Relational question

Q: How many objects have the shape of the orange object?

A: 3

Sort-of-CLEVR

RN 모델의 경우 non-relational / relational Q에서 모두 94% 이상의 정확도
CNN augmented with MLP의 경우 relational Q에서 63%

bAbI

Succeeded on 18/20 tasks.

basic induction task에서 Sparse DNC(54%), DNC(55.1%), EntNet(52.1%)에 비해 RN의 경우 2.1%의 error rate를 보인 점이 유의미함.

Dynamic Physical System

connection을 유추하는 문제에서 93%, counting 문제에서 95%의 정확도를 보임.

MLP에 비해 성능이 월등히 좋다.

학습한 모델에 motion capture data를 넣었더니 RN이 예측한 connection이 사람의 형상이 됨.

[Youtube 동영상](#)

Simple CNN, LSTM based VQA architecture에서 RN을 통해 성능을 끌어올렸음.

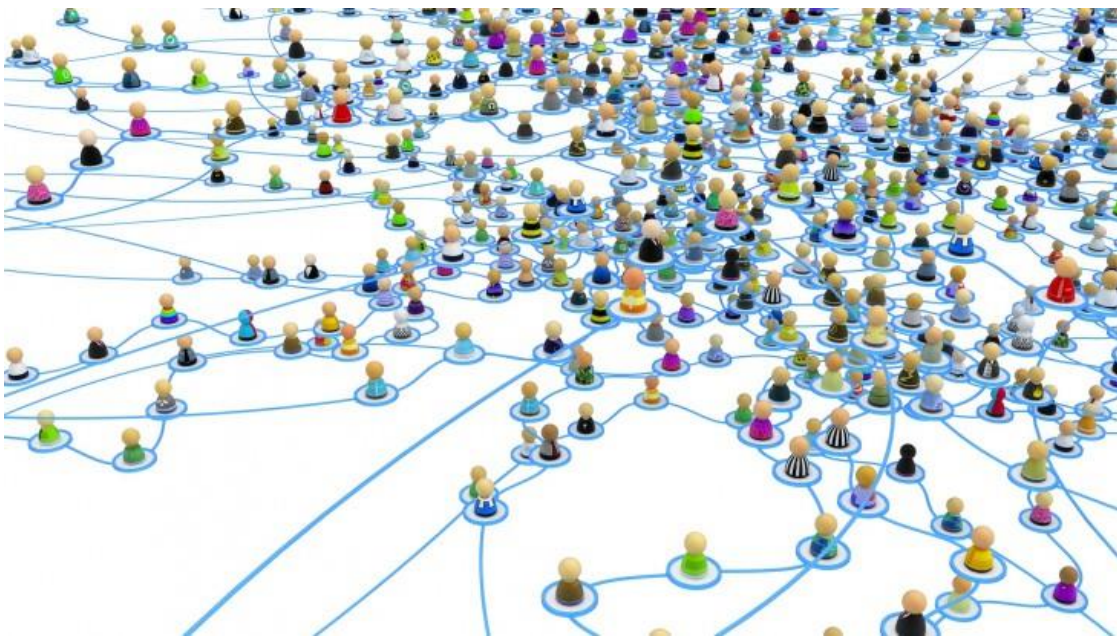
관계형 추론에 관한 학습만을 담당하는 RN을 통해 CNN이 image processing에 집중할 수 있도록 함. Processing과 reasoning을 하나의 복잡한 모델(ResNet 등)로 학습하기 보다는 구분해 학습하는 것이 더 좋을 수 있음.

“object-like”, 즉 unstructured inputs and outputs에 대해 학습해 relation을 추론할 수 있는 능력이 RN의 장점.

향후 과제

RN을 modeling social networks, abstract problem solving 등의 영역에 적용해보면 어떨까?

RN의 계산 속도를 향상시키는 방법은 뭐가 있을까? Object set에 attention의 적용방법 등



Solving abstract problems

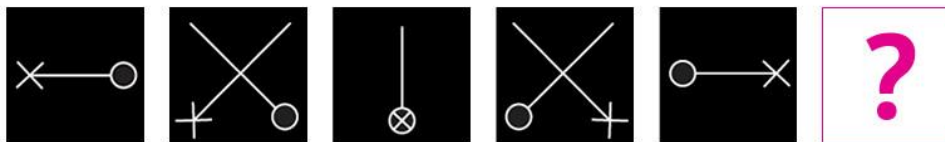
Completed

12%

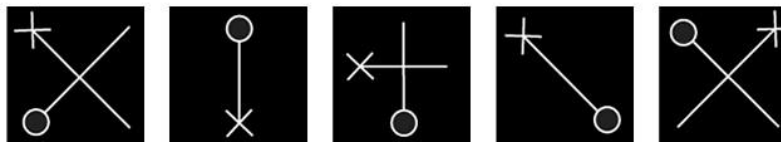
Question 3

Please look at the series. Then decide what should be the next image, out of the five options.

Series



Options



☐ A

☐ B

☐ C

☐ D

☐ E

Save and exit >

Skip question >

Next question >