HOW TO REPRESENT RELATIONS

2018. 11. 14 Naver TechTalk
SNU Datamining Laboratory
Sungwon, Lyu
lyusungwon@dm.snu.ac.kr

CONTENTS

DEEPEST

DEEPEST

SNU (Based) Deep Learning Society

Research, Project, Study, Competition, Discussion

EE, CS, MD, IE & Naver, Kx, Sx etc

Every Saturday 3PM

http://deepest.ai/

DEEPEST

Projects

- Bayesian DeepLearning
- Disentangled Representation in Audio
- Language generation using discrete latent variable
- RL Start
- Video Super Resolution
- Trends in RNN
- PRML Study

Hosting Topics

Neural Architecture Search

Flow-based generative model (NICE, Real NVP, Glow)

Breaking Illusion on 'PSNR'

Engineering Reinforcement Learning

ICML Review

High Resolution Variational Auto Encoder: Beyond Pixelwise Loss

Weakly-supervised Semantic Segmentation

Co-Training of Audio and Video Representations

Python Optimization Methods

unsupervised domain adaptation

3 Issues on Current Neural Networks

Speaker recognition

An overview of image enhancement

Introducing Magenta

Neo-backpropagation, Part 2

my painful climb to score>0.80

Image to Image translation

poisoning attack

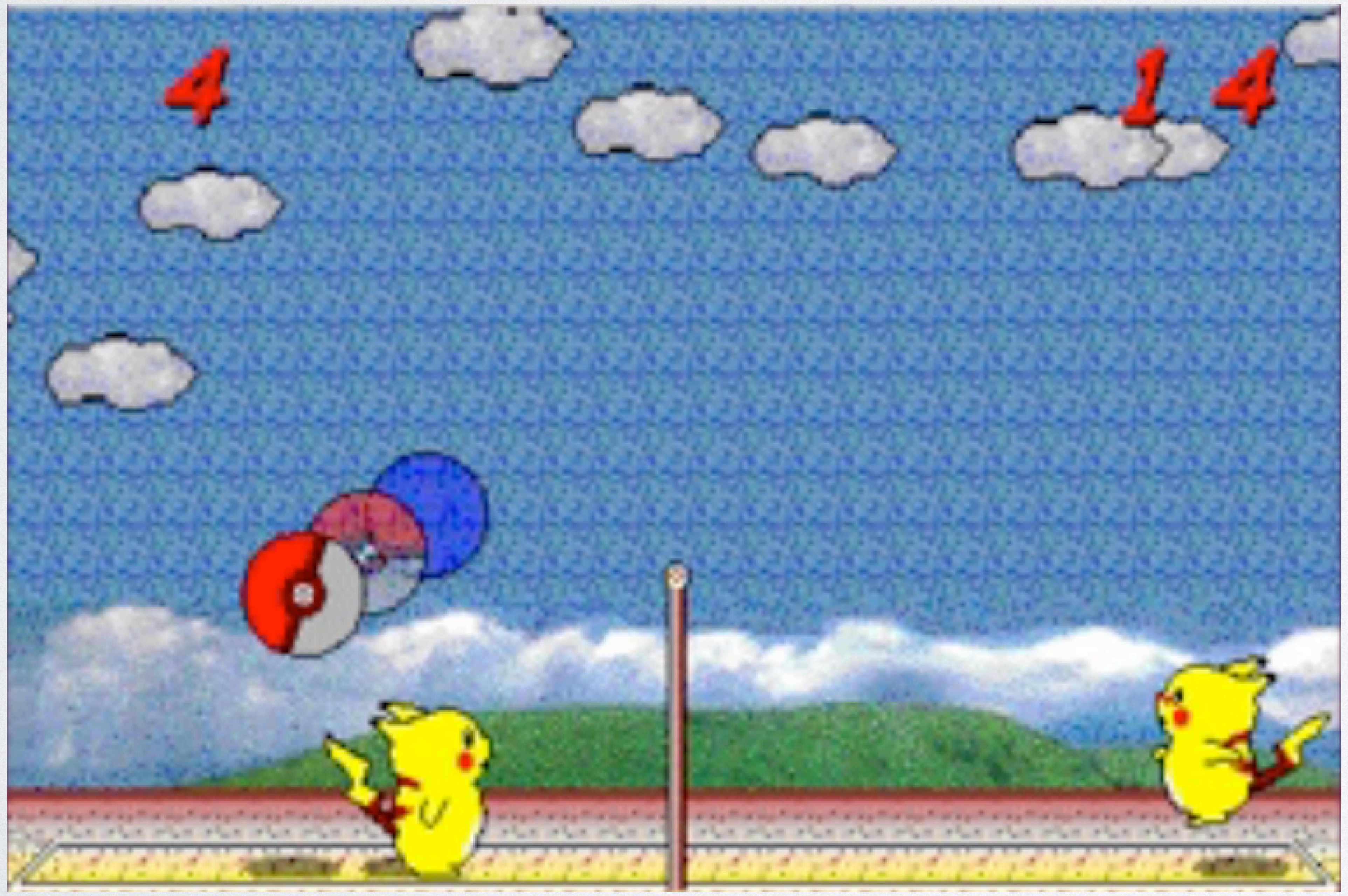
Visual Domain Adaptation

FloWaveNet

Music Generation using MIDI

DEEPEST

- **My Projects**
  - Training Pickachu Volleyball with Reinforcement Learning



  - DeepClear (2018 Digital Health Hackathon)

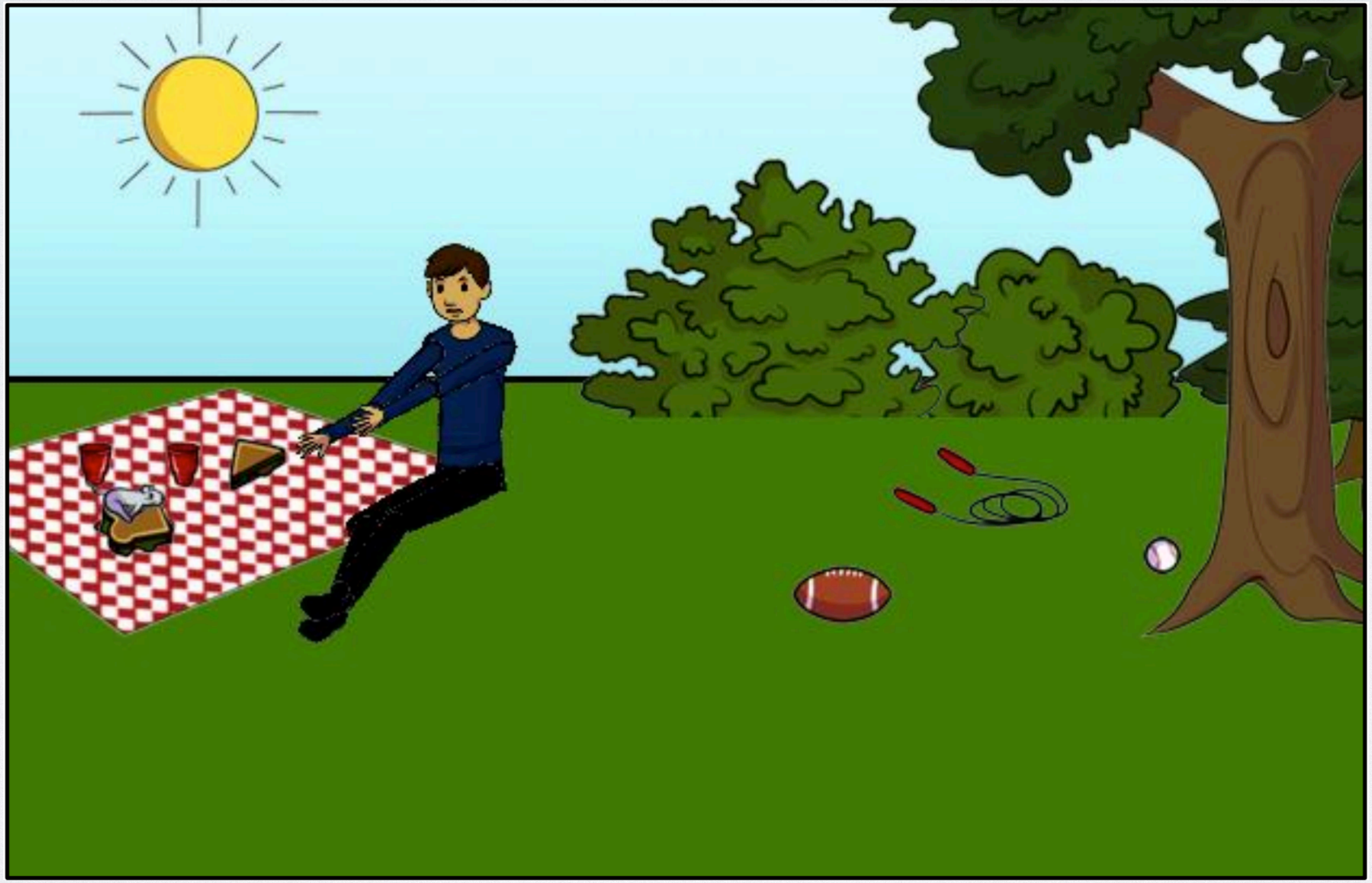
SPEAKER

- **Sungwon Lyu**
  - SNU IE Data-Mining Laboratory
  - https://lyusungwon.github.io/
- Interested Field
  - Deep Learning Engineering
  - **Representation Learning** with deep learning

REPRESENTATION

- **Representation**
  - Vector form (for Neural Network)
  - Task Specific
  - Examples
    - Image(C-H-W) : The last block of Classifier (Imagenet), latent Variable of (beta) VAE…
    - Audio(Raw Audio) : STFT, MFCC…
    - Text?
    - Relation?

RELATION



Source: Agrawal, Aishwarya, et al. "Vqa: Visual question answering." *arXiv preprint arXiv:1505.00468* (2015).

RELATIONAL REASONING

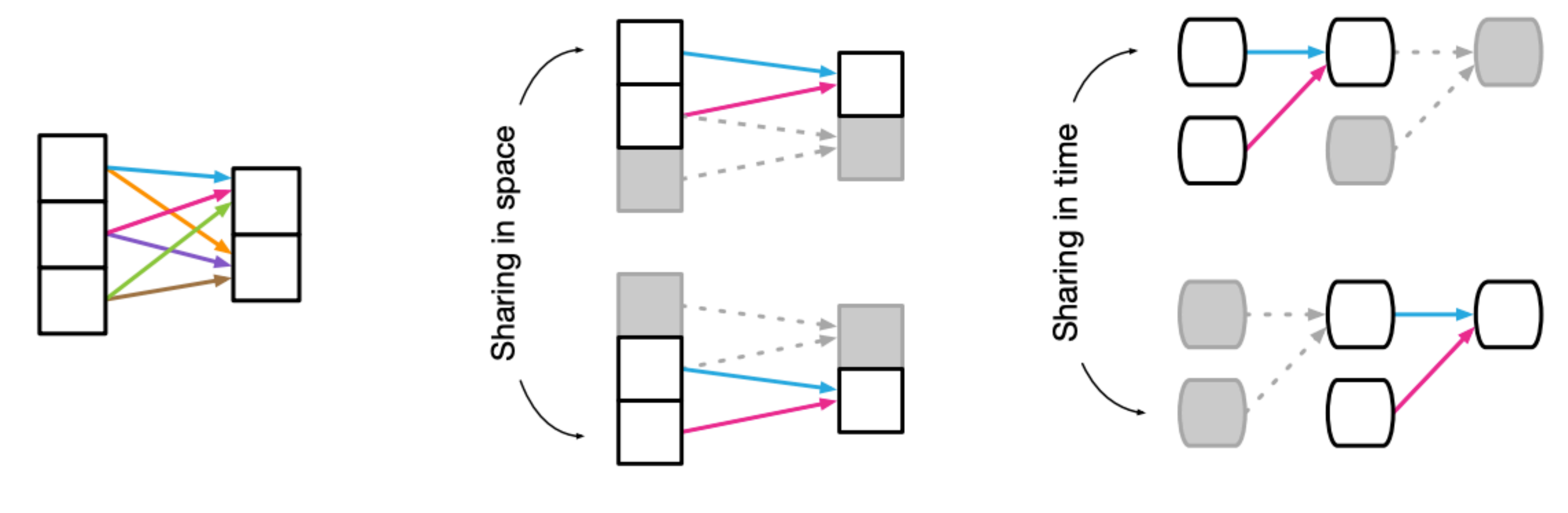
- **Relational Reasoning**
  - Relational reasoning involves manipulating structured representations of entities and relations, using rules for how they can be composed.
  - **Entity**: An element with Attributes
    - Physical objects with a size and mass
  - **Relation**: A property between entities
    - Same size as, heavier than, distance from…
  - **Rule**: Function that maps entities and relations to other entities and relations
    - Is entity X heavier than entity Y?

Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

INDUCTIVE BIAS

- An **inductive bias** allows a learning algorithm to prioritize one solution (or interpretation) over another, independent of the observed data.

Component	Entities	Relations	Rel. inductive bias	Invariance
Fully connected	Units	All-to-all	Weak	-
Convolutional	Grid elements	Local	Locality	Spatial translation
Recurrent	Timesteps	Sequential	Sequentiality	Time translation
Graph network	Nodes	Edges	Arbitrary	Node, edge permutations

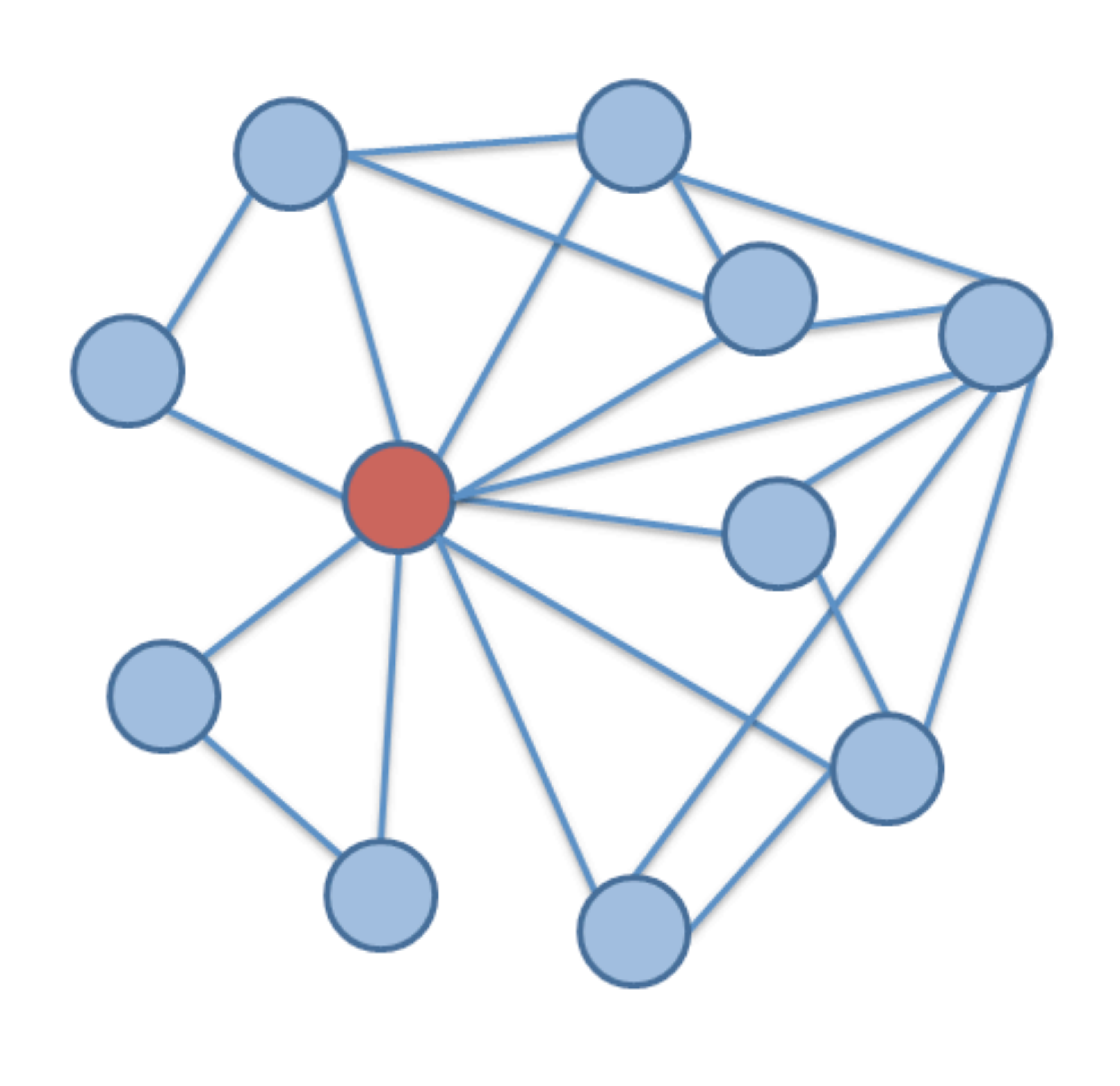


(a) Fully connected (b) Convolutional (c) Recurrent

Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

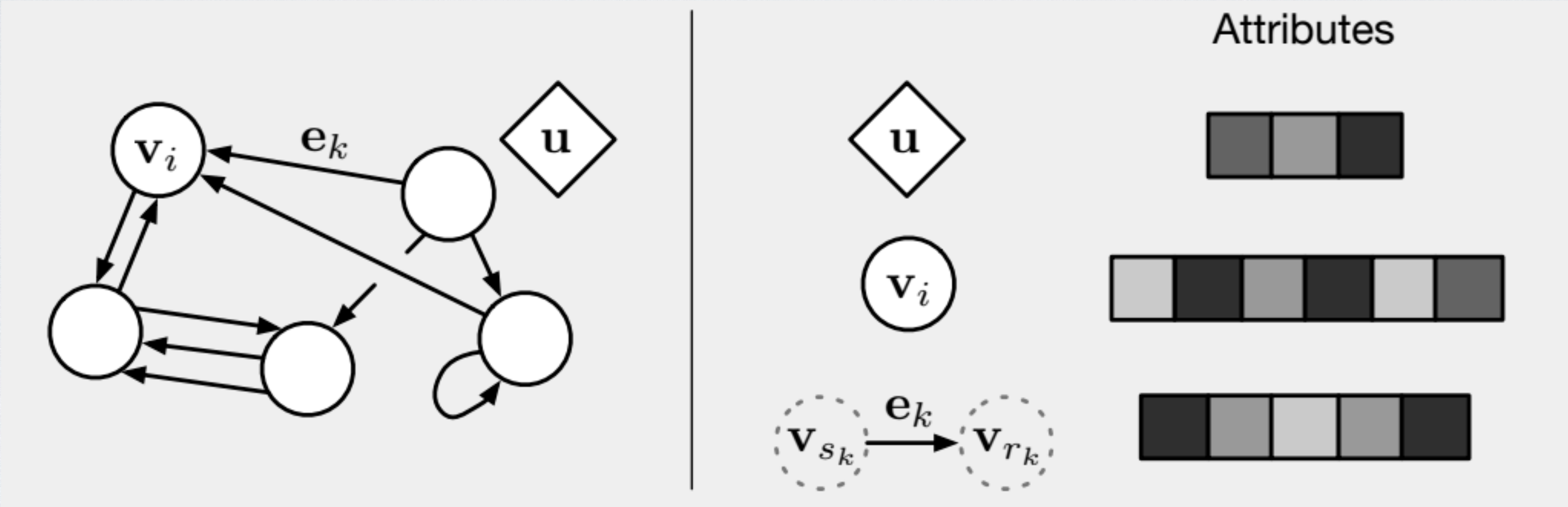
GRAPHS

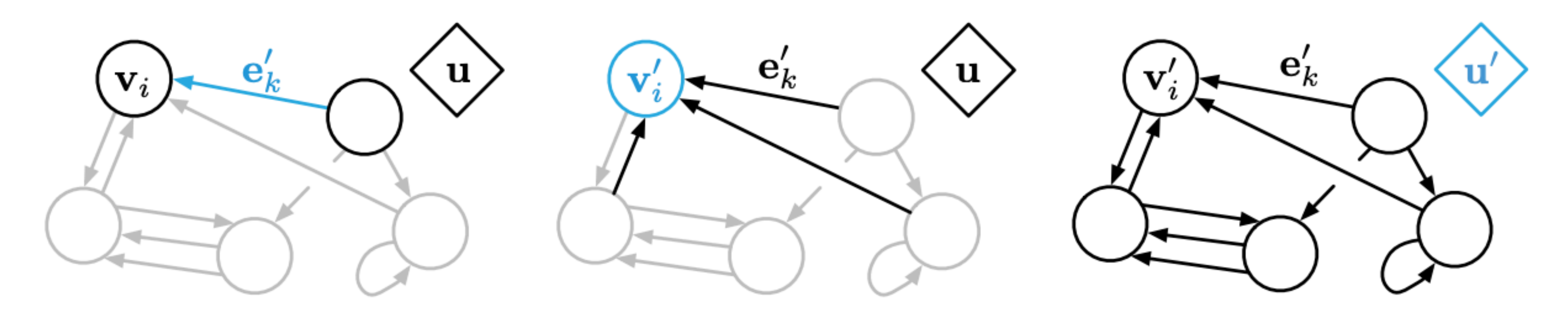
- **Graphs**
  - Visual Representation for (clearly defined) entities and relations
  - **REUSE** of entities and relations (Combinatorial Generalization)



Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

GRAPH NETWORKS

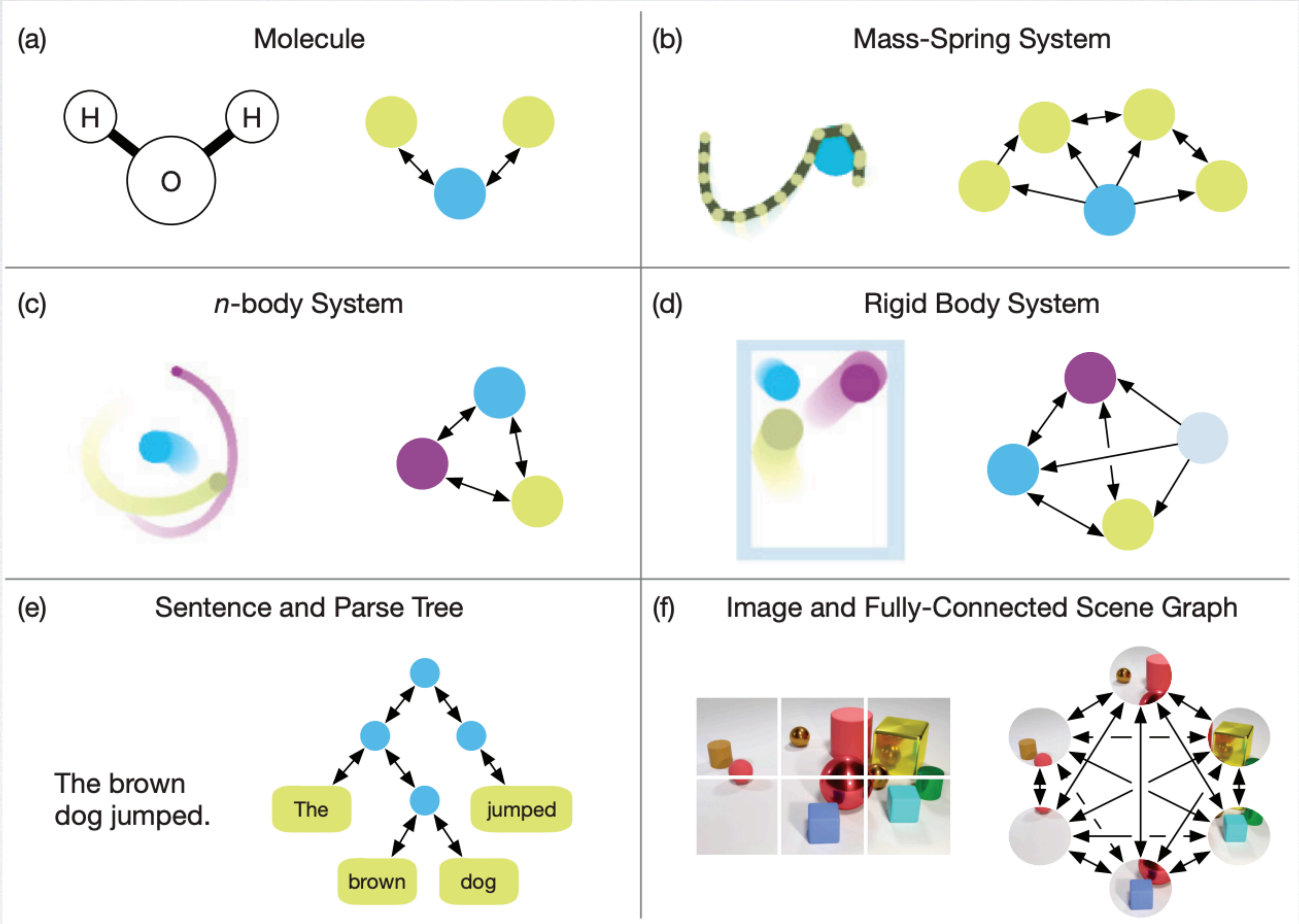




(a) Edge update (b) Node update (c) Global update

Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

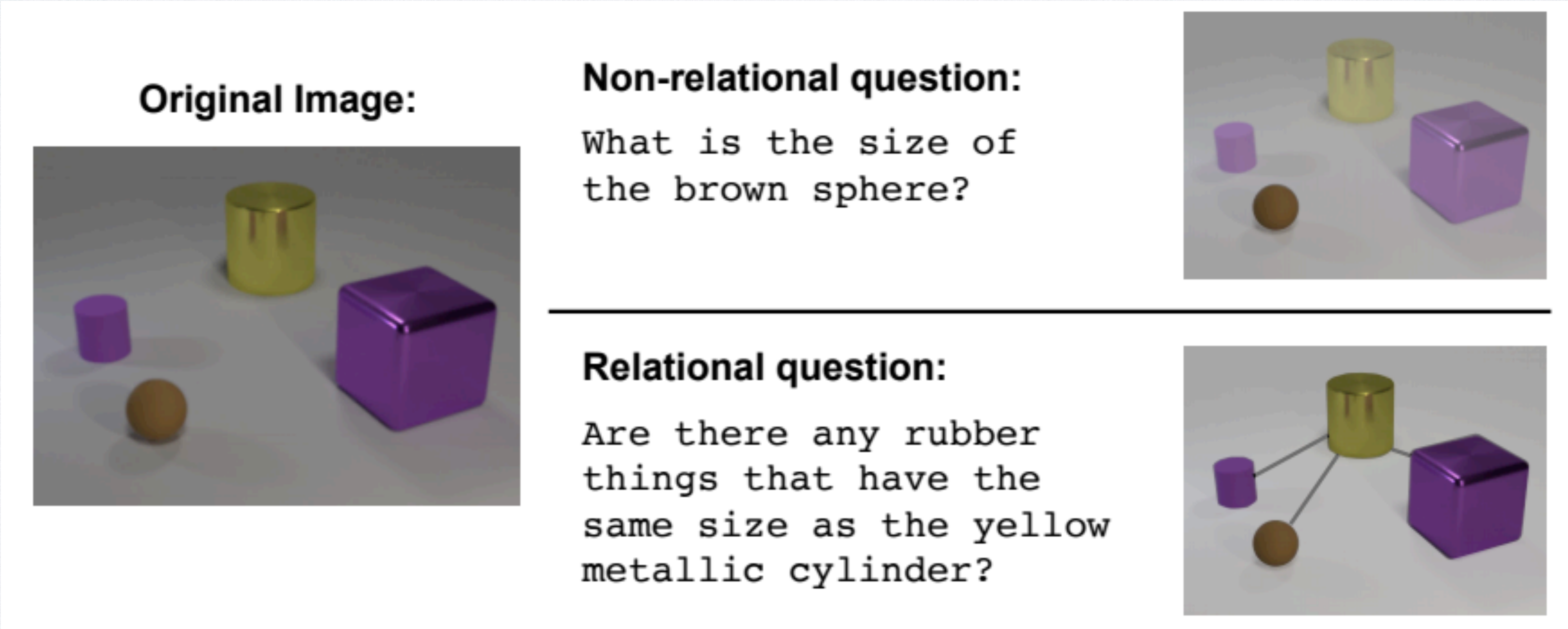
GRAPH NETWORKS



Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

CLEVR

- **CLEVR**
    - Cubes are gray, blue, brown, or yellow
    - Cylinders are red, green, purple, or cyan
    - Spheres can have any color



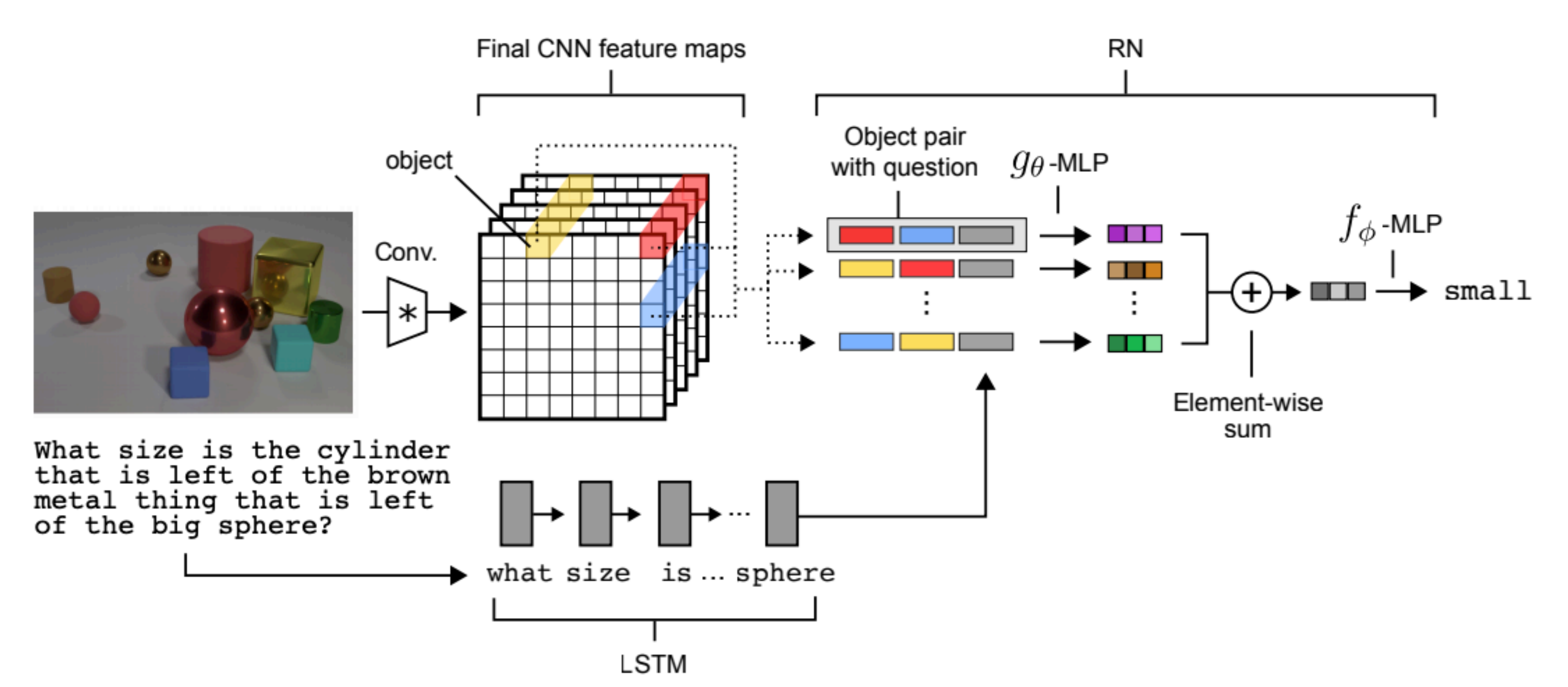
Source: Johnson, Justin, et al. "CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning." *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017.

RELATIONAL NETWORK

- **Relational Network**

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

  - Objects: each channel of middle layer of Conv
  - g-theta(relations), f-phi: MLP
  - Order Invariance among relations
  - Capture all possible relations
  - Reuse of relations



Source: Santoro, Adam, et al. "A simple neural network module for relational reasoning." *Advances in neural information processing systems*. 2017.

RELATIONAL NETWORK

- **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.

Source: Santoro, Adam, et al. "A simple neural network module for relational reasoning." *Advances in neural information processing systems*. 2017.

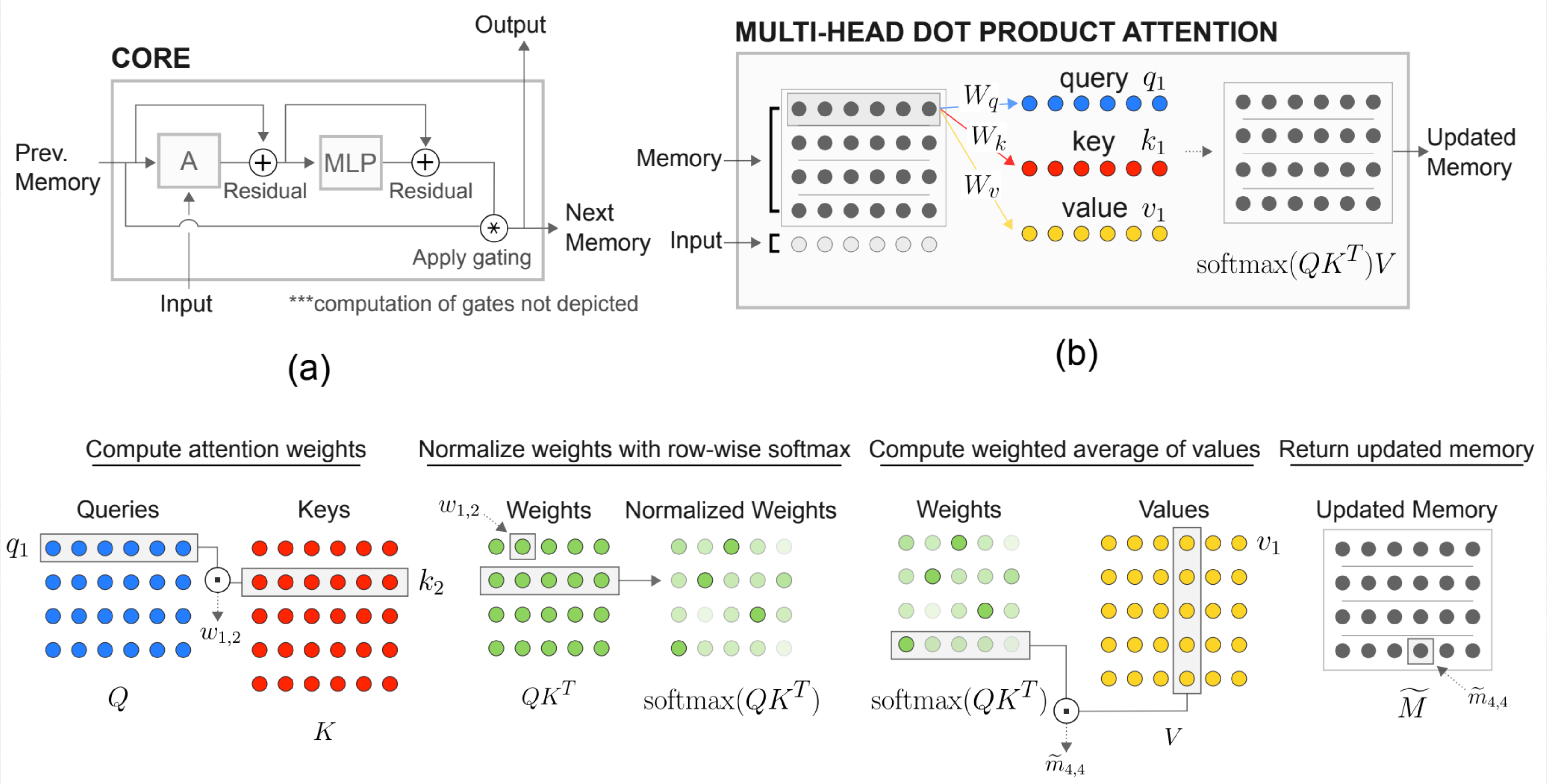
RELATIONAL NETWORK

- Questions:
    - "There is a cube that is on the left side of the large shiny object that is on the right side of the big red ball; what number of cubes are to the right of it?''
- All possible relations
    - A-B, A-C, A-D, B-C, B-D, C-D
    - A->C->B->A->D

Source: Santoro, Adam, et al. "A simple neural network module for relational reasoning." *Advances in neural information processing systems*. 2017.

RELATIONAL NETWORK - FOLLOW UPS (I)

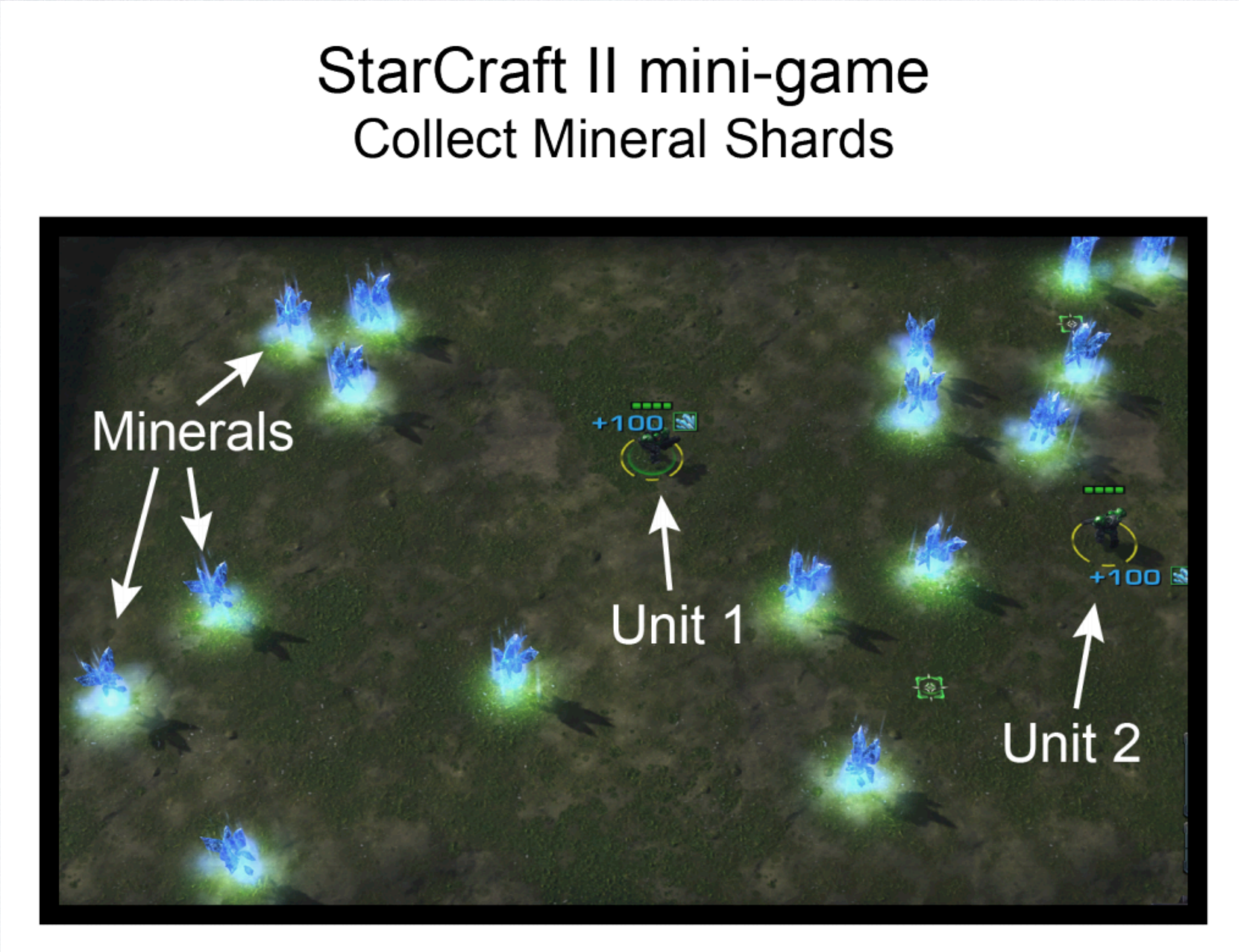
- Relational Recurrent Neural Network
  - MHDPA module for relation
  - Relations among memory slots in memory augmented neural network

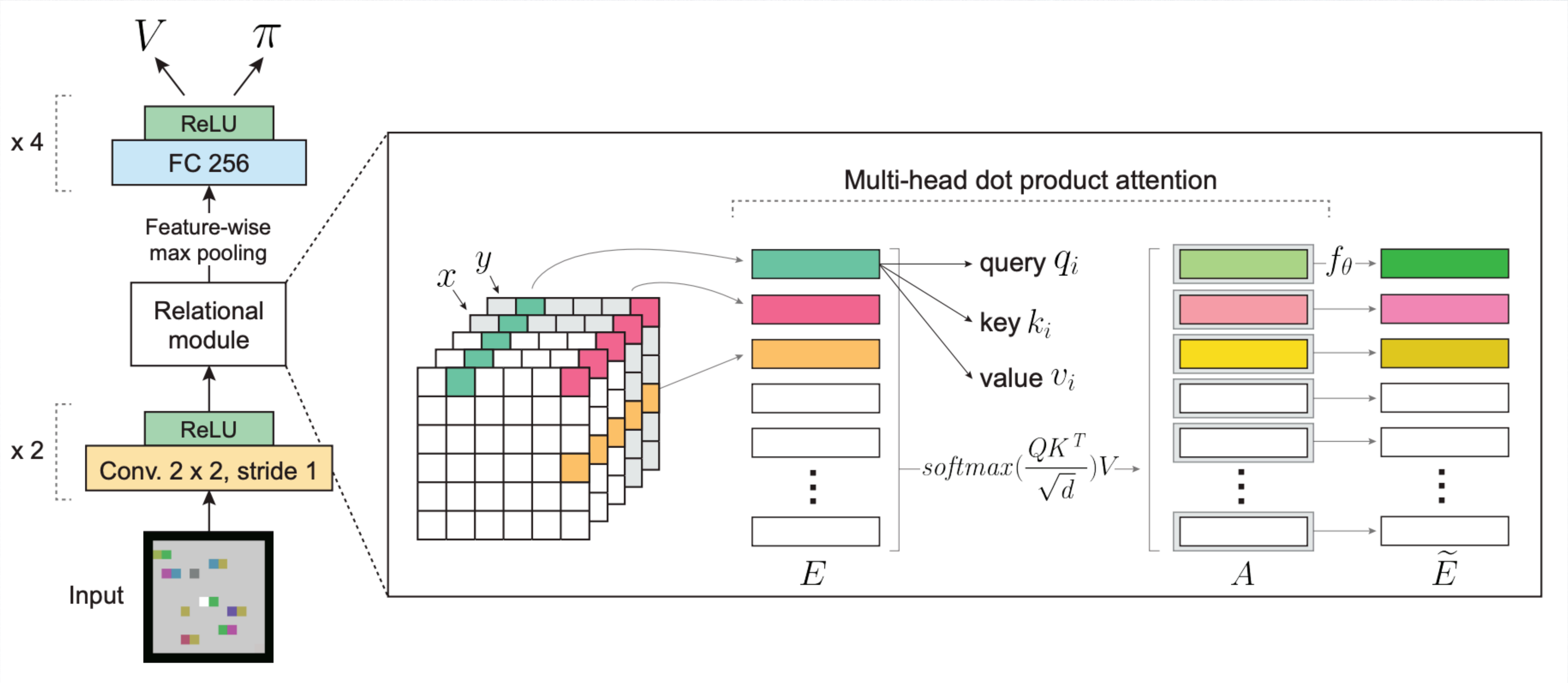


Source: Santoro, Adam, et al. "Relational recurrent neural networks." *arXiv preprint arXiv:1806.01822* (2018).

RELATIONAL NETWORK - FOLLOW UPS (2)

- Relational Deep Reinforcement Learning
  - MHDPA module for relation
  - Relational Module for reinforcement learning

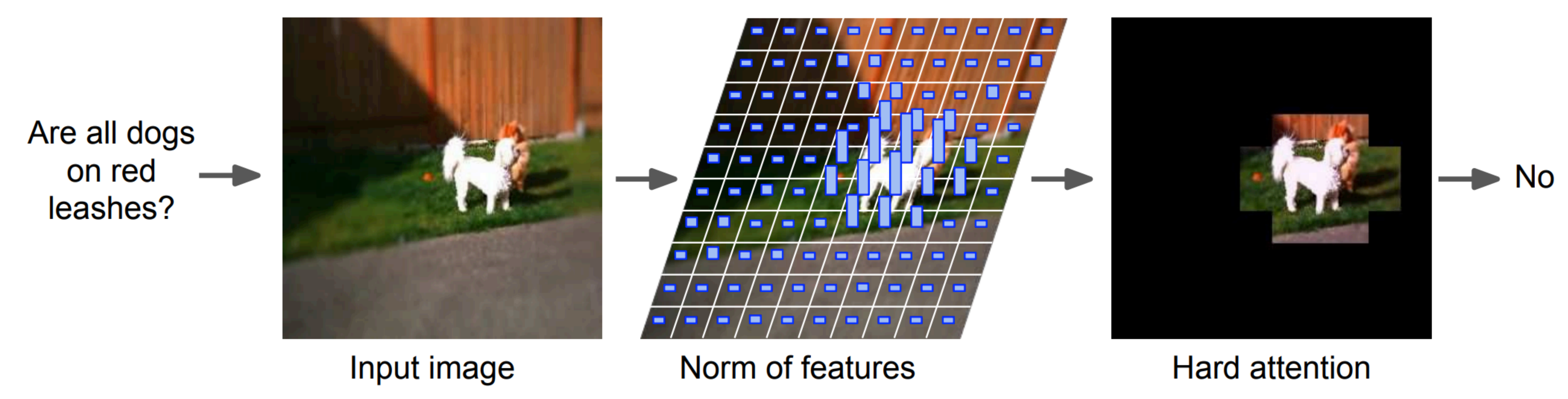




Source: Zambaldi, Vinicius, et al. "Relational Deep Reinforcement Learning." *arXiv preprint arXiv:1806.01830* (2018).

RELATIONAL NETWORK - FOLLOW UPS (3)

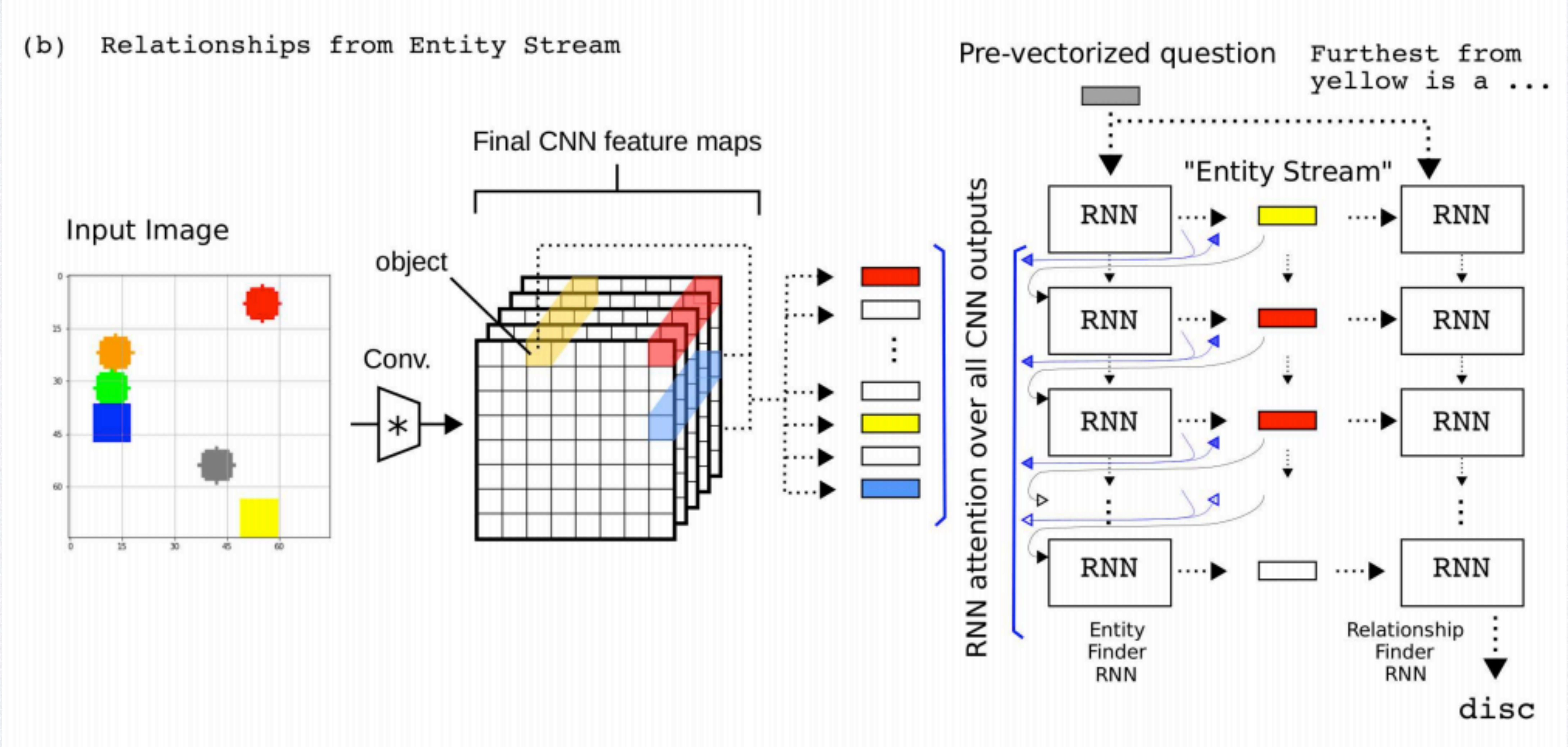
- Learning Visual Question Answering by Bootstrapping Hard Attention
  - MHDPA module for relation
  - Reduce the number of objects with hard attention



Source: Malinowski, Mateusz, et al. "Learning visual question answering by bootstrapping hard attention." *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

RELATIONAL NETWORK - FOLLOW UPS (4)

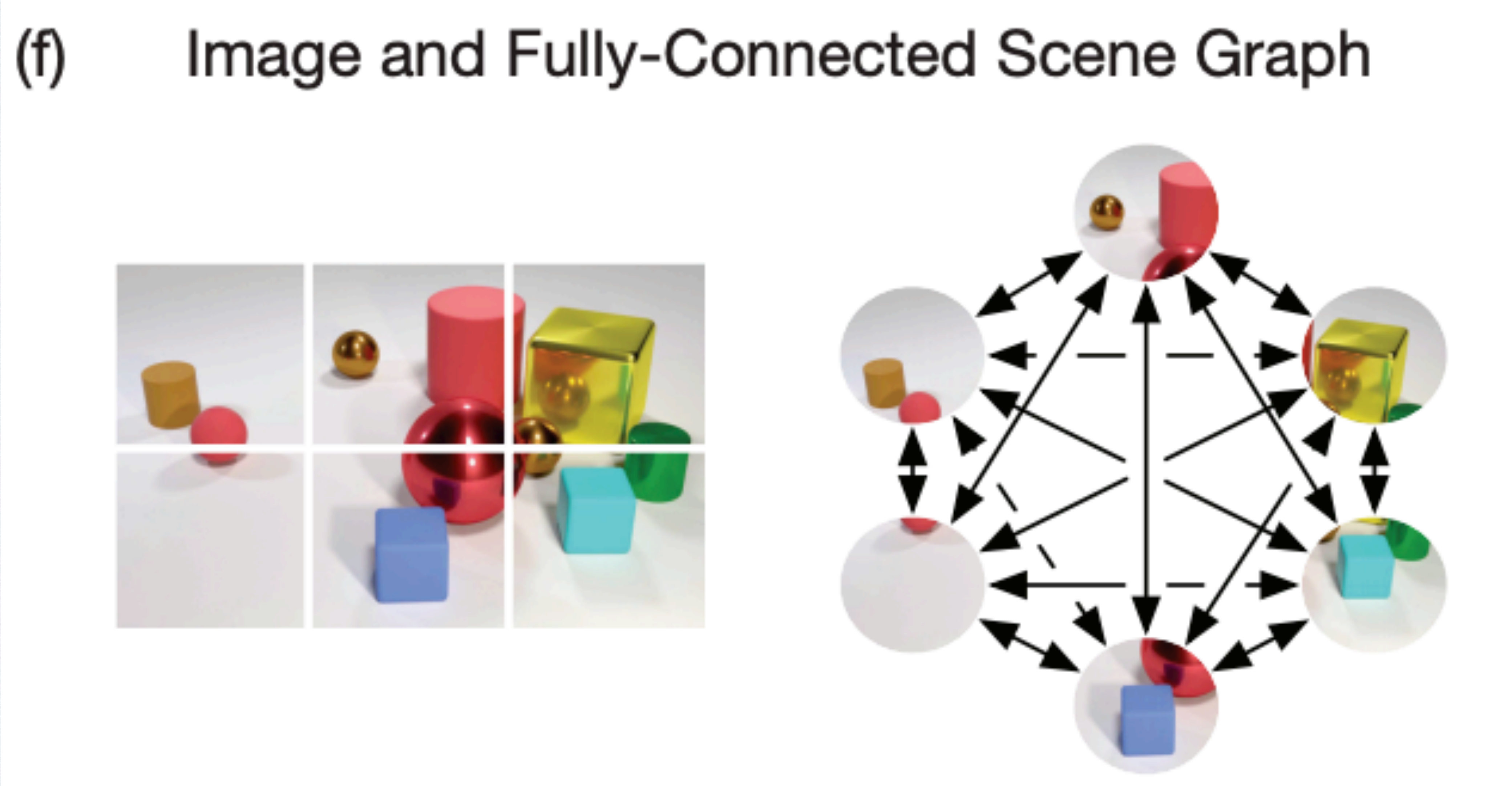
- Relationships from Entity Stream
  - LSTM to select Entity
  - LSTM to find Relationships
  - Reduced the number of pairings



Source: Andrews, Martin, Red Dragon AI, and Sam Witteveen. "Relationships from Entity Stream."

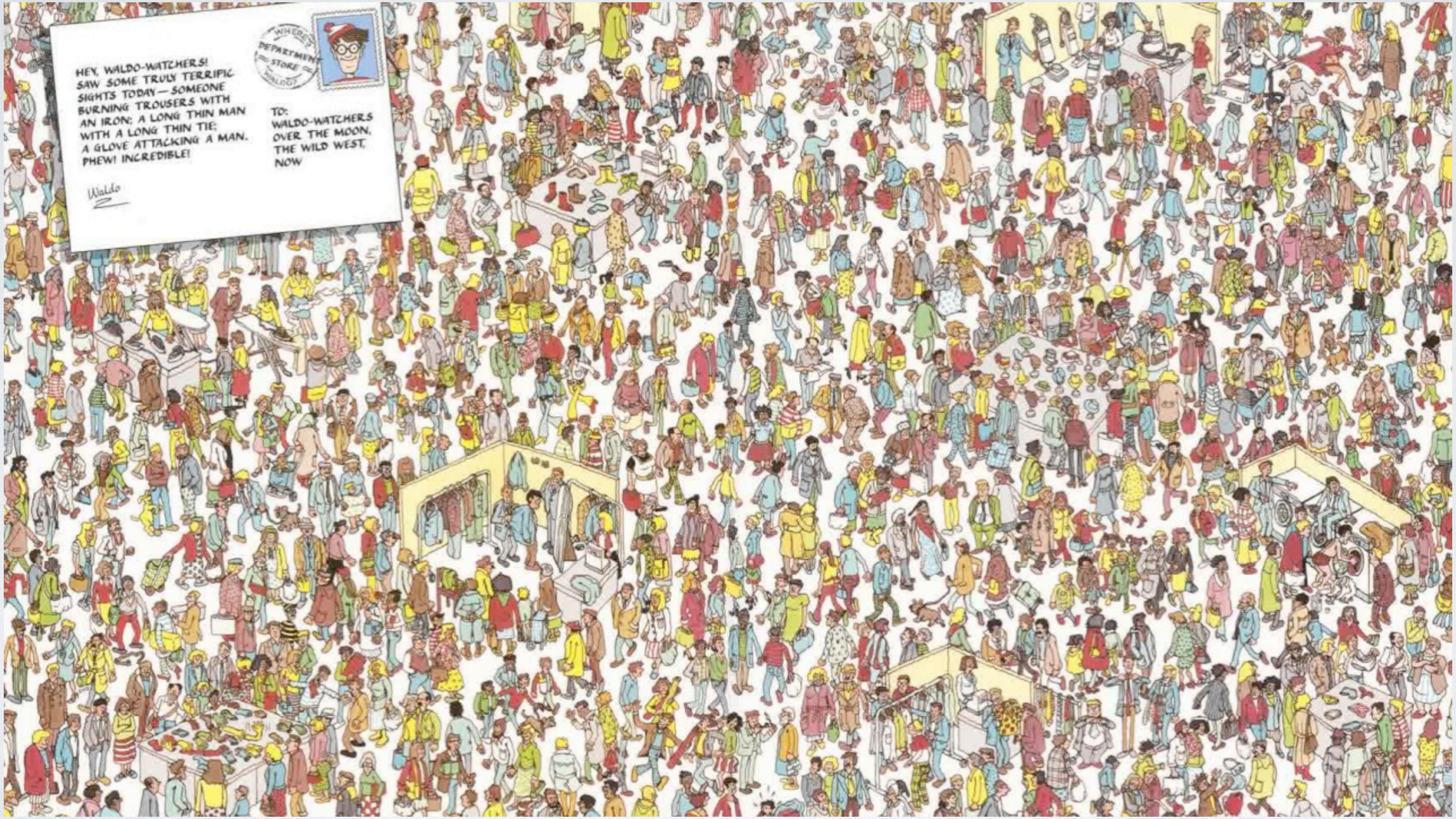
LIMITATION OF RN

- Are they good representations of relations?
  - Objects?
    - Fragmented / Number not matched
  - Fully Connected?
    - n^2
  - Interpretable?
  - *Relational inductive bias does not come from the presence of something, but rather from the absence.*



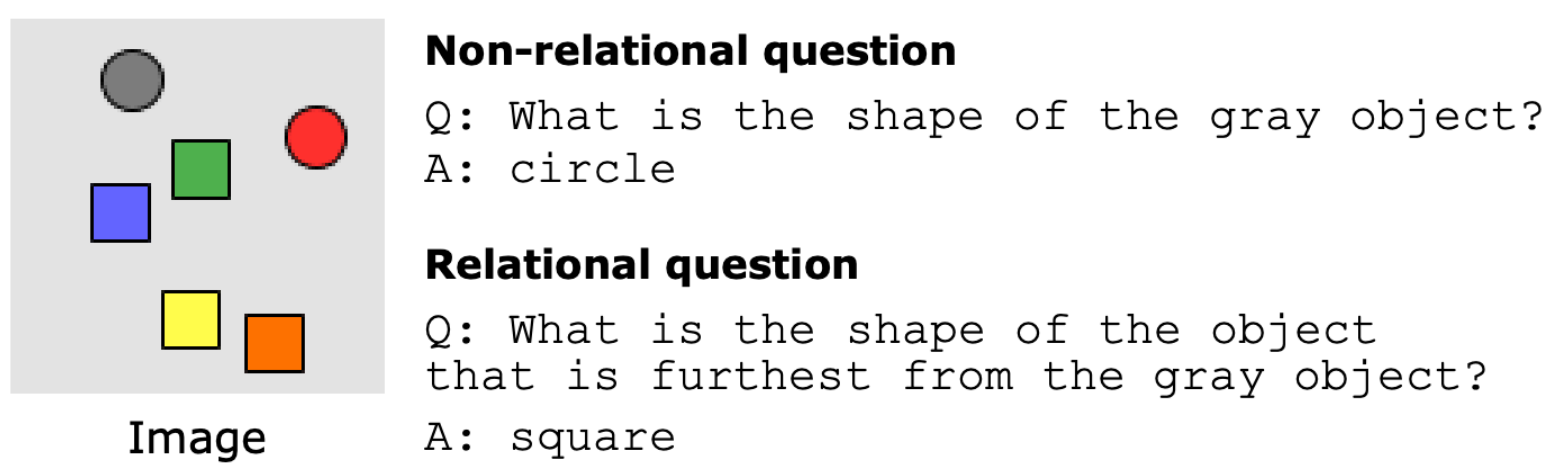
Source: Battaglia, Peter W., et al. "Relational inductive biases, deep learning, and graph networks." *arXiv preprint arXiv:1806.01261* (2018).

LIMITATION OF RN



SORT-OF-CLEVR

- **Sort of CLEVR**
  - 6 Objects with unique color of red, blue, green, orange, yellow, gray
  - A randomly chosen shape (square or circle).
  - Relational question
    - Color / shape of closest / furthest object from certain color
    - Number of object of the same shape with certain color



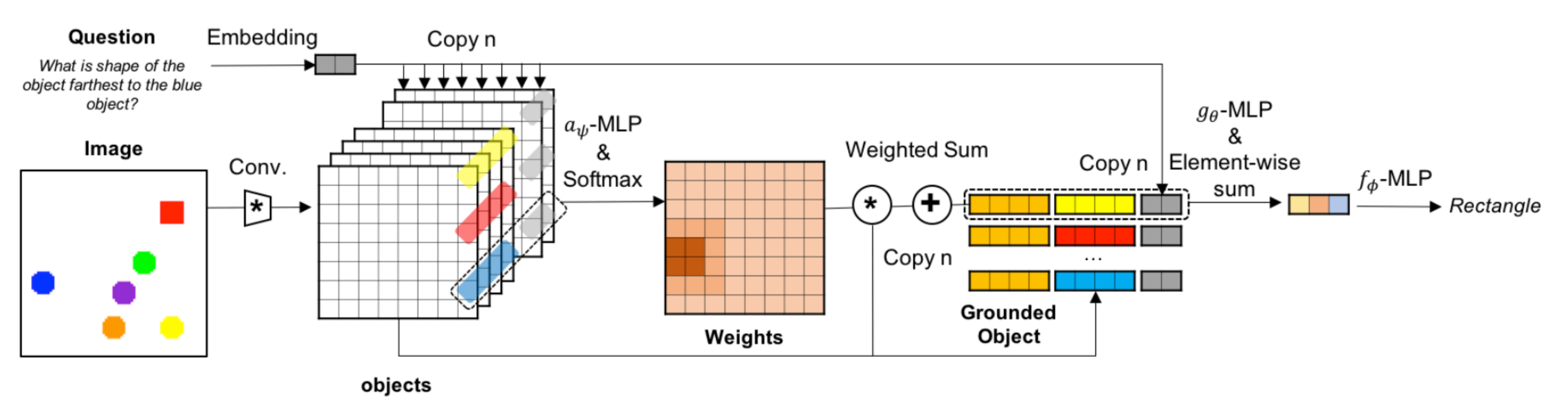
SARN

- **SARN**: Sequential Attention Relational Network

$$a_i = a_\psi(o_i, q) \quad i = 1, \cdots, n \tag{1}$$

$$ro = \sum_{i=1,\cdots,n} a_i * o_i \tag{2}$$

$$g_{\theta output} = \sum_{i=1,\cdots,n} g_\theta(o_i, ro, q) \tag{3}$$



SARN

- Result
  - Sort-of-clevr

Table 1: Test accuracy

model	overall	non-rel	rel
SARN	**96.73**	**99.84**	**94.88**
RN	93.56	99.81	89.83
base line	89.07	97.58	83.97

Table 2: Test accuracy: non-relational questions

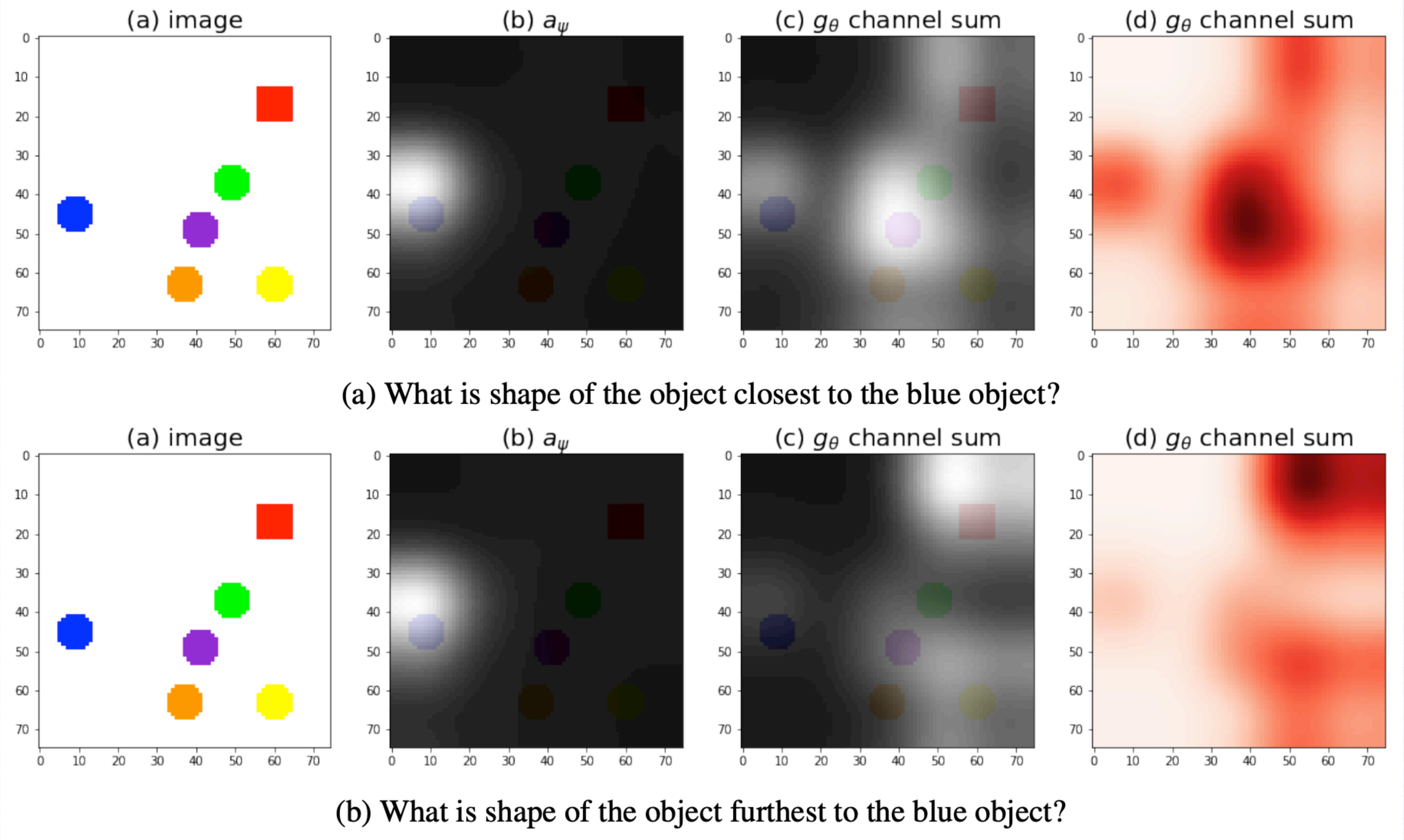
model	horizontal	vertical	shape	non-rel
SARN	**99.92**	**99.67**	**99.92**	**99.84**
RN	**99.92**	**99.67**	99.83	99.81
base line	96.33	96.58	99.83	97.58

Table 3: Test accuracy: relational questions

model	cl_col	cl_sh	fur_col	fur_sh	count	rel
SARN	**90.75**	**93.92**	**93.75**	**96.33**	99.67	**94.88**
RN	86.33	88.42	84.17	90.25	**100**	89.83
base line	84.92	88.50	67.83	79.25	99.33	83.97

SARN

- Result

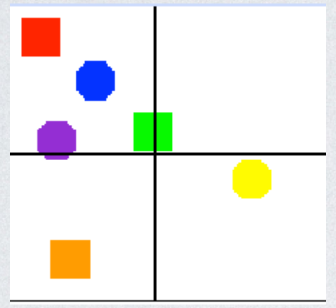


(a) What is shape of the object closest to the blue object?

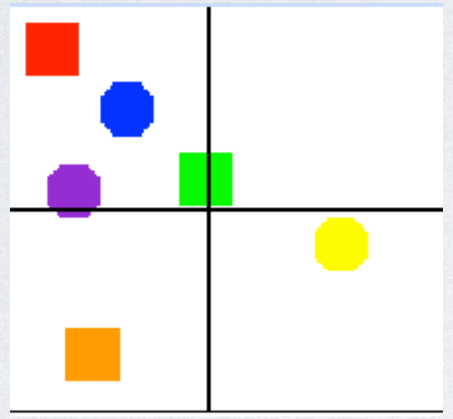
(b) What is shape of the object furthest to the blue object?

SARN

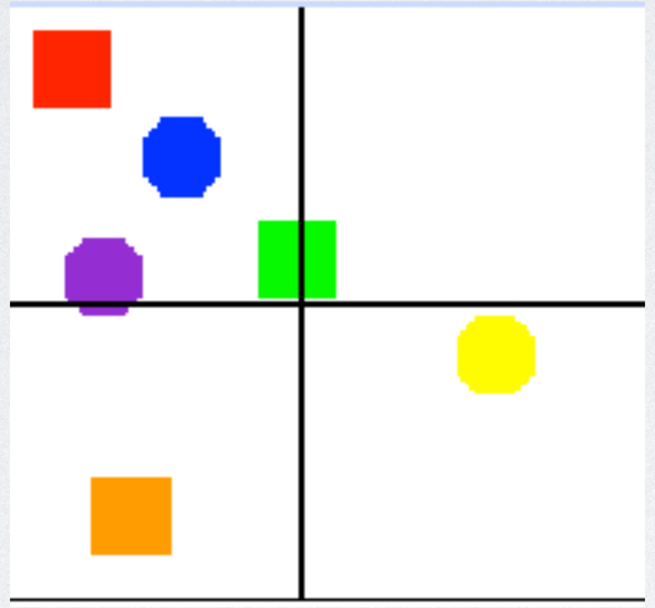
- Robustness on image size and object sparsity



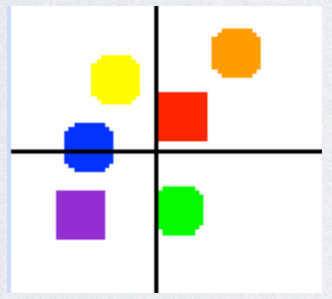
64_4



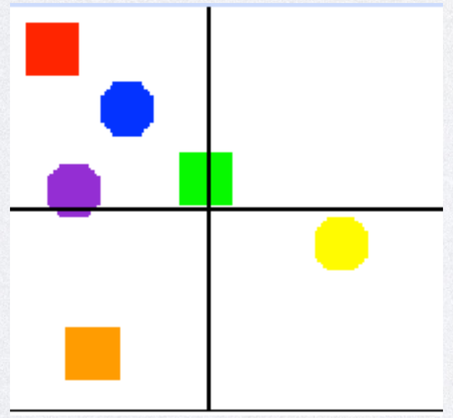
75_5



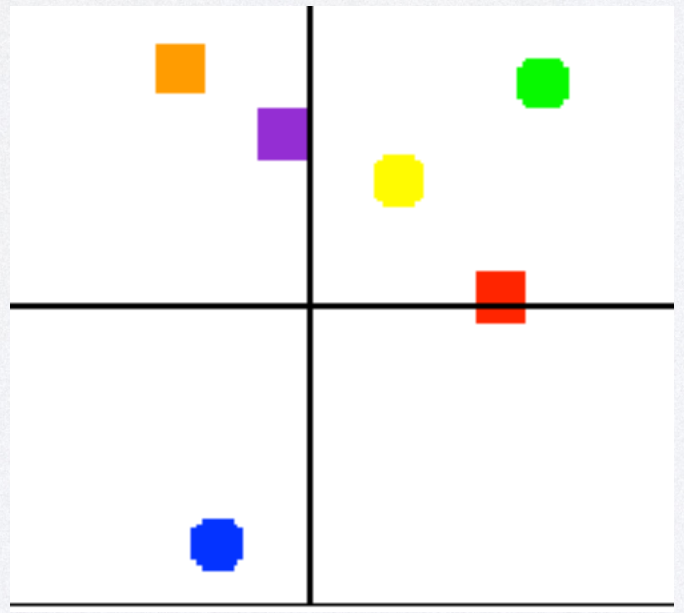
128_8



64_5



75_5



128_5

Table 4: image size-object size

		64_4	128_8	64_5	128_5
SARN	non-rel	0.9970	0.9999	0.9948	0.9988
	rel	0.8949	0.9440	0.8370	0.8669
	total	0.9345	0.9650	0.8970	0.9163
RN	non-rel	0.9944	0.9981	0.9964	0.9931
	rel	0.8415	0.8207	0.8430	0.7719
	total	0.8989	0.8872	0.9005	0.8555
baseline	non-rel	0.9941	0.9972	0.9933	0.9978
	rel	0.8120	0.8625	0.8163	0.8532
	total	0.8803	0.9130	0.8827	0.9074

STRENGTH OF SARN

1. **Computation efficiency**
    - n^2 -> n
2. **Better Performance**
3. **Interpretability**

FUTURE WORKS

- Lack of Chaining (yet!)
  - Memory
- Reuse of entities
  - A->C->B->**A**->D

CONCLUSION

- How to represent relations? = How to form a reasonable graph from image?
  - Identify **Entities** (Modularity)
    - Attention / Conditional CNN
  - **Relations** are defined from relational reasoning
    - MLP / Self-attention
  - **Chaining**
    - Sum / LSTM?