CSCI 599: Deep Learning and its Applications

Lecture 10

Spring 2019
Youngwoon Lee

Cognitive Science inspired Deep Learning

Image from http://clipart-library.com/clipart/695303.htm

Attention

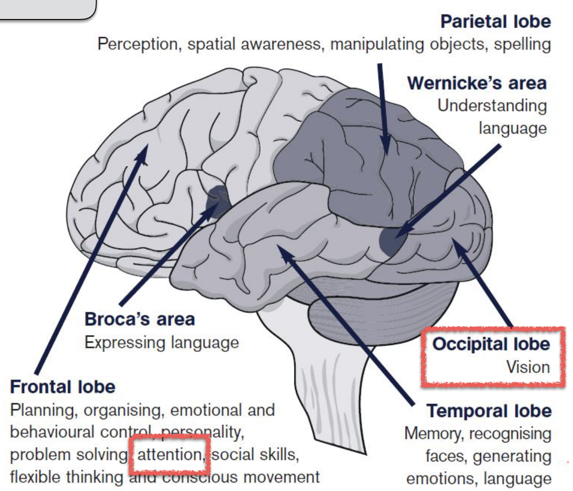


Image from http://headwaythamesvalley.org.uk/about-brain-injury-2/about-the-brain/

Memory

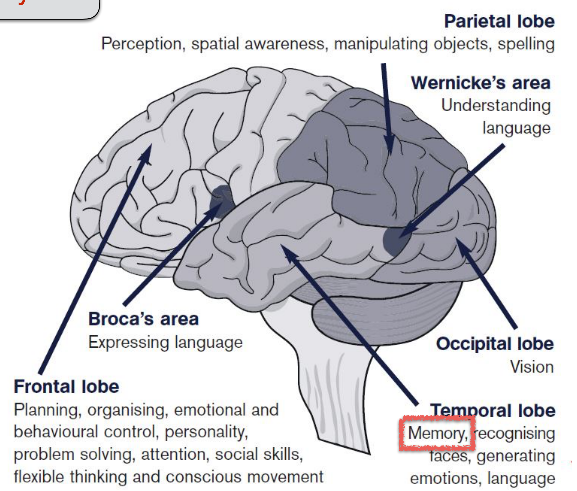


Image from http://headwaythamesvalley.org.uk/about-brain-injury-2/about-the-brain/

Relationship

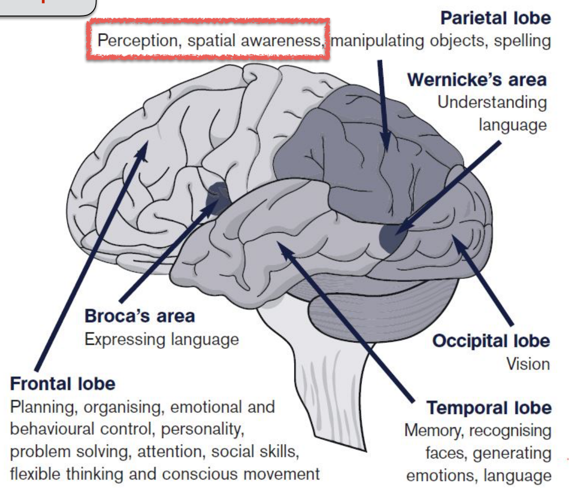


Image from http://headwaythamesvalley.org.uk/about-brain-injury-2/about-the-brain/

Today's agenda

- Part 1
  - Attention Model
- Part 2
  - Memory Network
- Part 3
  - Relation Network

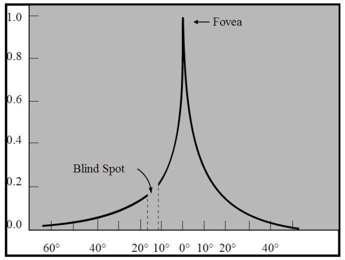
Attention Model

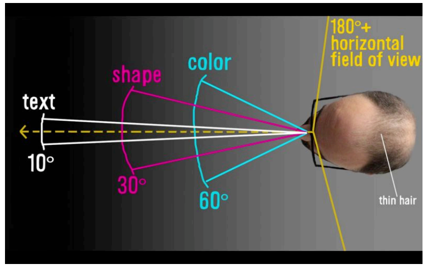
Image from http://clipart-library.com/clipart/695303.htm

Cocktail party effect

Image from https://www.sciencenews.org/article/new-ai-can-focus-one-voice-crowd

Human eyes cannot see everywhere in high resolution





Spatial attention

A woman is throwing a frisbee in a park.

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

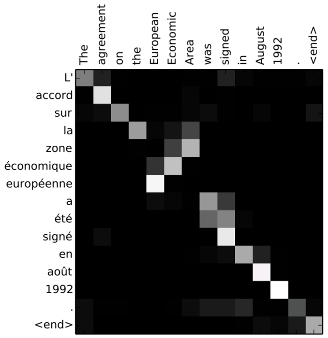
Spatial attention

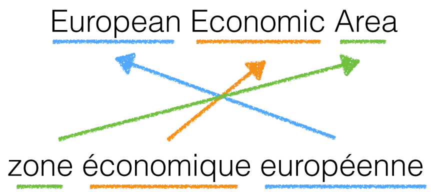
A woman is throwing a frisbee in a park.

- Humans only see specific regions in a higher resolution

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

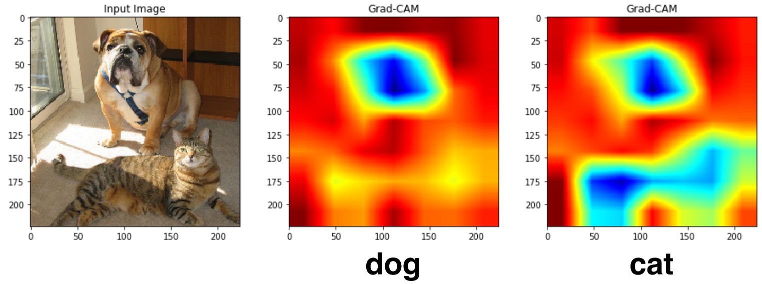
Temporal attention





Bahdanau et. al. Neural Machine Translation by Jointly Learning to Align and Translate. 2015.

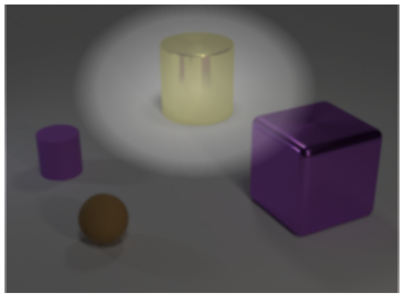
Implicit attention in neural networks



- Deep neural networks naturally learn to focus more on some parts of the data than others

From assignment #1

Implicit attention in neural networks



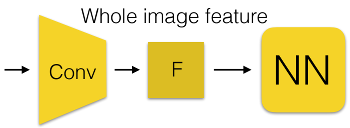
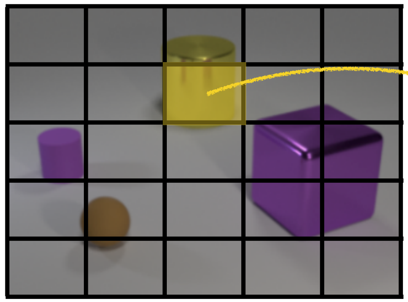
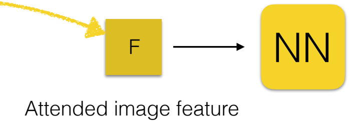


Image from Santoro et. al. A simple neural network module for relational reasoning. 2017.

Explicit attention in neural networks



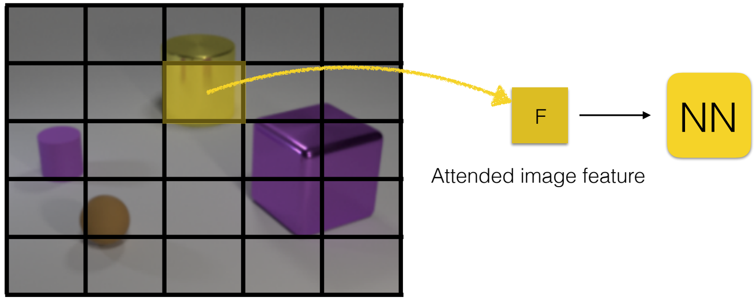


- An explicit attention mechanism provides an inductive bias to a neural network

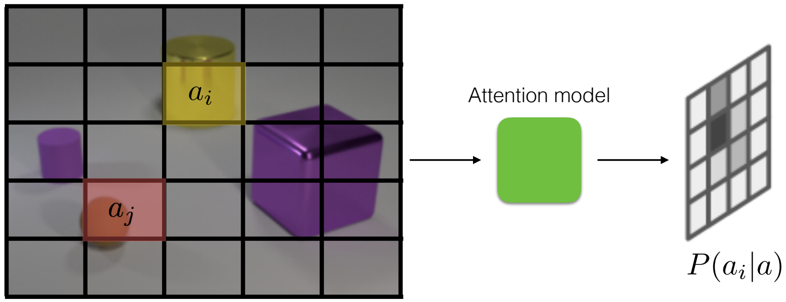
Advantages of explicit attention

- Efficient computation
  - Extract important information and discard useless information
- Scalable to variable length inputs
- Can be applied to temporal domain (language, video)
- Easier to interpret where to focus

Explicit attention in neural networks

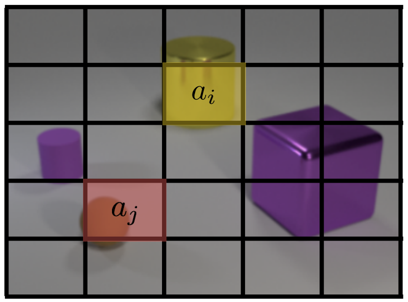


Attention model



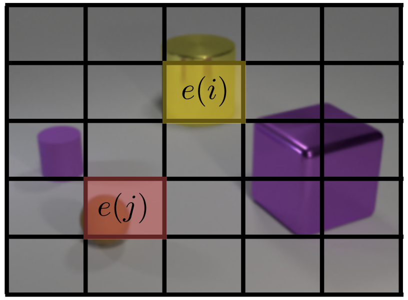
- Attention models predict a probability distribution over glimpses from an input **a**: $P(a_i|a)$

Attention model



a_i: CNN feature of the cell ***i***

Attention model

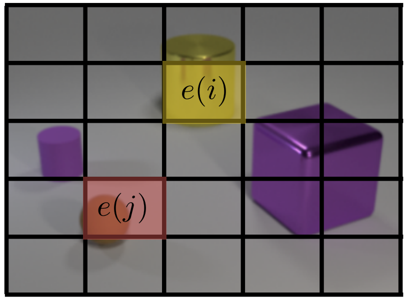


Importance of each cell

$$e(i) = f_{\text{att}}(a_i, a)$$

a_i: CNN feature of the cell ***i***

Attention model



Importance of each cell

$$e(i) = f_{\text{att}}(a_i, a)$$

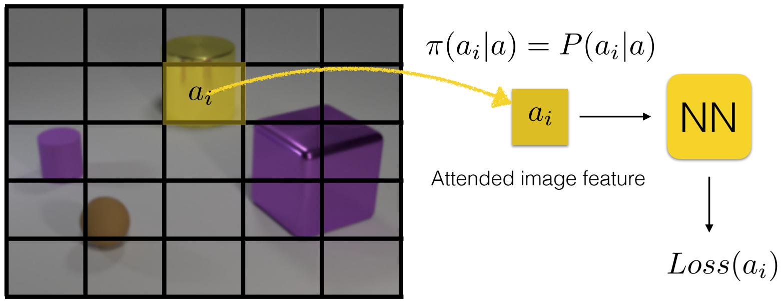
Attention probability of each cell

$$P(\cdot|a) = \text{softmax}(e(\cdot))$$

a_i: CNN feature of the cell ***i***

Attention model with RL

$$\pi(a_i|a) = P(a_i|a)$$



- We can learn a supervised task with RL by setting a task loss as a reward

Attention model with RL

$$\pi(a_i|a) = P(a_i|a)$$

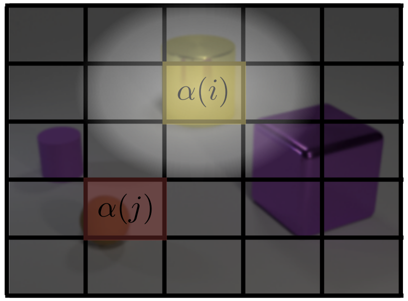
Hard attention is not differentiable!

Make it differentiable (Soft attention)

$a_i)$

- We can learn a supervised task with RL by setting a task loss as a reward

Soft attention



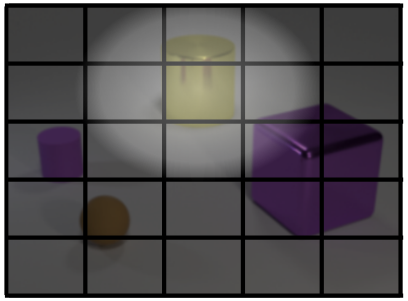
Importance of each cell

$$e(i) = f_{\text{att}}(a_i, a)$$

Attention probability of each cell

$$P(\cdot|a) = \text{softmax}(e(\cdot))$$

Soft attention



Importance of each cell

$$e(i) = f_{\text{att}}(a_i, a)$$

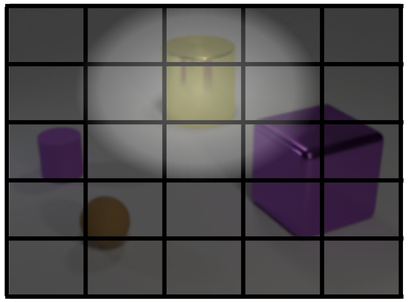
Attention probability of each cell

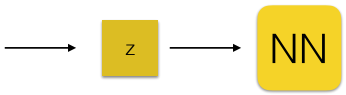
$$P(\cdot|a) = \text{softmax}(e(\cdot))$$

Attended feature

$$z = \sum_i P(a_i|a) \cdot a_i$$

Soft attention



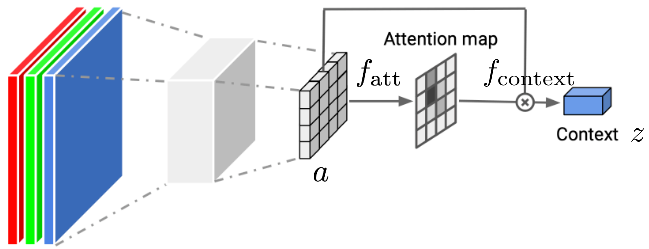


Weighted sum of features

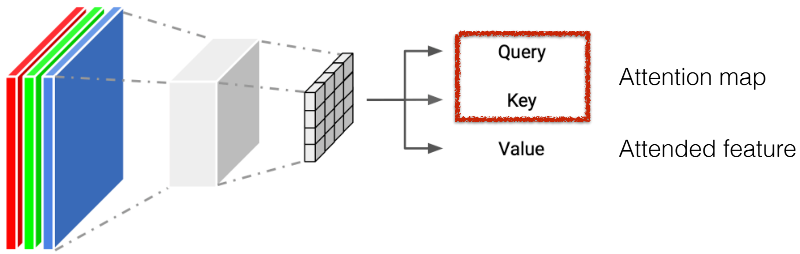
Attended feature

$$z = \sum_i P(a_i|a) \cdot a_i$$

Attention model

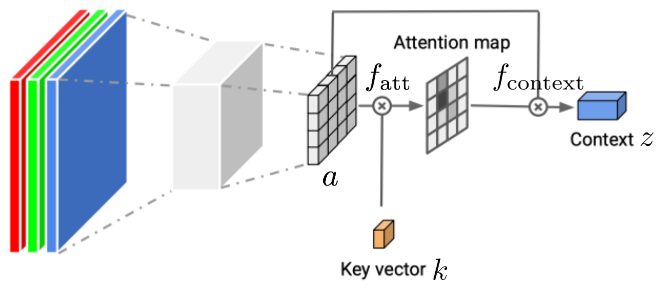


Associative Attention: attending by content



Importance score = Similarity(Query, Key)

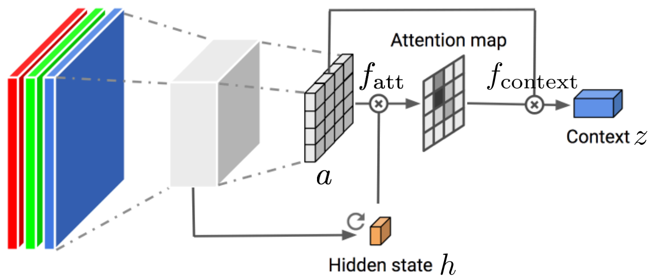
Associative Attention: attending by content



$$e(i) = f_{\text{att}}(a_i) \longrightarrow e(i) = f_{\text{att}}(a_i, a, k)$$

Similarity between the content and key

Attention in recurrent networks



$$e(i) = f_{\text{att}}(a_i) \longrightarrow e(i) = f_{\text{att}}(a_i, h_{t-1})$$

Applications of attention models

- Image captioning
- Image generation
- Language model and machine translation
- RL

Image captioning with Attention

A woman is throwing a frisbee in a park.

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention

A woman

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention

A woman **is throwing**

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention

A woman is throwing **a frisbee**

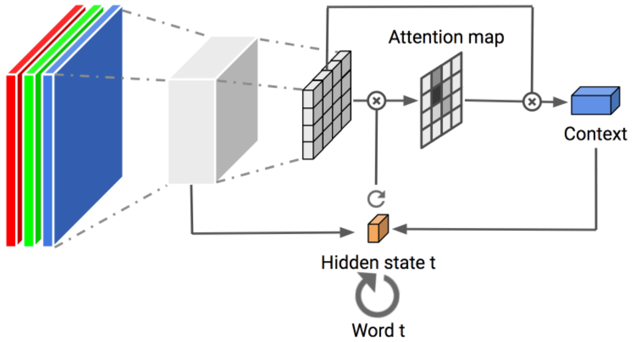
Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention

A woman is throwing a frisbee **in a park**

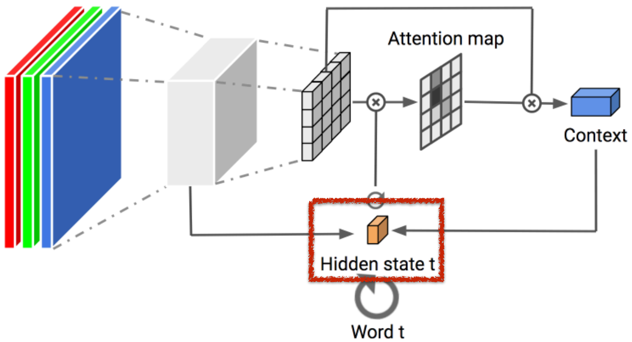
Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention



Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

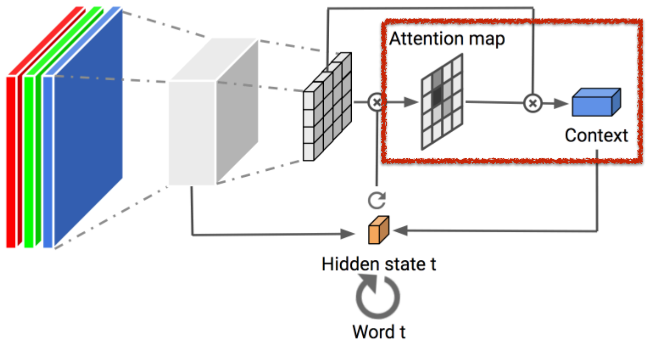
Image captioning with Attention



Compute a hidden state based on the image

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

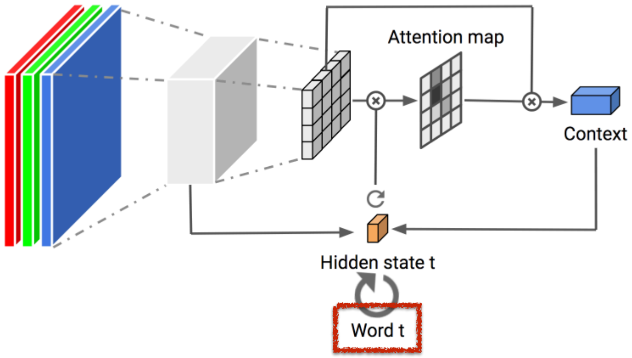
Image captioning with Attention



Compute an attention map and context

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

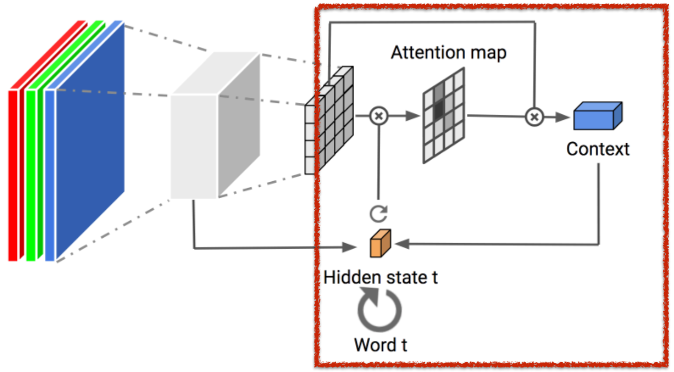
Image captioning with Attention



Generate a word

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Image captioning with Attention



Repeat this process to generate a sentence

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

Results

A woman is throwing a frisbee in a park. A dog is standing on a hardwood floor. A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear. A group of people sitting on a boat in the water. A giraffe standing in a forest with trees in the background.

Xu et. al., Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. 2015.

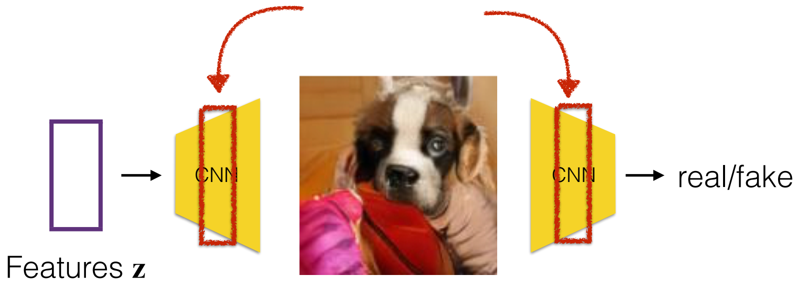
Image generation

A convolutional network focuses on local features.
It is hard to preserve long-range dependency.

Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

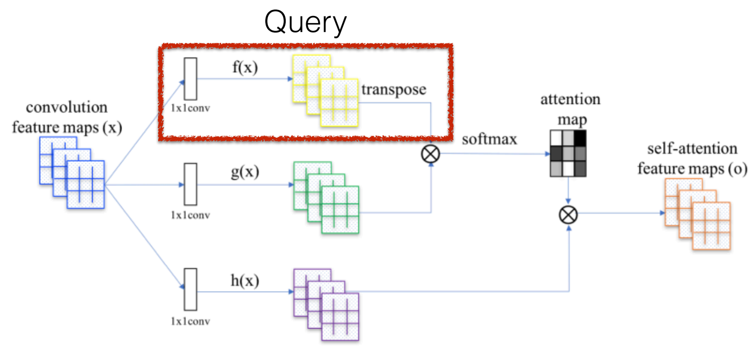
Image generation

Add attention models to get details from **all feature locations**



Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

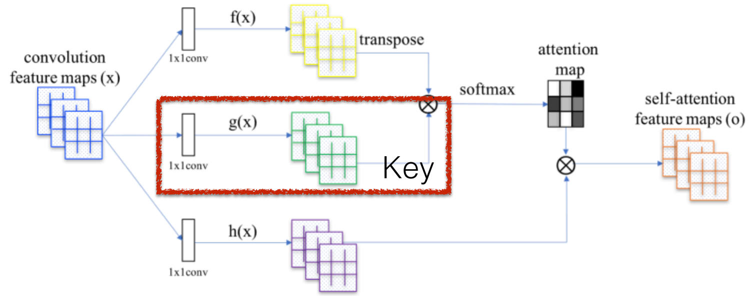
Self-Attention GAN



Use attention to share local features over the whole image

Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

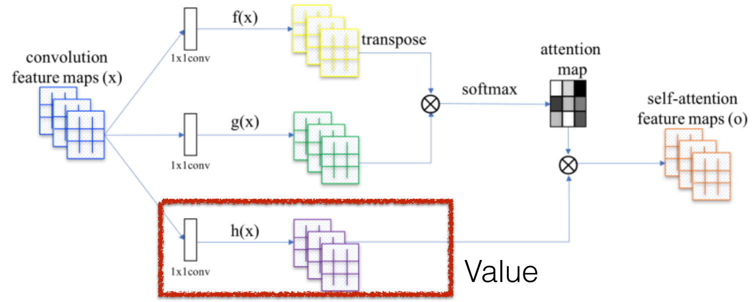
Self-Attention GAN



Use attention to share local features over the whole image

Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

Self-Attention GAN

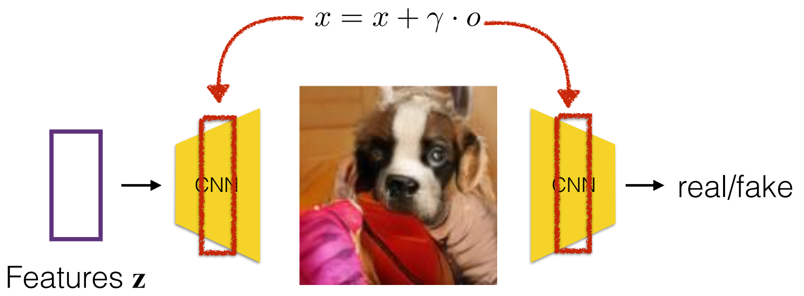


Use attention to share local features over the whole image

Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

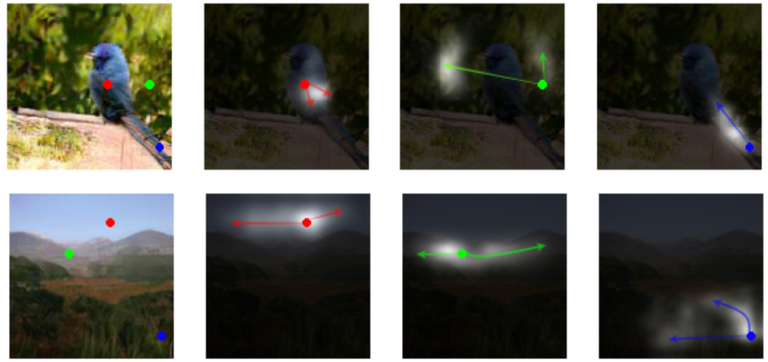
Self-Attention GAN

Add attention models to get details from **all feature locations**



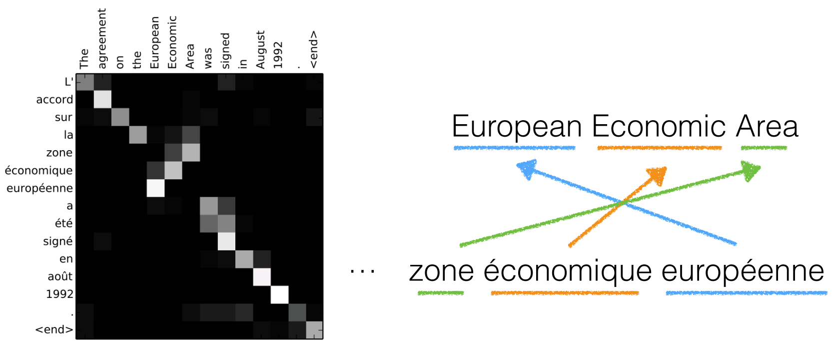
Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

Results



Zhang et. al., Self-Attention Generative Adversarial Networks. 2018.

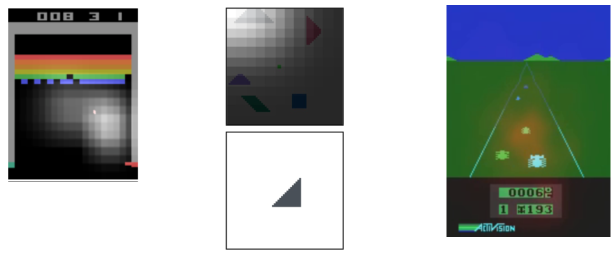
Neural machine translation



Bahdanau et. al. Neural Machine Translation by Jointly Learning to Align and Translate. 2015.

RL with Attention

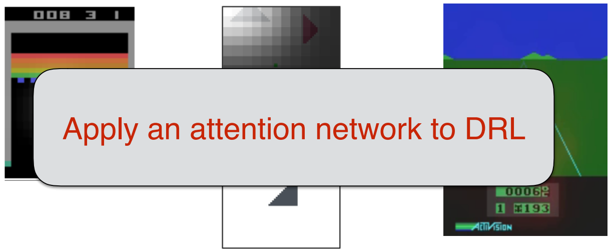
When we play games, we only focus on important things



Sorokin et. al., Deep Attention Recurrent Q-Network. 2015.

RL with Attention

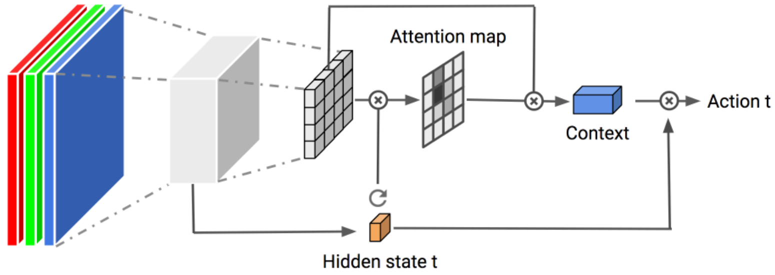
When humans play game, we only focus on important things



Apply an attention network to DRL

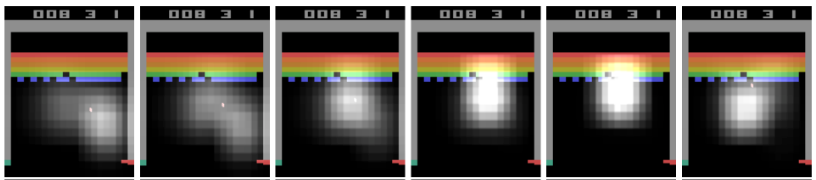
Sorokin et. al., Deep Attention Recurrent Q-Network. 2015.

RL with Attention



Sorokin et. al., Deep Attention Recurrent Q-Network. 2015.

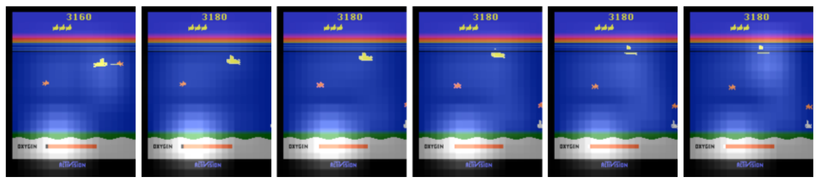
Results



Focus on the ball

Sorokin et. al., Deep Attention Recurrent Q-Network. 2015.

Results



Focus on the Oxygen gauge

Sorokin et. al., Deep Attention Recurrent Q-Network. 2015.

Summary of attention model

- Attention models help deep neural networks to focus on important information
- Interpretability
- End-to-end training is possible using soft attention
- Many different attention models are possible depending on applications

Today's agenda

- Part 1
  - Attention Model
- Part 2
  - Memory Network
- Part 3
  - Relation Network

Memory Network

Slide credit: Graves et al. http://people.idsia.ch/~rupesh/rnnsymposium2016/slides/graves.pdf
Graves et. al. Hybrid computing using a neural network with dynamic external memory. 2016.

Image from http://clipart-library.com/clipart/695303.htm

Memory network

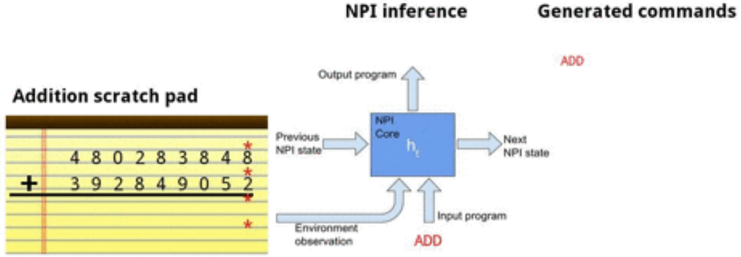


Image from https://medium.com/near-ai/review-of-neural-programmer-interpreters-854a14a494fb

Memory network

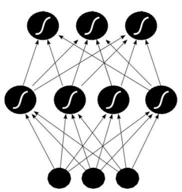
- A neural network is hardly storing data over long time

Memory network

- A neural network is hardly storing data over long time
- Explicit memory can help to
  - Remember past events
  - Manipulate complex data (graph)
  - Construct a map (for navigation)

Memory network

- Augment an external memory to a neural network



'CPU'

+

Memory

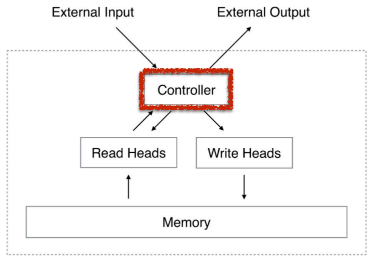
=

Computer that learns from examples

(*or* neural net that separates computation from memory)

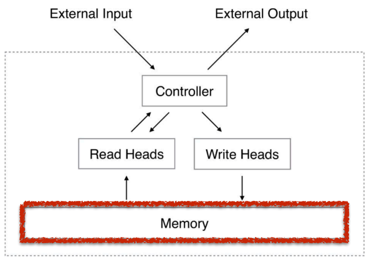
Memory network overview

- The **Controller** is a main module



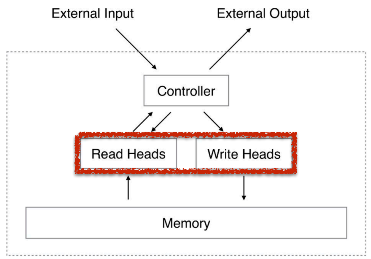
Memory network overview

- The **Controller** is a main module
- The **Memory** can be
  - a list of values
  - key, value pairs
  - a map (2d grid, 3d grid)



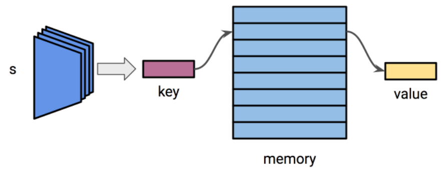
Memory network overview

- The **Controller** is a main module
- The **Memory** can be
  - a list of values
  - key, value pairs
  - a map (2d grid, 3d grid)
- The controller can access the memory through **Read/Write Heads**

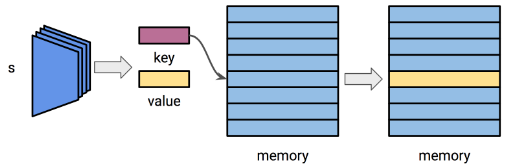


Architecture

Read



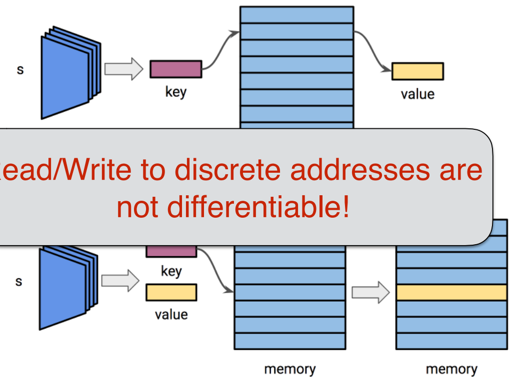
Write



Architecture

Read

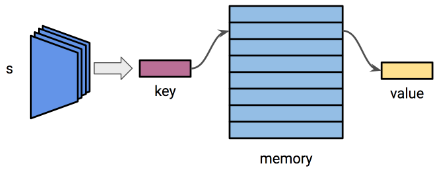
Write



Read/Write to discrete addresses are not differentiable!

Differentiable memory network

Read

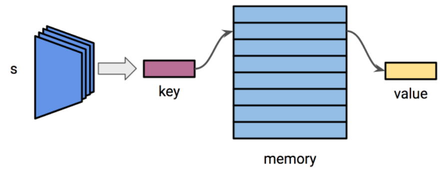


Apply a soft attention when read

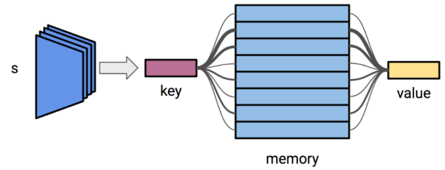
$$r = \sum_i w^r[i] \cdot \mathrm{Memory}[i, \cdot]$$

Differentiable memory network

Read

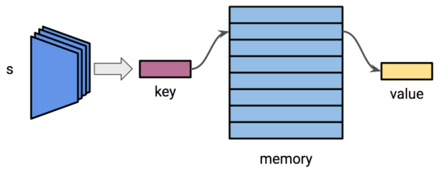


Attention

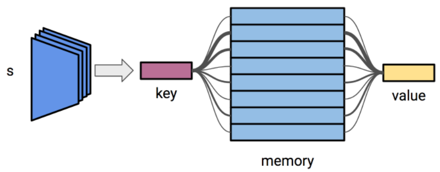


Differentiable memory network

Read



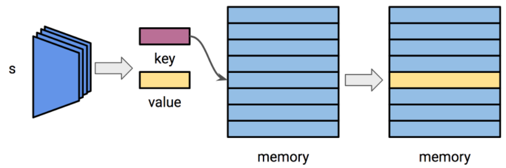
Attention



Differentiable!

Differentiable Memory Network

Write

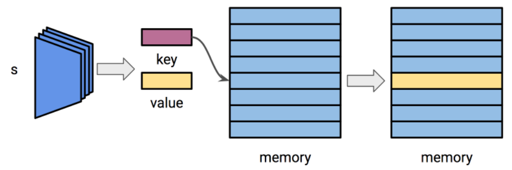


Apply a soft attention when write

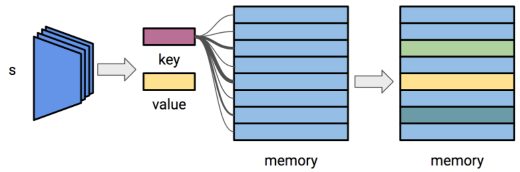
$$\mathrm{Memory}[i, \cdot] = (1 - \gamma \cdot w^w[i]) \cdot \mathrm{Memory}[i, \cdot] + w^w[i] \cdot \mathrm{NewValue}$$

Differentiable Memory Network

Write

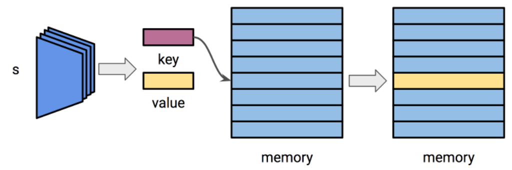


Attention

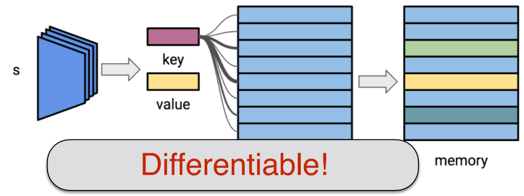


Differentiable Memory Network

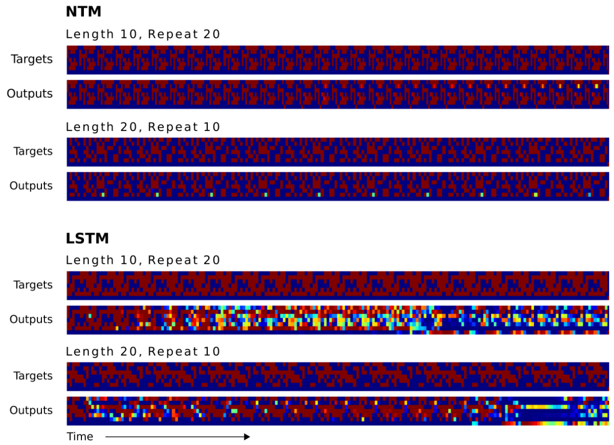
Write



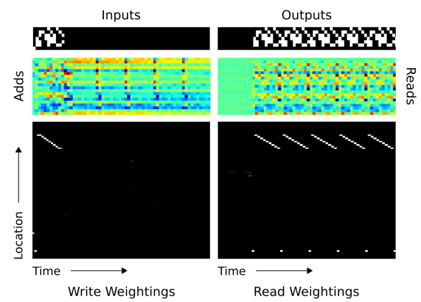
Attention



Neural Turing Machine



Neural Turing Machine

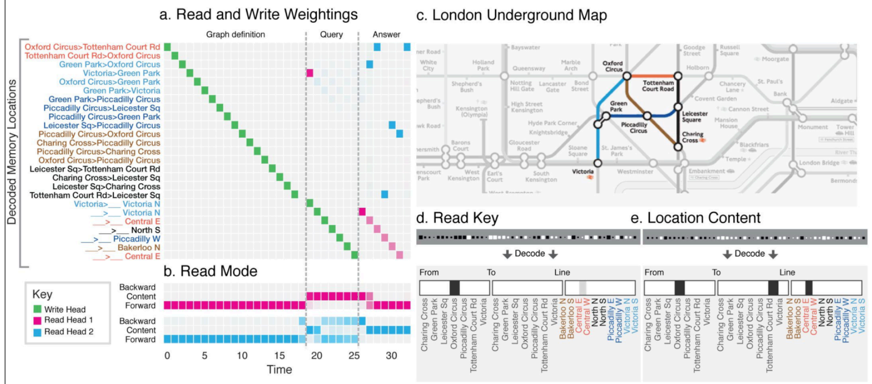


Q&A task

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

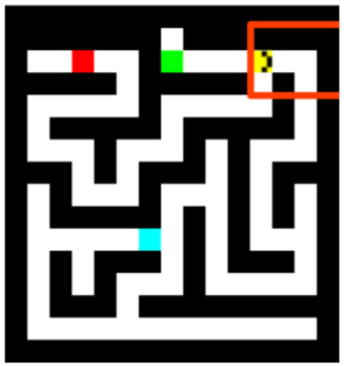
Sukhbaatar et. al. End-To-End Memory Networks. 2015.

Differentiable Neural Computers



Graves et. al. Hybrid computing using a neural network with dynamic external memory. 2016.

Neural Map



(a) Maze

(b) Observation

Parisotto et. al. Neural Map: Structured Memory for Deep Reinforcement Learning. 2017.

Summary of memory network

- The external memory can preserve data longer
- Soft attention can make a memory differentiable

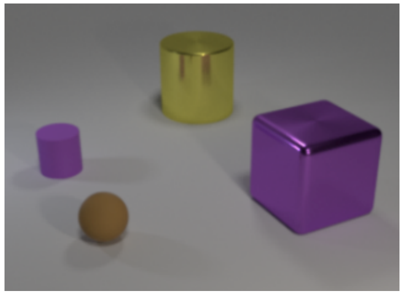
Today's agenda

- Part 1
  - Attention Model
- Part 2
  - Memory Network
- Part 3
  - Relation Network

Relation Network

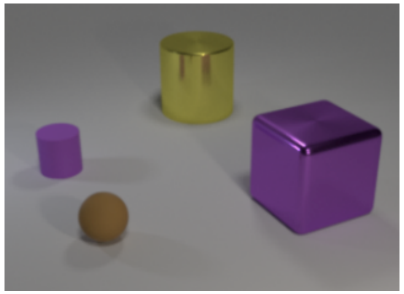
Image from http://clipart-library.com/clipart/695303.htm

Relational Reasoning



Santoro et. al. A simple neural network module for relational reasoning. 2017.

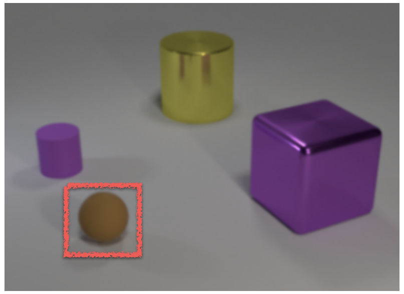
Relational Reasoning



Question:
What is the shape of the object that is farthest from the rubber sphere?

Santoro et. al. A simple neural network module for relational reasoning. 2017.

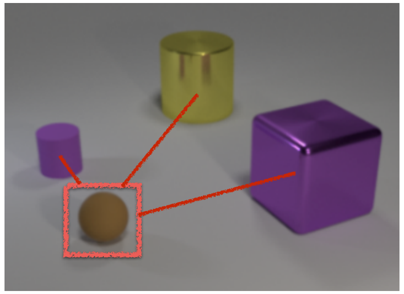
Relational Reasoning



Question:
What is the shape of the object that is farthest from the rubber sphere?

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational Reasoning

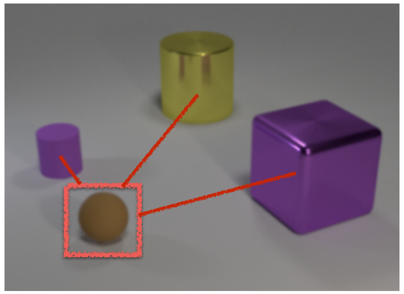


Question:
What is the shape of the object that is farthest from the rubber sphere?

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Not efficient

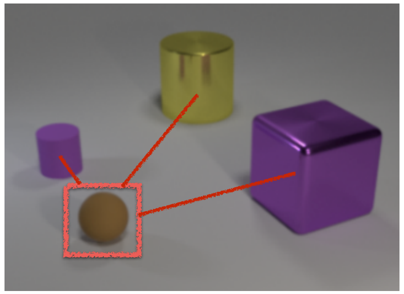
Naive CNN



Question:
What is the shape of the object that is farthest from the rubber sphere?

Santoro et. al. A simple neural network module for relational reasoning. 2017.

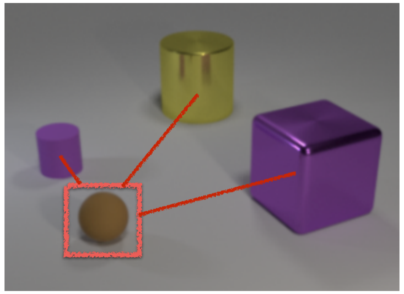
Relational Reasoning



How to understand spatial relationship between objects?

Santoro et. al. A simple neural network module for relational reasoning. 2017.

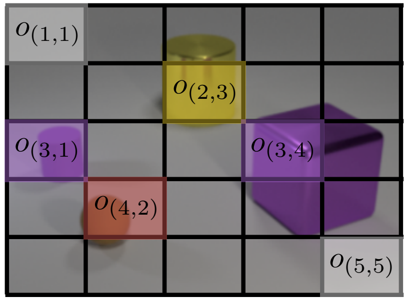
Relational Reasoning



Object- and relation-centric state representation and relational reasoning need to be learned.

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational network

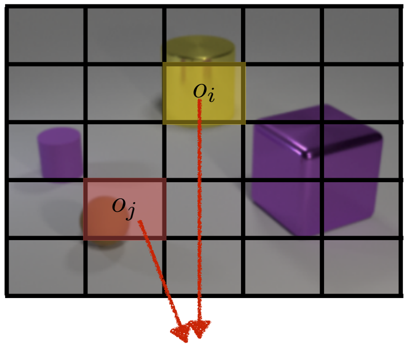


5 x 5 objects

Each cell is considered as an “object”

Santoro et. al. A simple neural network module for relational reasoning. 2017.

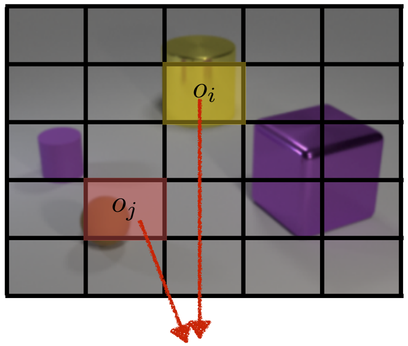
Relational network



Infer relation between two objects

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational network

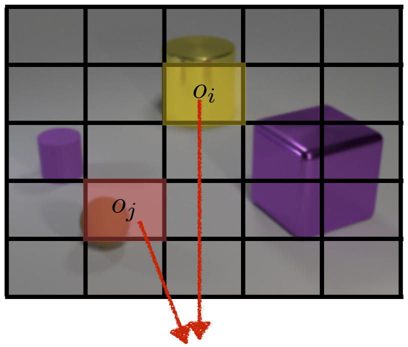


Infer relation between two objects

$g_\theta(o_i, o_j)$

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational network

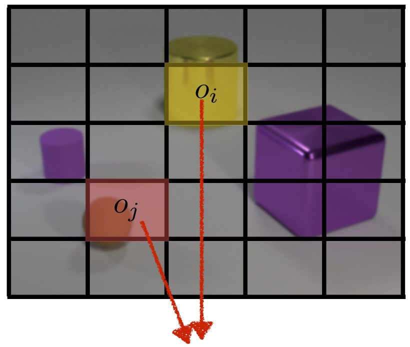


Infer relation between two objects

$$\sum_{i,j} g_\theta(o_i, o_j)$$

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational network

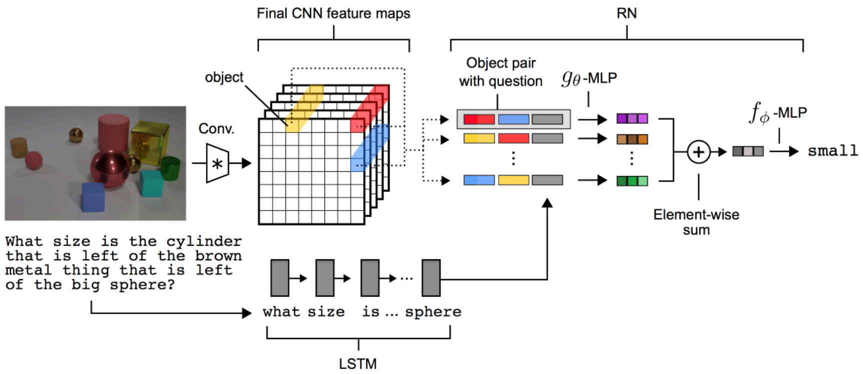


Infer relation between two objects w.r.t. a **question**

$$\sum_{i,j} g_\theta(o_i, o_j, q)$$

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational Network

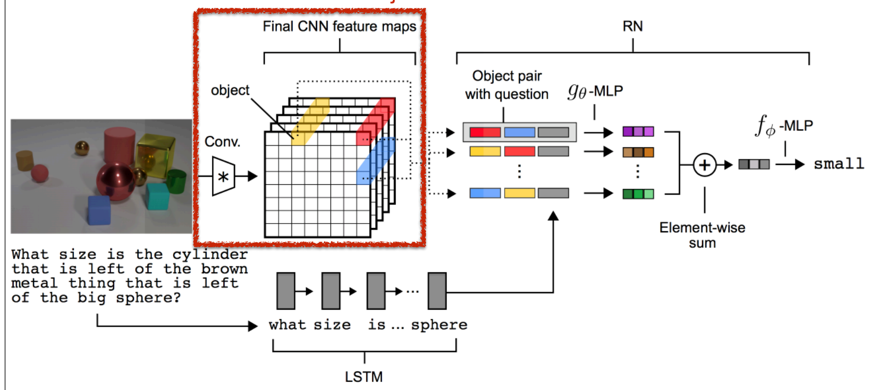


Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational Network

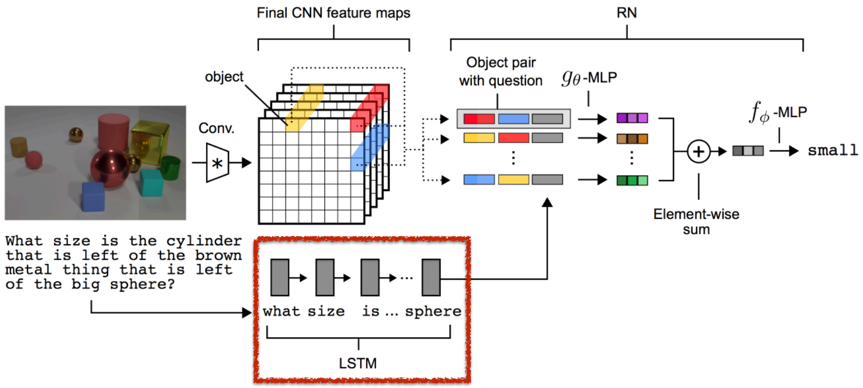
Features of objects



Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational Network

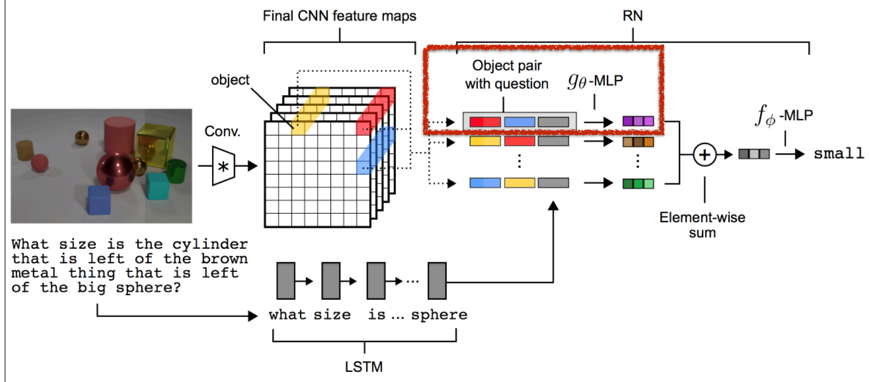


Question embedding

Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow
Santoro et. al. A simple neural network module for relational reasoning. 2017.

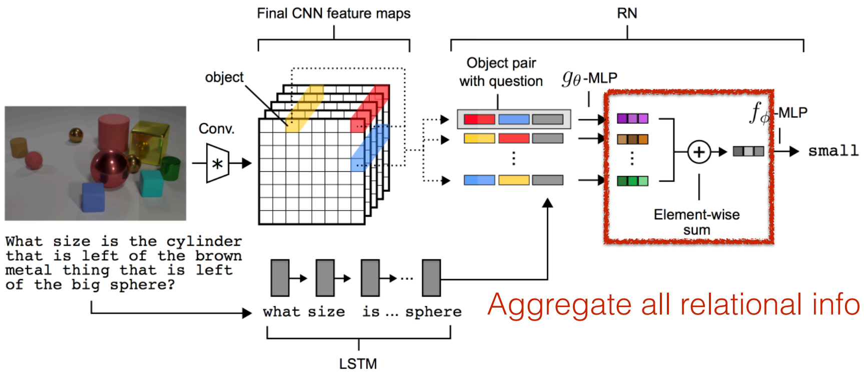
Relational Network

Relationship between objects



Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow
Santoro et. al. A simple neural network module for relational reasoning. 2017.

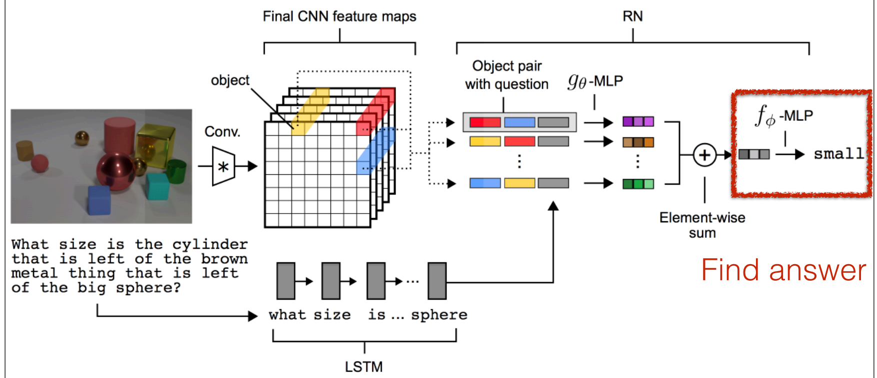
Relational Network



Aggregate all relational info

Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow
Santoro et. al. A simple neural network module for relational reasoning. 2017.

Relational Network



Find answer

Code is available at https://github.com/gitlimlab/Relation-Network-Tensorflow
Santoro et. al. A simple neural network module for relational reasoning. 2017.

Results

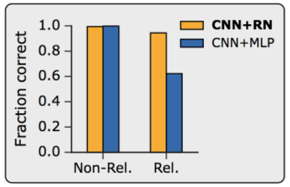
Sort-of-CLEVR

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

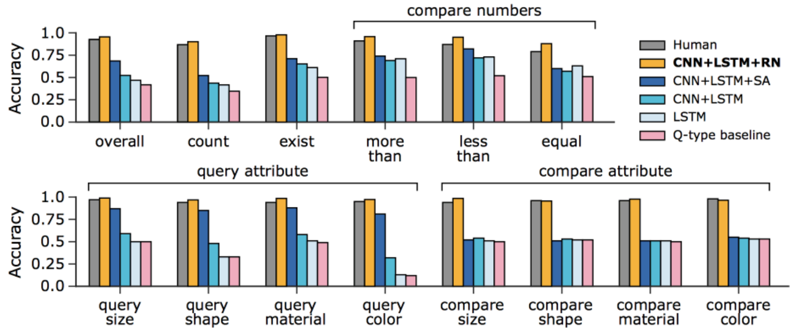
Relational question

Q: How many objects have the shape of the orange object?
A: 3

Santoro et. al. A simple neural network module for relational reasoning. 2017.

Results

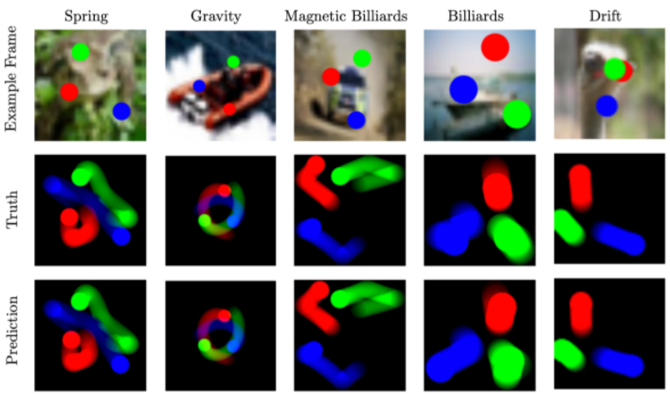
CLEVR



Super-human performance

Santoro et. al. A simple neural network module for relational reasoning. 2017.

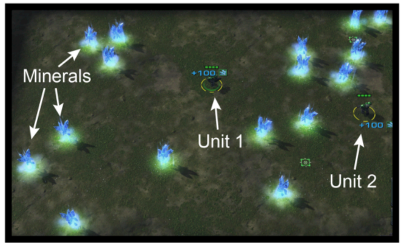
Results



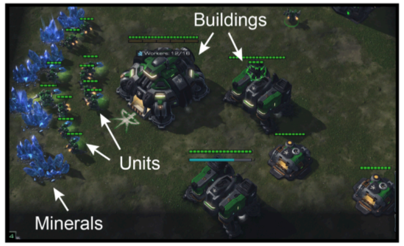
Watters et. al. Visual Interaction Networks. 2017.

Iterative relational reasoning

StarCraft II mini-game
Collect Mineral Shards



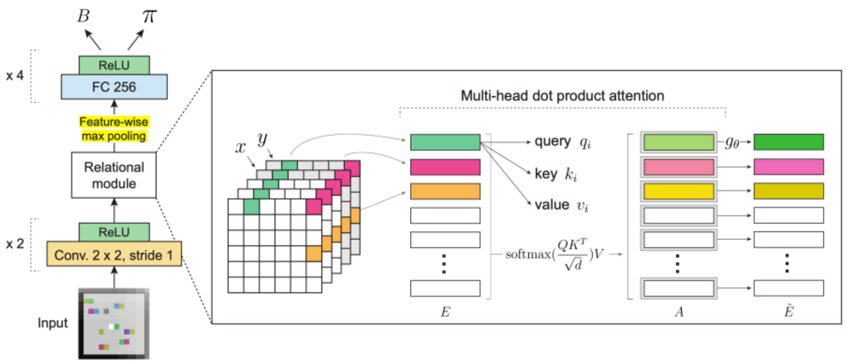
StarCraft II mini-game
Build Marines



One-step relational reasoning is not enough to solve complicated tasks

Zambaldi et. al. Deep reinforcement learning with relational inductive biases. 2019.

Relation network + Attention



Zambaldi et. al. Deep reinforcement learning with relational inductive biases. 2019.

Zambaldi et. al. Deep reinforcement learning with relational inductive biases. 2019.

Summary of relation network

- A relation network can efficiently represent relational information by sharing parameters over different pairs
- Relational reasoning is necessary to simplify a complexity of some problems with structures

Cognitive Science inspired Deep Learning

Image from http://clipart-library.com/clipart/695303.htm

Next lecture

- Resume Reinforcement Learning next class