Faculté
des Sciences
Aix*Marseille Université

MacNetwork

Tiago Cortinhal
Professor: Thierry Artières
University Of Aix-Marseille

Outline

1. **Reasoning**
   - *The task at hand*
2. **The Datasets**
3. **End to End Memory Network**
   - *How to solve this?*
4. **MAC Cell**
5. **Some Results**
   - *What we did*
   - *CLEVR*
   - *Babi*

Reasoning

What is relational reasoning?

- Finding who is the murder on an Agatha Christie novel
- Figure out spacial relation from textual clues
- Understanding which number comes next on a sequence of numbers

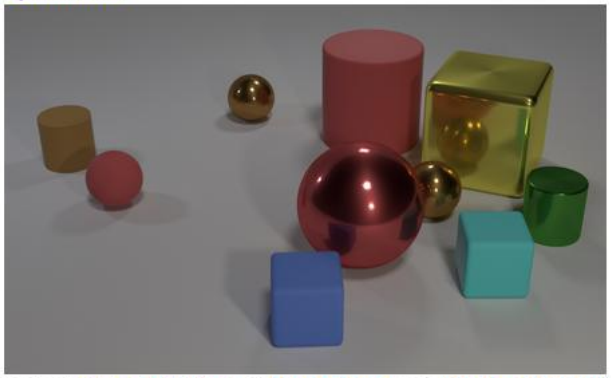
Reasoning

What is relational reasoning?

- Manipulating previous knowledge in order to answer a new question
- Composition rules to guide and address new questions

Example of reasoning datasets

Questions in CLEVR test various aspects of visual reasoning including **attribute identification**, **counting**, **comparison**, **spatial relationships**, and **logical operations**.



Q: Are there an **equal number** of **large things** and **metal spheres**?
Q: **What size** is the **cylinder** **that is left of** the **brown metal** thing **that is left of** the **big sphere**?
Q: There is a **sphere** with the **same size as** the **metal cube**; is it **made of the same material as** the **small red sphere**?
Q: **How many** objects are **either small cylinders** or **red** things?

Johnson et al, 2017

Task 1: Single Supporting Fact
Mary went to the bathroom.
John moved to the hallway.
Mary travelled to the office.
Where is Mary? A:office

Task 2: Two Supporting Facts
John is in the playground.
John picked up the football.
Bob went to the kitchen.
Where is the football? A:playground

Task 3: Three Supporting Facts
John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A:office

Task 4: Two Argument Relations
The office is north of the bedroom.
The bedroom is north of the bathroom.
The kitchen is west of the garden.
What is north of the bedroom? A: office
What is the bedroom north of? A: bathroom

Task 5: Three Argument Relations
Mary gave the cake to Fred.
Fred gave the cake to Bill.
Jeff was given the milk by Bill.
Who gave the cake to Fred? A: Mary
Who did Fred give the cake to? A: Bill

Task 6: Yes/No Questions
John moved to the playground.
Daniel went to the bathroom.
John went back to the hallway.
Is John in the playground? A:no
Is Daniel in the bathroom? A:yes

Task 7: Counting
Daniel picked up the football.
Daniel dropped the football.
Daniel got the milk.
Daniel took the apple.
How many objects is Daniel holding? A: two

Task 8: Lists/Sets
Daniel picks up the football.
Daniel drops the newspaper.
Daniel picks up the milk.
John took the apple.
What is Daniel holding? milk, football

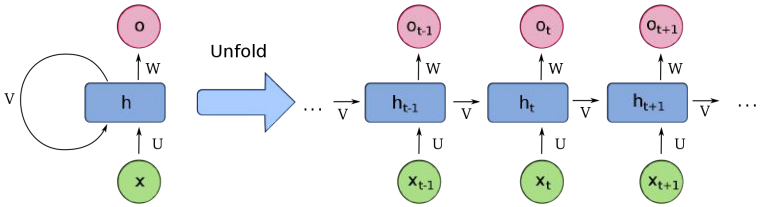
Task 9: Simple Negation
Sandra travelled to the office.
Fred is no longer in the office.
Is Fred in the office? A:no
Is Sandra in the office? A:yes

Task 10: Indefinite Knowledge
John is either in the classroom or the playground.
Sandra is in the garden.
Is John in the classroom? A:maybe
Is John in the office? A:no

Weston et al, 2015

How to solve this?

Let's start with RNN



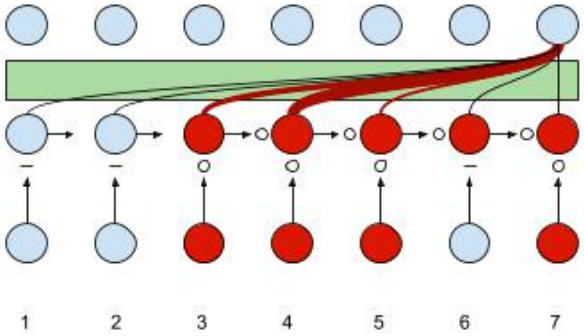
- Recurrent Neural Networks are good for temporal data (like audio, text, etc)
- Each node receives the ouput of the previous node alongside with the current data

Problems

- What about long term dependencies?
- And what if we have out of order sequences?

How to solve this?

Attention Model (Bahdanau et al. 2014)



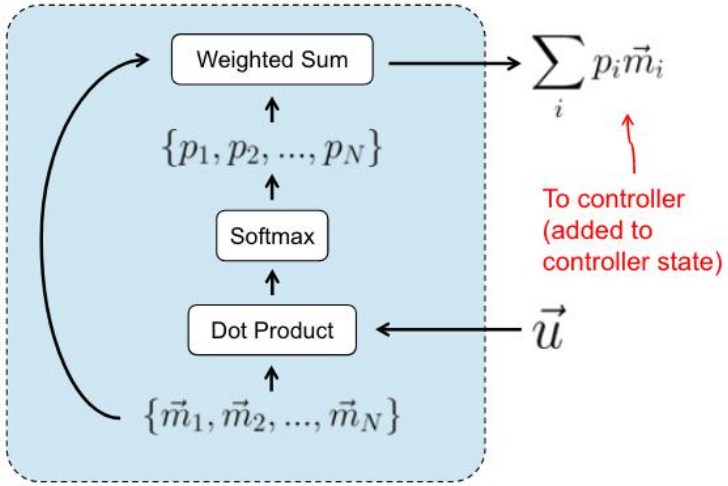
- We will have a new layer that tries to grasp the dependencies (attention) between the output and the previous states
- Started as a NLP solution but soon *invaded* other areas

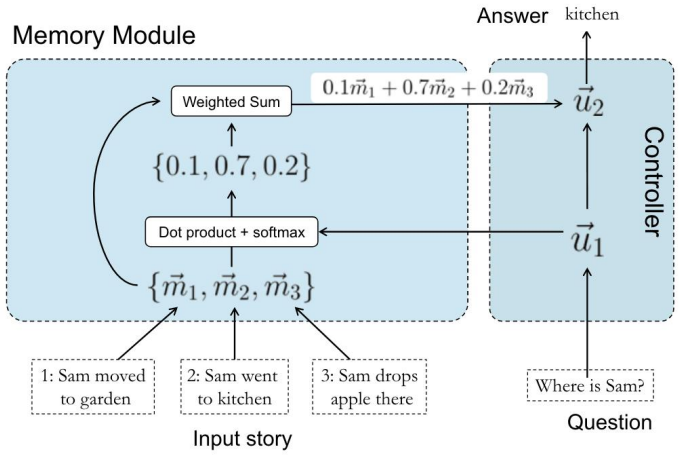
What does this solve?

- It allows the network to use the previous hidden states and the representation of the input to figure out what it should pay attention to
- Can help with long term dependencies

End to End Memory Network

Sukhbaatar et al. 2015





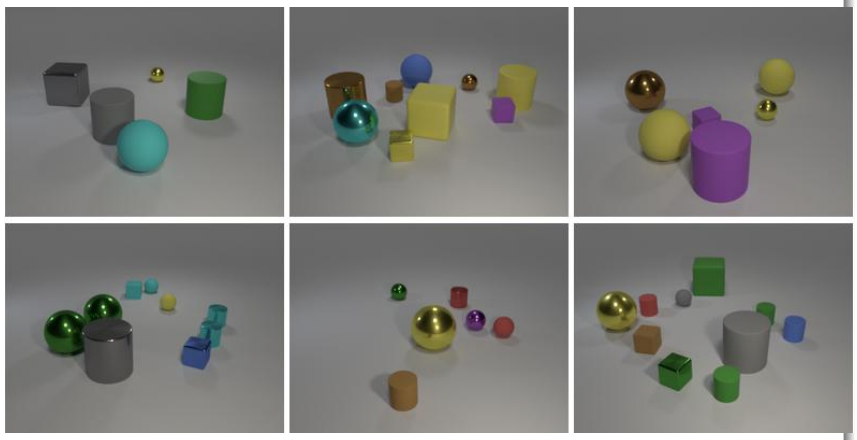
Main concepts

- Two different Modules: Memory and Control
- Memory controls what will be fed into our Controller
- Controller will help us guide the query until we reach the answer

Intuitions and Task at Hand

VQA and CLEVR

- Querying attributes
  - Two different sizes
  - Two materials
  - Three Shapes
  - Eight Colors
- Comparing
- Existence
- Counting
- Integer comparison



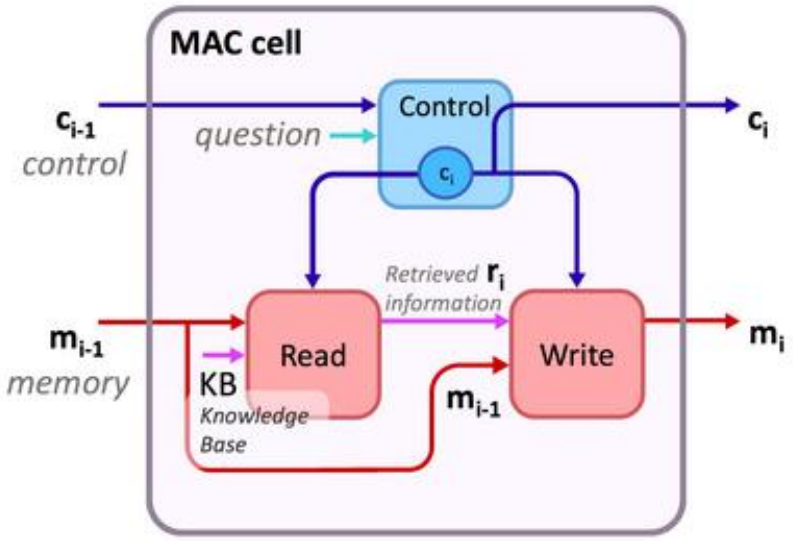
Intuitions and Task at Hand

Why was CLEVR created?

- The idea behind CLEVR is to have a Visual Questioning dataset with reduced bias and superficial reasoning
- By doing so it allows a deeper analysis of the performance and the system itself
- Artificial Dataset: images and questions are generated

Mac Cell

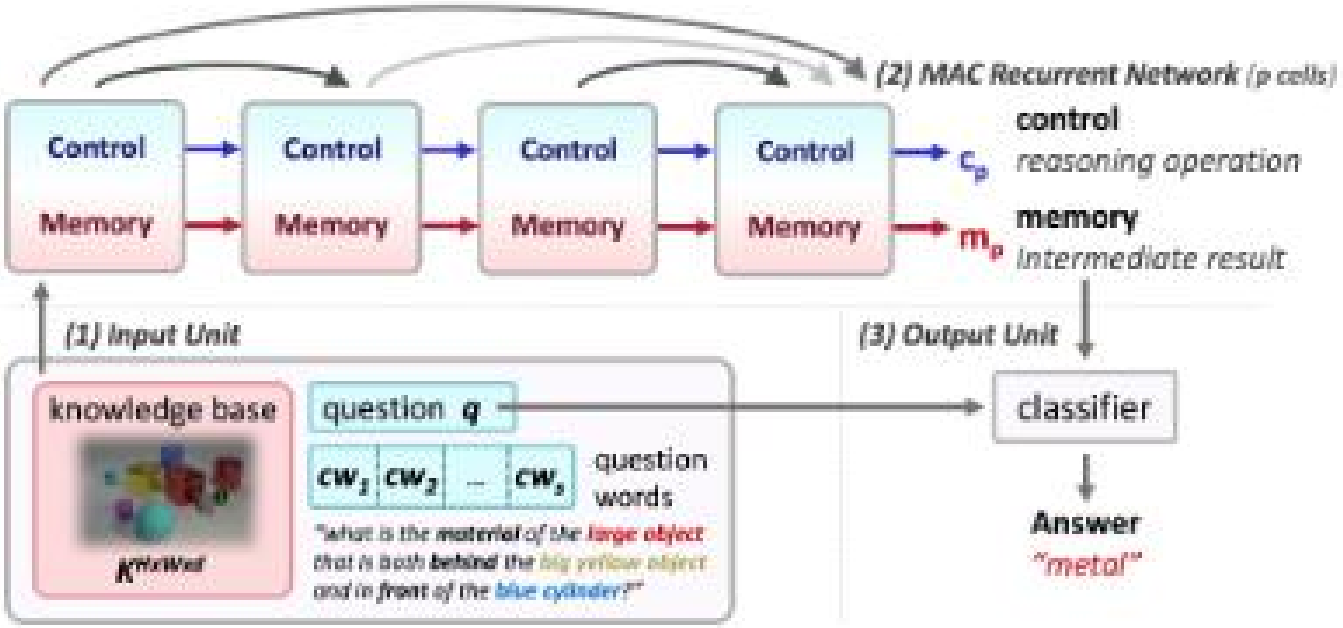
Mac Cell (Hudson et al, 2018)



- Three modules: Control, Read and Write
- Input
  - String of length S that will be processed by a d-dimensional *biLSTM* which will give the question representation $q = [\overleftarrow{cw_1}, \overrightarrow{cw_S}]$
  - Image is processed by a feature extractor trained on ResNet101 (outputs conv4) and passed through two CNN Layers with d output channels yielding a $KB_v = [H, W, d]$
- It has two recurrent states: Control and Memory (sounds familiar?)

Mac Cell

Example of Use

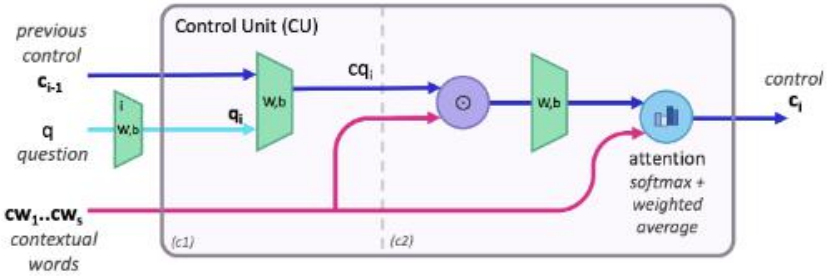


How does it work

- The Mac cell is constructed to work in a sequential manner, i.e., several cells connected together to add extra steps of reasoning
- We actually have 28 possible answers (from colors to numbers,etc), for the classification we use a softmax over the last memory and the question itself

Control Unit

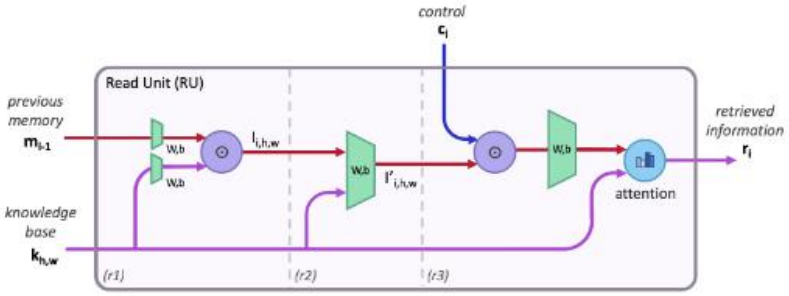
The Structure



- Computes a control state, extracting an instruction that focuses on some aspect of the query
- Input
  - Previous control state
  - Query
  - Contextual Words (from *BiLSTM*)
- First it calculates a time specific query representation
- Then it combines it with the preceding control
- Finally we project it down to the actual words of the query with an attention mechanism

Read Unit

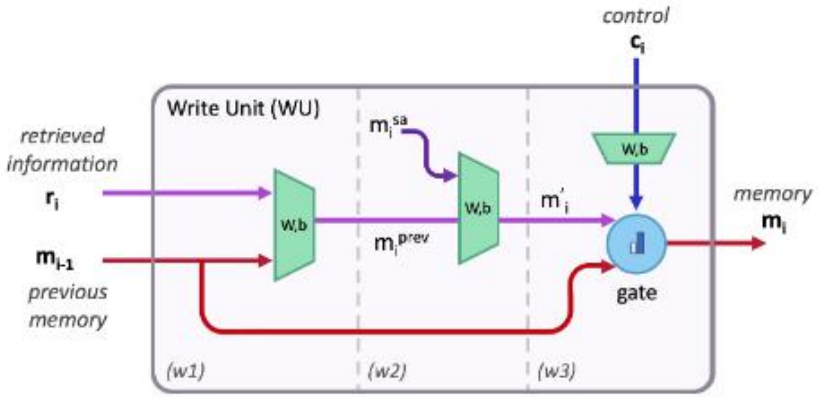
The Structure



- Retrieves information from the Knowledge base given the current control state and previous memory state
- Input
  - Previous memory
  - Knowledge Base
  - Current Control
- Will consider each item from the KB (the features of the image)
- We will then consider the interaction between the previous memory state (through a Hadamard product)
- The last step allows us do transitive reasoning (previous memory will guide you to that)
- After we consider the information in light of the current control
- We finalize by doing attention

Write Unit

The structure



- Update the new memory state, merging old and new information
- Input
  - Retrieved Information (from the Read Unit)
  - Previous Memory
  - Current Control
- We combine the previous memory with the new memory
- A self-attention mechanism (which gives access to all the previous memories) to help it guide the construction of the new memory
- We will have a gate that will allow us to decide if we want to update the memory or not

Our Attempt

What we did

- We first finished the Multi-GPU section ofthe original code (missing the gradient calculation between GPUs)
- We then tried to implement the MAC network on Keras (original works in tensorflow and torch), due to lack of time we didn't finish it
- Our goal was to train on Babi to confirm the hypothesis that MAC isn't only for VQA

- You can find both the improvement and the attempt here: TiagoCortinhal

CLEVR

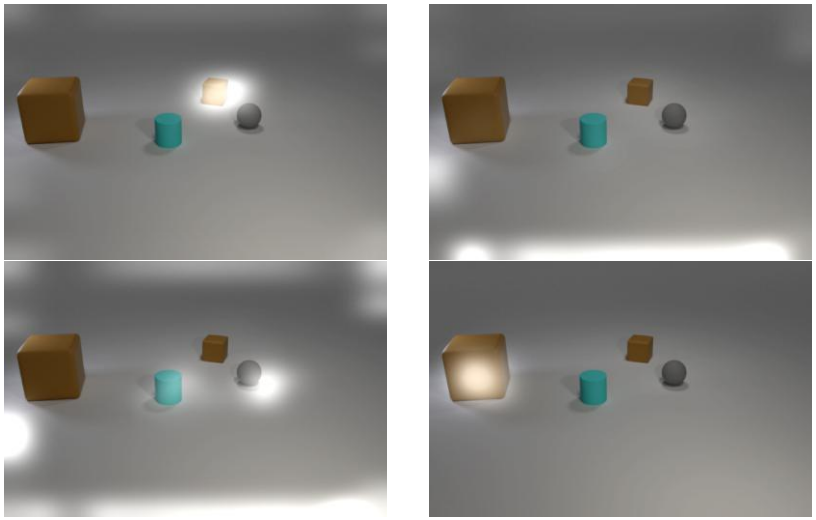
Implementation used

- Implementation of Hudson et al, 2018
- As presented in the paper

How training was done

- We used all CLEVR dataset
- trained over 25 epochs and 4 reasoning steps
- We achieved results of 97%

CLEVR



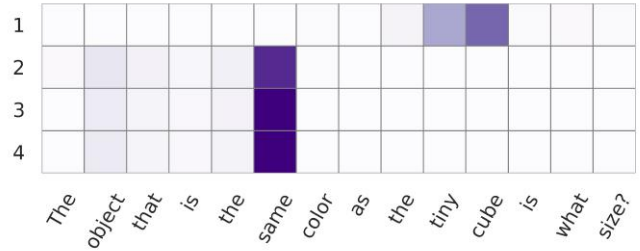
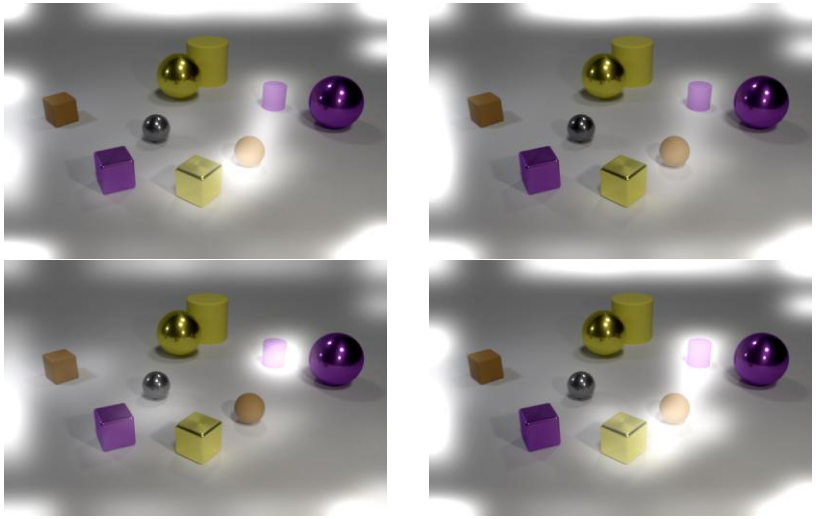


Figure: G:Large P:Large

CLEVR



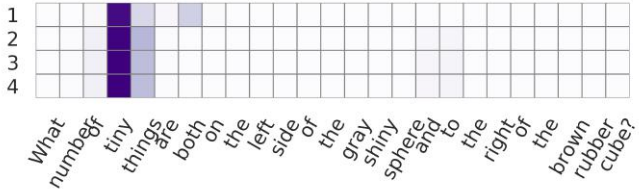
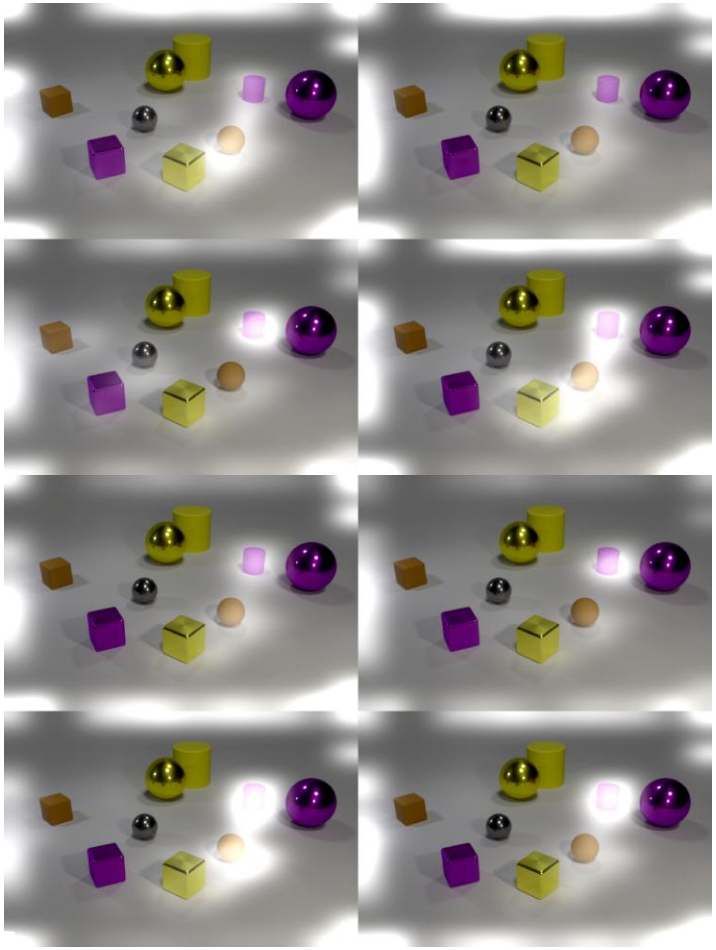


Figure: G:0 P:1

CLEVR



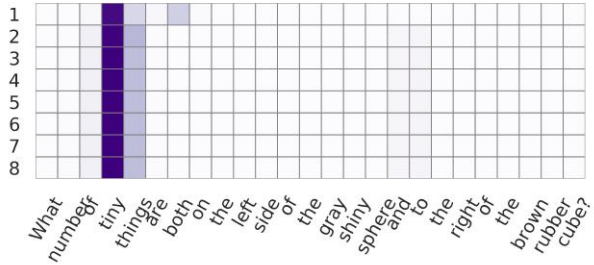
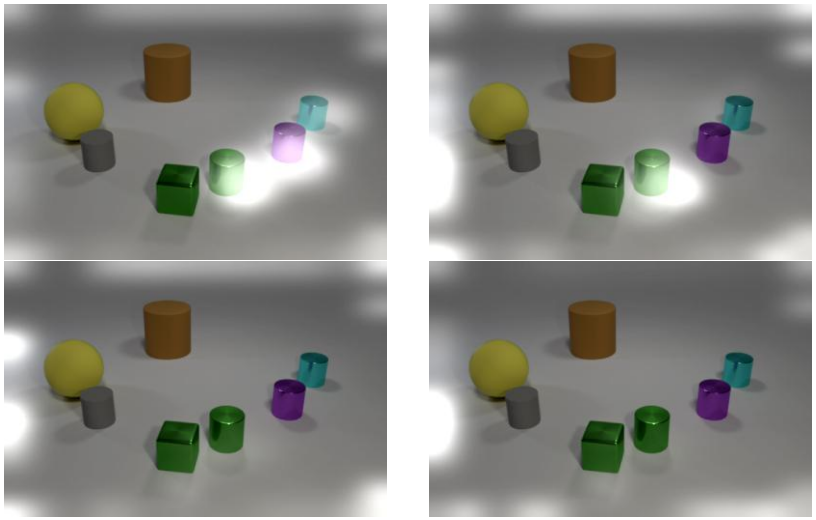


Figure: G:0 P:0

CLEVR



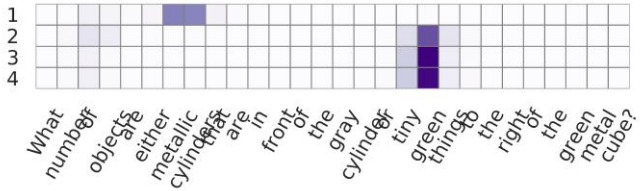
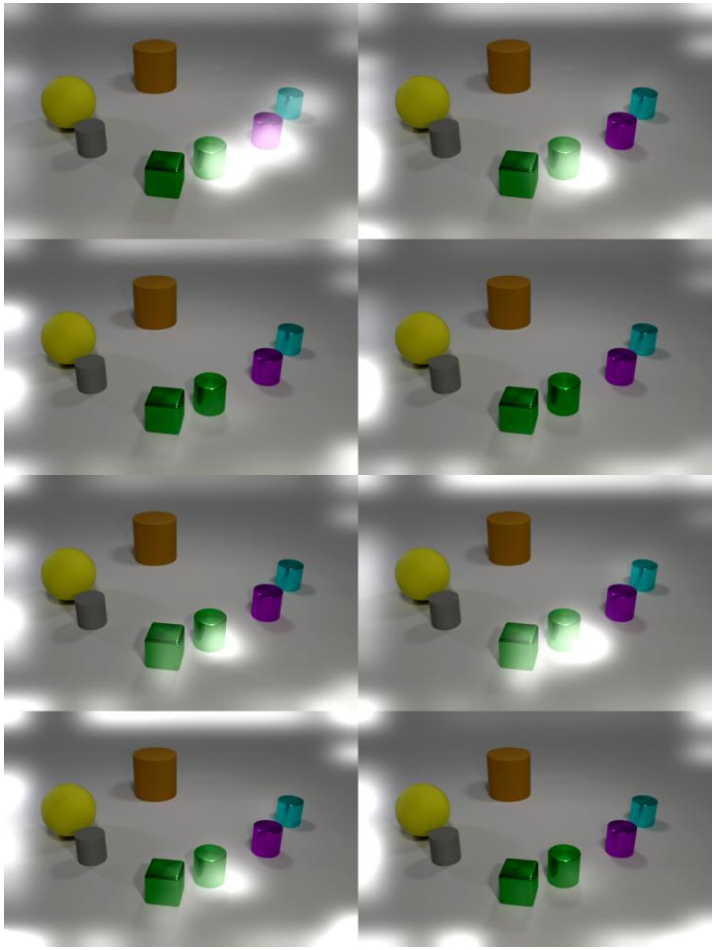


Figure: G:1 P:2

CLEVER



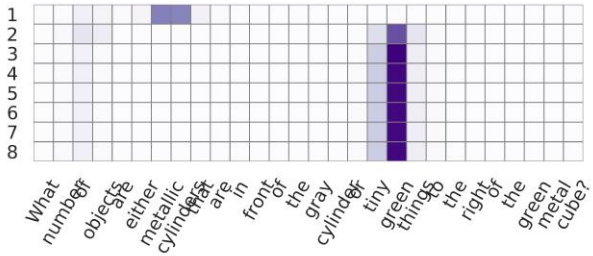


Figure: G:1 P:1

Structure of Babi

Babi

```
1 John travelled to the hallway.
2 Mary journeyed to the bathroom.
3 Where is John?        hallway 1
4 Daniel went back to the bathroom.
5 John moved to the bedroom.
6 Where is Mary?        bathroom        2
7 John went to the hallway.
8 Sandra journeyed to the kitchen.
9 Where is Sandra?      kitchen 8
10 Sandra travelled to the hallway.
11 John went to the garden.
12 Where is Sandra?     hallway 10
13 Sandra went back to the bathroom.
14 Sandra moved to the kitchen.
15 Where is Sandra?     kitchen 14
```

```
1 Jason is thirsty.
2 Where will jason go?  kitchen 1
3 Antoine is bored.
4 Where will antoine go?        garden  3
5 Jason went to the kitchen.
6 Why did jason go to the kitchen?      thirsty 1
7 Antoine journeyed to the garden.
8 Why did antoine go to the garden?     bored   3
9 Sumit is hungry.
10 Where will sumit go? kitchen 9
11 Antoine grabbed the football there.
12 Why did antoine get the football?    bored   3
13 Sumit went to the kitchen.
14 Why did sumit go to the kitchen?     hungry  9
15 Jason took the milk there.
16 Why did jason get the milk?  thirsty 1
17 Sumit took the apple there.
18 Why did sumit get the apple? hungry  9
19 Yann is thirsty.
20 Where will yann go?  kitchen 19
21 Yann travelled to the kitchen.
22 Why did yann go to the kitchen?      thirsty 19
```

- Was we have seen before, we will have several tasks:
  - Single/Two/Three supporting fact
  - Counting
  - Listing
  - Size reasoning
  - etc
- Story size varies, and we can have stories with several questions

Difference between CLEVR and Babi

- We have 20 different tasks ranging from simple questions to more intricate ones like path-finding or time reasoning

Babi

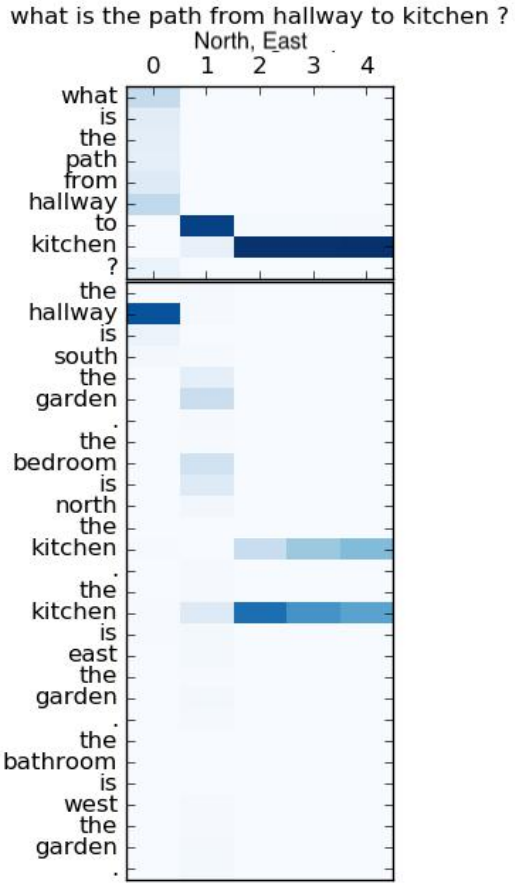
Implementation used

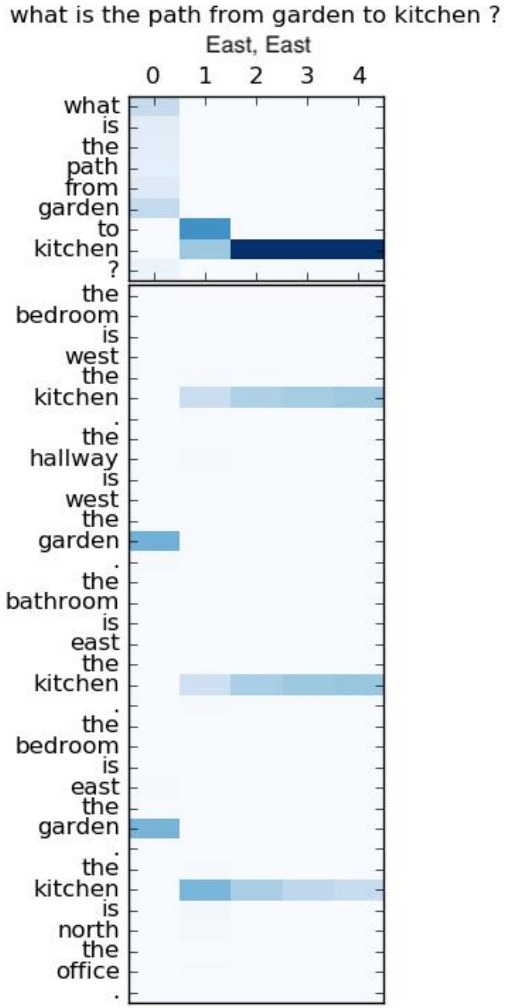
- Implementation of Selvakumar et al, 2018.
- Follows the implementation of the original paper

How training was done

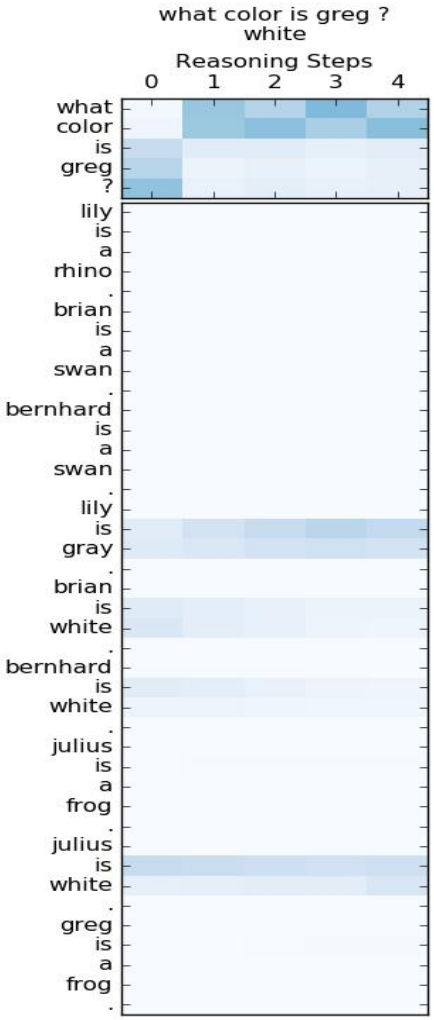
- We had actually to regenerate the pathfinder task due to low accuracy (from 10k examples to 50k)
- trained over 10 epochs and 5 reasoning steps
- we obtained similar scores to the article (95%)

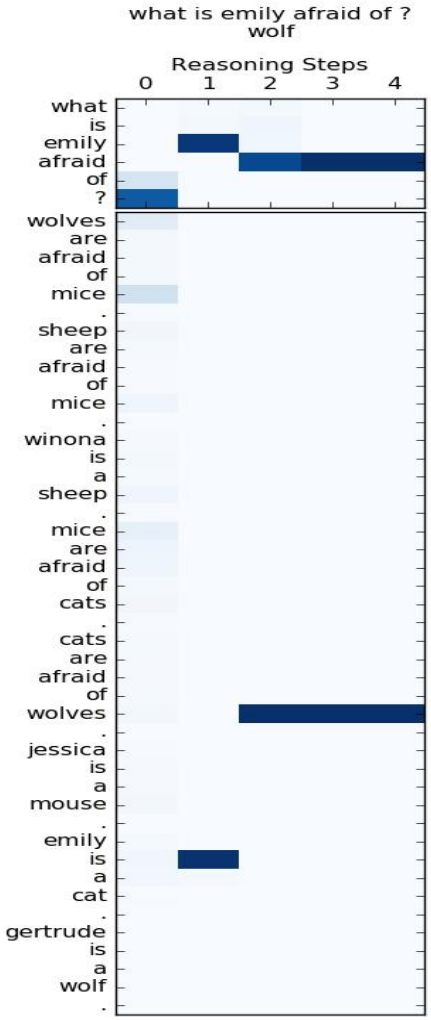
Babi Results





Babi Results





Faculté
des Sciences
Aix*Marseille Université

The End!!