Attention & Memory

Nov 24th, 2016

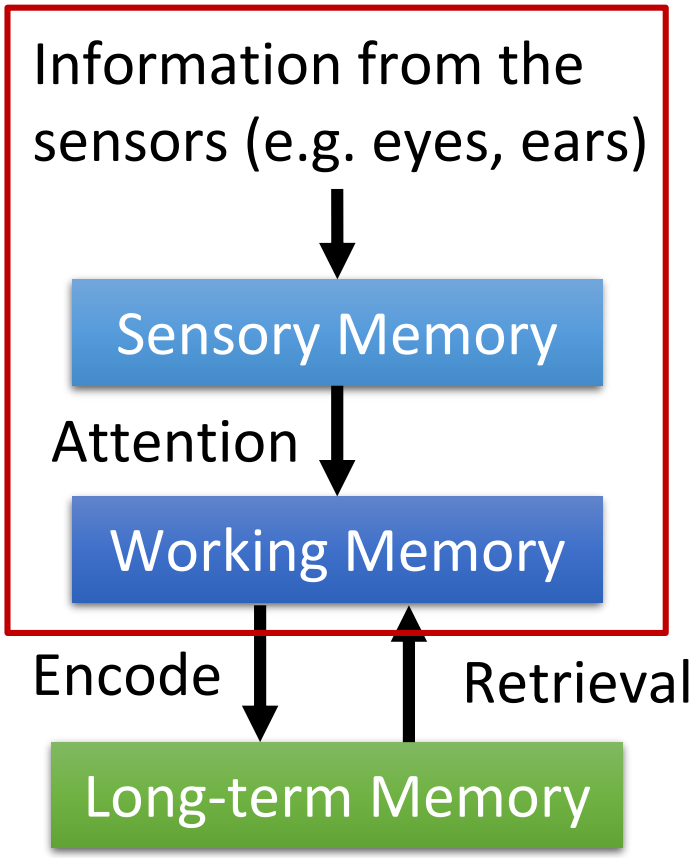
Applied Deep Learning

YUN-NUNG (VIVIAN) CHEN WWW.CSIE.NTU.EDU.TW/~YVCHEN/F105-ADL

National Taiwan University

Slide credit from Hung-Yi Lee

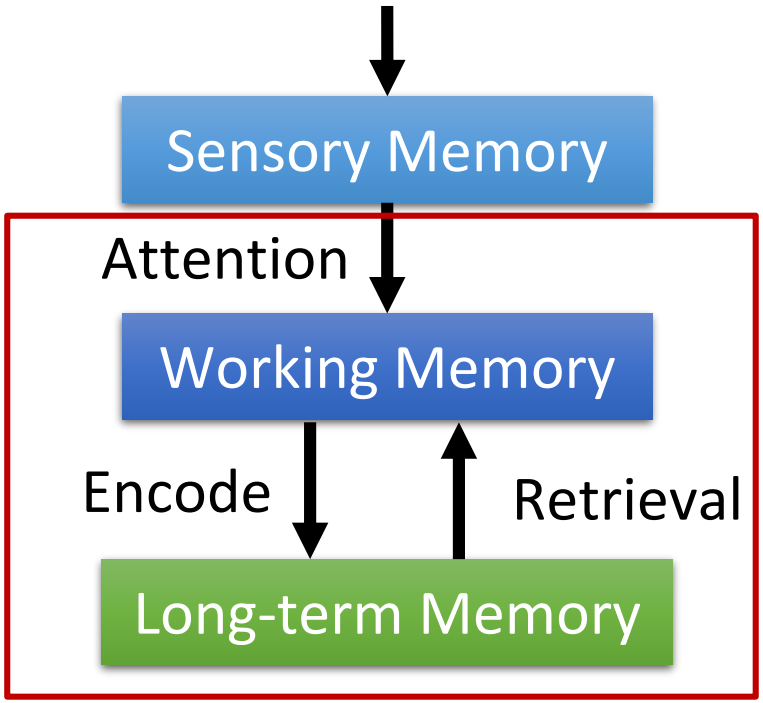
Attention and Memory



When the input is a very long sequence or an image

→ Pay attention on partial of the input object each time

Attention and Memory



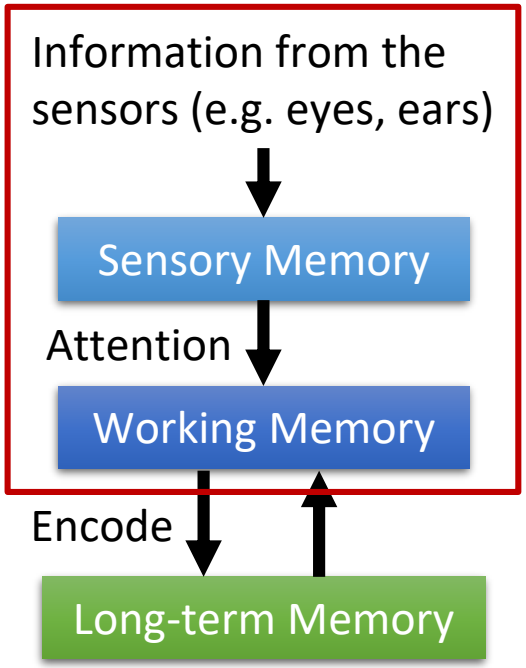
When the input is a very long sequence or an image

→ Pay attention on partial of the input object each time

In RNN/LSTM, larger memory implies more parameters

→ Increasing memory size will not increasing parameters

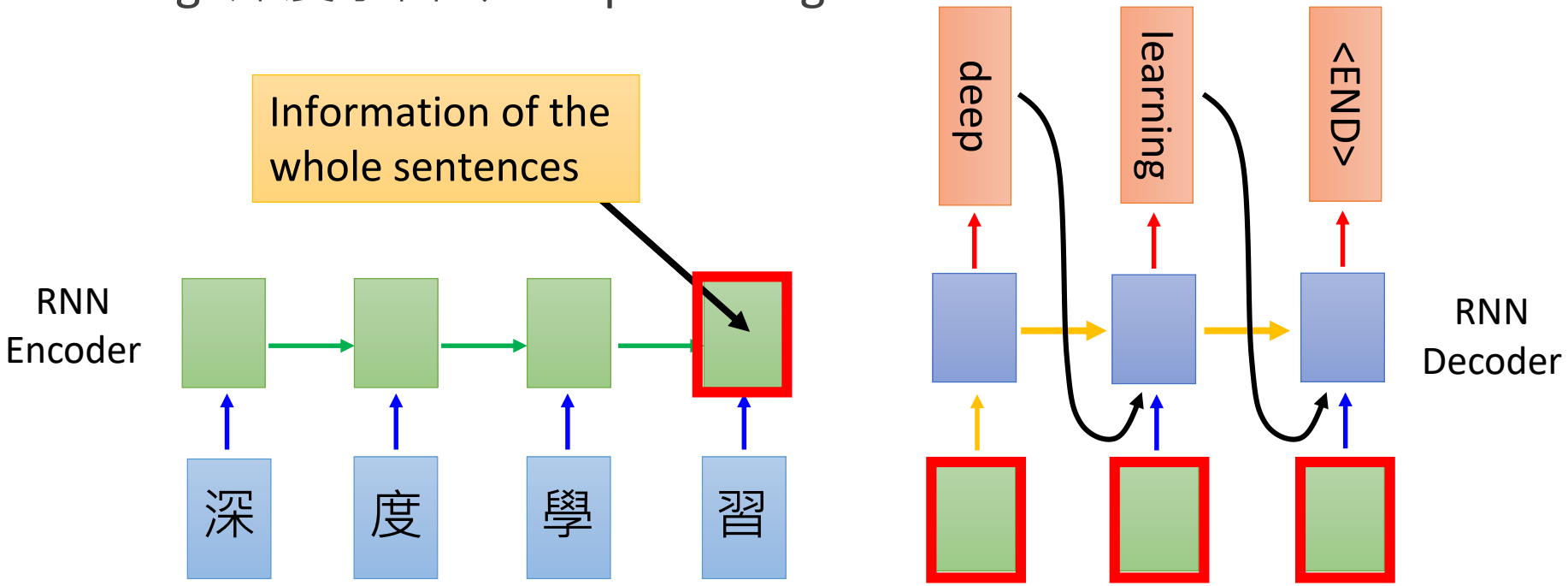
Attention on Sensory Info



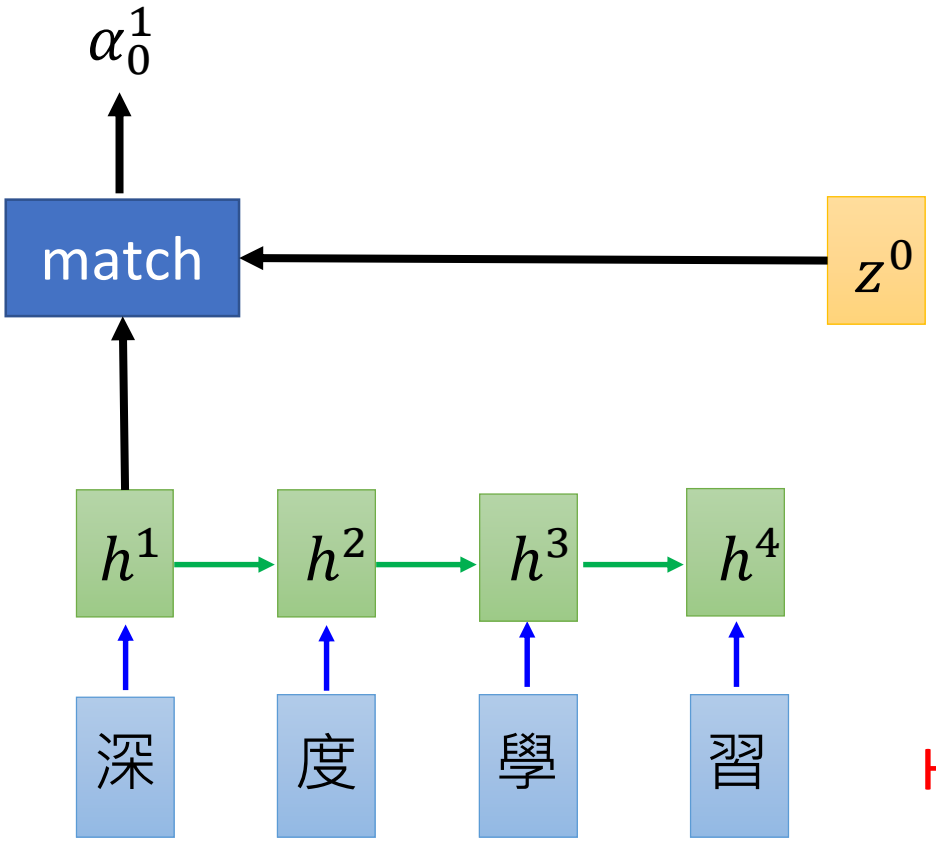
Machine Translation

Sequence-to-sequence learning: both input and output are both sequences ***with different lengths***.

E.g. 深度學習 → deep learning



Machine Translation with Attention

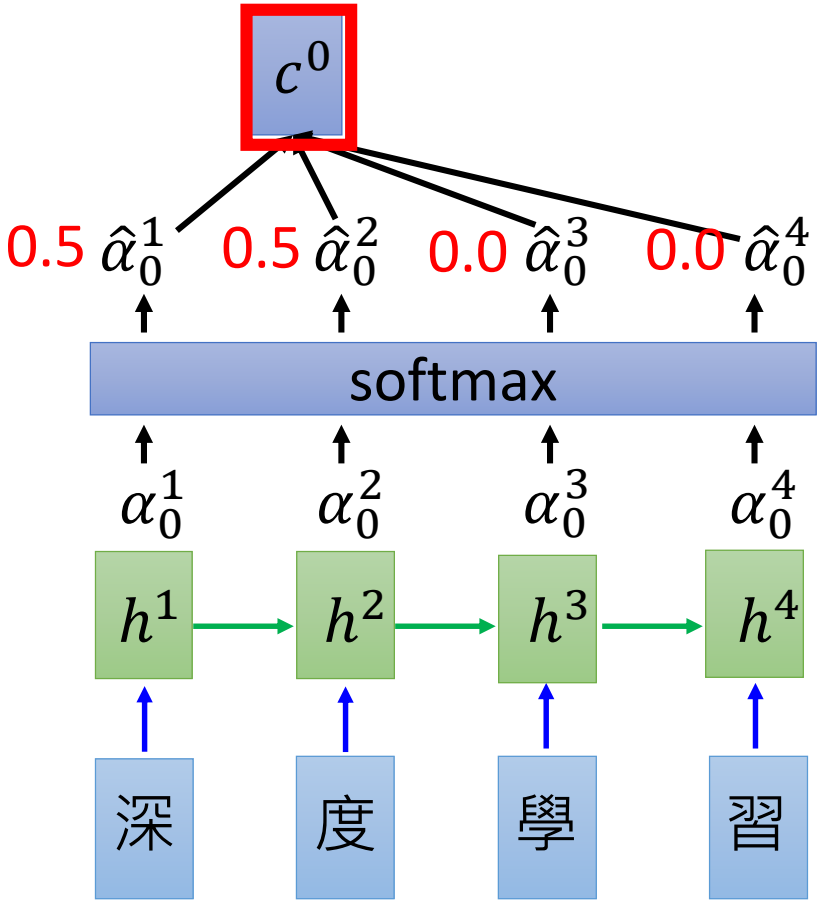


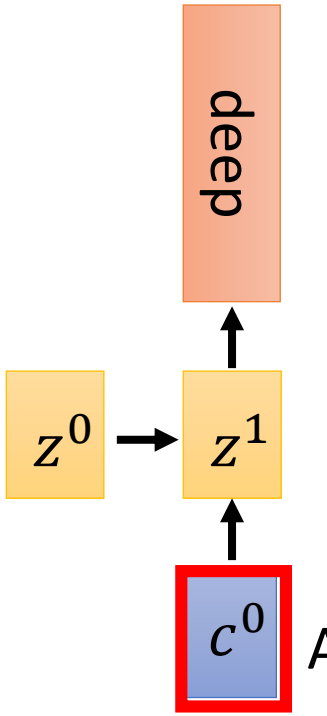
What is match ?

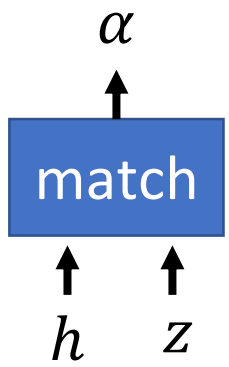
- Cosine similarity of z and h
- Small NN whose input is z and h, output a scalar
- $\alpha = h^T W z$

How to learn the parameters?

Machine Translation with Attention





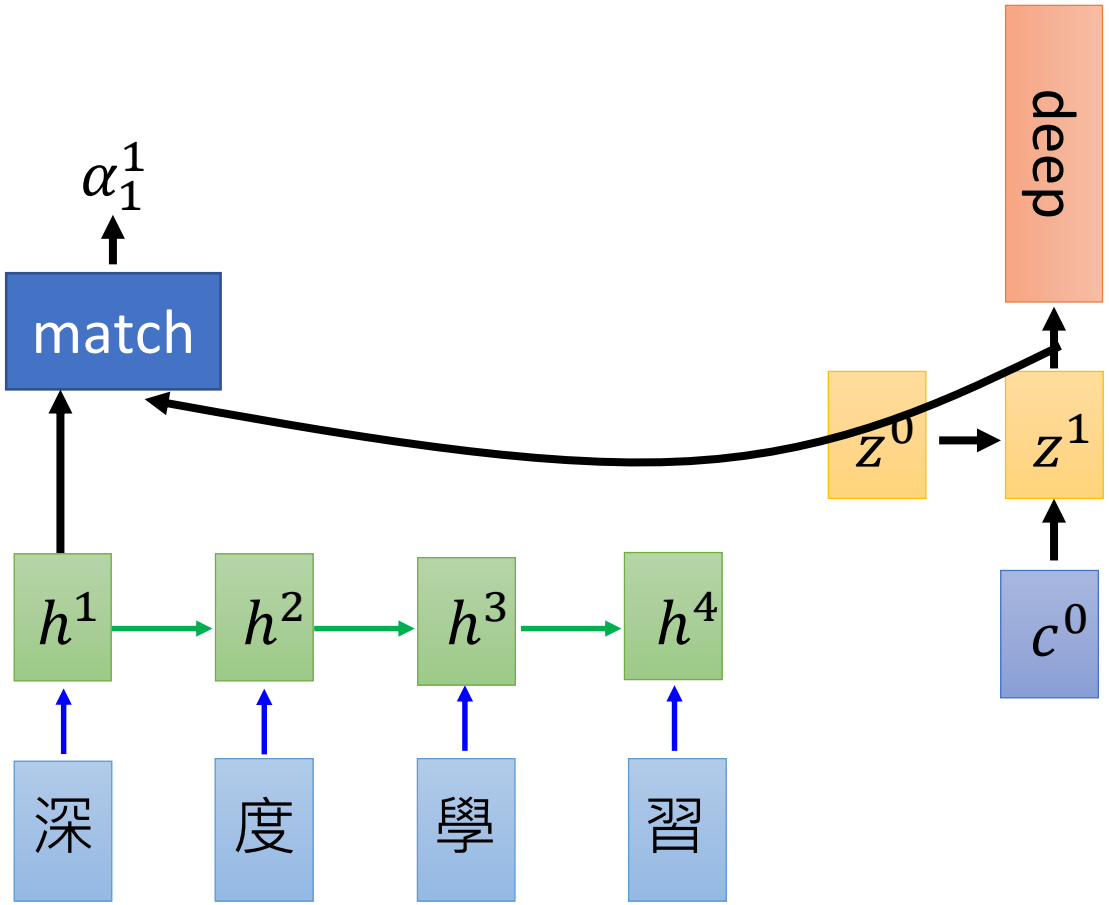


How to learn the parameters?

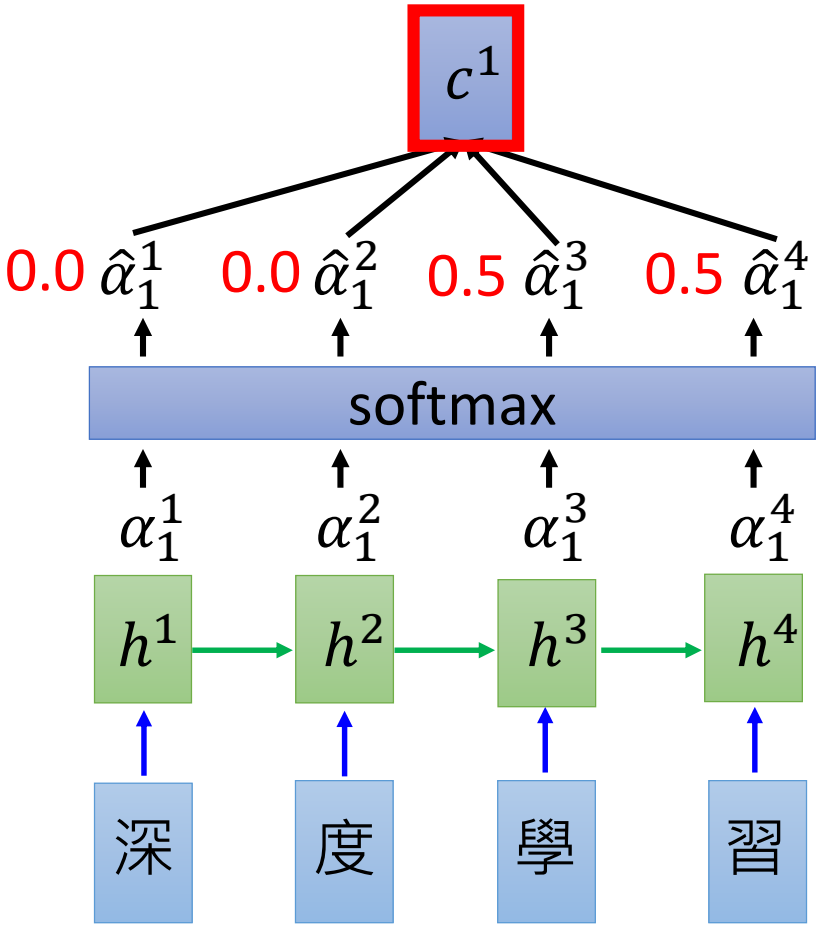
As RNN input

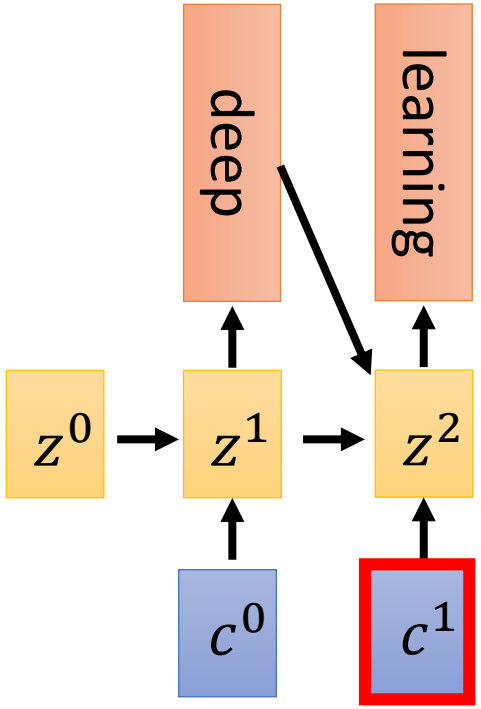
$$c^0 = \sum \hat{\alpha}_0^i h^i = 0.5h^1 + 0.5h^2$$

Machine Translation with Attention



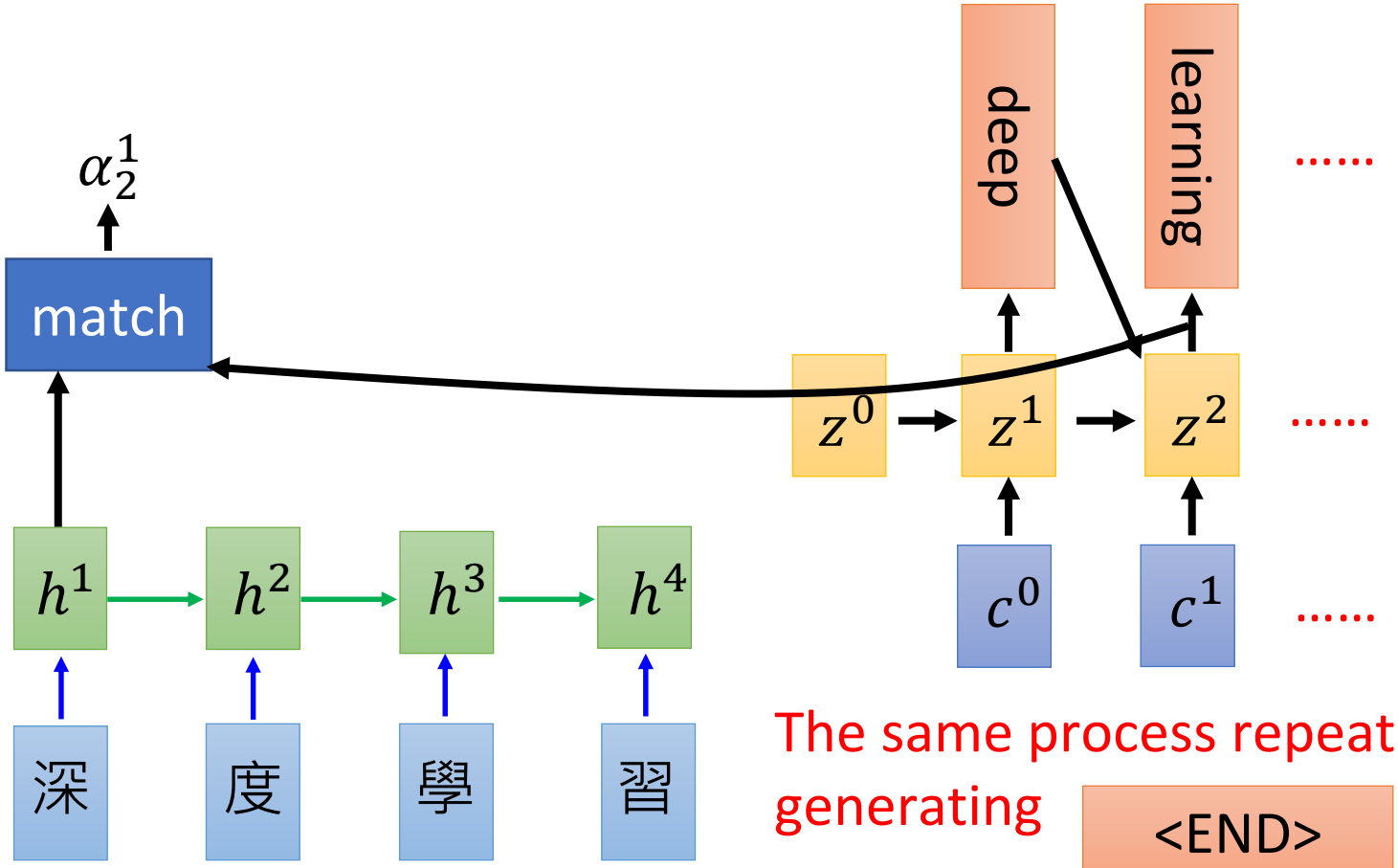
Machine Translation with Attention





$$c^1 = \sum \hat{\alpha}_1^i h^i = 0.5h^3 + 0.5h^4$$

Machine Translation with Attention



Speech Recognition with Attention

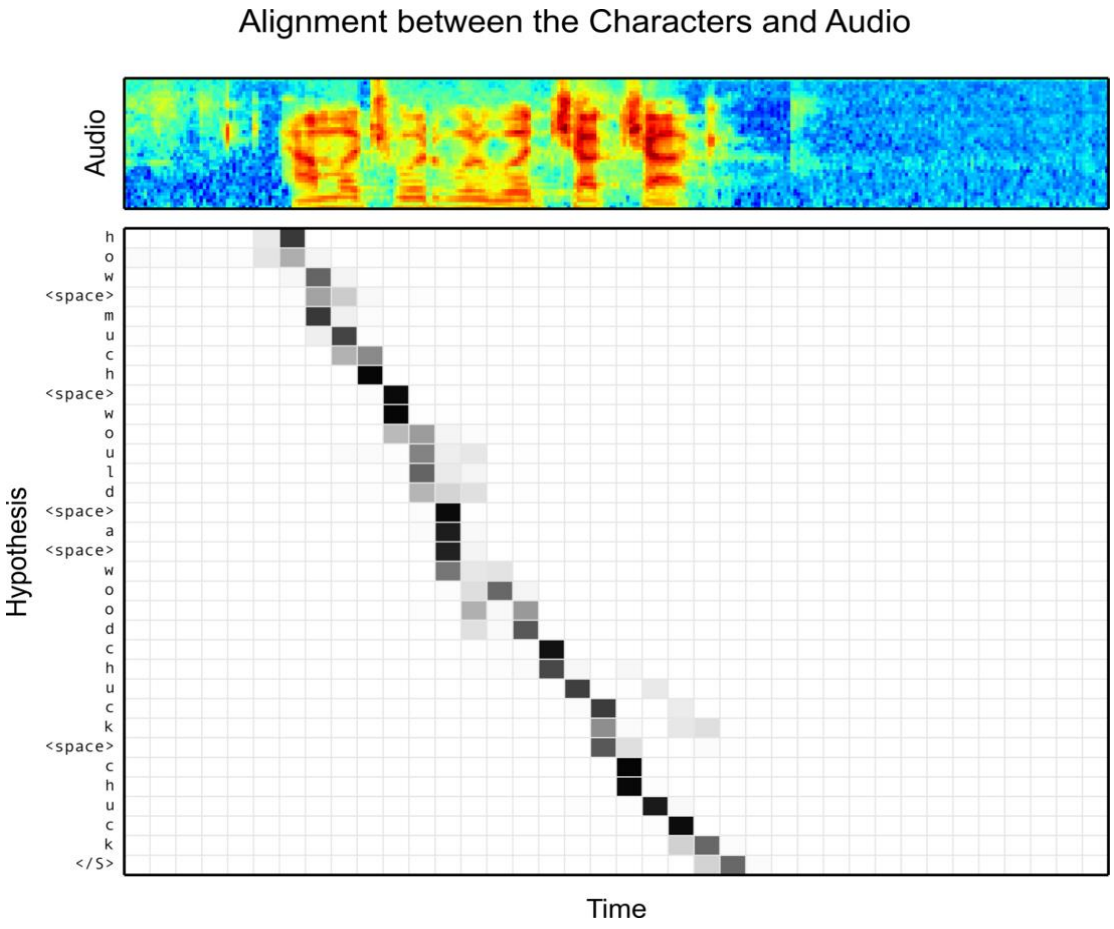
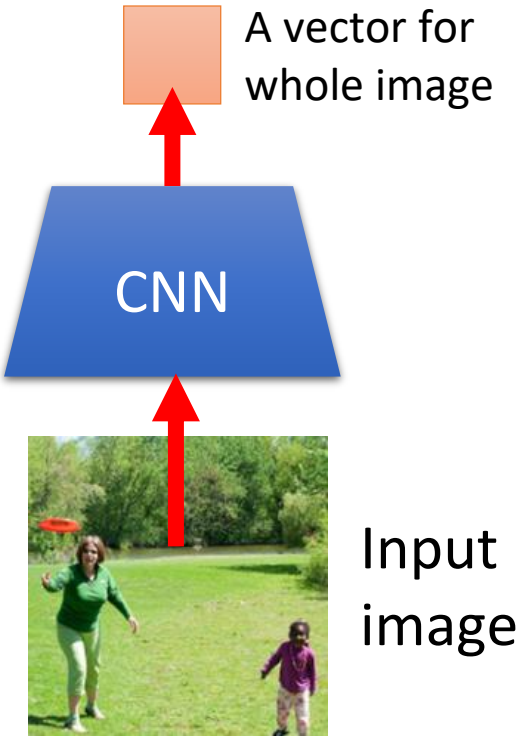


Image Captioning

Input: image

Output: word sequence



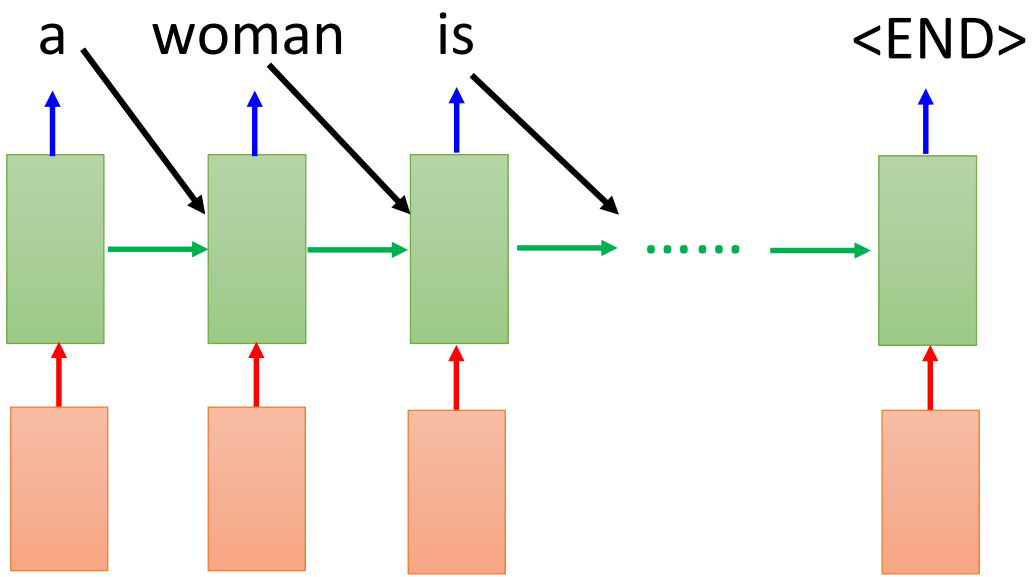


Image Captioning with Attention

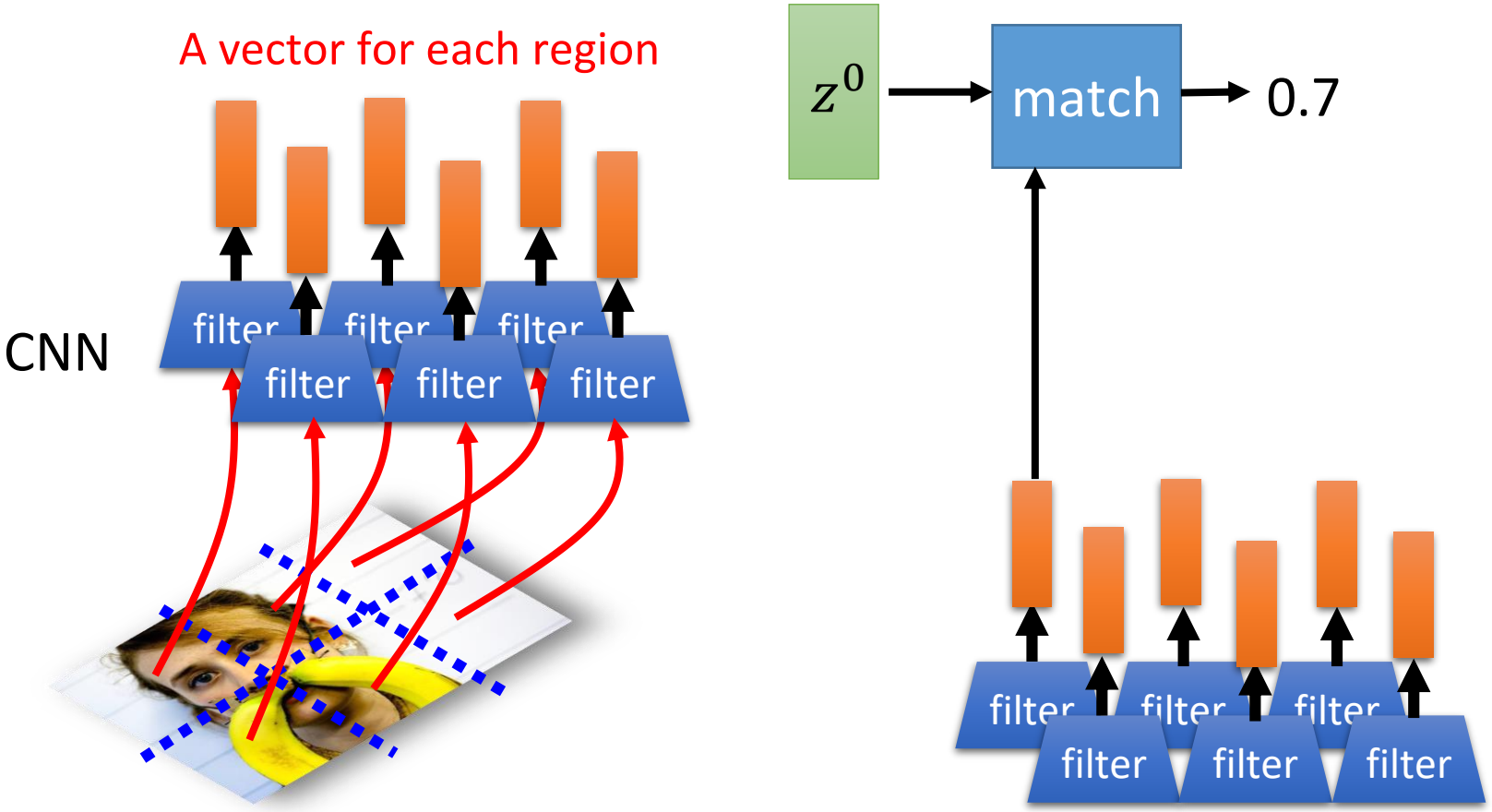


Image Captioning with Attention

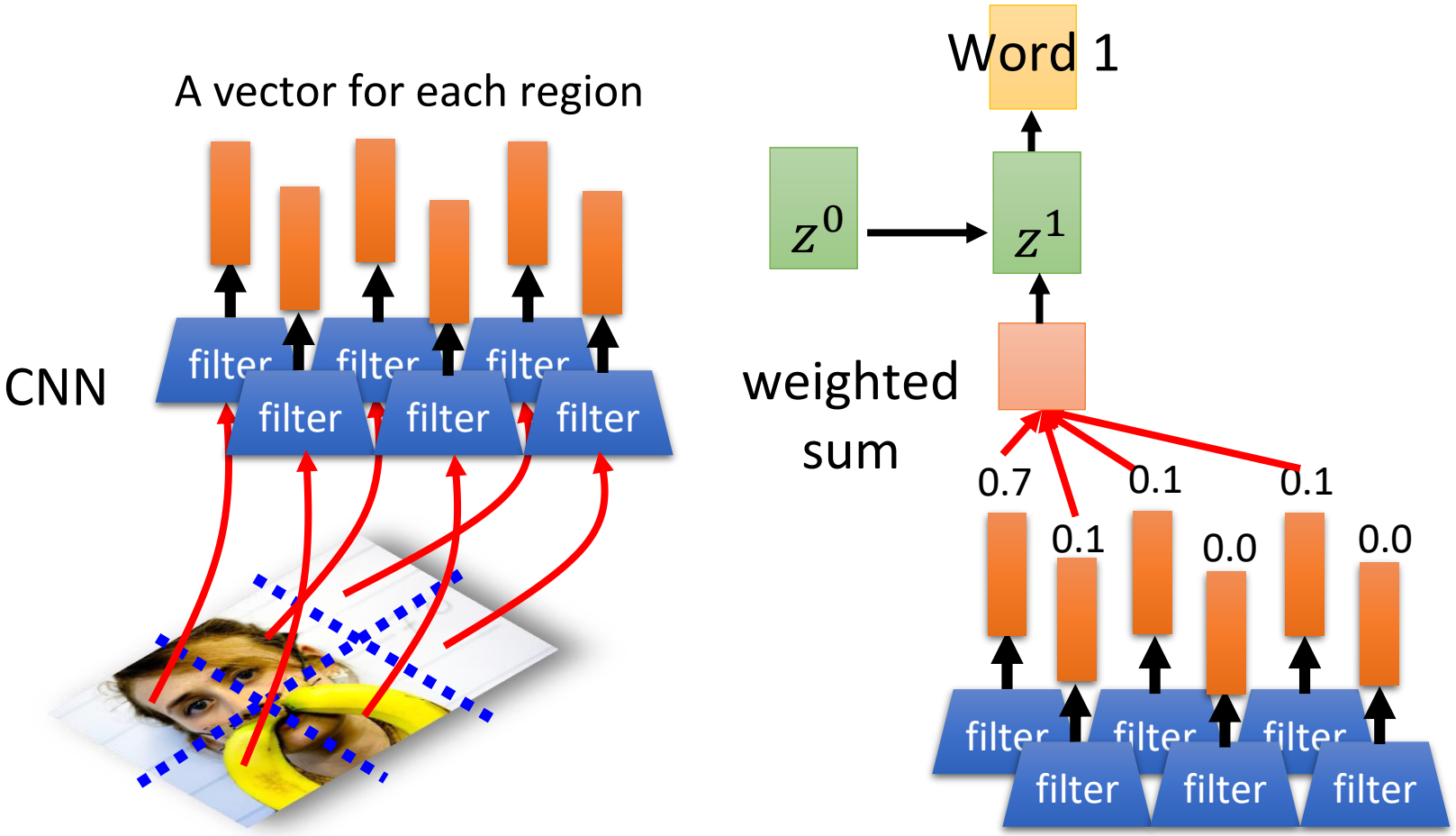


Image Captioning with Attention

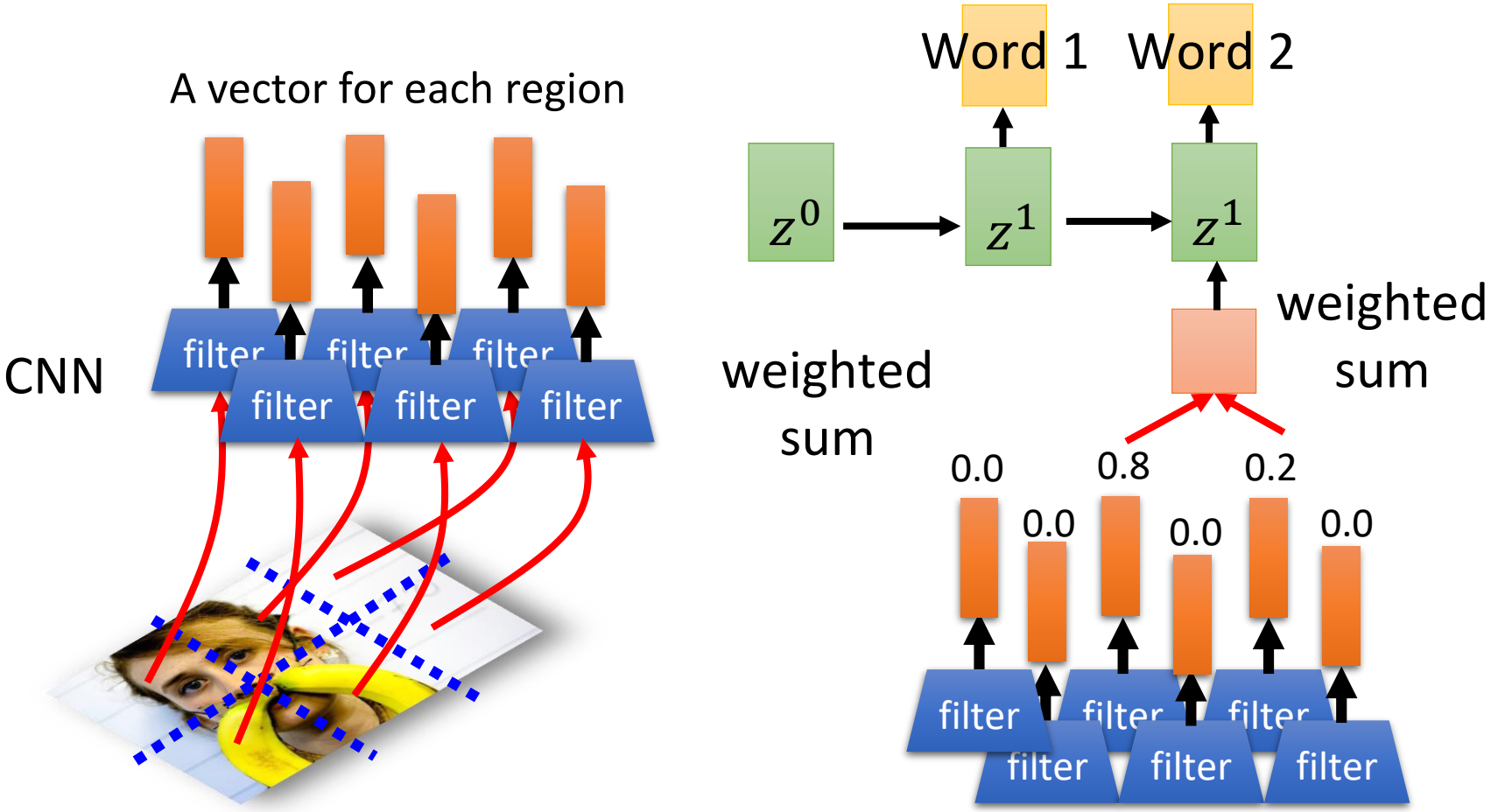


Image Captioning

Good examples

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.

Image Captioning

Bad examples

A large white bird standing in a forest.

A woman holding a clock in her hand.

A man wearing a hat and a hat on a skateboard.

A person is standing on a beach with a surfboard.

A woman is sitting at a table with a large pizza.

A man is talking on his cell phone while another man watches.

Video Captioning

Ref: A man and a woman ride a motorcycle

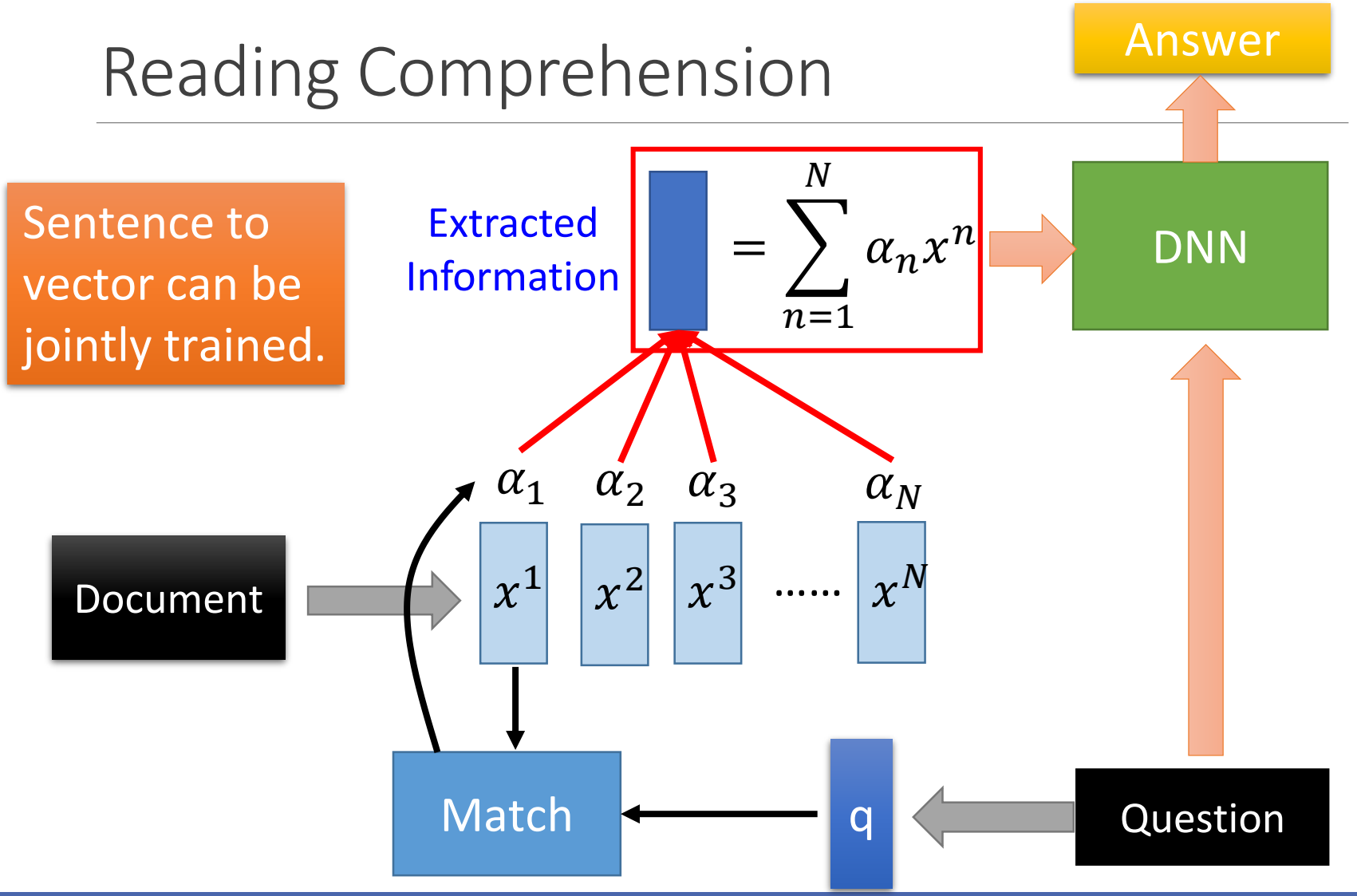
A **man** and a **woman** are **talking** on the **road**

Video Captioning

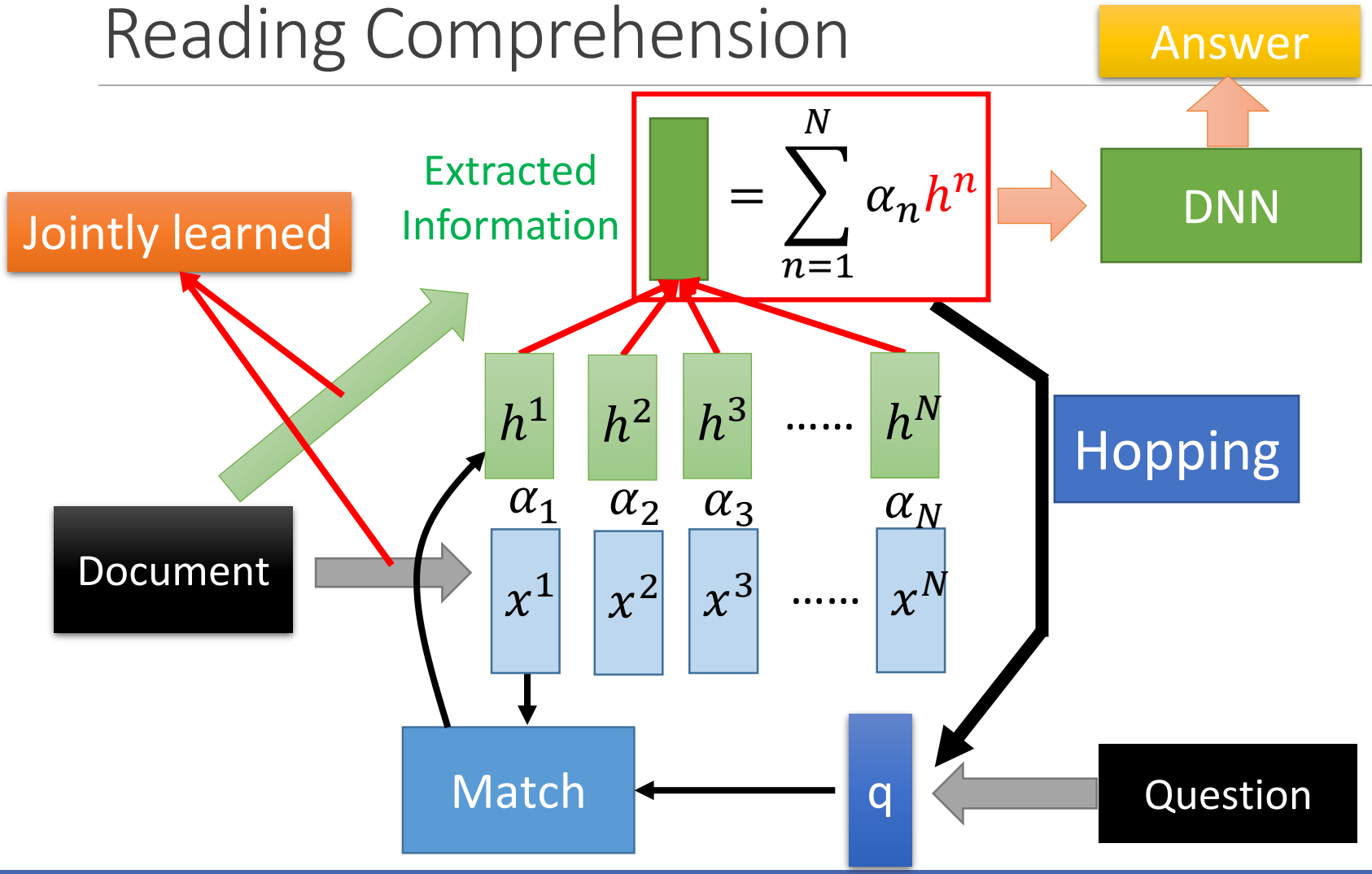
Ref: A woman is frying food

Someone is **frying** a **fish** in a **pot**

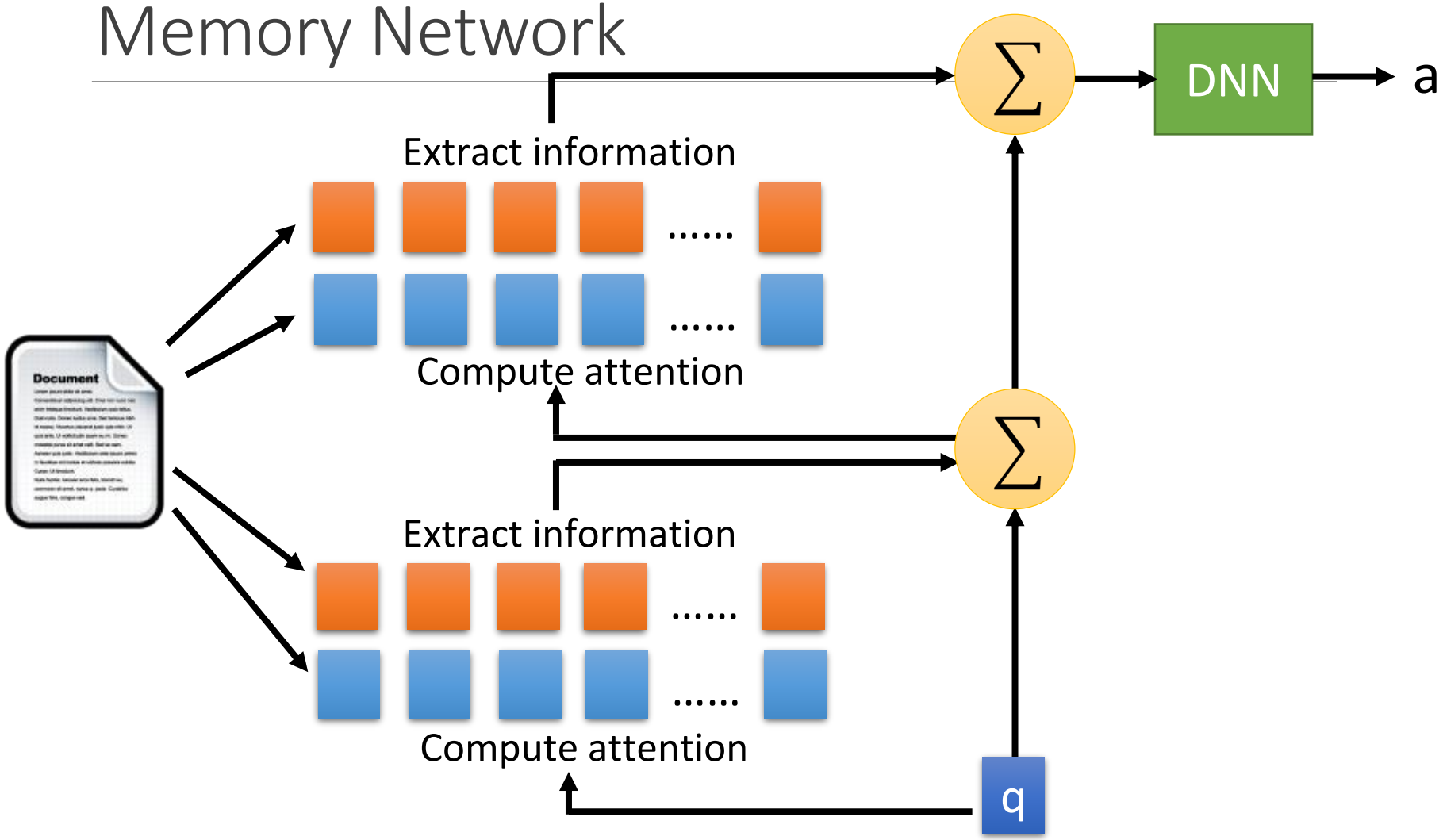
Reading Comprehension



Reading Comprehension



Memory Network

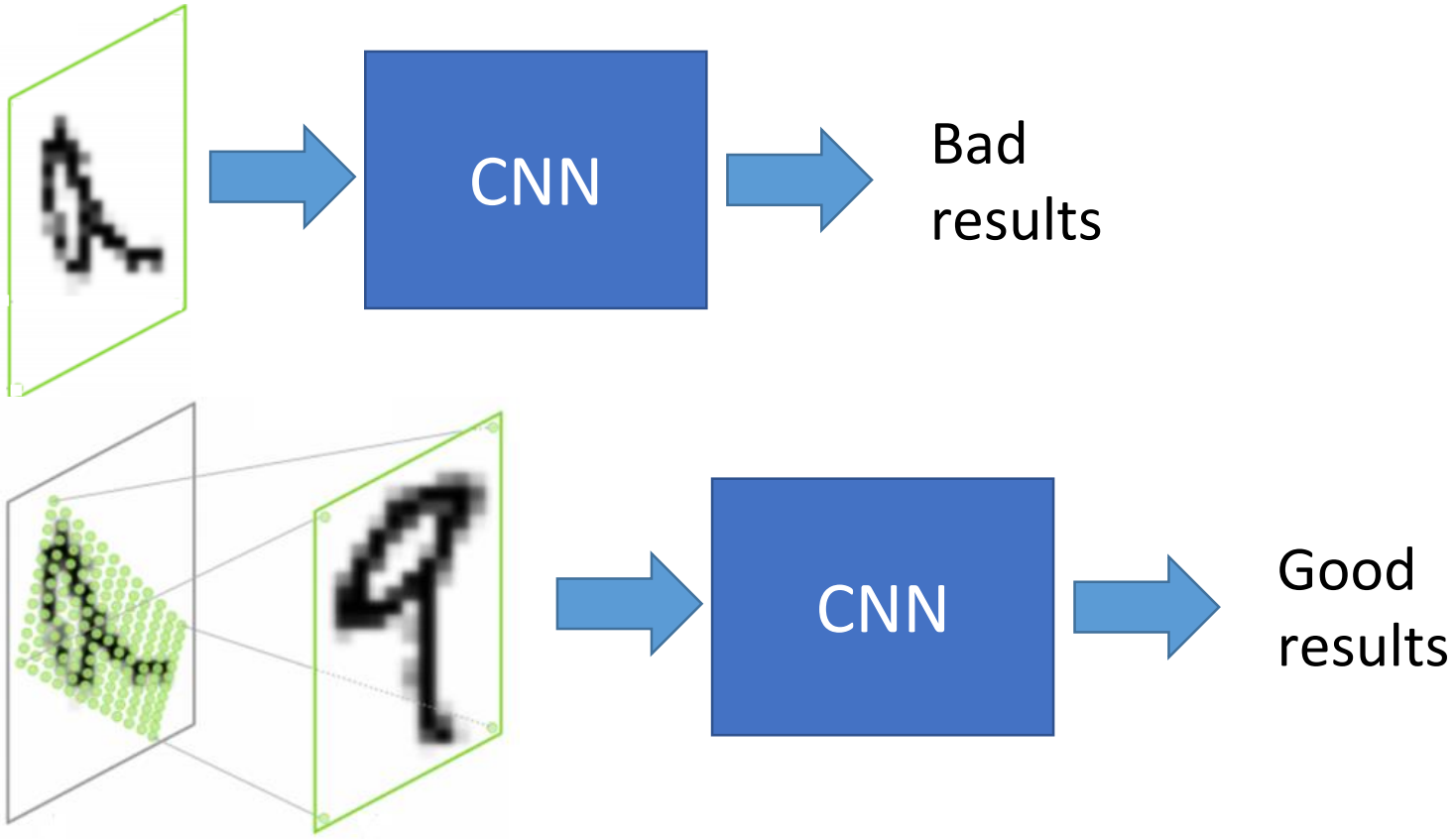


Memory Network

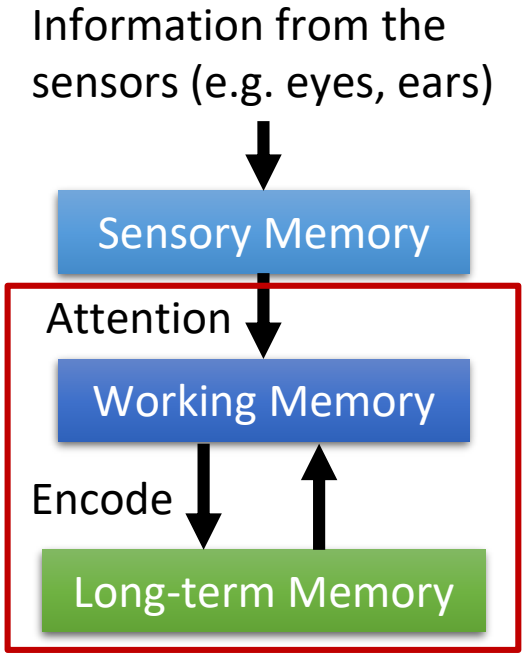
Muti-hop performance analysis

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow	**Prediction: yellow**			

Special Attention: Spatial Transformers



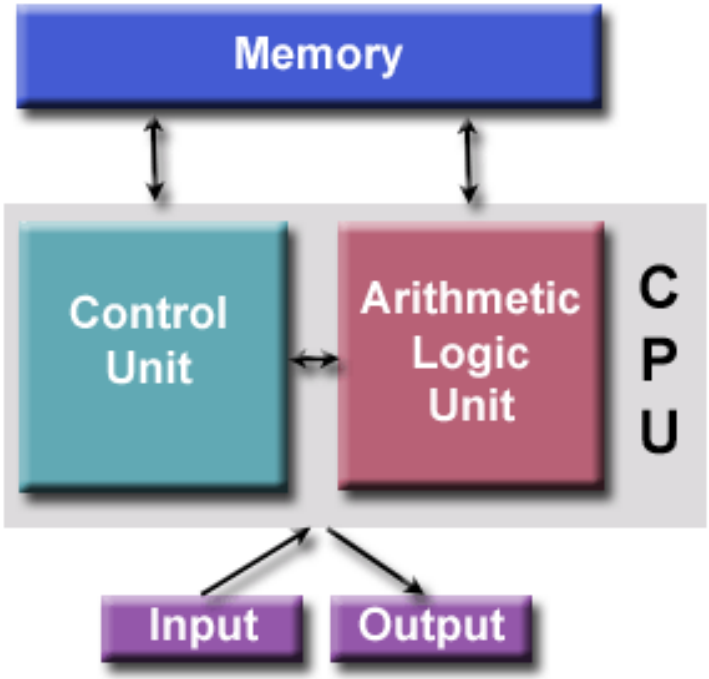
Attention on Memory



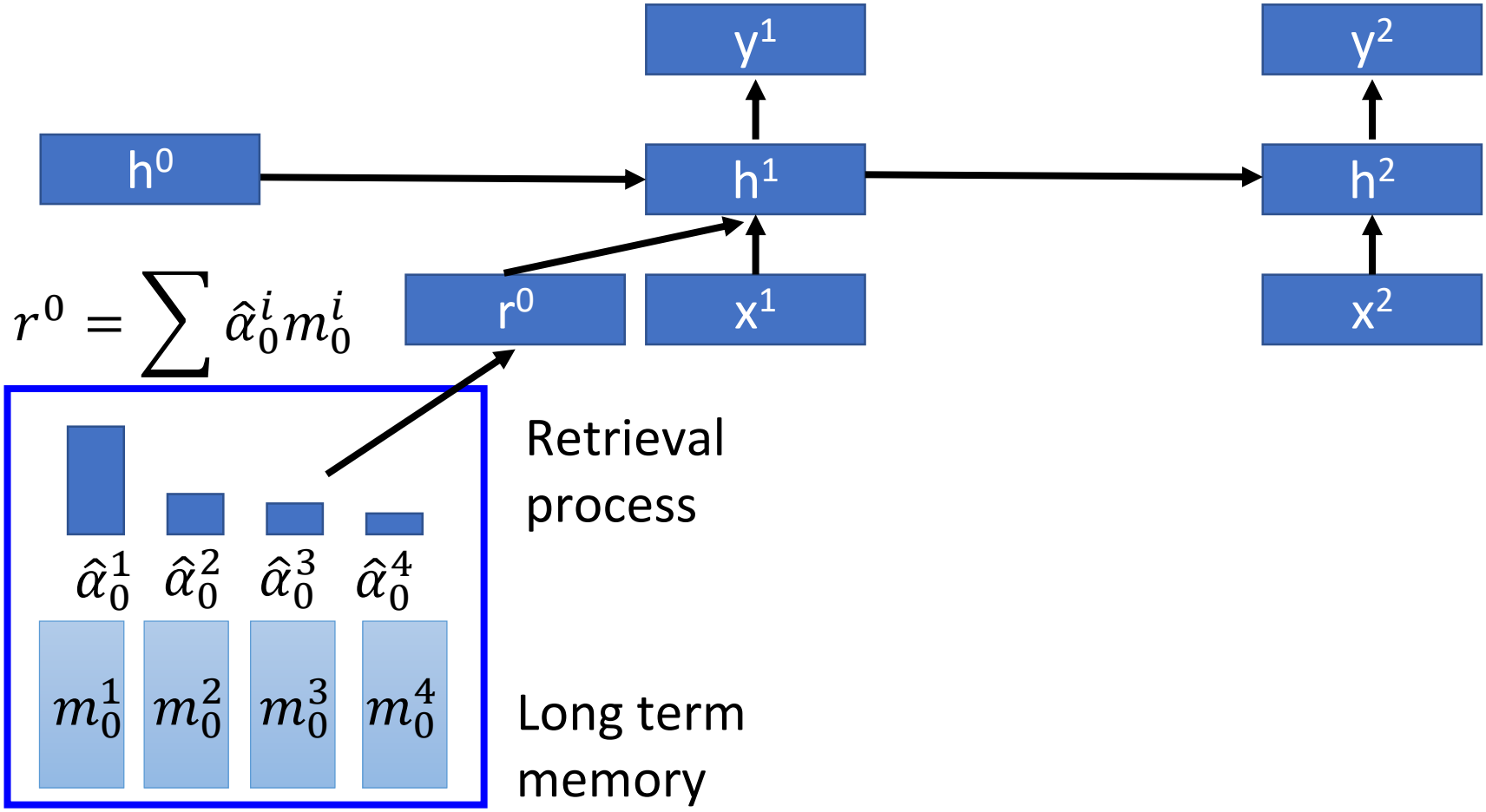
Neural Turing Machine

Von Neumann architecture

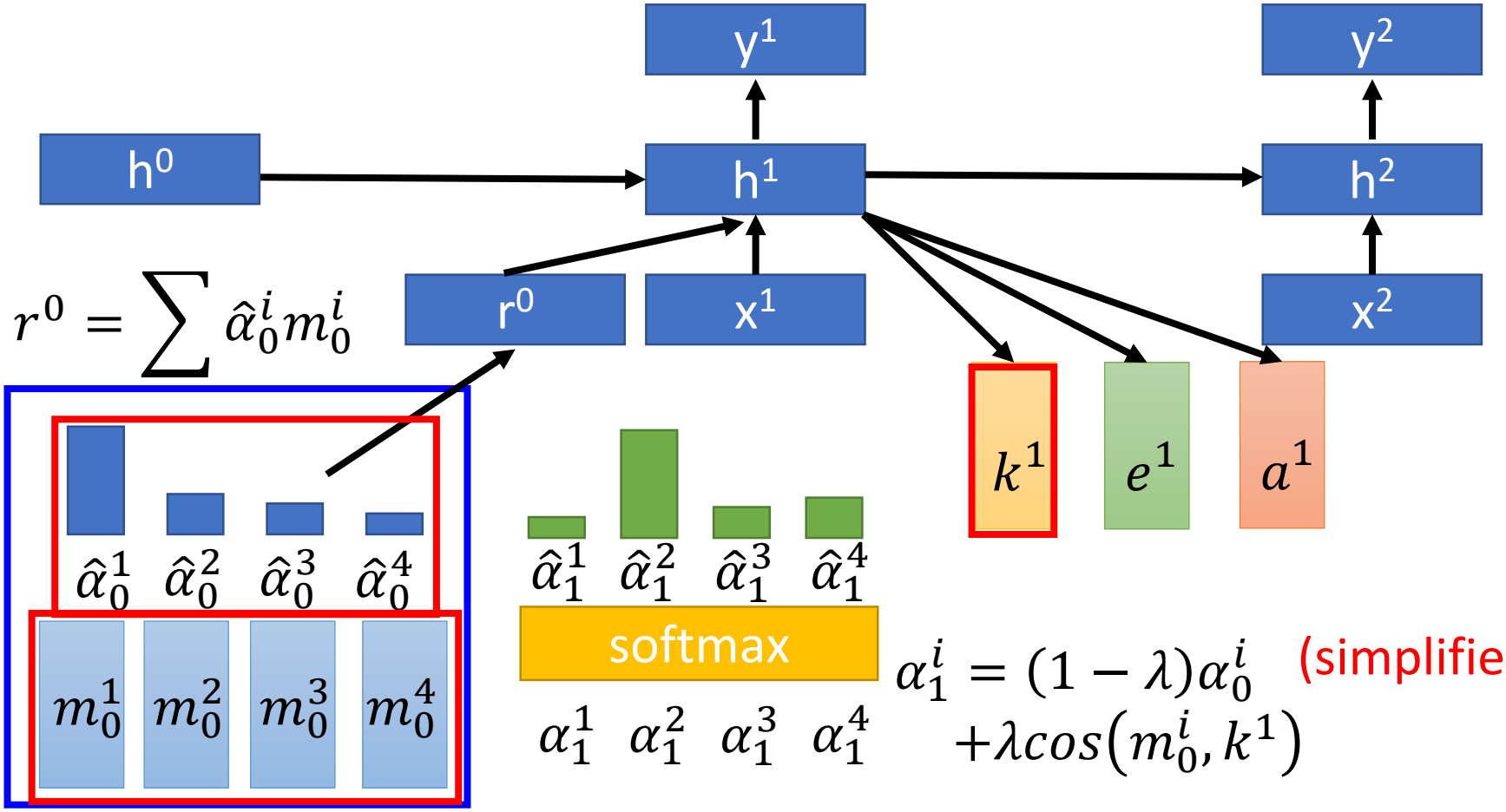
Neural Turing Machine is an advanced RNN/LSTM.



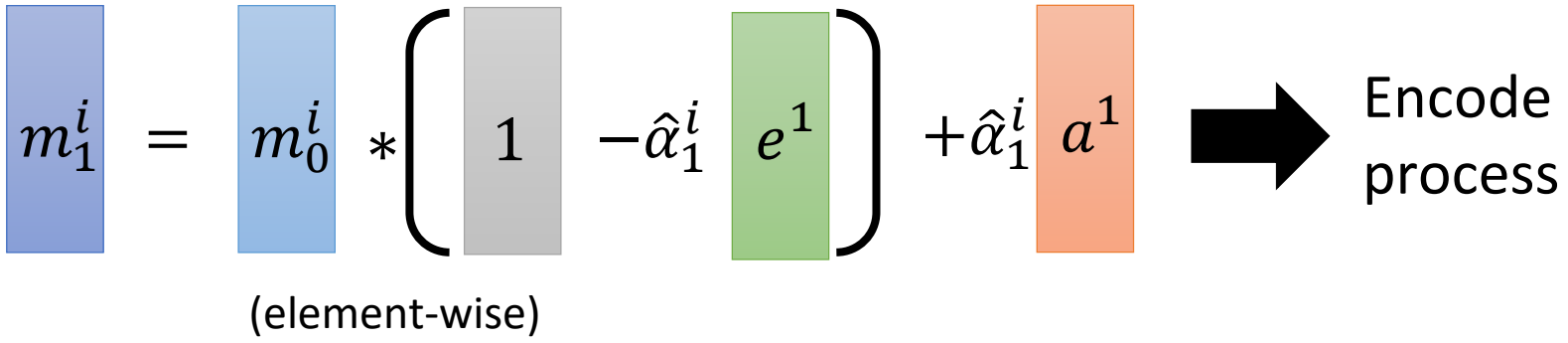
Neural Turing Machine

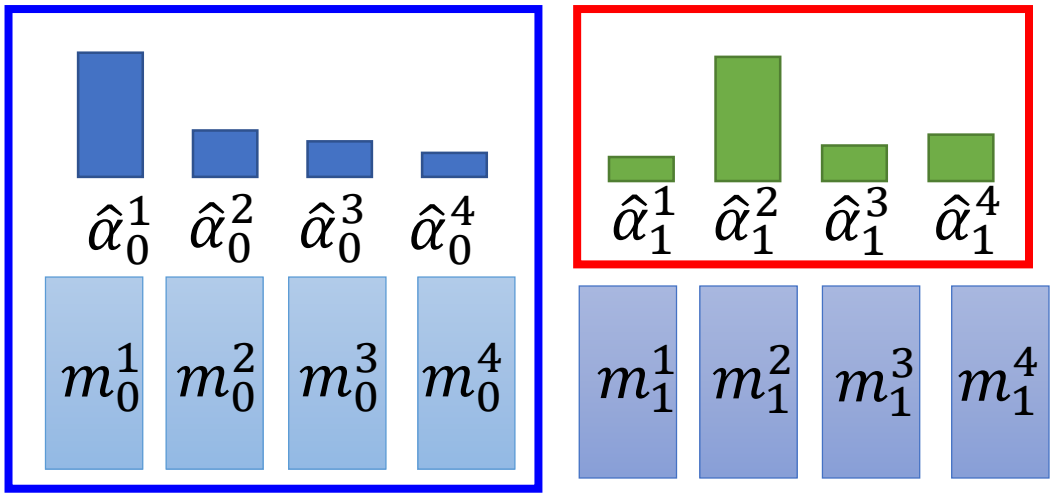


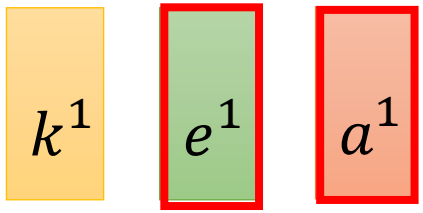
Neural Turing Machine



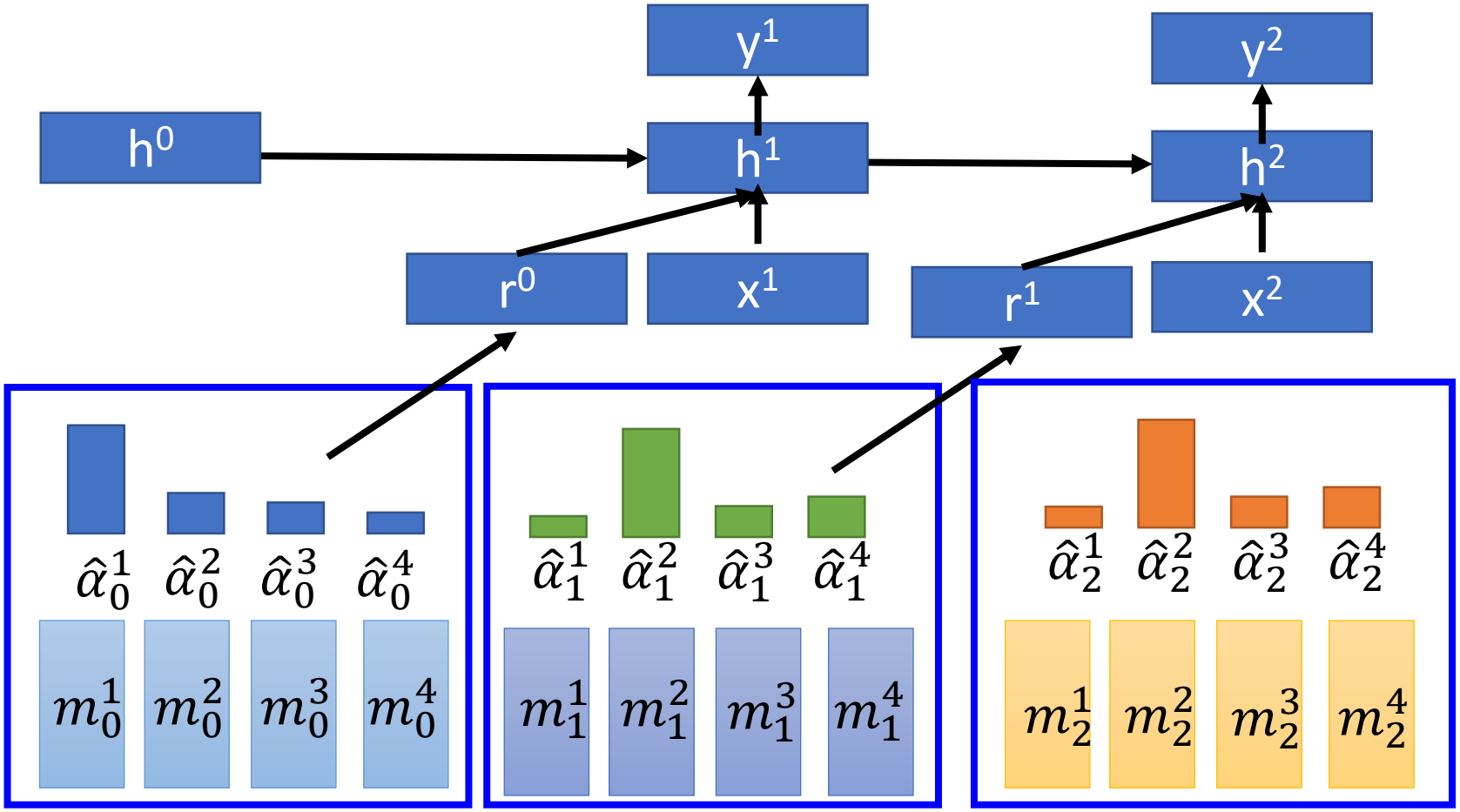
Neural Turing Machine







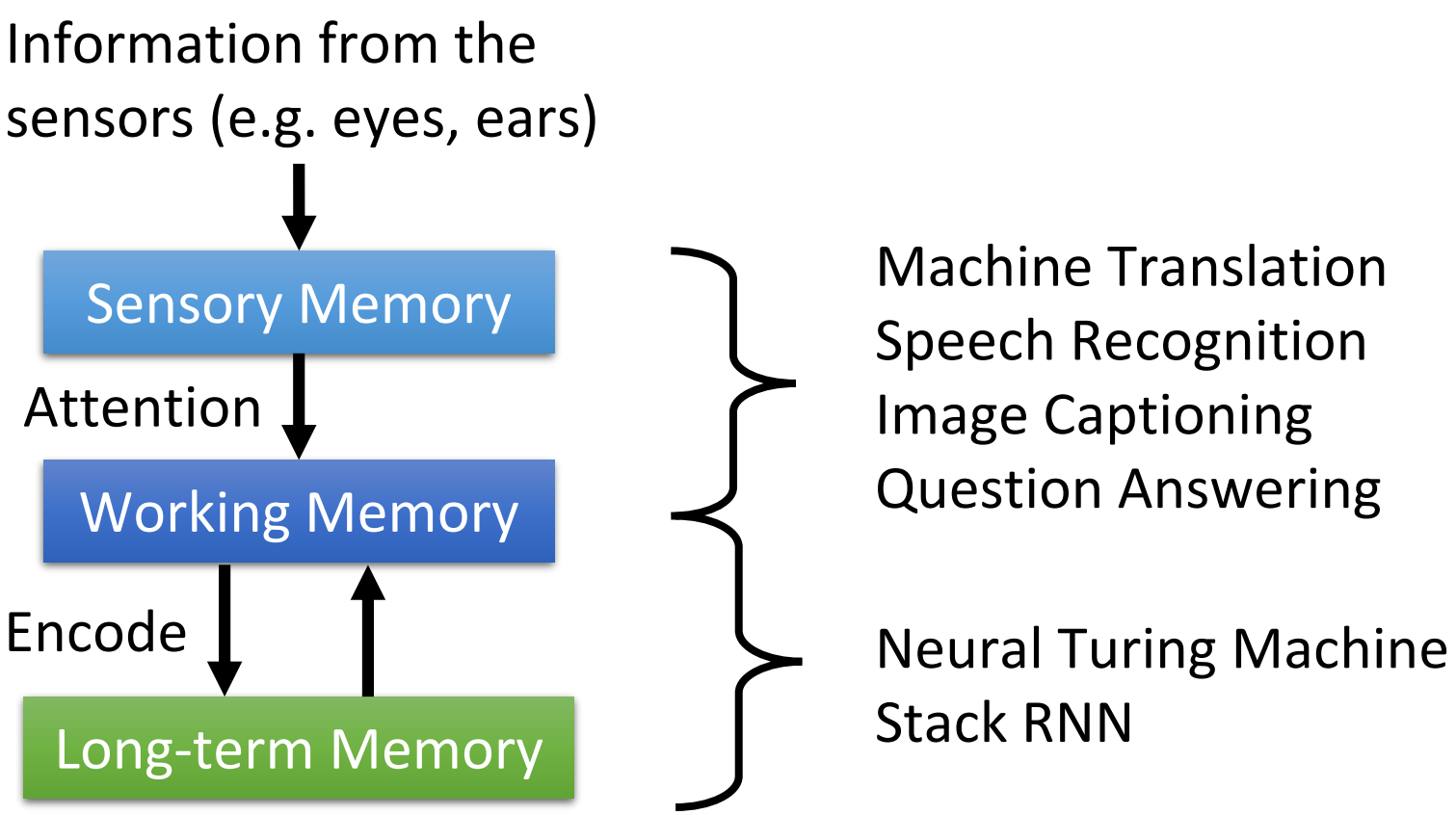
Neural Turing Machine



Stack RNN


stack
0.7
0.2
0.1
Push, Pop, Nothing
Information
to store
y^t
h^t
x^t
Push
Pop
Nothing
-1
-1
-1
X0.7
+
X0.2
+
X0.1

Concluding Remarks



Reference

End-To-End Memory Networks. S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus. arXiv Pre-Print, 2015.

Neural Turing Machines. Alex Graves, Greg Wayne, Ivo Danihelka. arXiv Pre-Print, 2014

Ask Me Anything: Dynamic Memory Networks for Natural Language Processing. Kumar et al. arXiv Pre-Print, 2015

Neural Machine Translation by Jointly Learning to Align and Translate. D. Bahdanau, K. Cho, Y. Bengio; International Conference on Representation Learning 2015.

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. Kelvin Xu et. al.. arXiv Pre-Print, 2015.

Attention-Based Models for Speech Recognition. Jan Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, Yoshua Bengio. arXiv Pre-Print, 2015.

A Neural Attention Model for Abstractive Sentence Summarization. A. M. Rush, S. Chopra and J. Weston. EMNLP 2015.