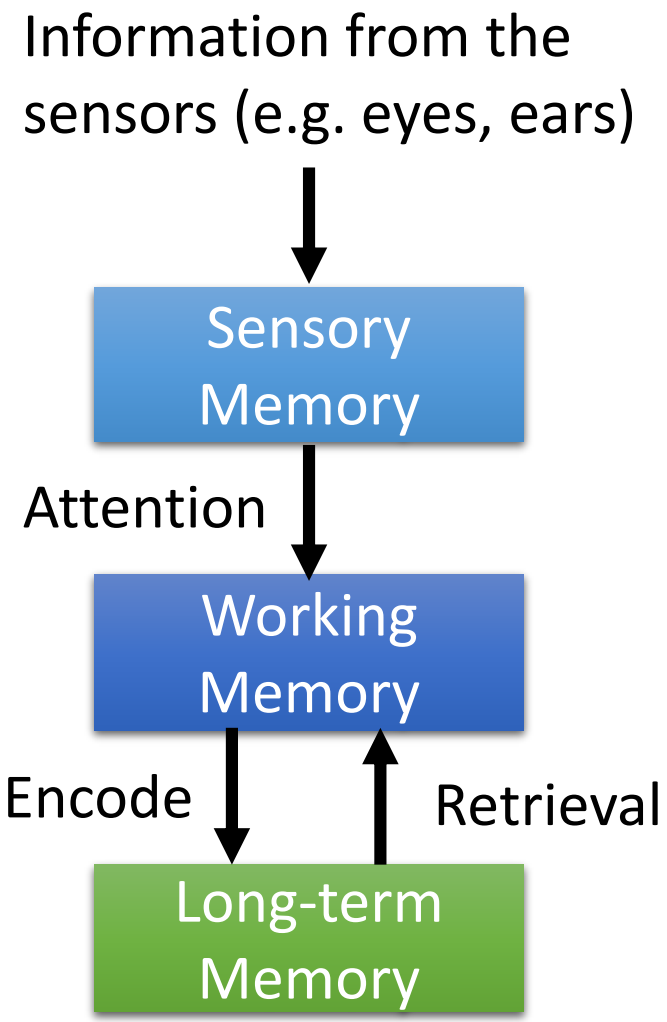
Attention-based Model

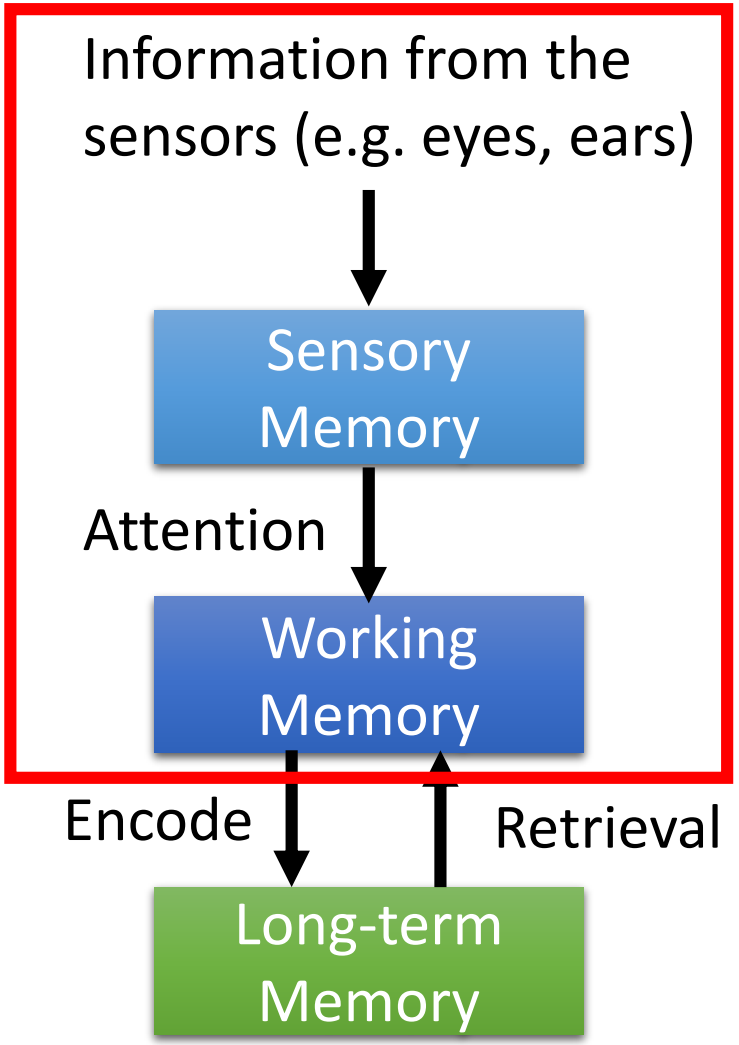
Hung-yi Lee

Attention

- Reasoning, memory
- Human's memory
- Working memeoyA

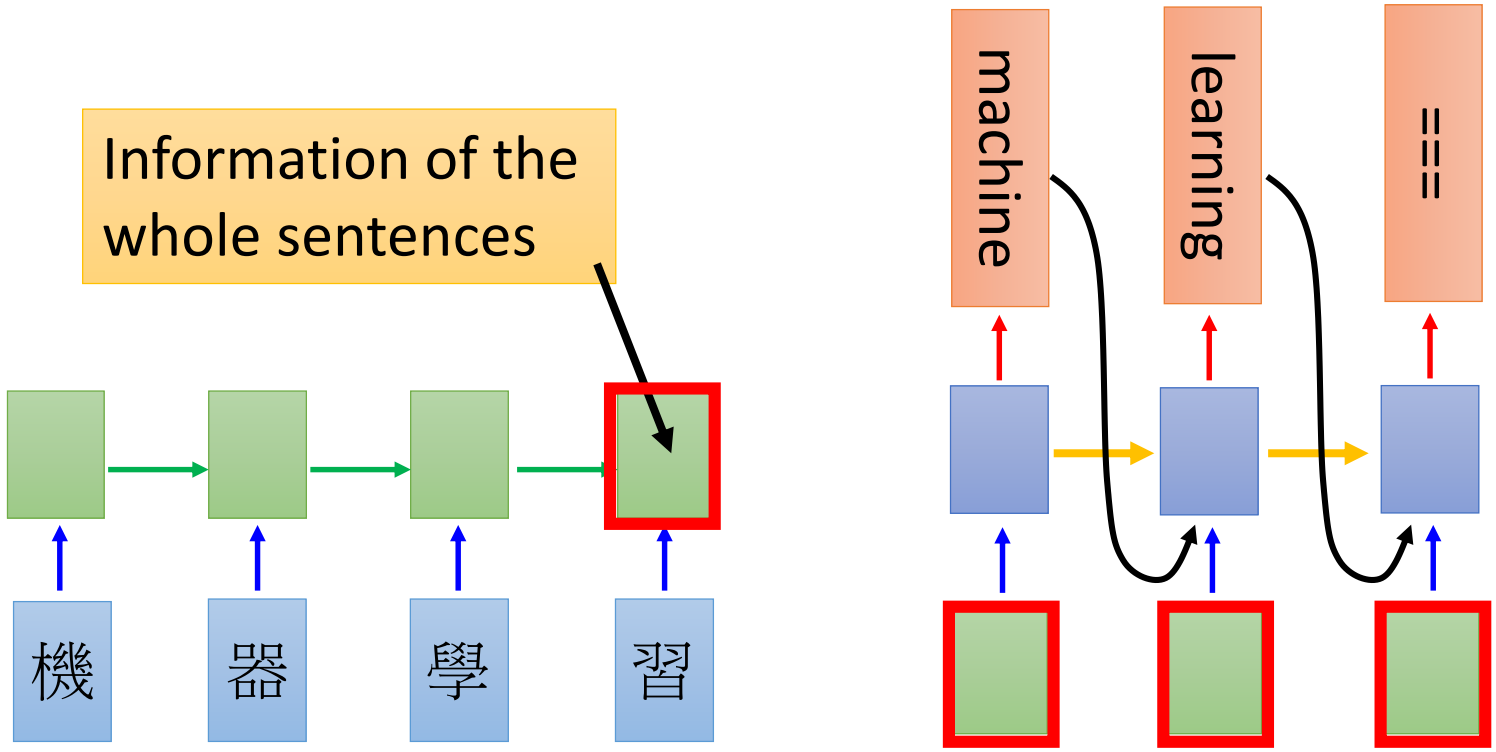


Attention on Sensory Information



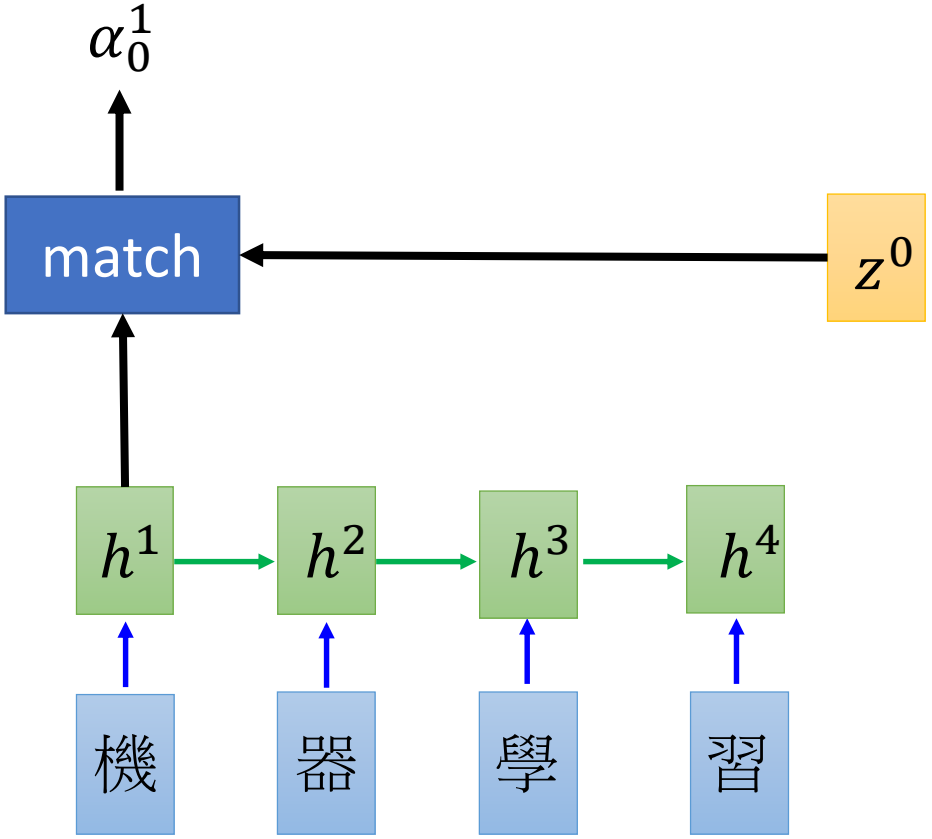
Machine Translation

- Sequence to sequence learning: Both input and output are both sequences ***with different lengths***.
- E.g. 機器學習→machine learning



Machine Translation

- Attention-based model



What is match ?

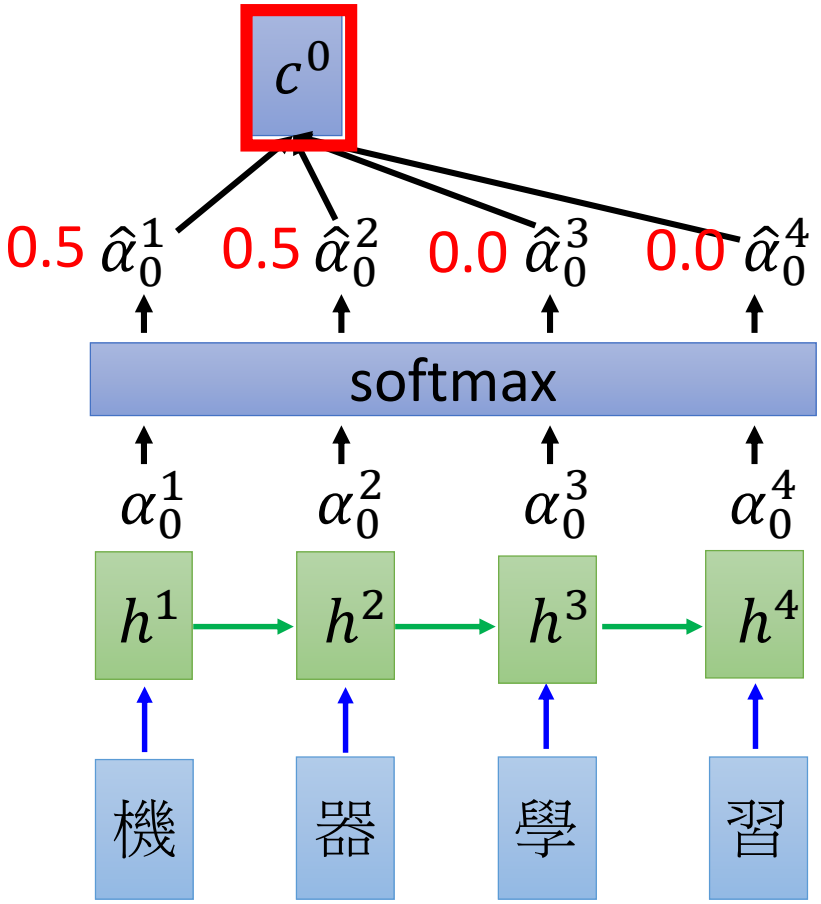
- Cosine similarity of z and h
- Small NN whose input is z and h, output a scalar
- $\alpha = h^T W z$

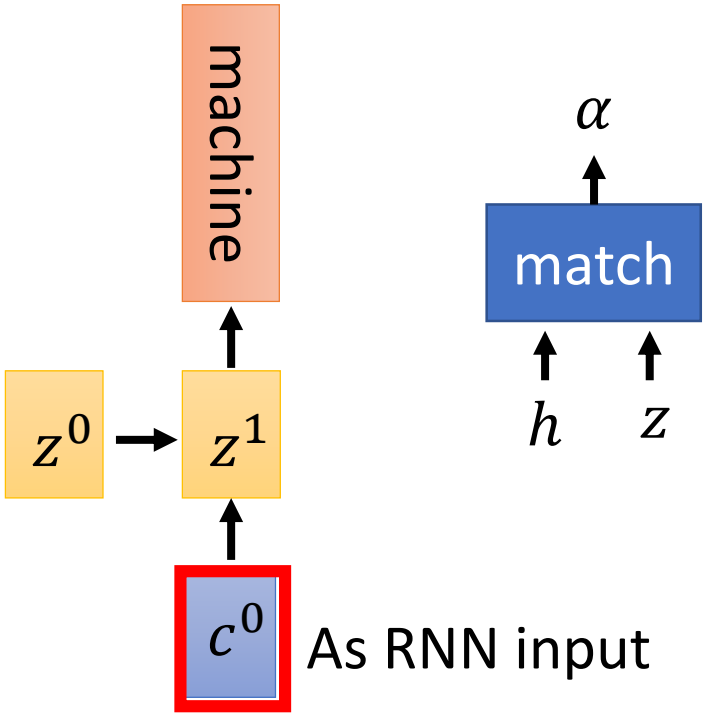
How to learn the parameters?

Machine Translation

How to learn the parameters?

- Attention-based model



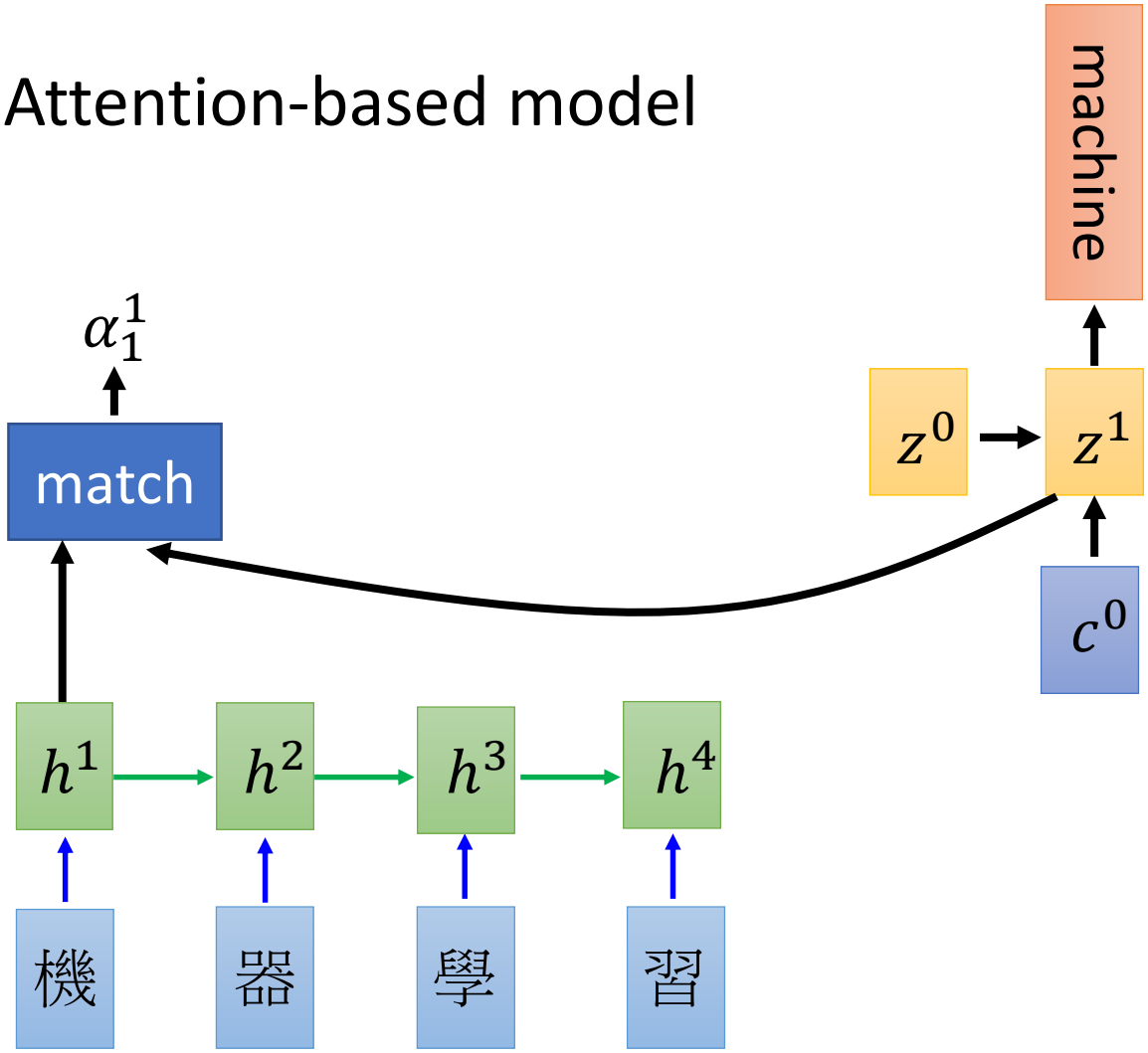


$$c^0 = \sum \hat{\alpha}_0^i h^i$$

$$= 0.5h^1 + 0.5h^2$$

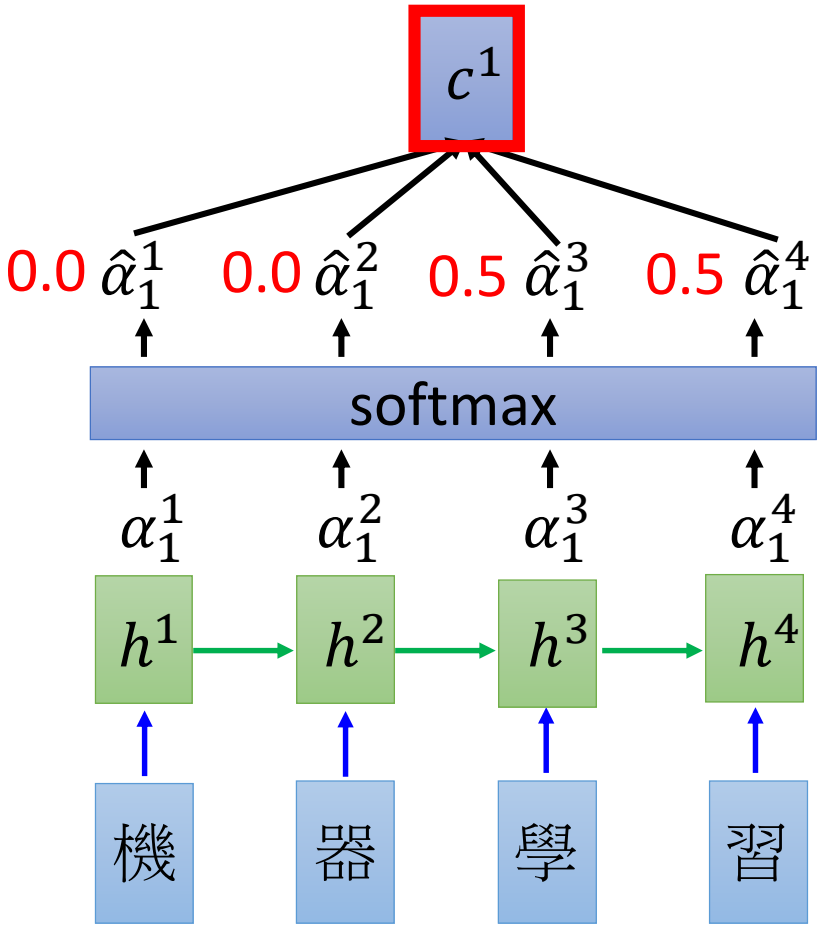
Machine Translation

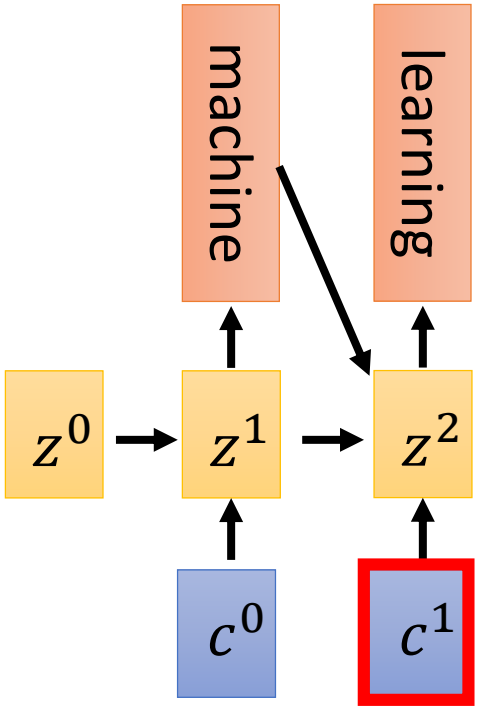
- Attention-based model



Machine Translation

- Attention-based model



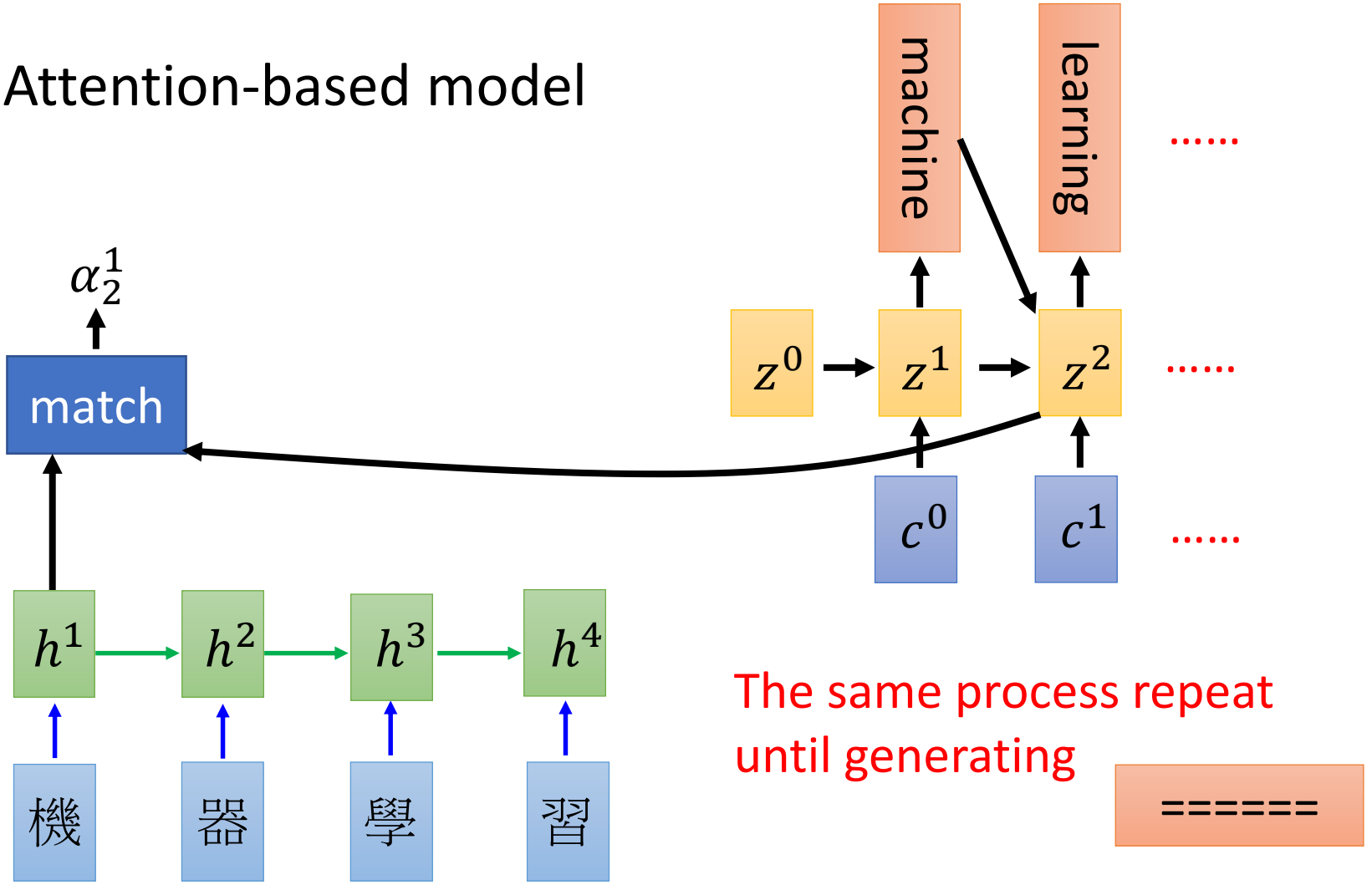


$$c^1 = \sum \hat{\alpha}_1^i h^i$$

$$= 0.5h^3 + 0.5h^4$$

Machine Translation

- Attention-based model



Speech Recognition

William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals, "Listen, Attend and Spell", arXiv'15

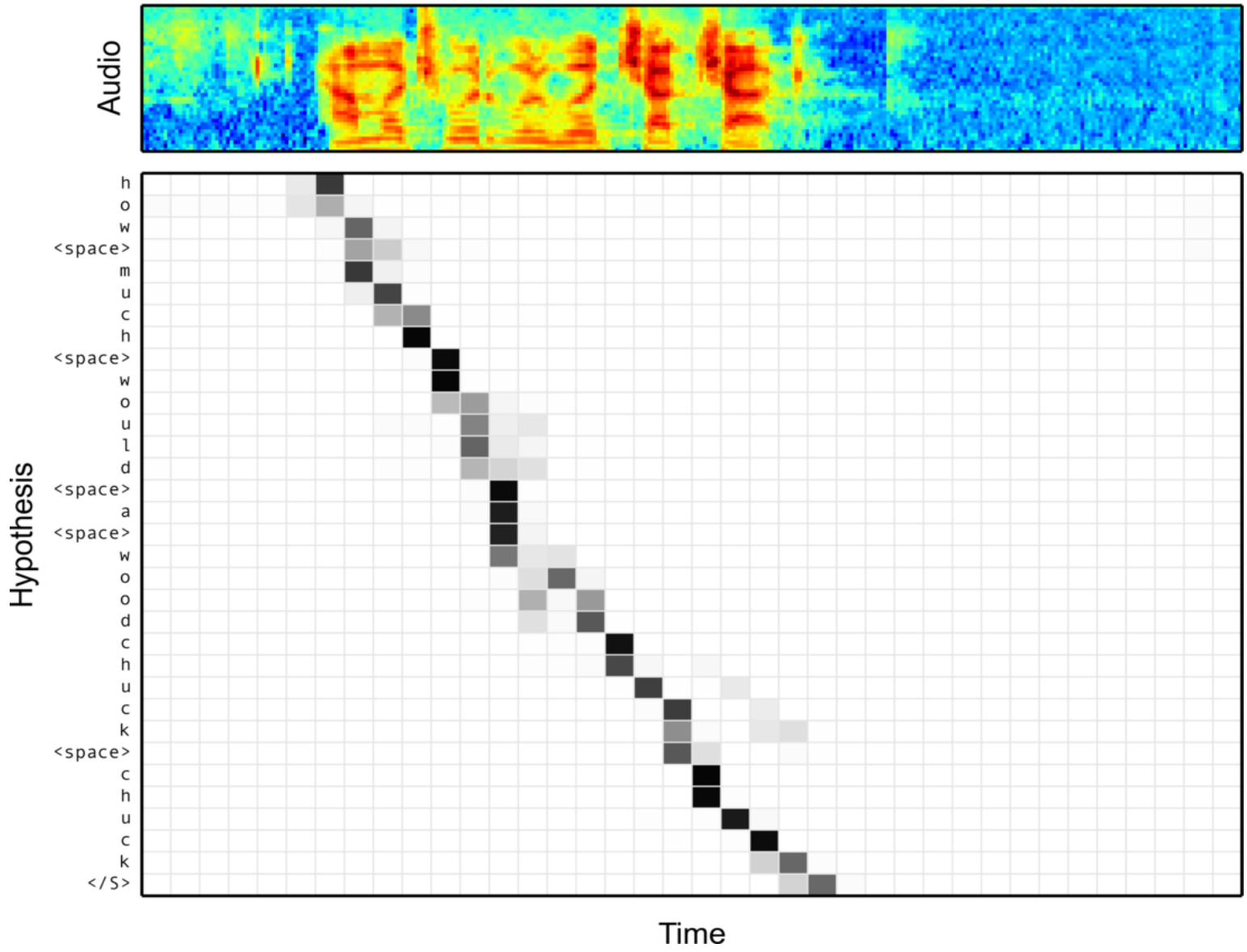


Image Caption Generation

- Input an image, but output a sequence of words

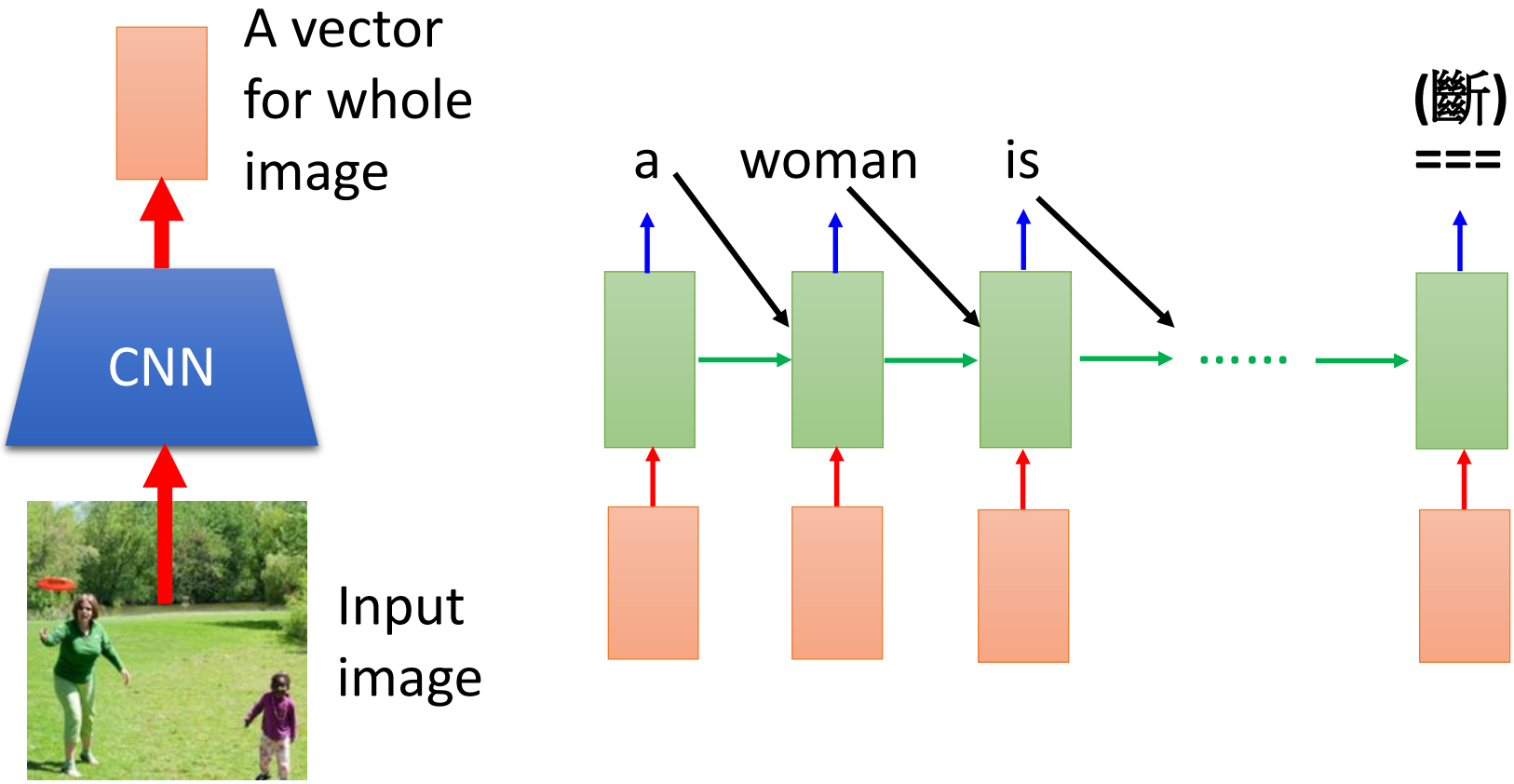


Image Caption Generation

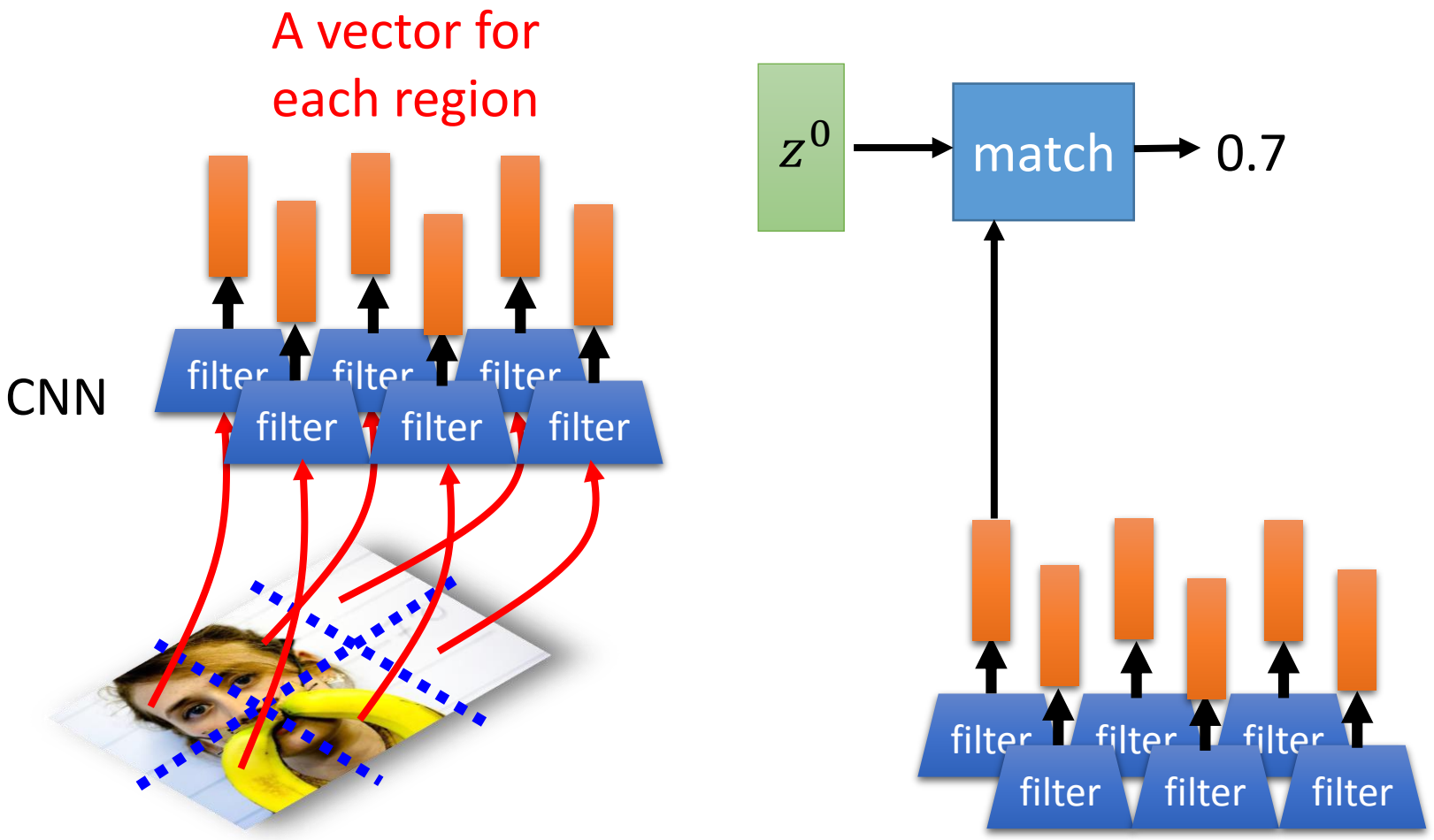


Image Caption Generation

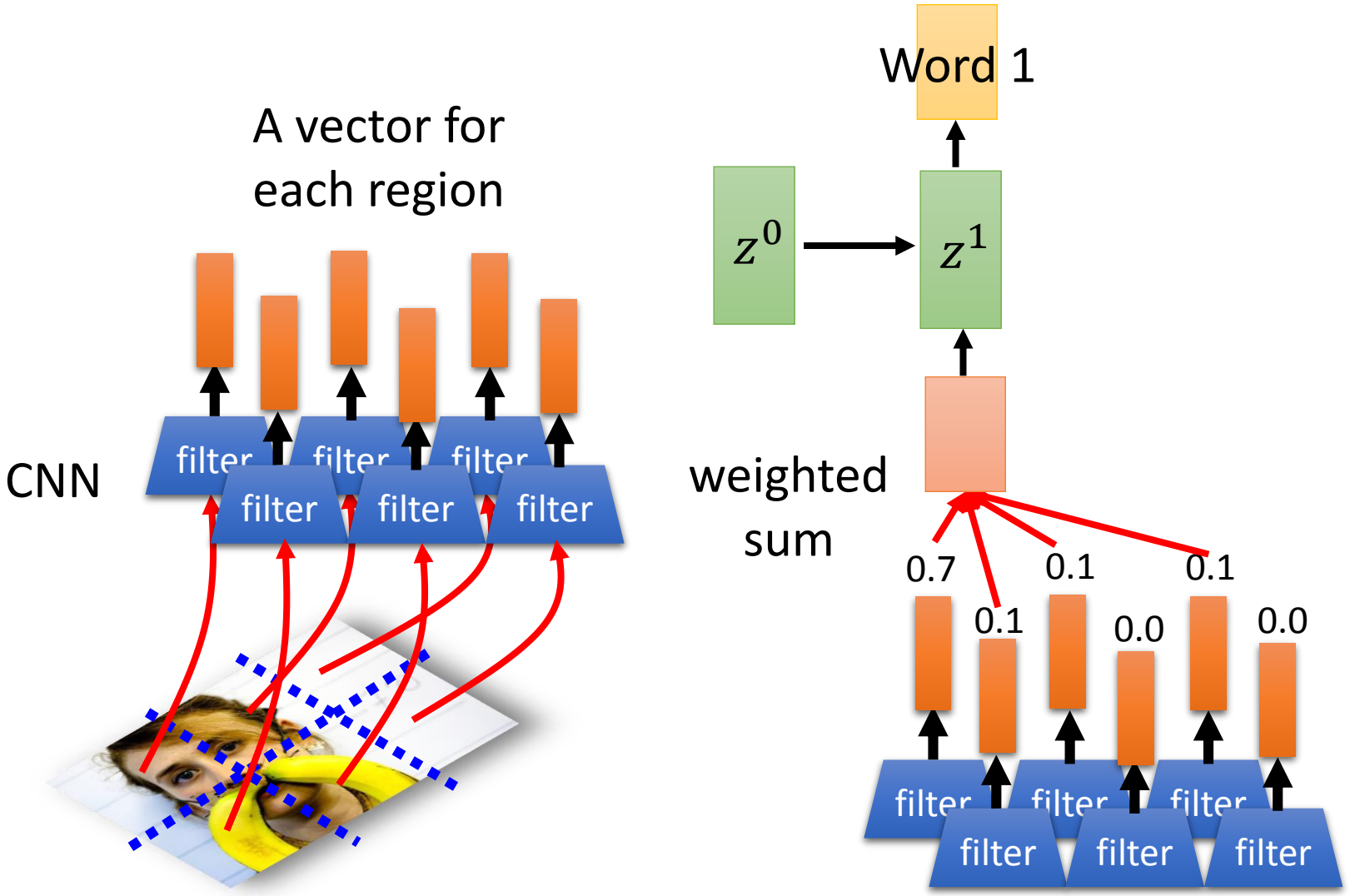


Image Caption Generation

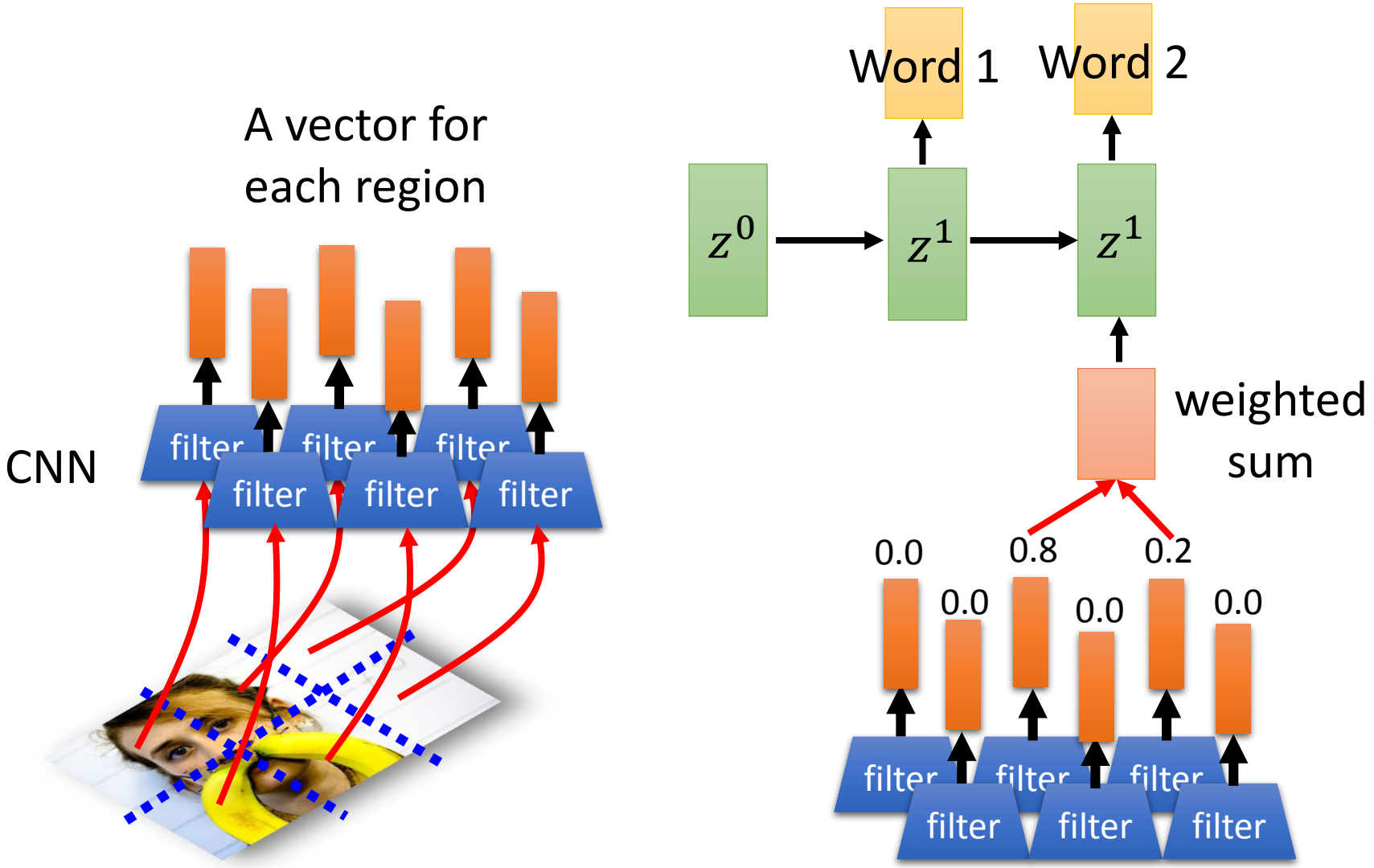


Image Caption Generation

- Good captions

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 Caption Generation

- Bad captions

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.

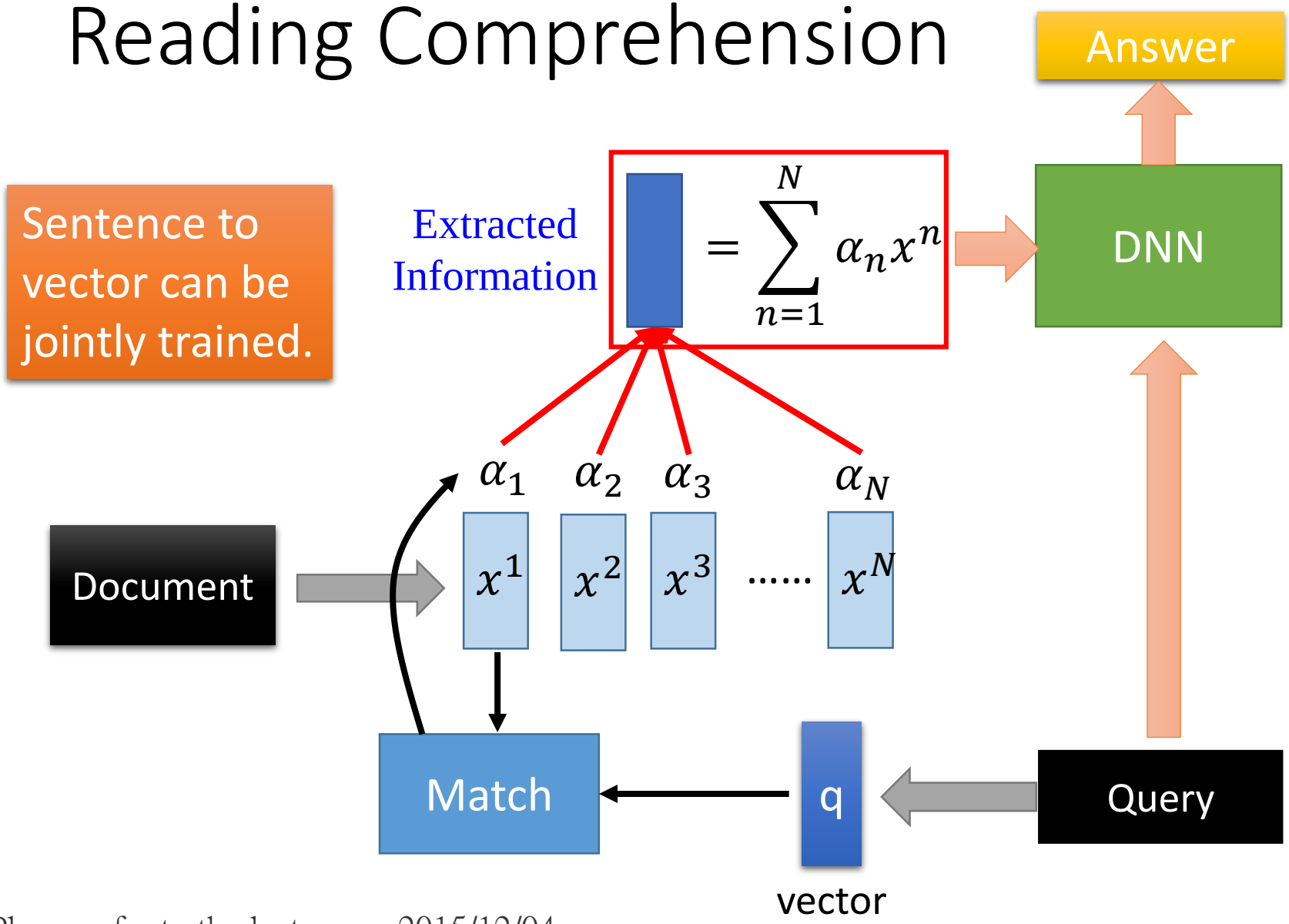
Ref: A man and a woman ride a motorcycle

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

Ref: A woman is frying food

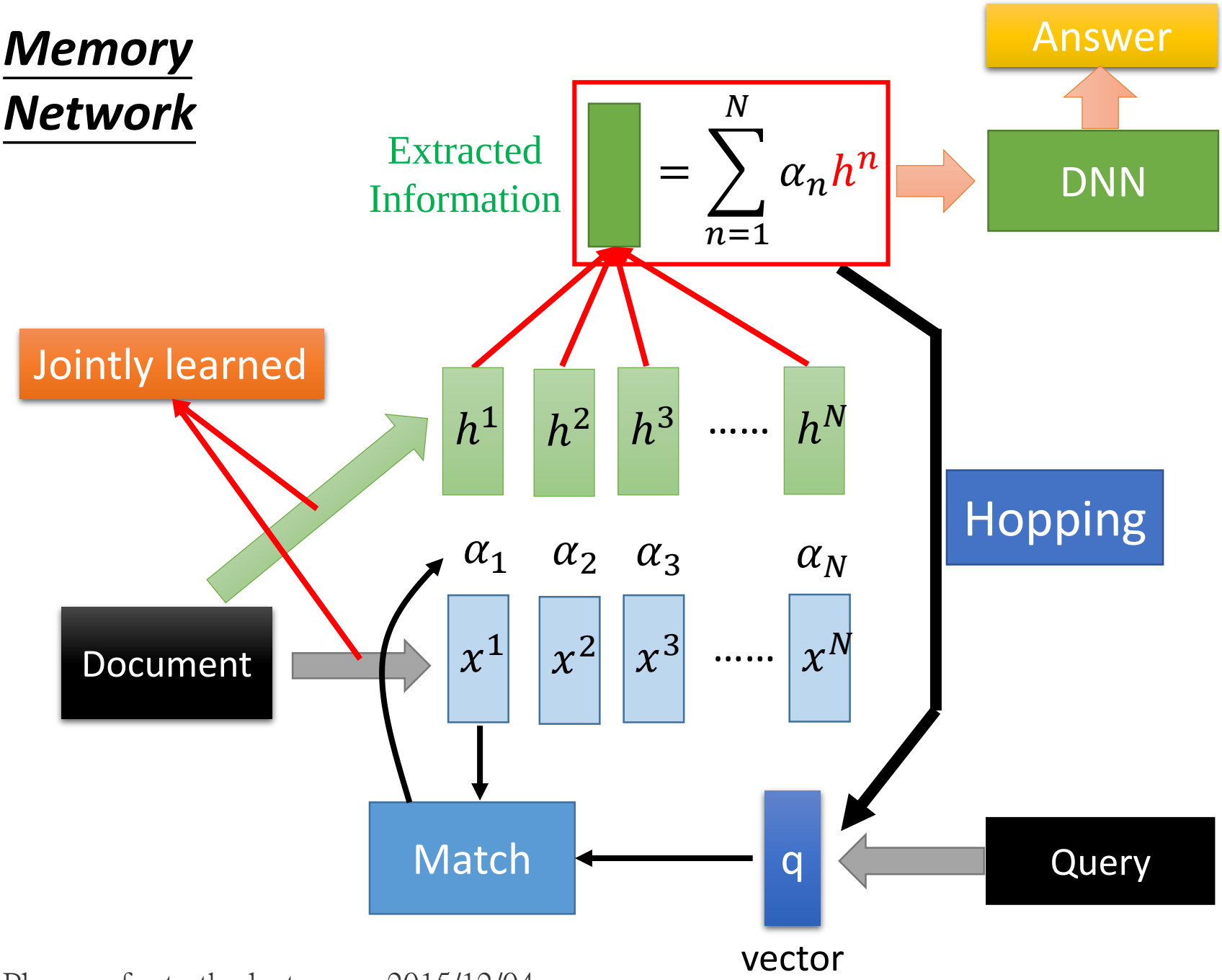
Someone is **frying** a **fish** in a **pot**

Reading Comprehension

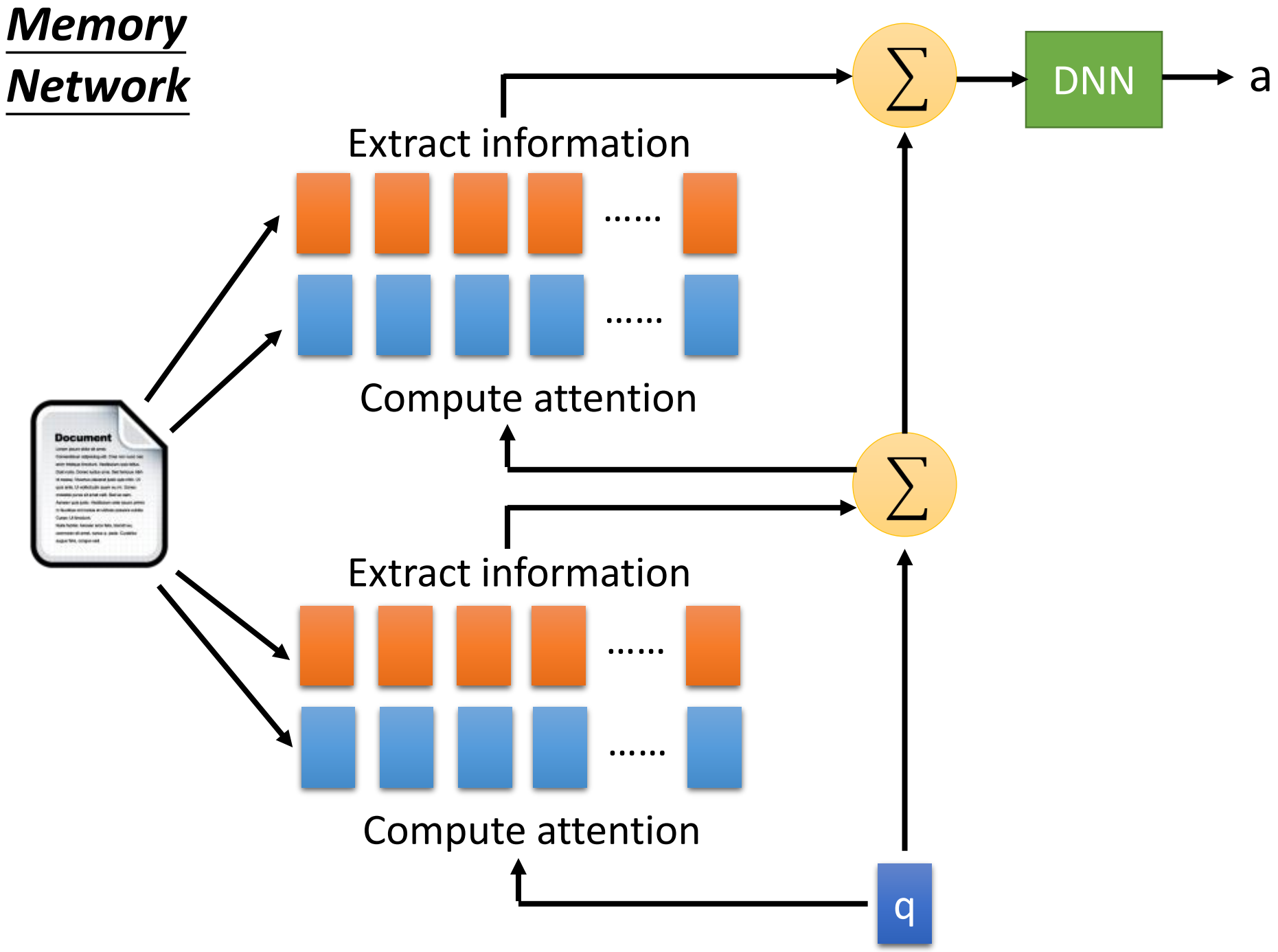


Please refer to the lecture on 2015/12/04

Memory Network



Please refer to the lecture on 2015/12/04


Memory Network
Extract information
Compute attention
Extract information
Compute attention
Document
DNN
a
q

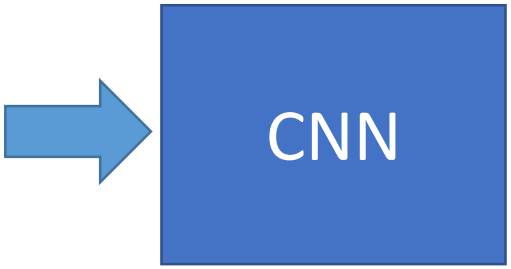
Memory Network

- Performance of Hopping

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**			

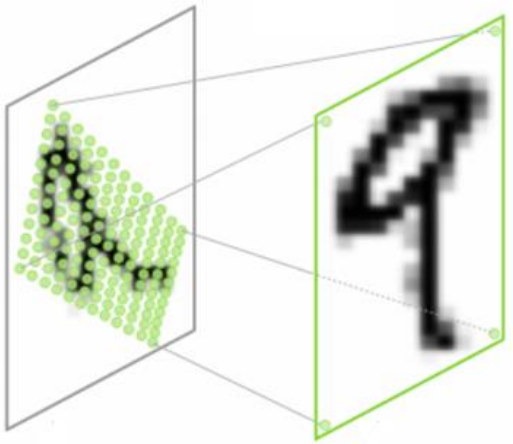
Demo video: https://www.facebook.com/Engineering/videos/10153098860532200/

Special Attention: Spatial Transformers



Good results

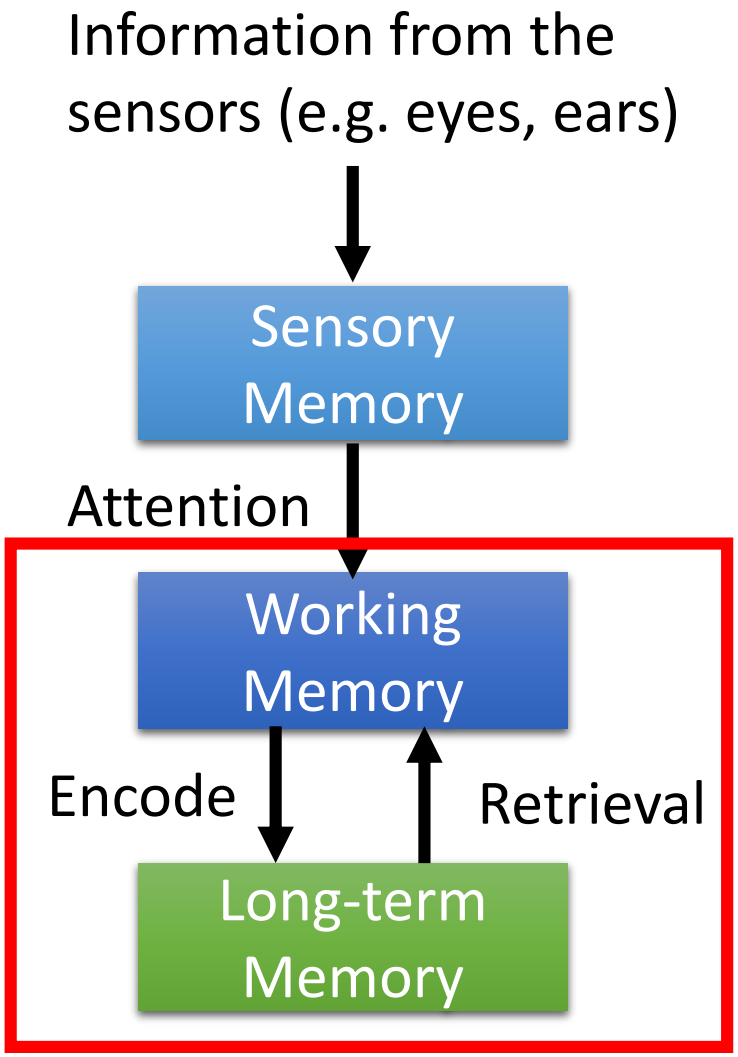
Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu, Spatial Transformer Networks, arXiv'15



Bad results

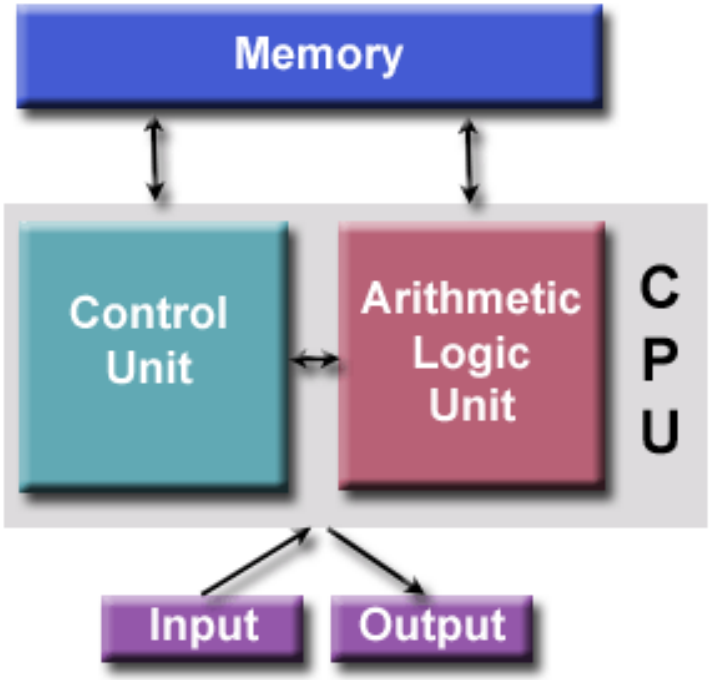
Jointly learned

Attention on Memory



Neural Turing Machine

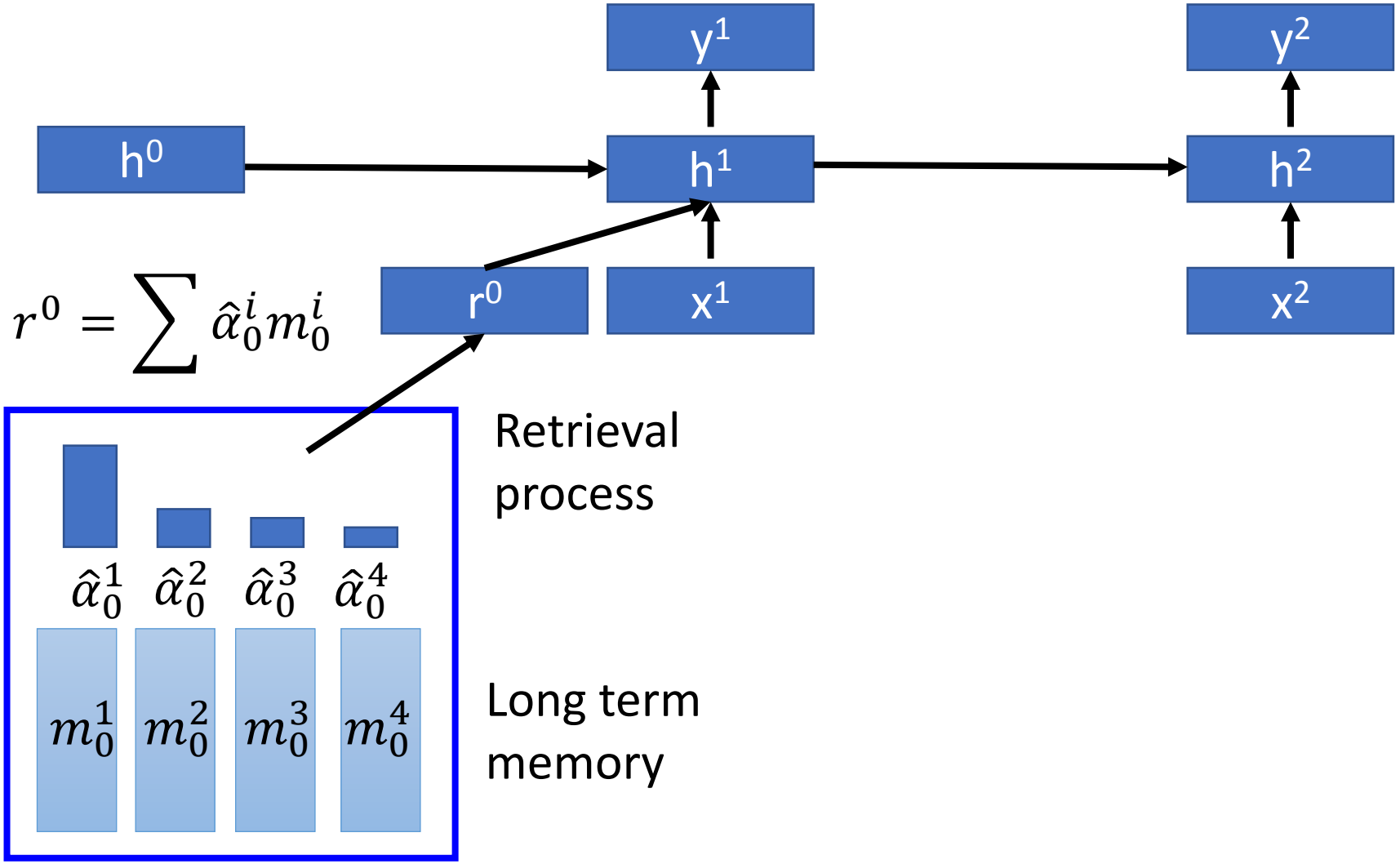
- von Neumann architecture



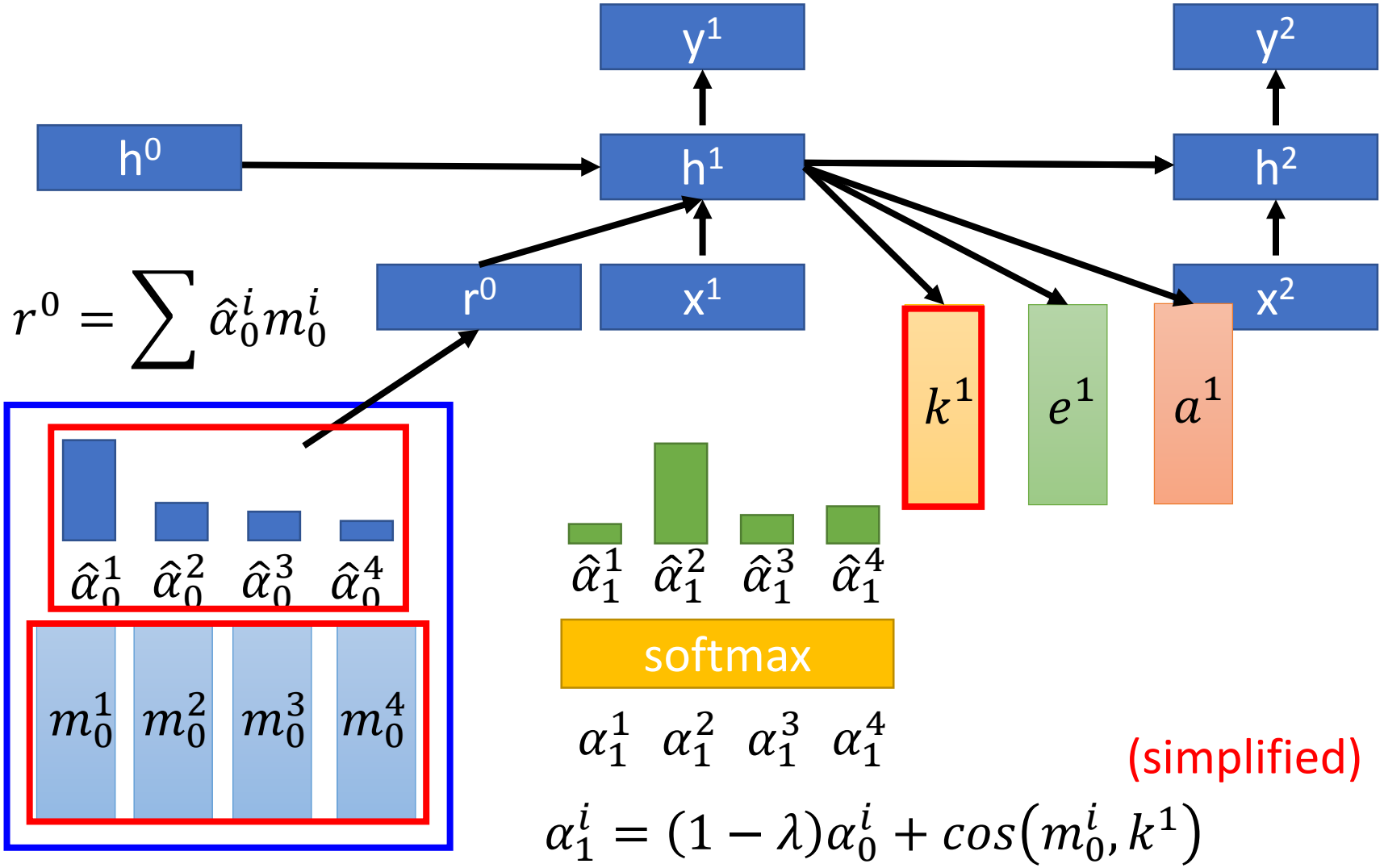
Actually, Neural Turing Machine is an advanced RNN/LSTM.

https://www.quora.com/How-does-the-Von-Neumann-architecture-provide-flexibility-for-program-development

Neural Turing Machine



Neural Turing Machine



(simplified)

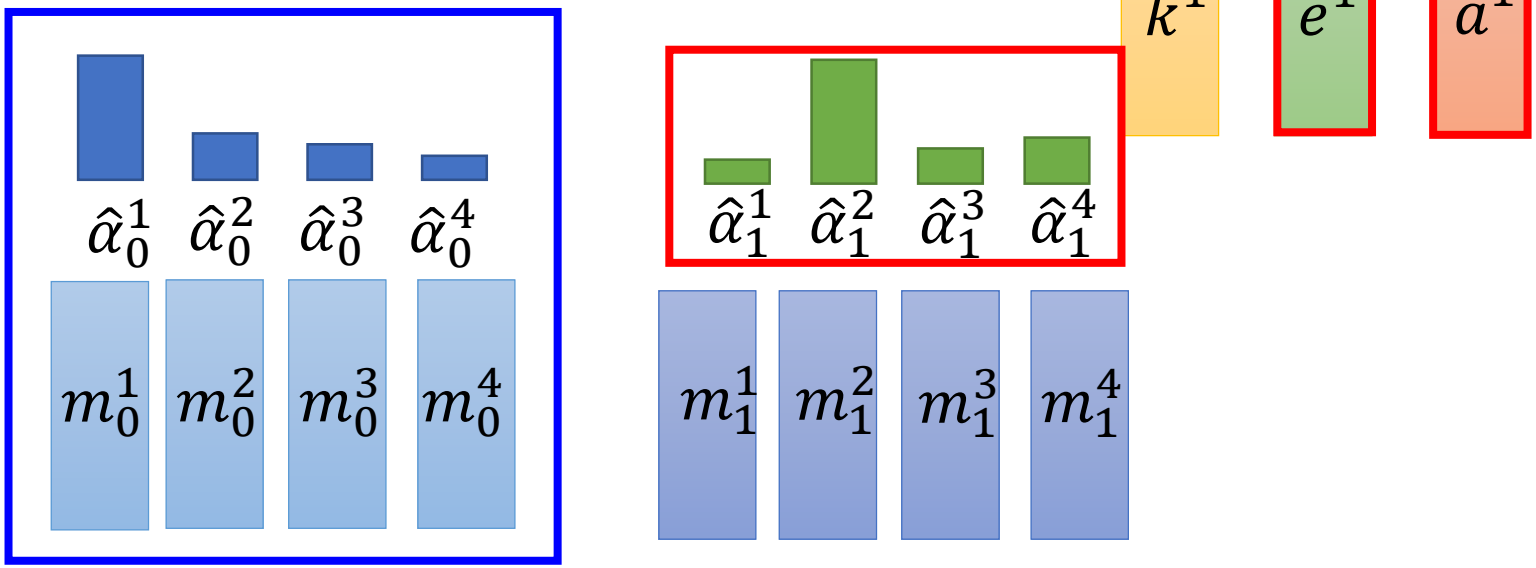
$$\alpha_1^i = (1-\lambda)\alpha_0^i + cos(m_0^i, k^1)$$

Neural Turing Machine

$$m_1^i = m_0^i * \left(1 - \hat{\alpha}_1^i \, e^1 \right) + \hat{\alpha}_1^i \, a^1$$

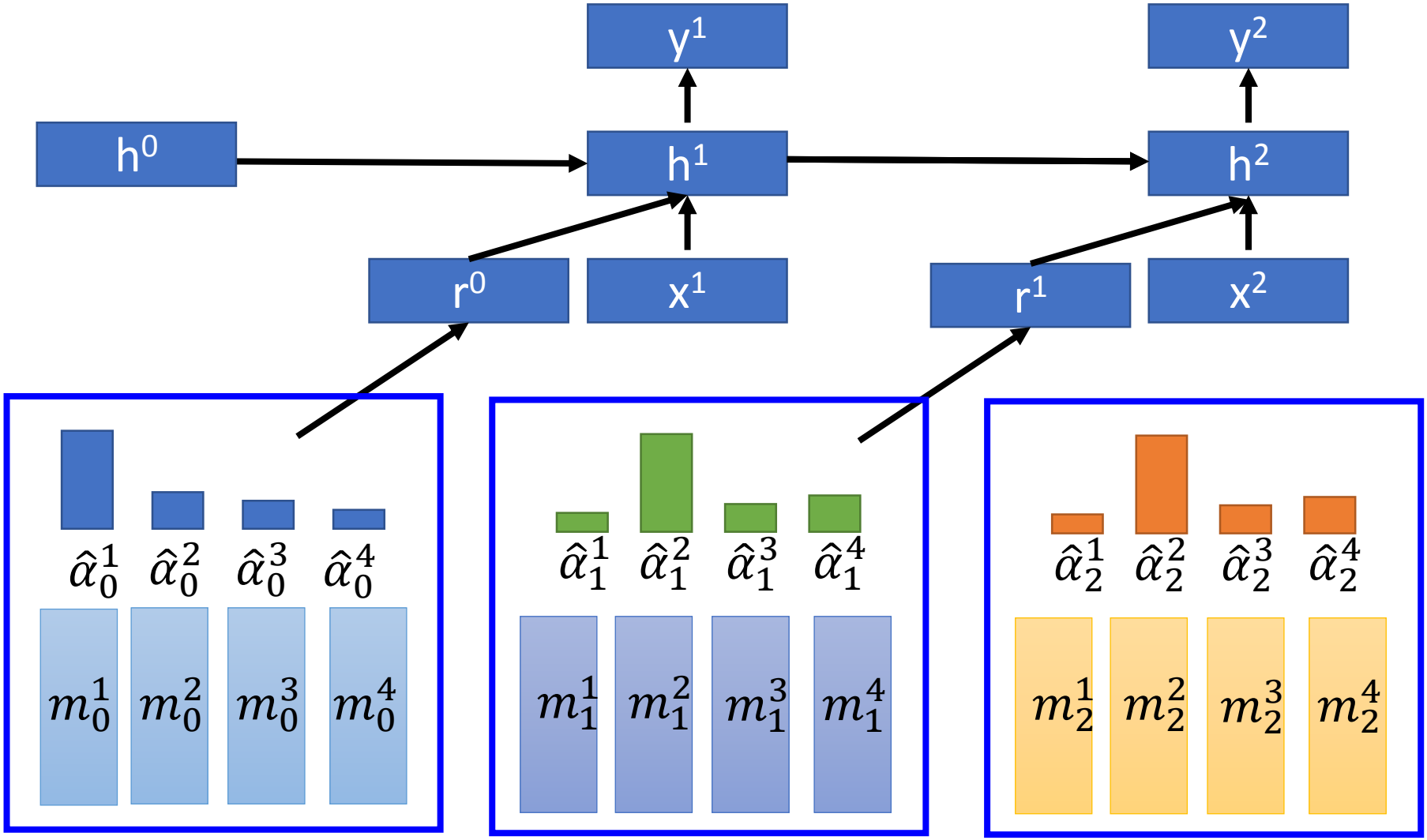
Encode process

(element-wise)



Neural Turing Machine

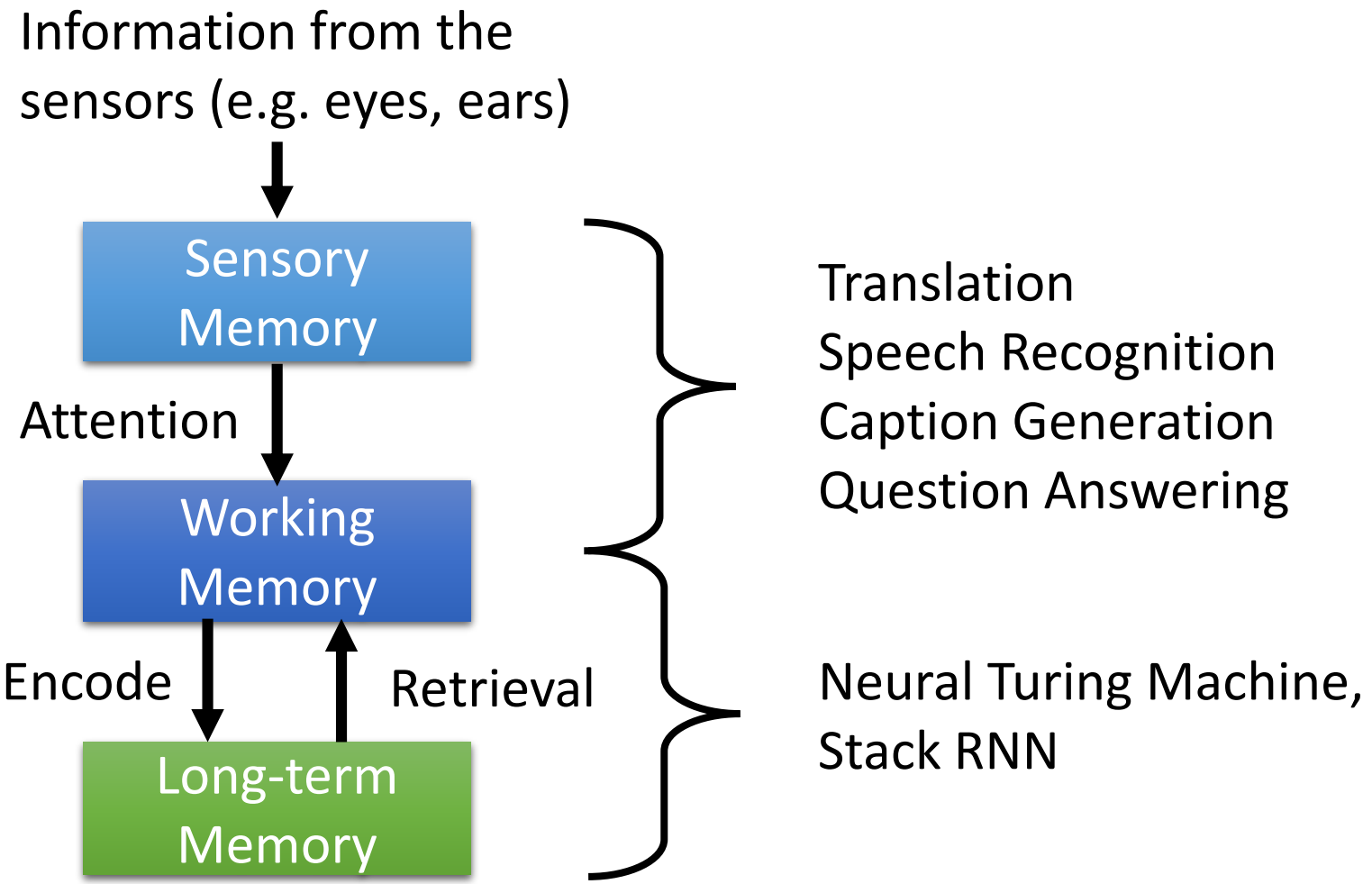
Wei Zhang, Yang Yu, Bowen Zhou, Structured Memory for Neural Turing Machines, arXiv'15



Stack RNN

Armand Joulin, Tomas Mikolov, Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets, 2015

Concluding Remarks

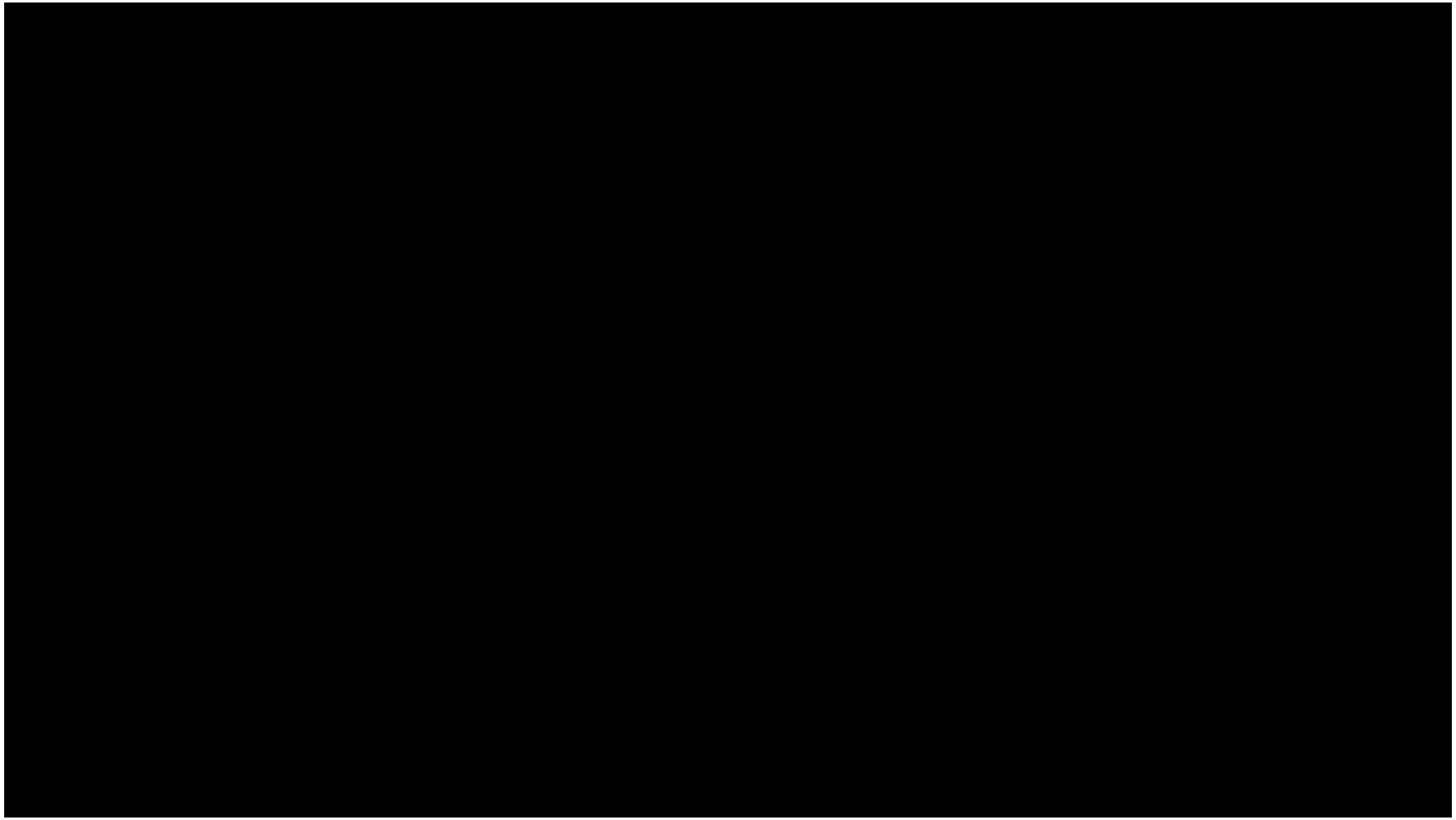


Reference

- 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.

Plan

- 1/8 (五) 23:59: Presentation team decided
- 1/13 (三) 23:59: Presentation slides deadline
- 1/15 (五)
  - 上課時間：Presentation
  - 返鄉投票
- 1/16 (六) : 投票
- 1/20 (三) 23:59: Report deadline



Teaching Machines to Read and Comprehend, Hermann et. al. (2015)