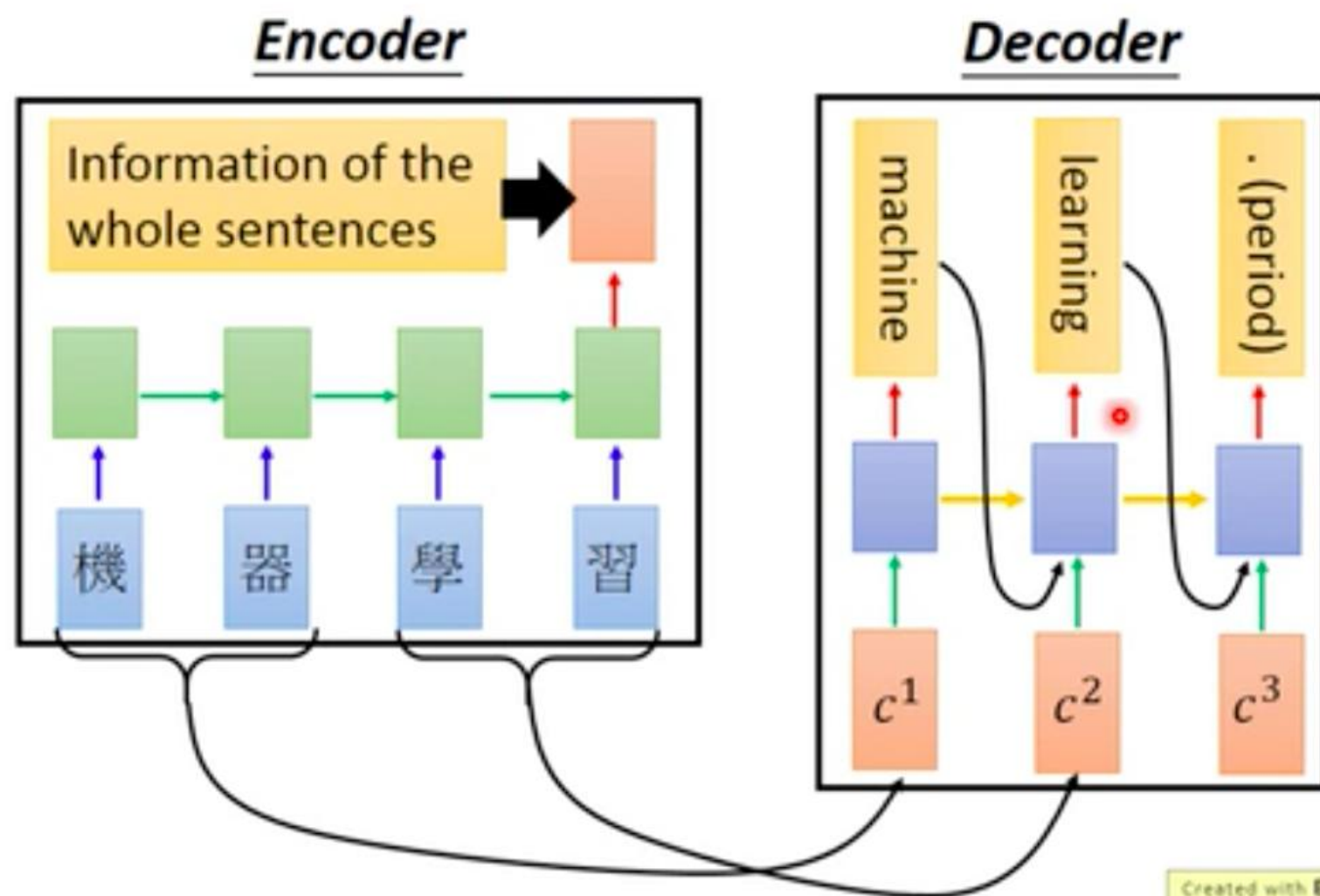
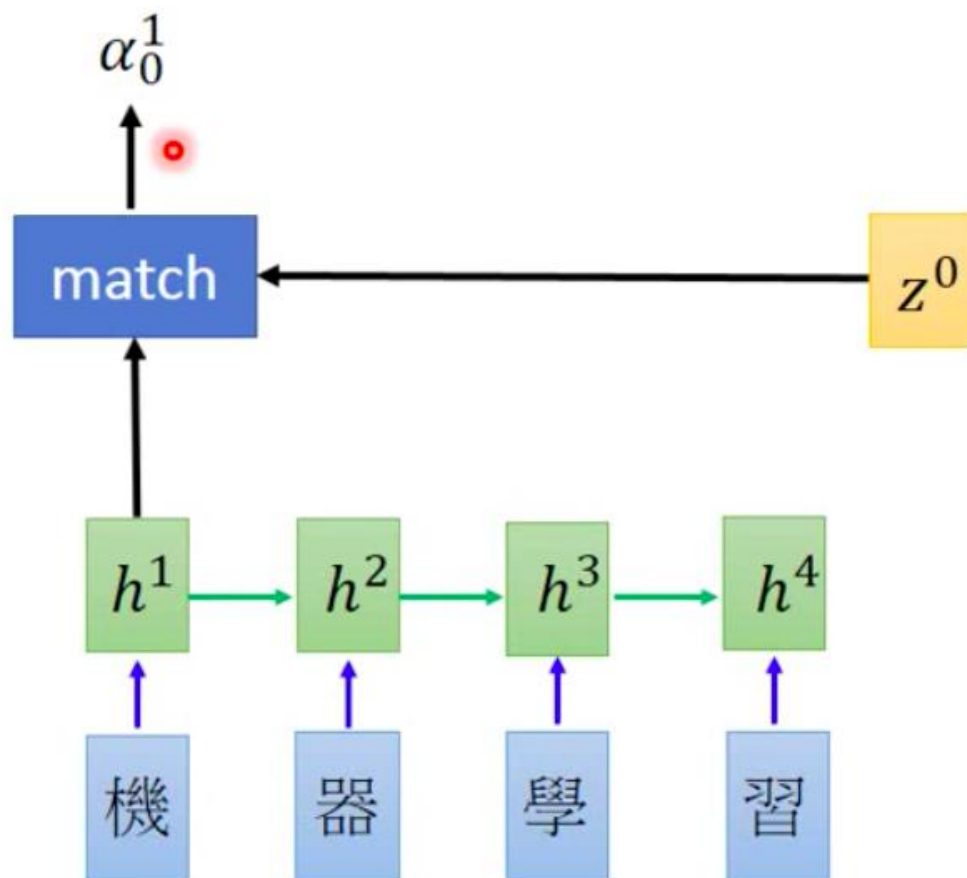


Dynamic Conditional Generation



Machine Translation

- Attention-based model



What is **match** ?

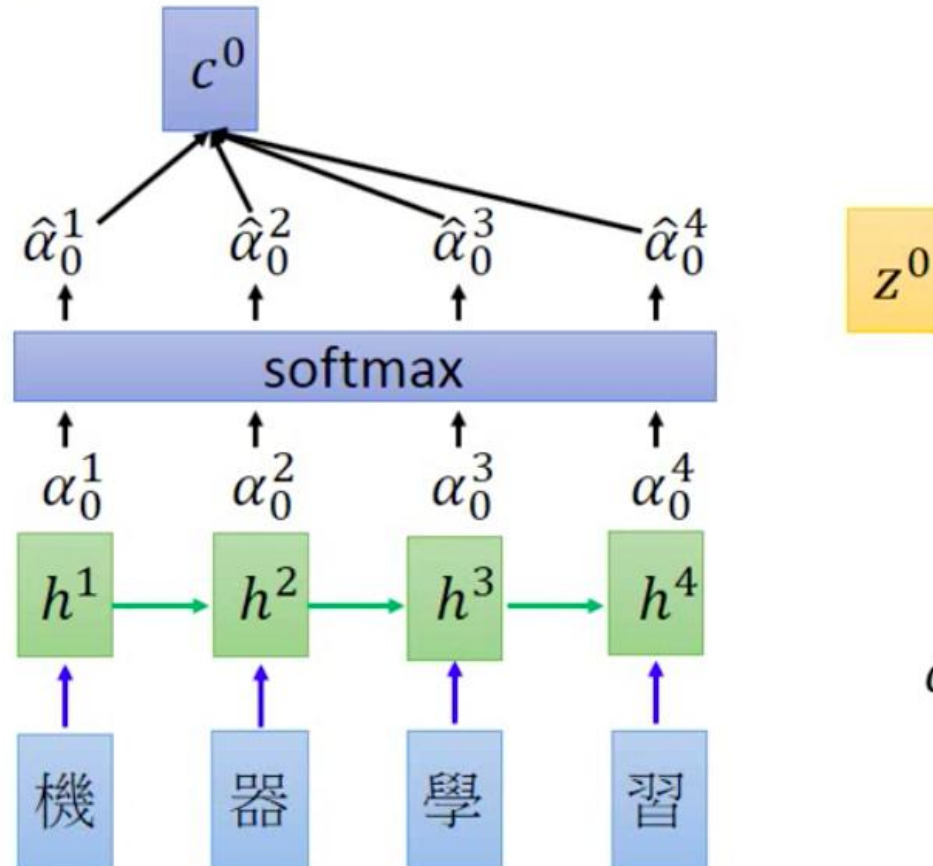
Design by yourself

- Cosine similarity of z and h
- Small NN whose input is z and h , output a scalar

➤ $\alpha = h^T W z$

Machine Translation

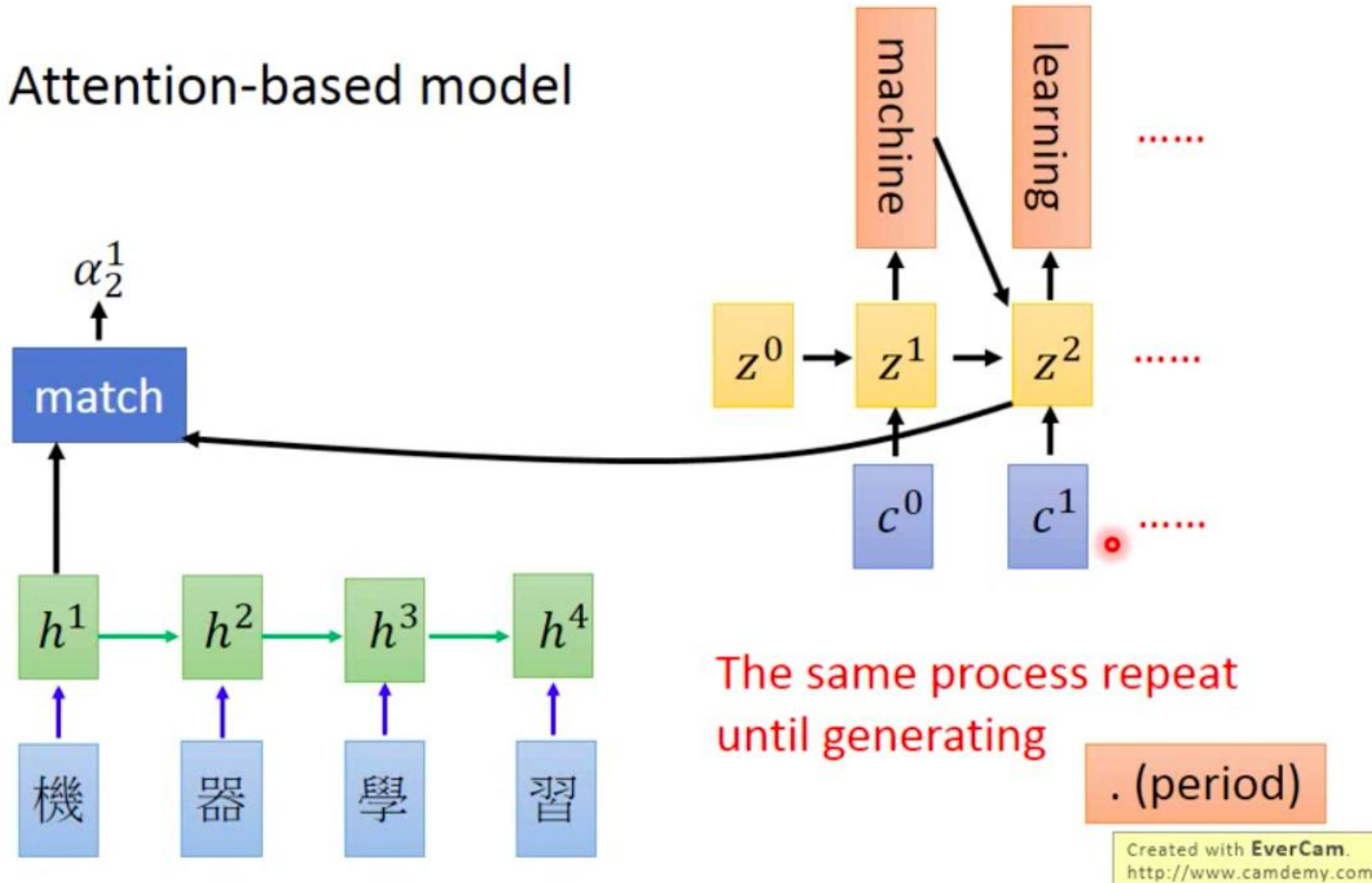
- Attention-based model



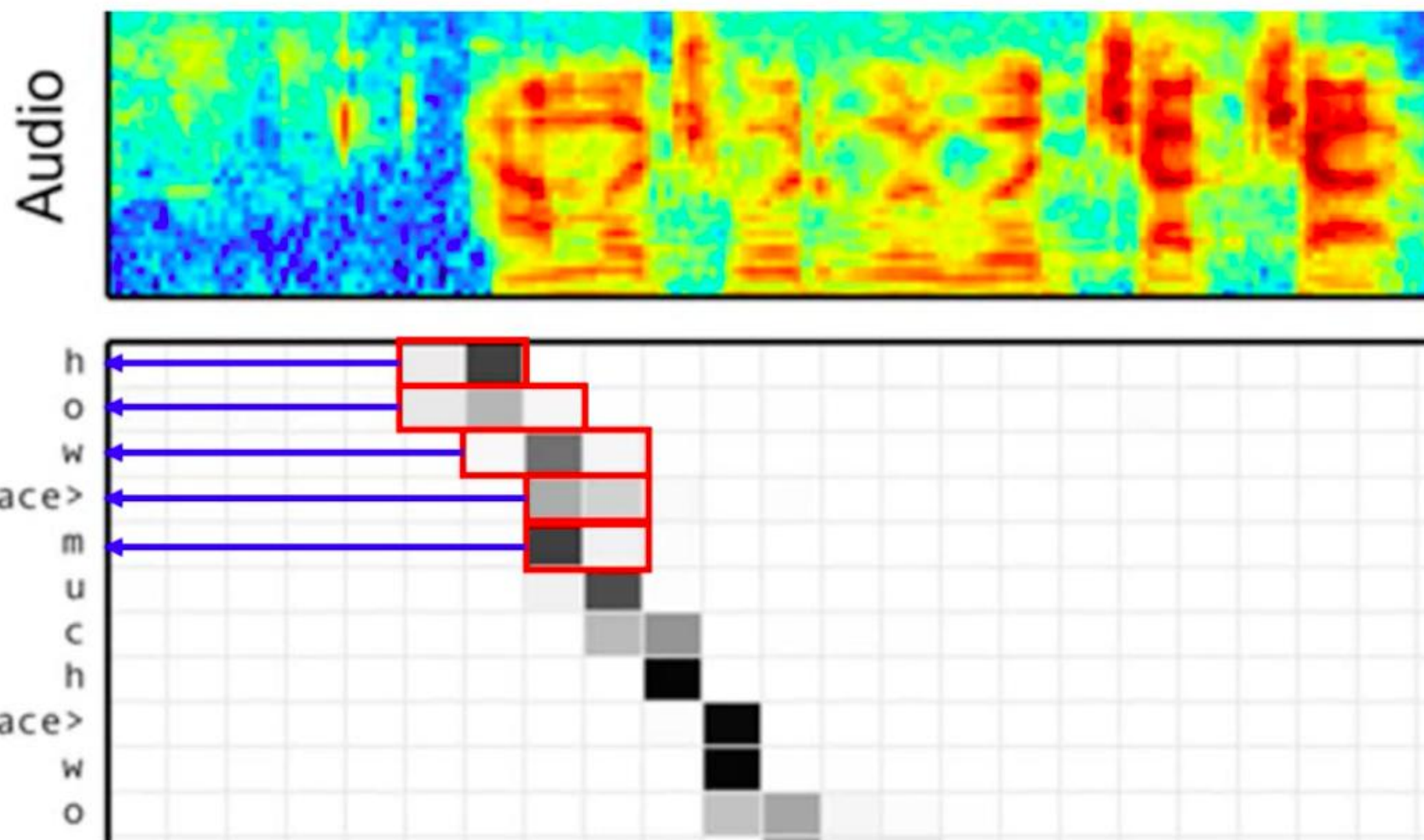
$$c^0 = \sum \hat{\alpha}_0^i h^i$$

Machine Translation

- Attention-based model



Speech Recognition



William Chan, Navdeep Jaitly, Quoc V. Le, Oriol Vinyals, "Listen, Attend and Spell", ICASSP, 2016

Image Caption Generation

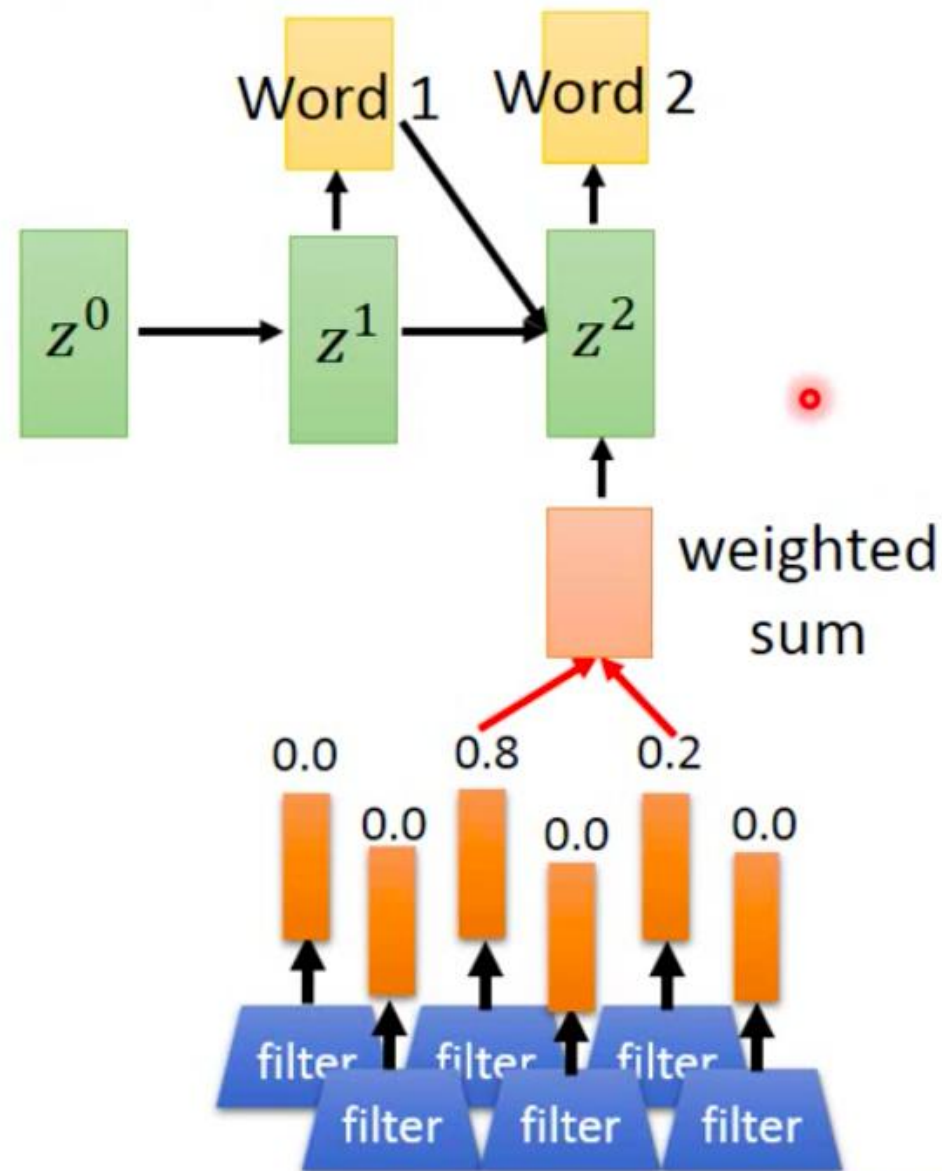
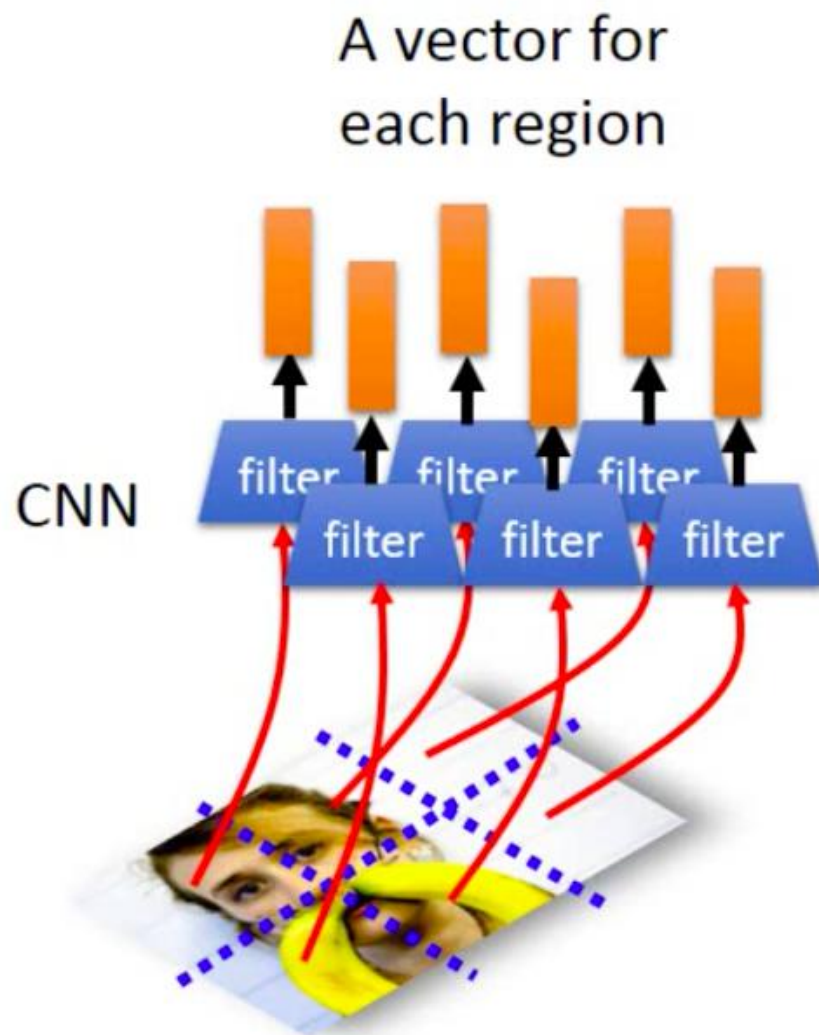


Image Caption Generation



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.



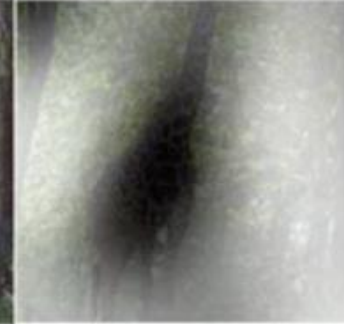
A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.



A giraffe standing in a forest with tree in the background.



Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML, 2015

Image Caption Generation



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.



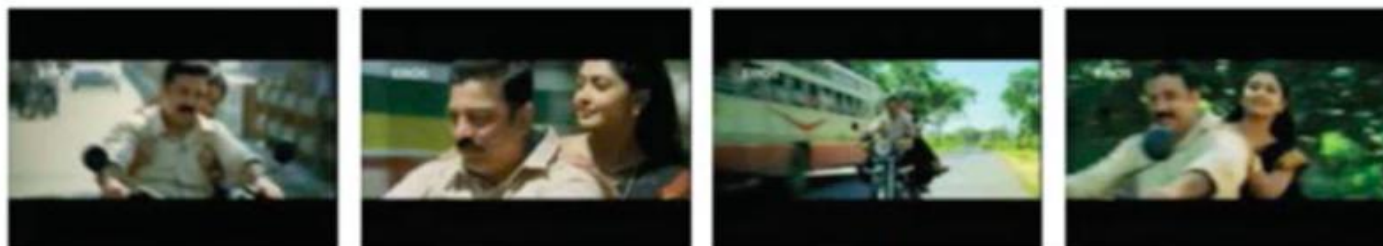
A woman is sitting at a table with a large pizza.



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



Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhutdinov, Richard Zemel, Yoshua Bengio, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention", ICML, 2015



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**