AI Solution for AI needs customer

Mr.MIND

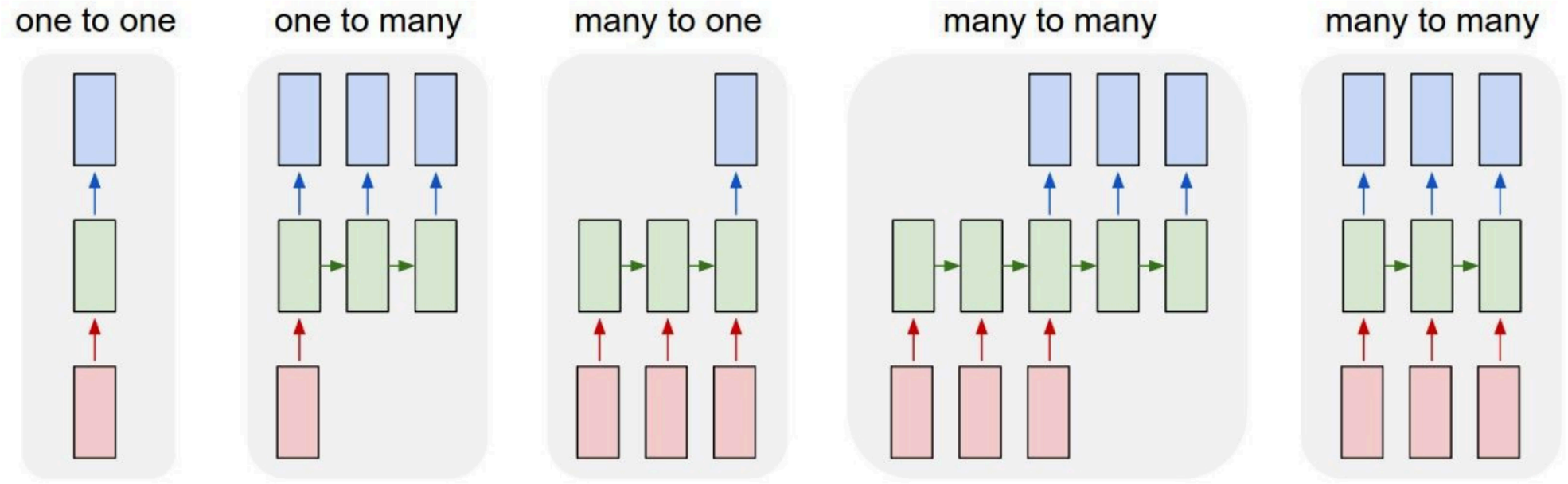
Sequence to Sequence와 Attention Mechanism

최태균 Mindmap.ai 선임연구원
| tgchoi03@gmail.com | 010-2004-1188

V1.0 | Released in 2018.01
Document by **Mr.MIND AI Consulting**

1. Sequence to Sequence Model

Sequence to Sequence Model

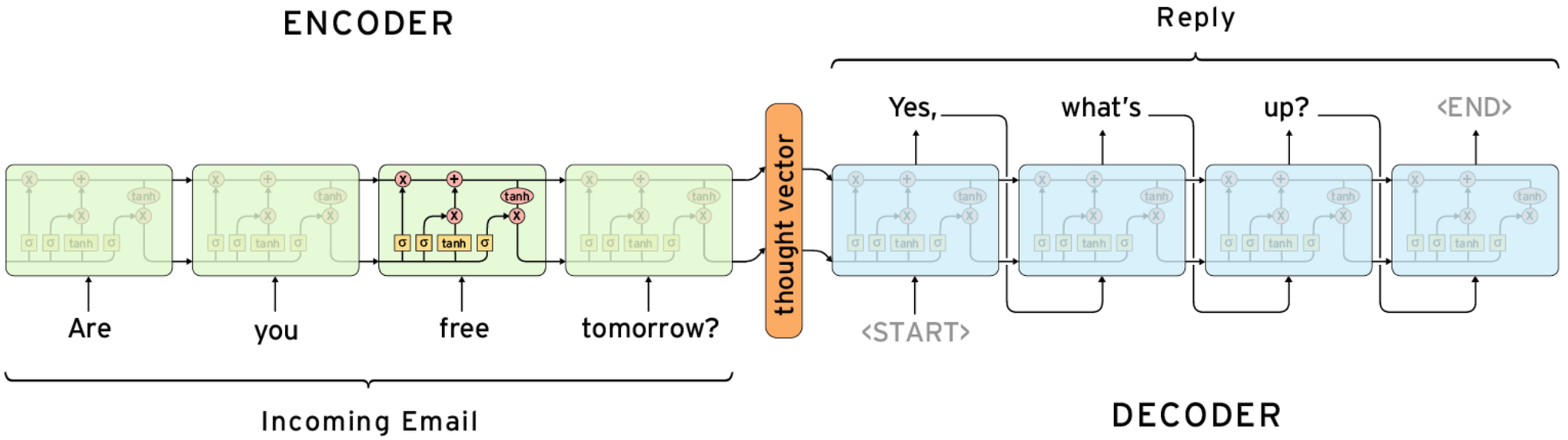


e.g. **Machine Translation**
seq of words -> seq of words

모델 소개

Sequence to Sequence Model

- Sequence 입력하여 Sequence를 출력하는 모델
- 많은 자연어처리에서 활용 중 i.e) 기계번역, 챗봇 등



Sequence to Sequence 모델

모델 소개

Sequence to Sequence Model

- 입력 sequence X에 대해 출력 sequence Y가 나타날 확률을 모델링

$$P(y_1, y_2, \dots, y_{t-1}, y_{T_a} | x_1, x_2, \dots, x_{t-1}, x_{T_b})$$

- Encoder RNN
  - 입력 sequence X의 정보에 대해 hidden state vector로 저장
  - h_{T_a}는 입력 sequence에 대한 정보를 종합적으로 담고 있다.

$$h_i = f_w(h_{t-1}, x_t), \quad z_t = f_{w'}(h_{T_a}, y_{t-1}, z_{t-1})$$

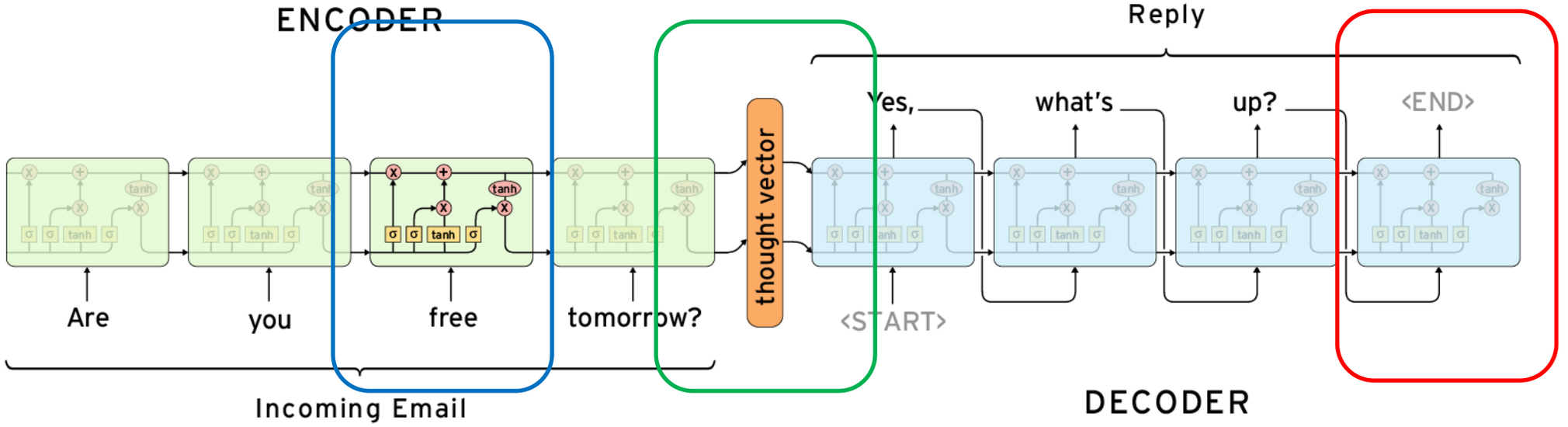
- Decoder RNN
  - 매 time step 마다 Encoder 정보를 summarize 한 z_t를 통해 출력 output y_t를 예측

$$P(y_t | t_{<t}, X) = g(z_{t-1})$$

모델 소개

Sequence to Sequence Model

$h_i = f_w(h_{t-1}, x_t)$ $z_t = f_{w'}(h_{T_a}, y_{t-1}, z_{t-1})$ $P(y_1, y_2, \dots, y_{t-1}, y_{T_a} | x_1, x_2, \dots, x_{t-1}, x_{T_b})$



Sequence to Sequence 모델

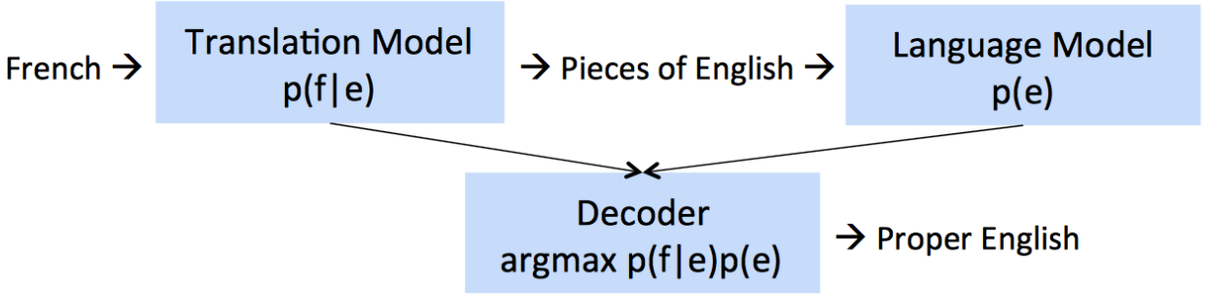
Traditional Machine Translation

- 기존의 방식은 Bayesian Rule과 같은 확률기반 모델을 활용

  - Source language f, e.g. French
  - Target language e, e.g. English
  - Probabilistic formulation (using Bayes rule)

$$\hat{e} = \text{argmax}_e p(e|f) = \text{argmax}_e p(f|e)p(e)$$

  - Translation model p(f|e) trained on parallel corpus
  - Language model p(e) trained on English only corpus (lots, free!)



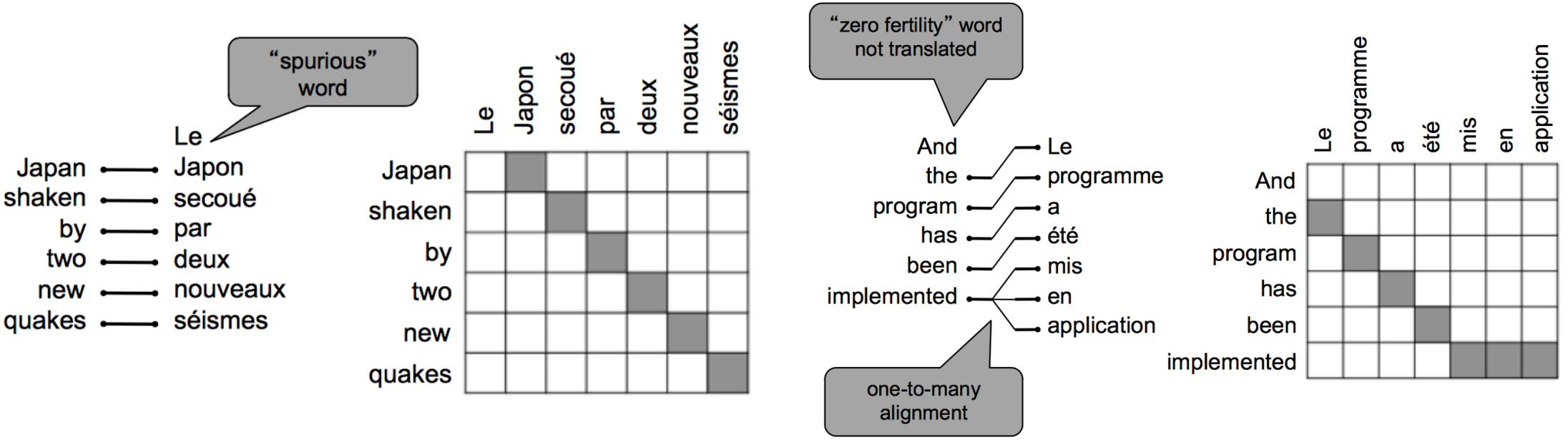
Bayes' Rule

$$p(\theta|x) = \frac{p(\theta)f(x|\theta)}{p(x)}$$

Machine Translation 사례

Traditional Machine Translation의 한계

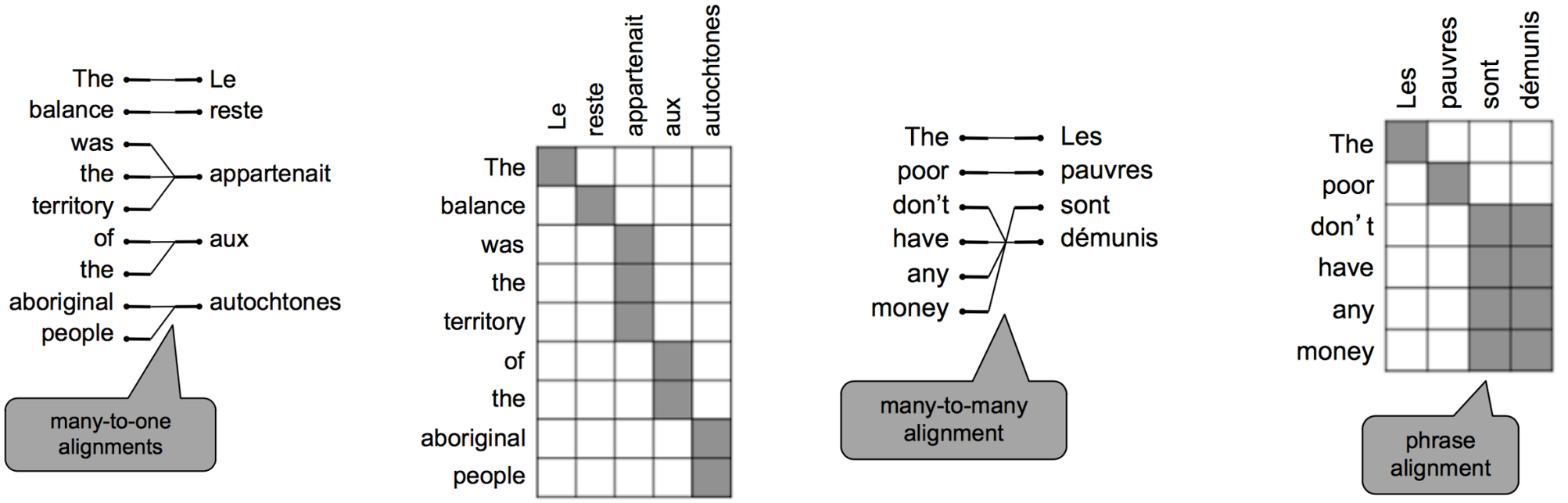
- 번역 시, 번역하고자 하는 언어에 대한 단어 또는 구에 대한 매칭이 쉽지 않다.



Machine Translation 사례

Traditional Machine Translation의 한계

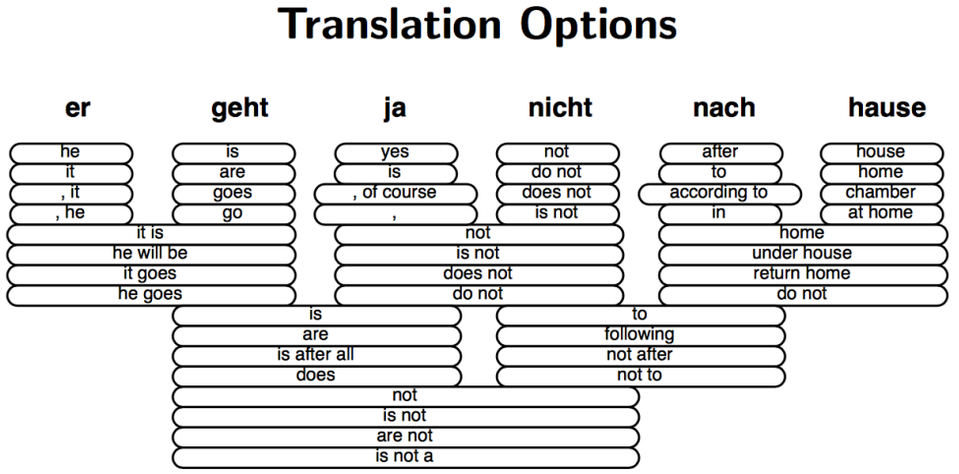
- 번역 시, 번역하고자 하는 언어에 대한 단어 또는 구에 대한 매칭이 쉽지 않다.

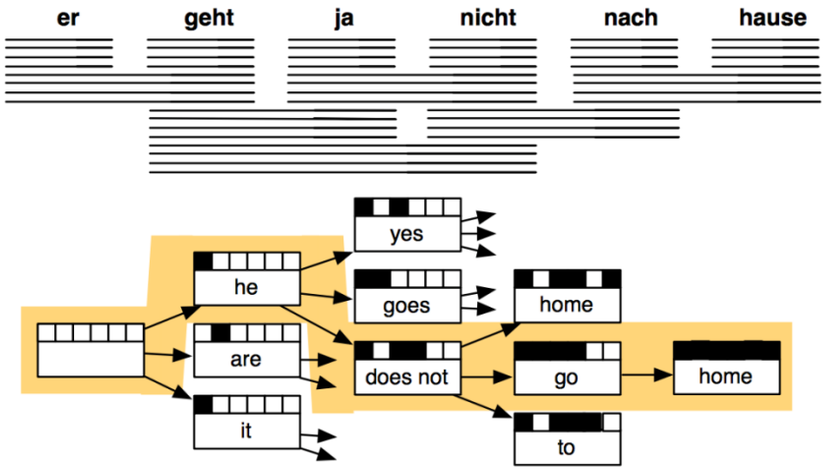


Machine Translation 사례

Traditional Machine Translation의 한계

- 가능한 번역 예상을 가능한 많이 해야 한다.
- 이는 정확한 번역을 어렵게 할 뿐만 아니라 연산 비용이 상당하다.
- 이러한 한계를 Recurrent Neural Network를 통해 해결하고자 한다.

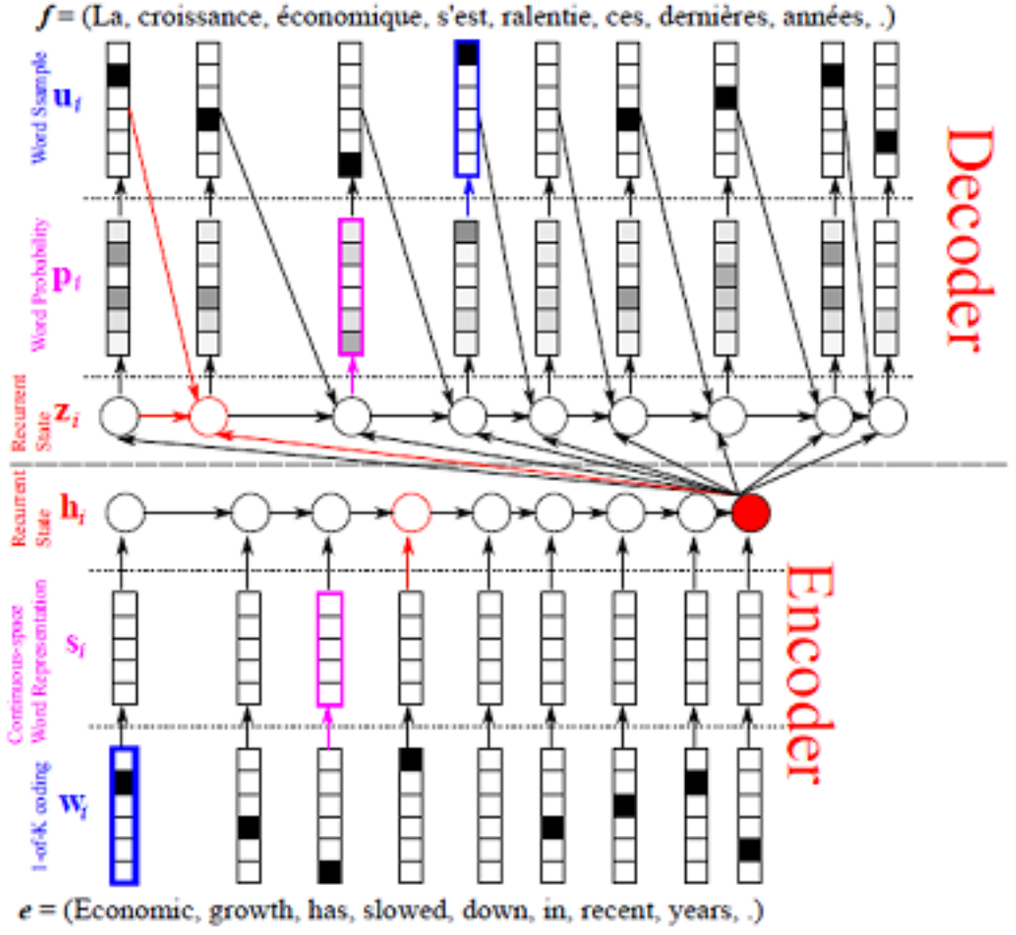




번역을 할 시 각 언어의 어휘나 구에 대한 정보에 대응하는 경우들의
조합을 가지고 탐색하는 것은 상당히 성능이 떨어질 것이다.

Machine Translation 사례

Neural Machine Translation (by Seq2Seq Model)



- Decoder

  (1)Recursively update the memory

  $$z_{t'} = f(z_{t'-1}, u_{t'-1}, h_T)$$

  (2)Compute the next word prob.

  $$p(u_{t'}|u_{<t'}) \propto \exp(R_{u_{t'}}^{\top} z_{t'} + b_{u_{t'}})$$

  (3)Sample a next word

  - *Beam search is a good idea*

- Encoder

  (1) 1-of-K coding of *source* words (One-hot Vector)

  (2)Continuous-space representation

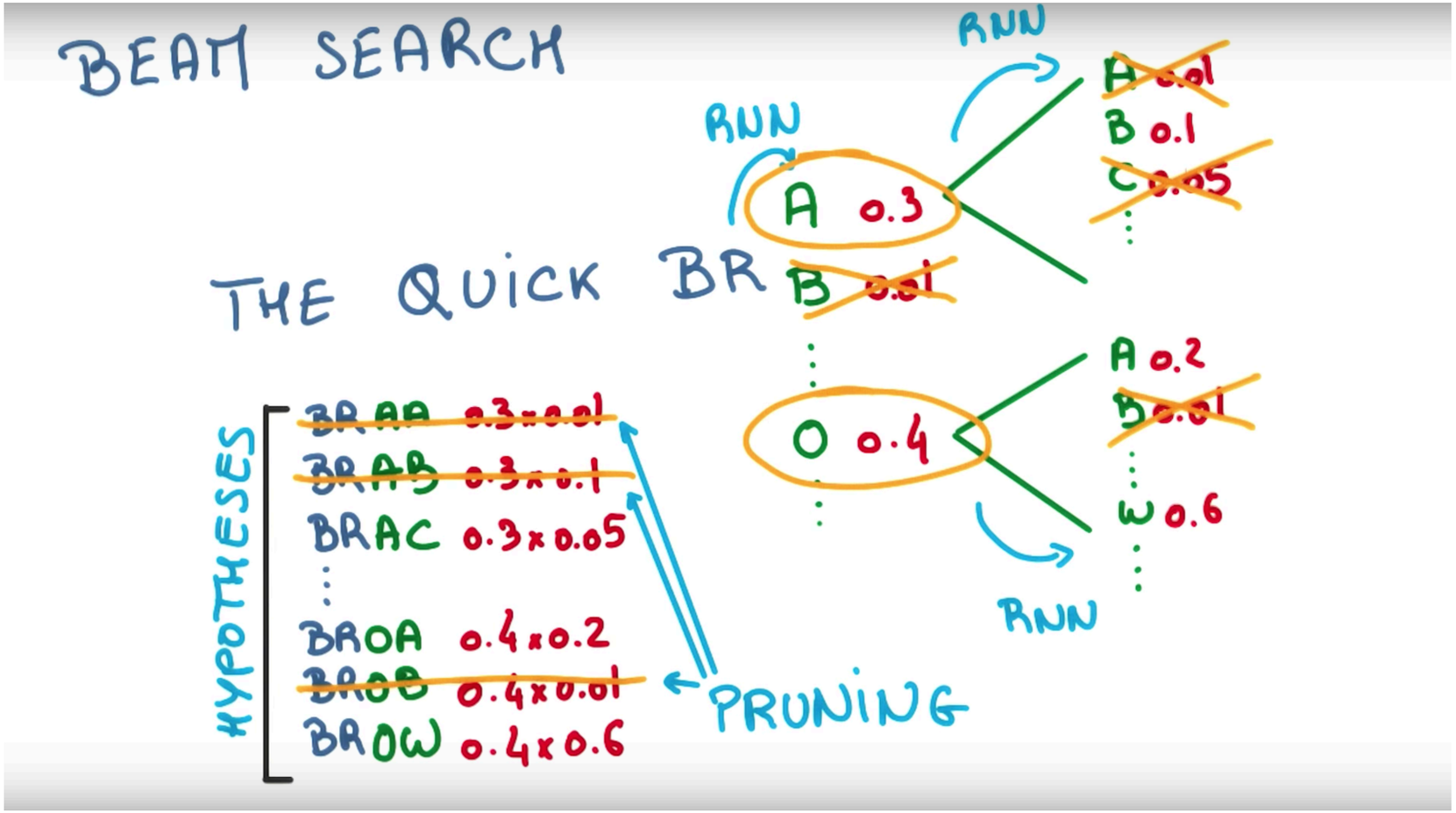
  $$s_{t'} = W^{\top} x_{t'}, \text{ where } W \in \mathbb{R}^{|V| \times d}$$

  (3)Recursively read words

  $$h_t = f(h_{t-1}, s_t), \text{ for } t = 1, \ldots, T$$

Machine Translation 사례

Beam Search



2. Attention Mechanism

Intuition

어떻게 현상을 보고 묘사 할 수 있을까?

- 현상을 묘사하기 위해선 전체 시야 전체가 아닌 일부에 집중을 해야 할 것이다.

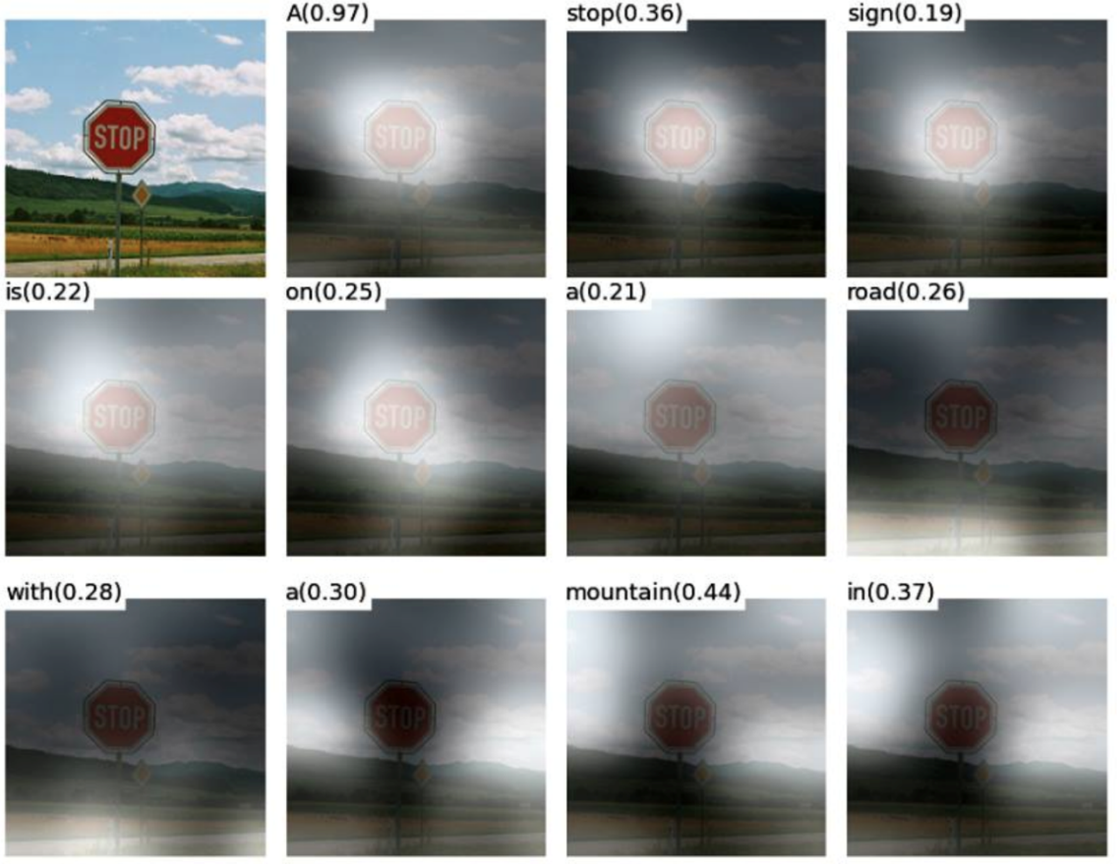
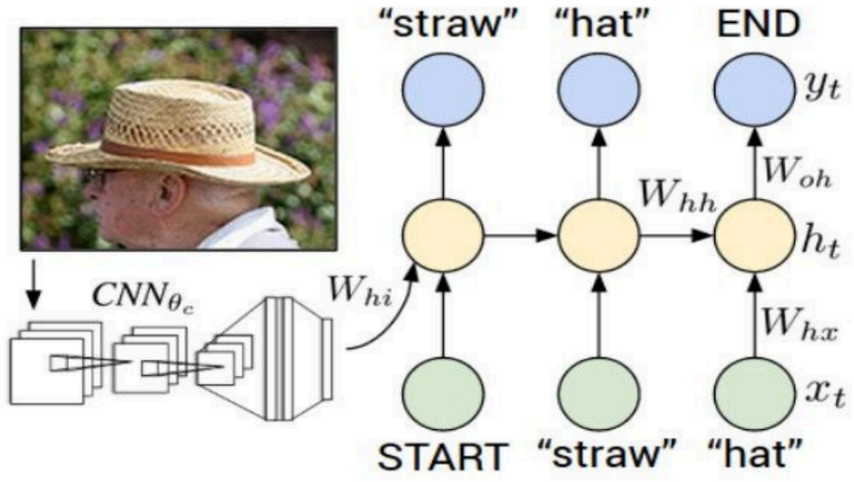


Image Captioning

- 이미지 정보를 바탕으로 Decoder RNN을 통해 이미지에 대한 내용을 설명하고자 하는 task이다.



Explain Images with Multimodal Recurrent Neural Networks, Mao et al.
Deep Visual-Semantic Alignments for Generating Image Descriptions, Karpathy and Fei-Fei
Show and Tell: A Neural Image Caption Generator, Vinyals et al.
Long-term Recurrent Convolutional Networks for Visual Recognition and Description, Donahue et al.
Learning a Recurrent Visual Representation for Image Caption Generation, Chen and Zitnick

Image Captioning

- Encoder CNN에서 이미지에 대한 summarize된 정보를 Decoder RNN에 전달한다고 볼 수 있다.

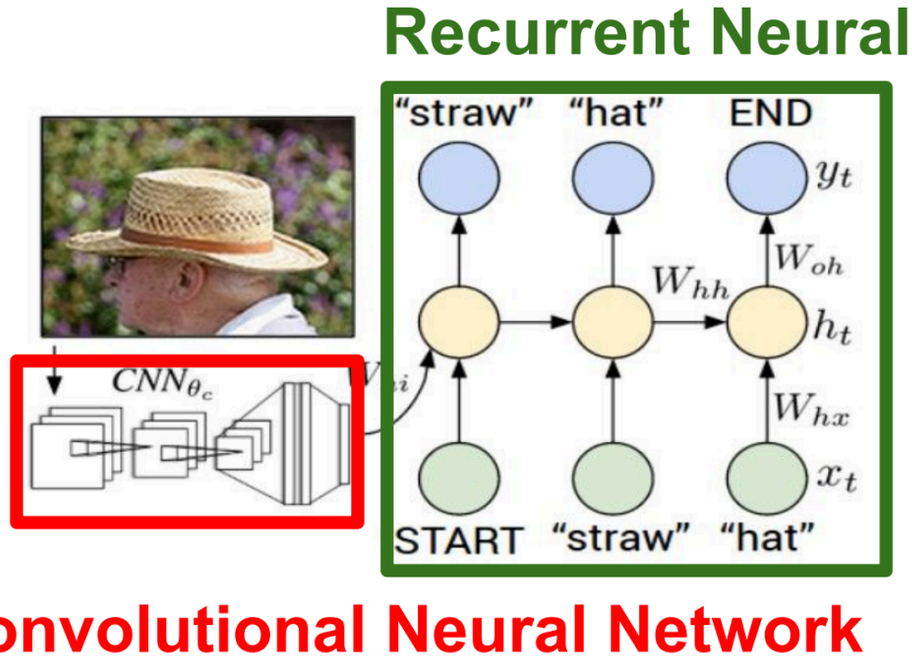
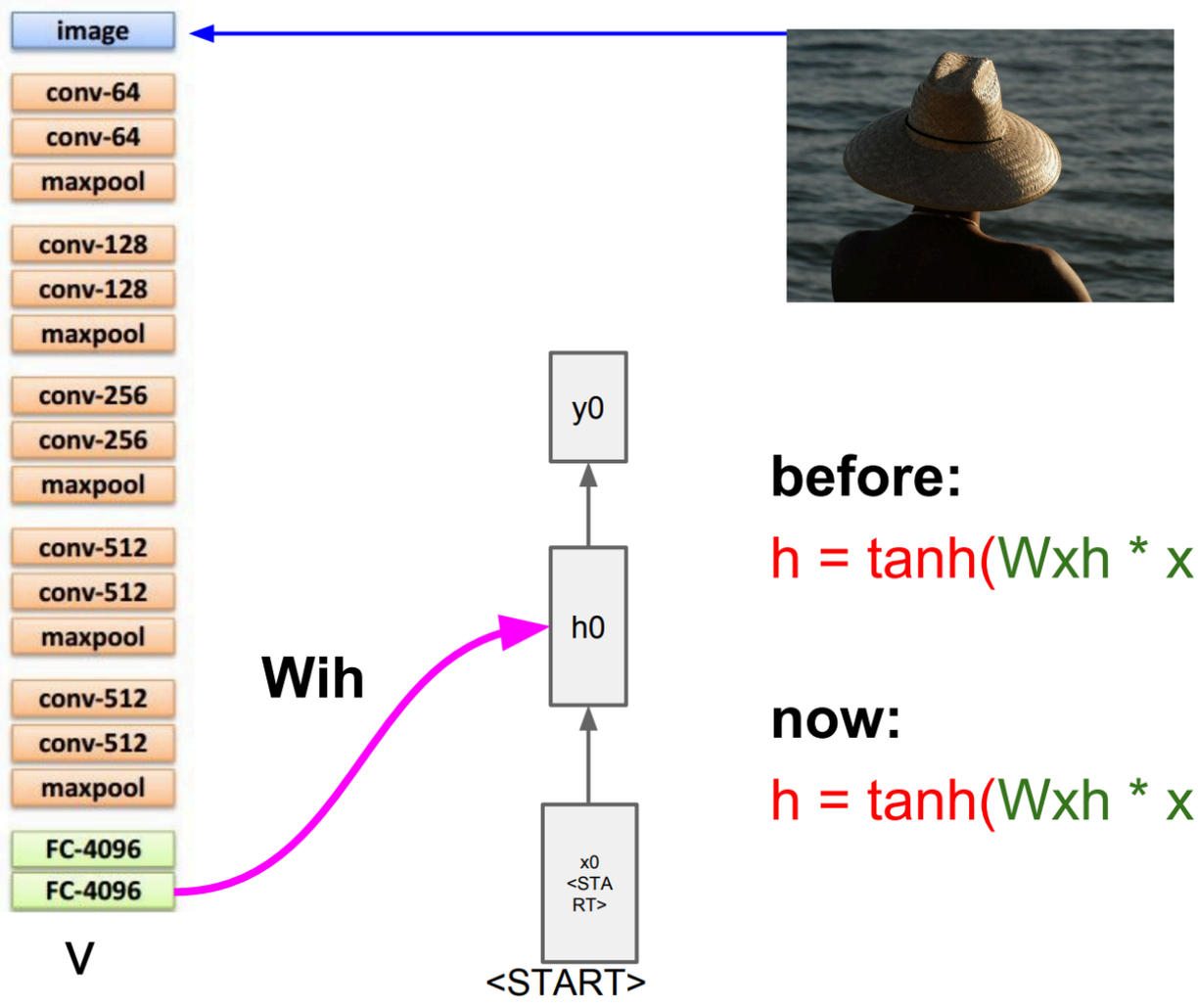


Image Captioning

Image Captioning



test image

before:

h = tanh(Wxh * x + Whh * h)

now:

h = tanh(Wxh * x + Whh * h + **Wih * v**)

Image Captioning

Image Captioning

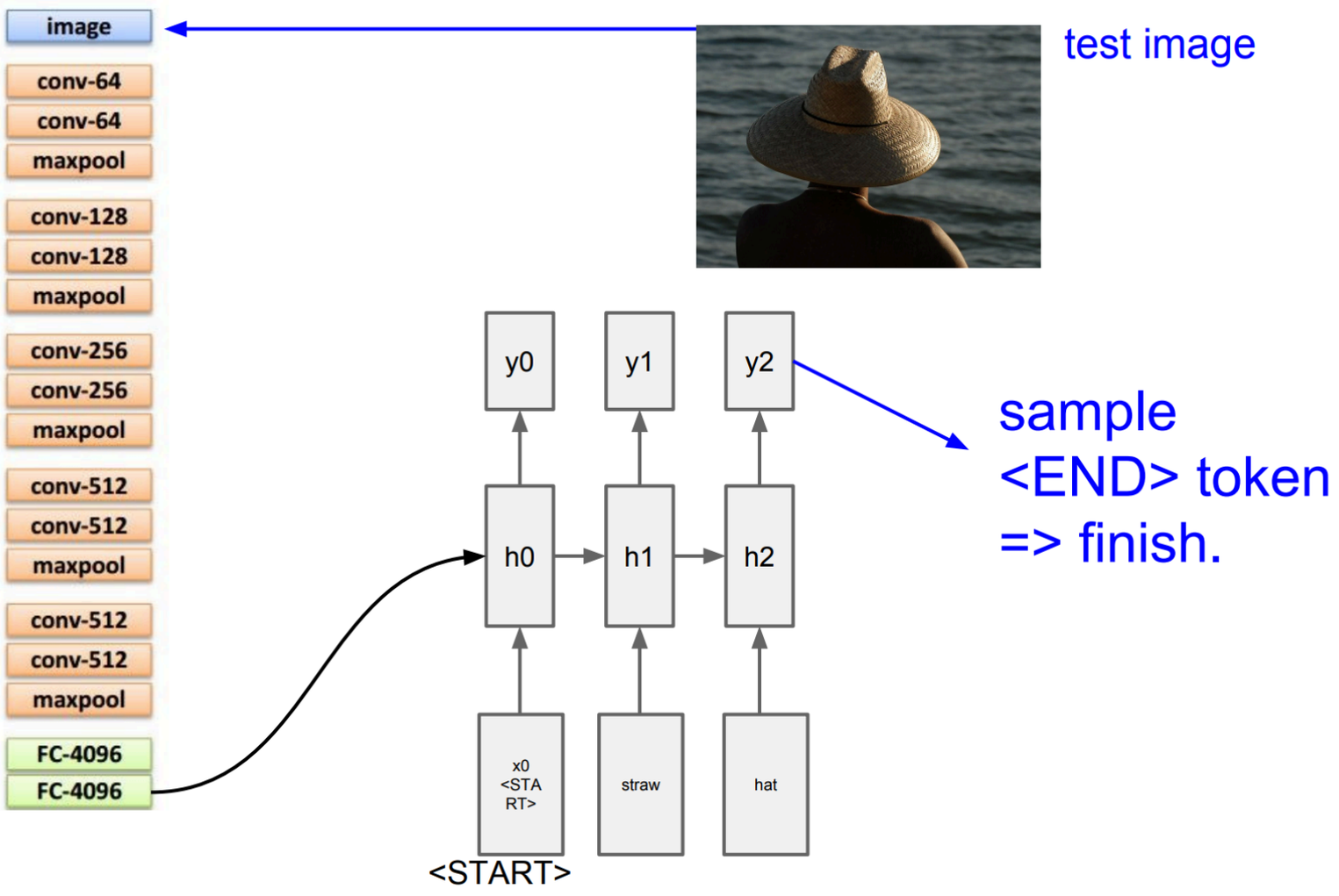


Image Captioning

Image Captioning

Image Captioning: Example Results

Captions generated using neuraltalk2
All images are CC0 Public domain:
cat suitcase, cat tree, dog, bear,
surfers, tennis, giraffe, motorcycle

A cat sitting on a suitcase on the floor

A cat is sitting on a tree branch

A dog is running in the grass with a frisbee

A white teddy bear sitting in the grass

Two people walking on the beach with surfboards

A tennis player in action on the court

Two giraffes standing in a grassy field

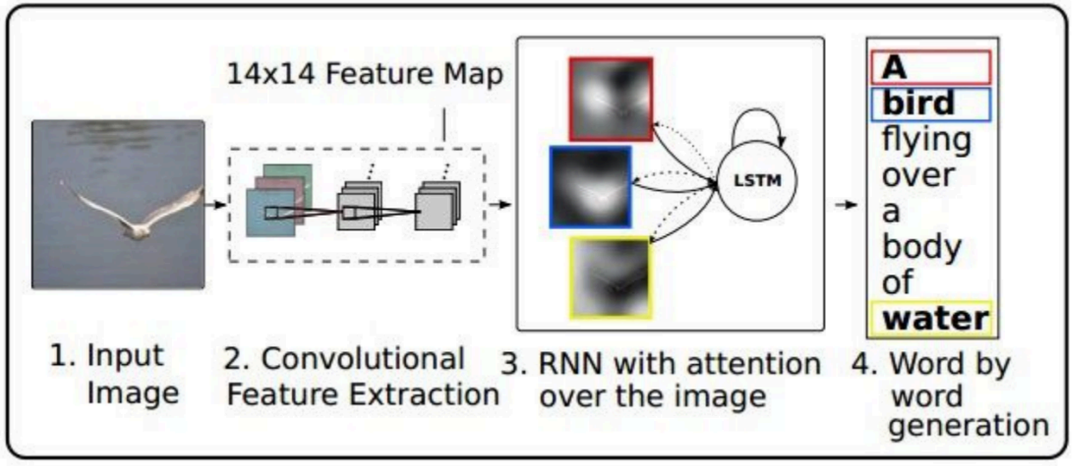
A man riding a dirt bike on a dirt track

Image Captioning with Attention

Image Captioning with Attention Mechanism

- 매 time step 마다 단어를 생성할 때 이미지에 어떠한 영역에 대해 더 가중치를 주어야 하는 지 Attention Weight을 주도록 한다.

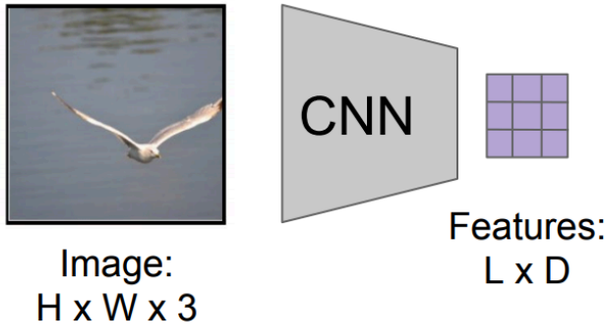
RNN attends spatially to different parts of images while generating each word of the sentence:



Show Attend and Tell, Xu et al., 2015

Image Captioning with Attention

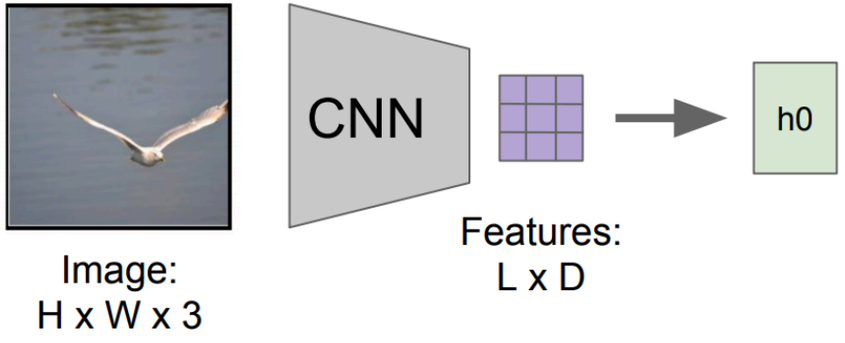
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

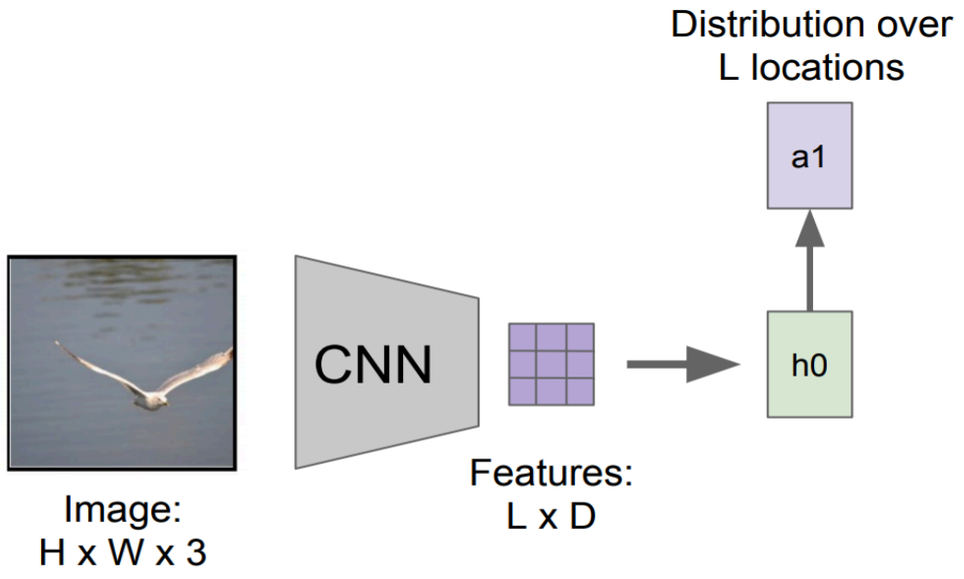
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

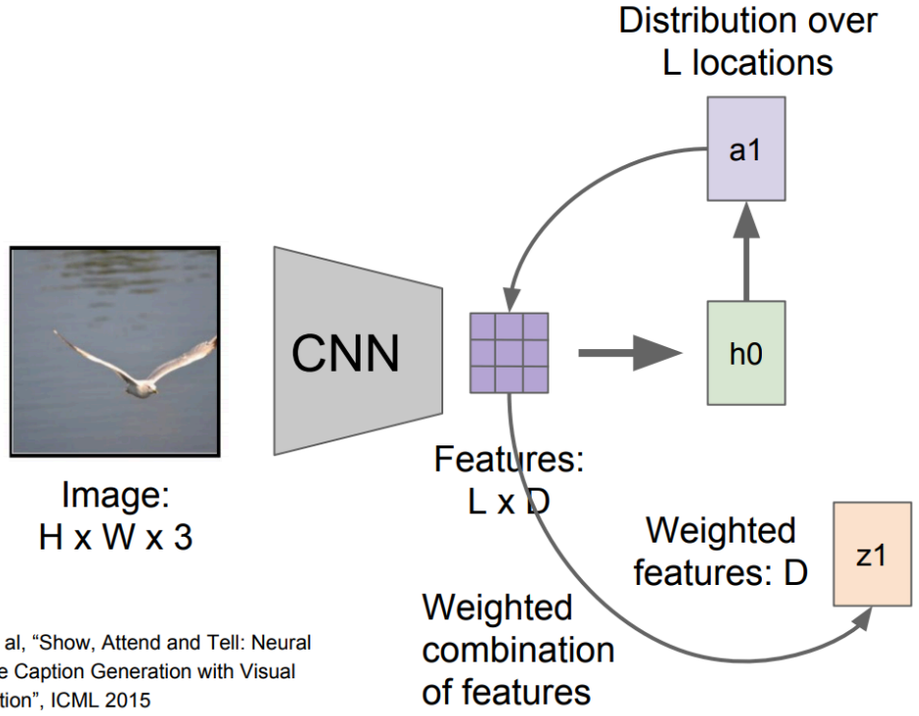
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

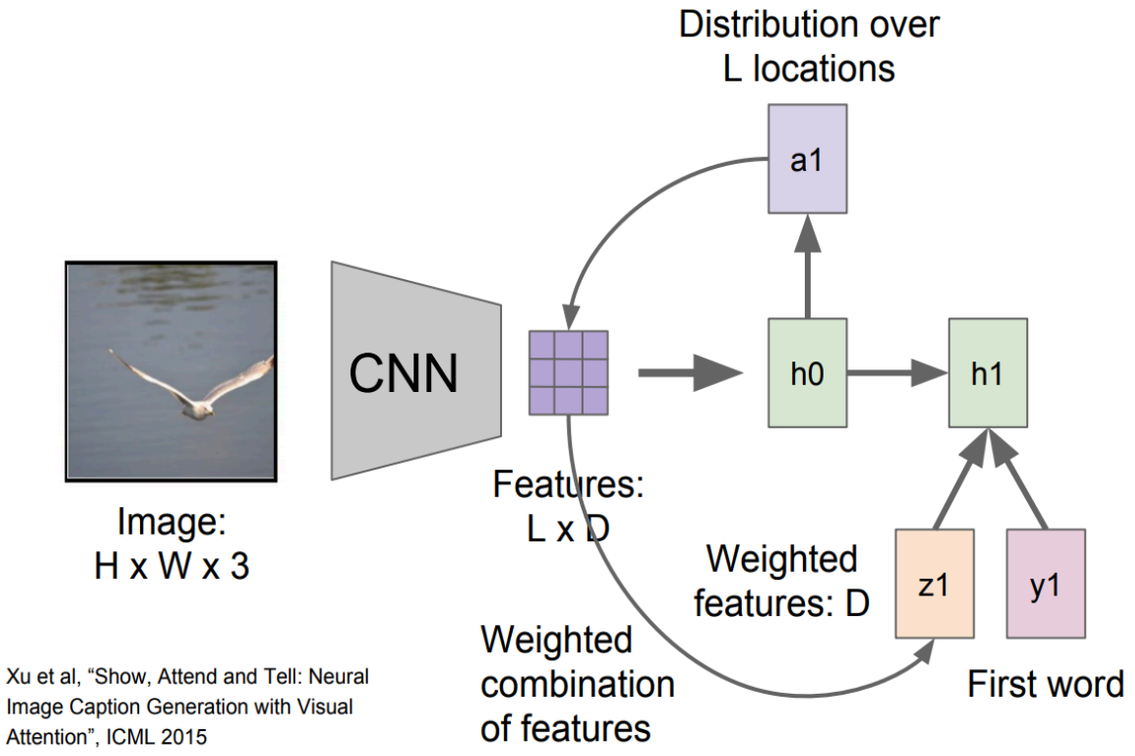
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

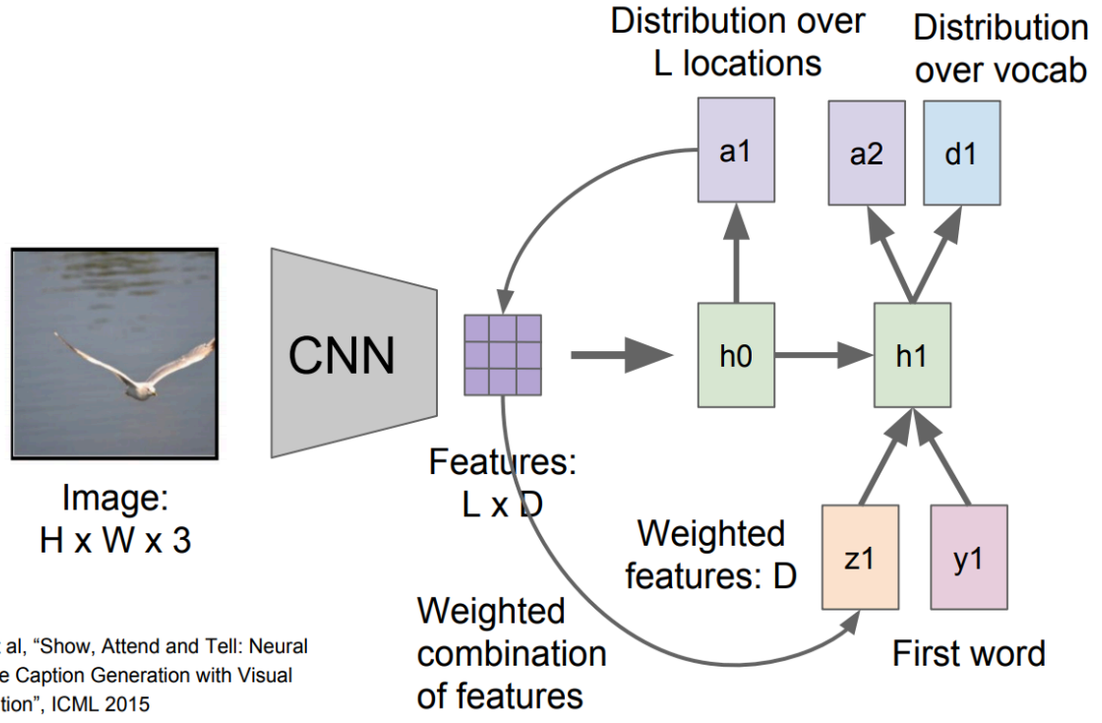
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

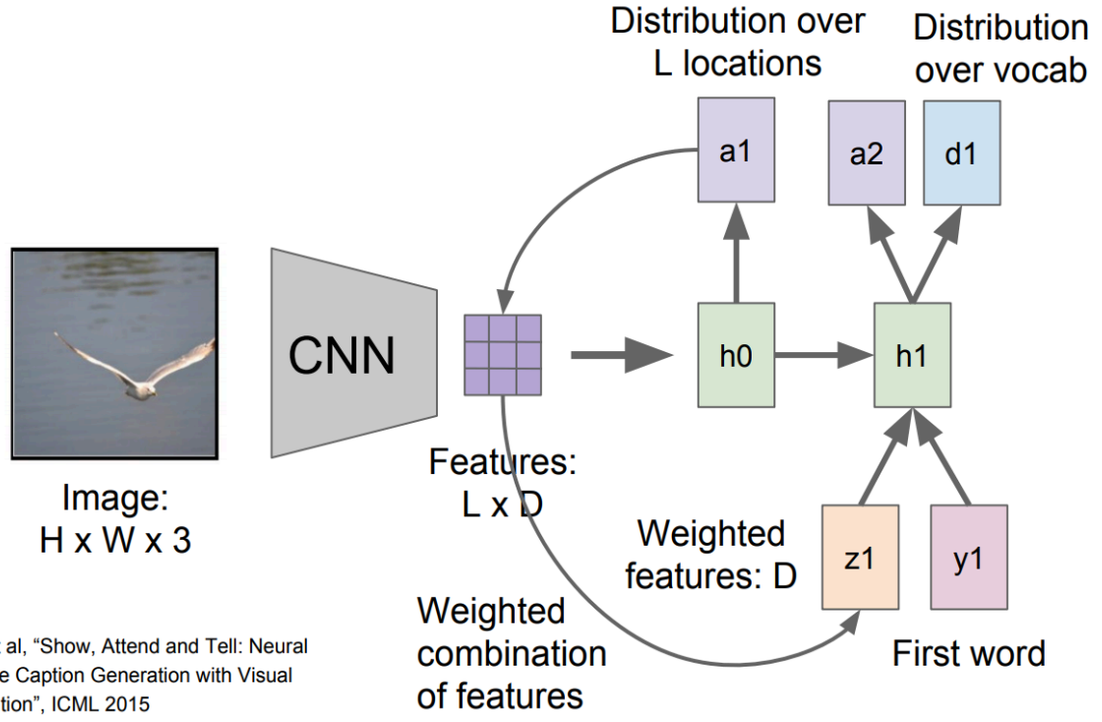
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

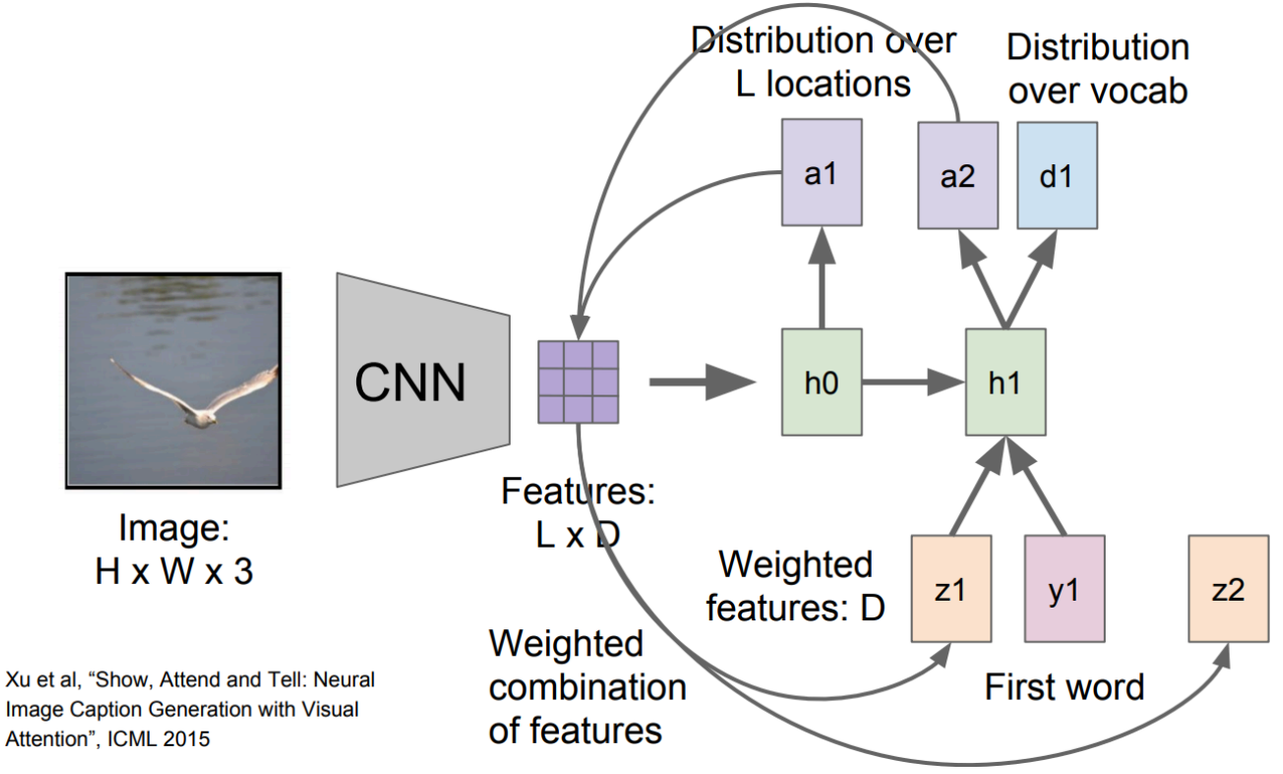
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

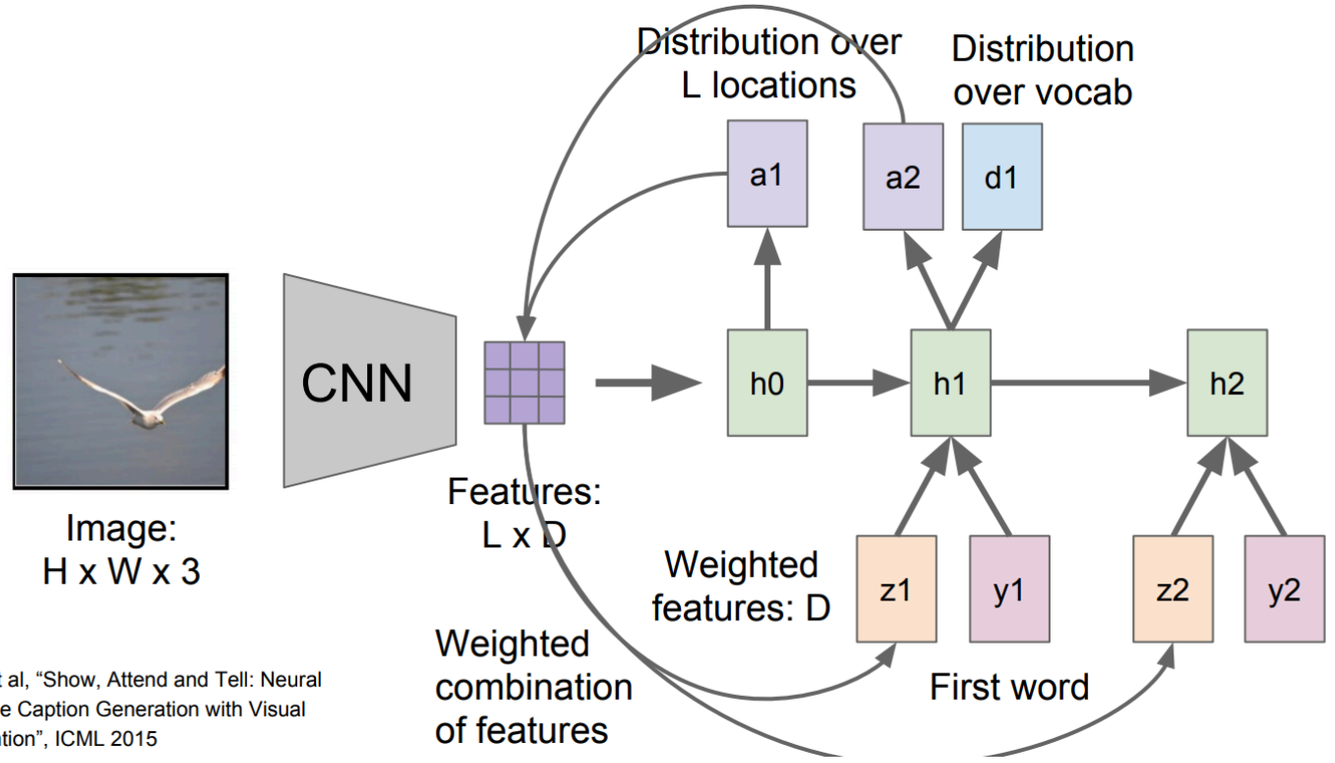
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

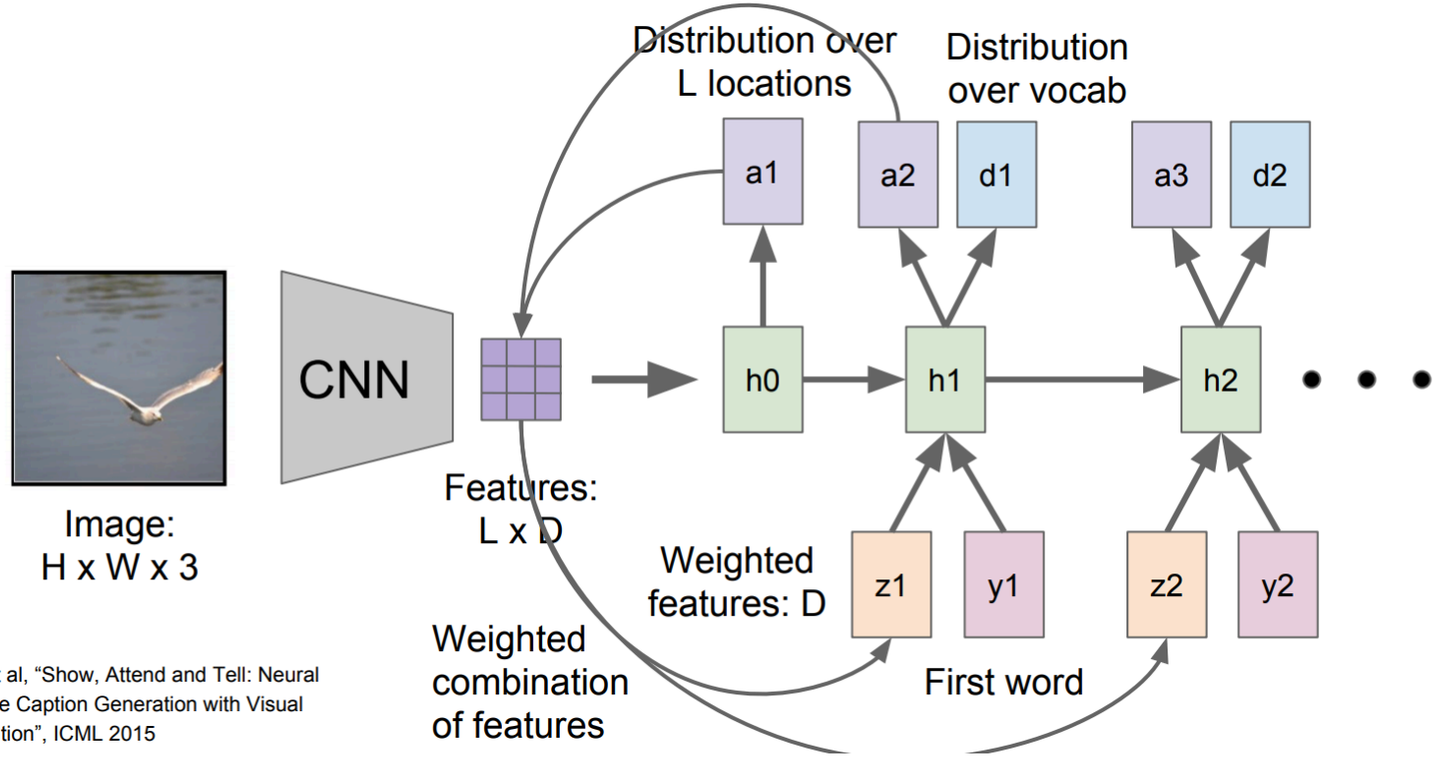
어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

어떻게 Attention이 작동되는가?



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

Image Captioning with Attention

Decoder에서 단어가 생성될 때마다 Attention은 어떻게 작동 될까?

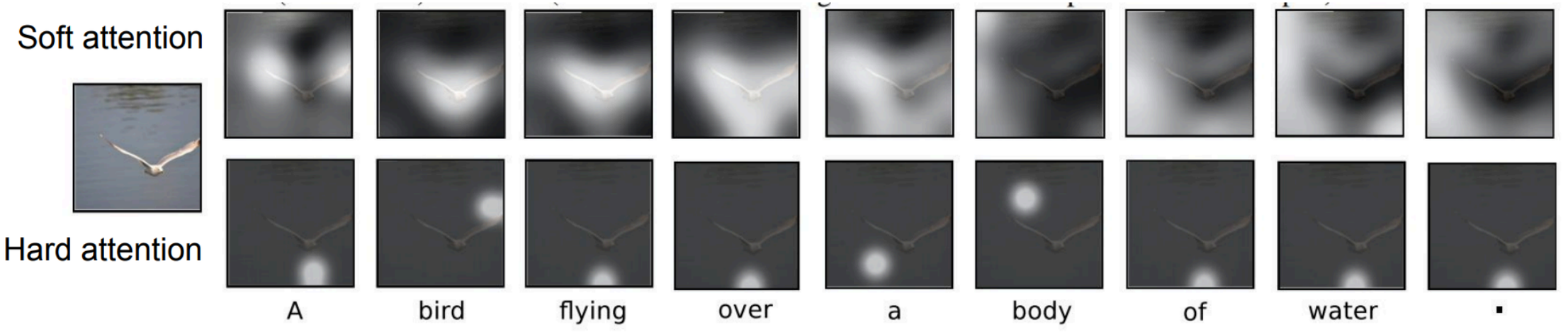
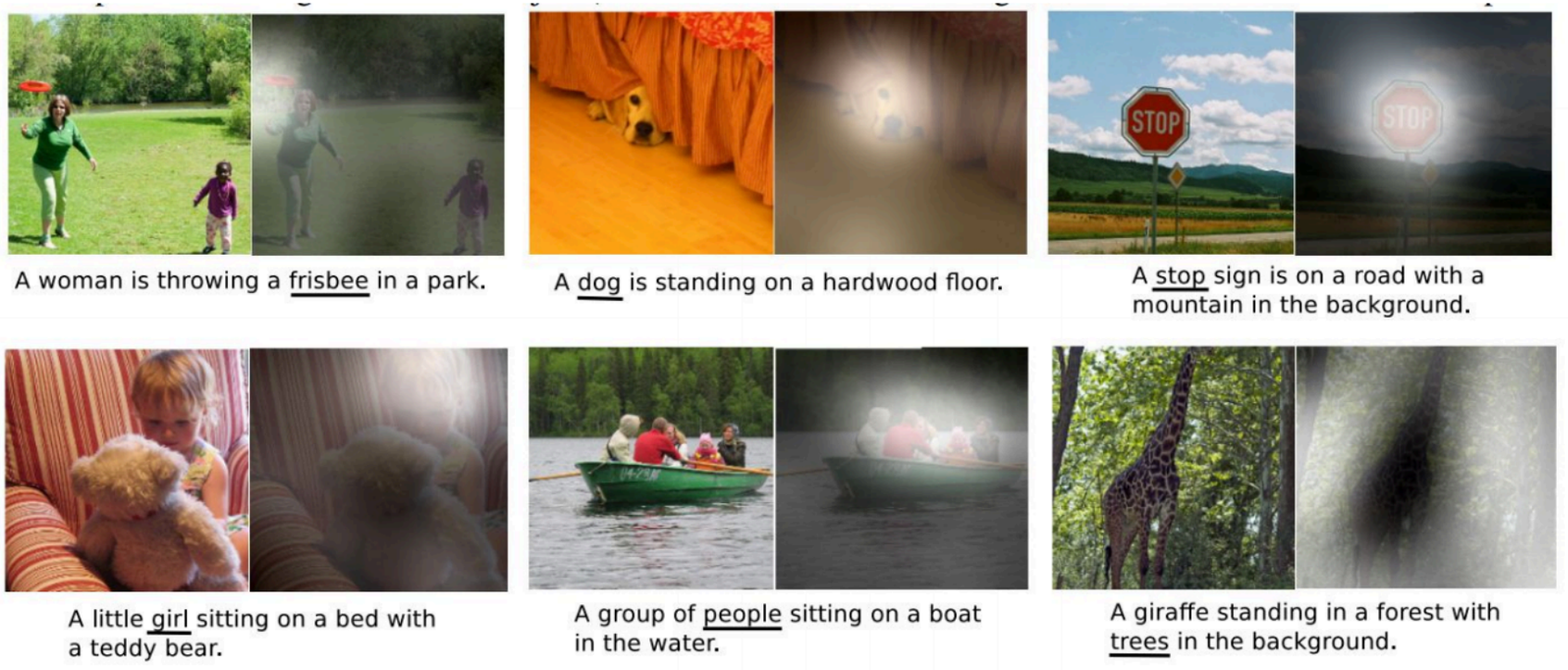


Image Captioning with Attention

Image Captioning with Attention Examples



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

Seq2Seq with Attention

Sequence to Sequence with Attention

- Image Captioning과는 달리 Encoder에 입력한 Input sequence의 각 위치에 대한 Attention Weight을 준다.
- Decoder의 hidden state 정보와 encoder에서 나온 각 time step에 대한 hidden state의 관계를 값을 통해 (i.e 유사도) 표현하여 어느 Encoder 값이 가장 관계가 있는지 표현하고, 이들을 하나의 Attention 정보로 Weighted Sum을 해준다.

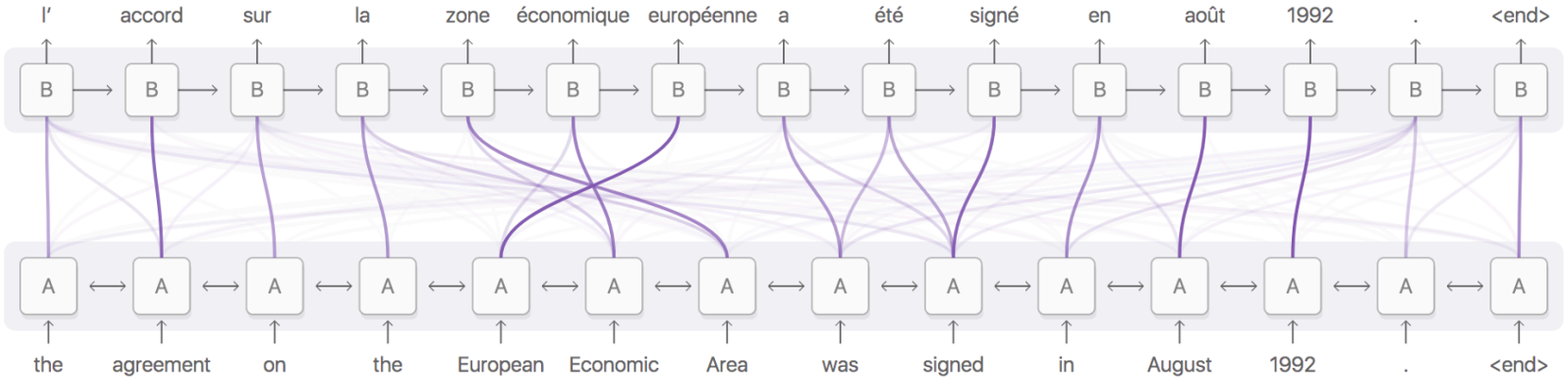
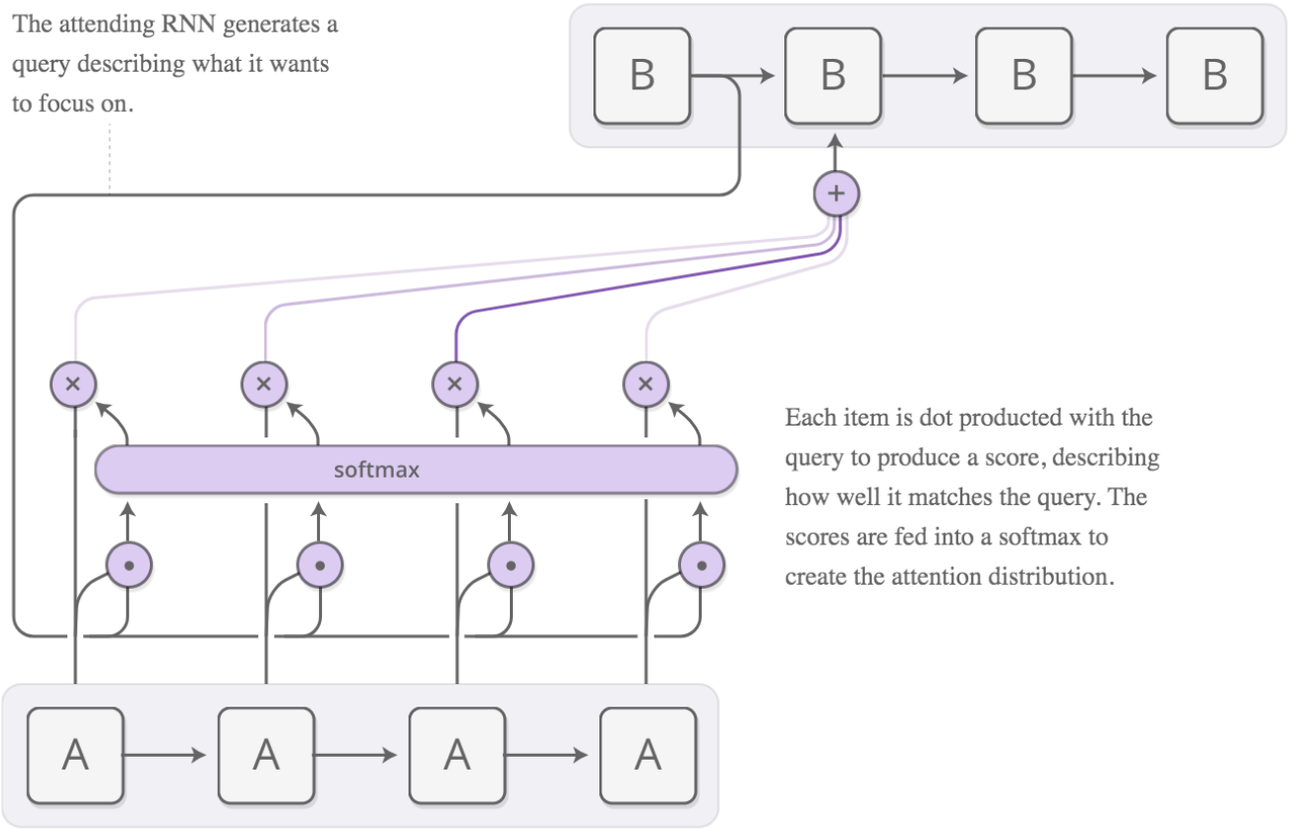


Diagram derived from Fig. 3 of Bahdanau, *et al.* 2014

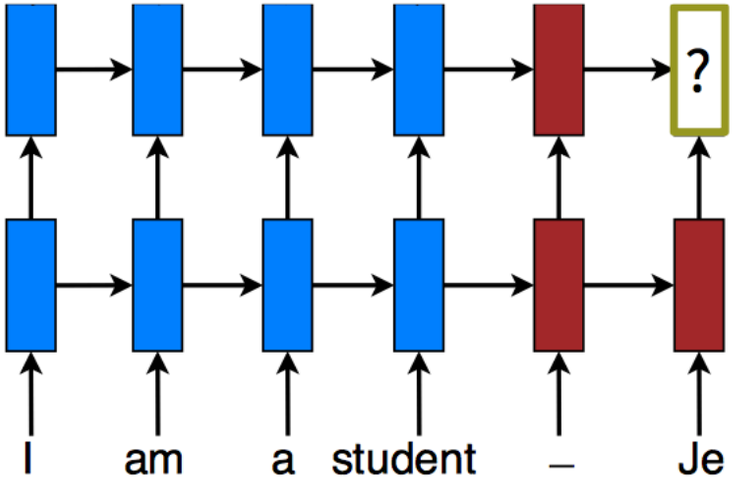
Seq2Seq with Attention

Sequence to Sequence with Attention



Seq2Seq with Attention

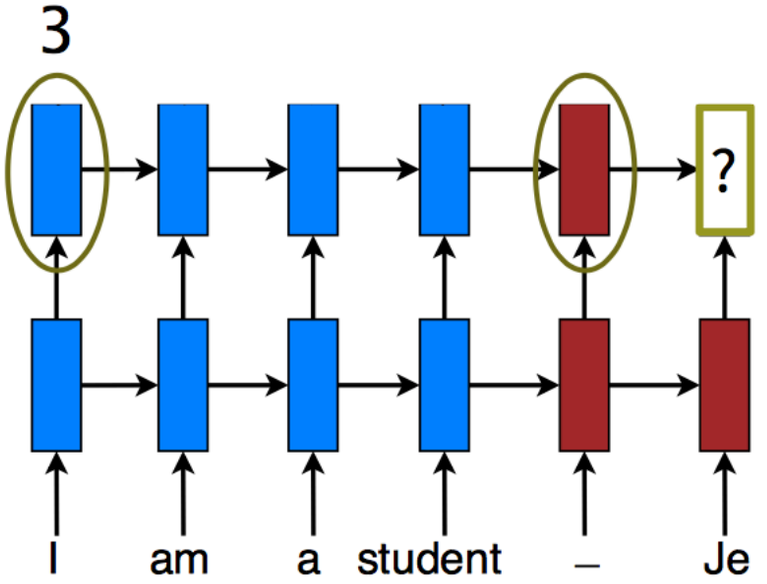
Seq2Seq Attention 작동 과정



Seq2Seq with Attention

Seq2Seq Attention 작동 과정

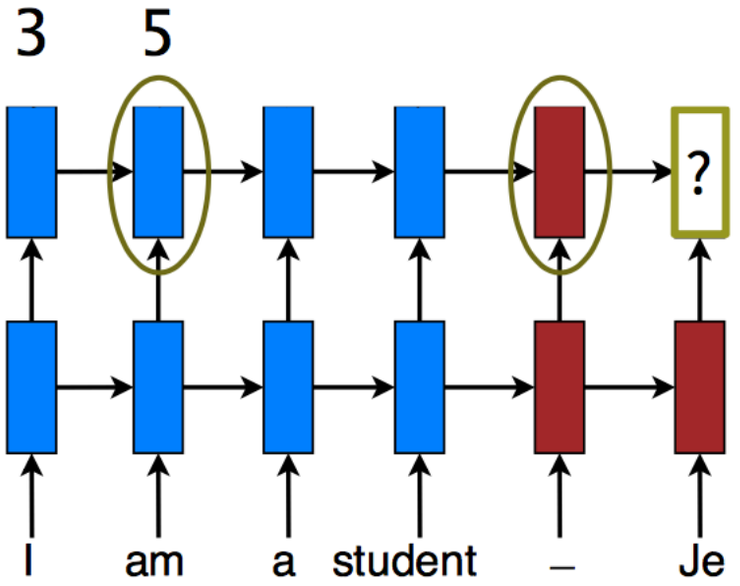
$$\text{score}(\boldsymbol{h}_{t-1}, \bar{\boldsymbol{h}}_s)$$



Seq2Seq with Attention

Seq2Seq Attention 작동 과정

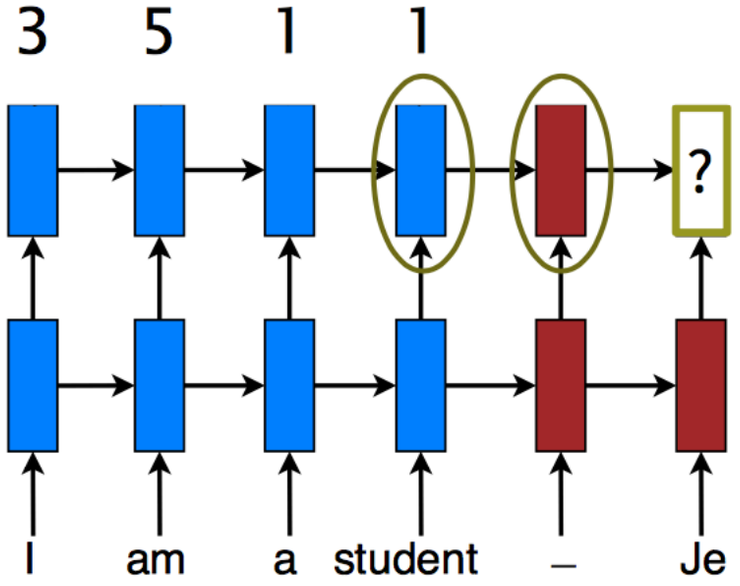
$$\text{score}(\boldsymbol{h}_{t-1}, \bar{\boldsymbol{h}}_s)$$



Seq2Seq with Attention

Seq2Seq Attention 작동 과정

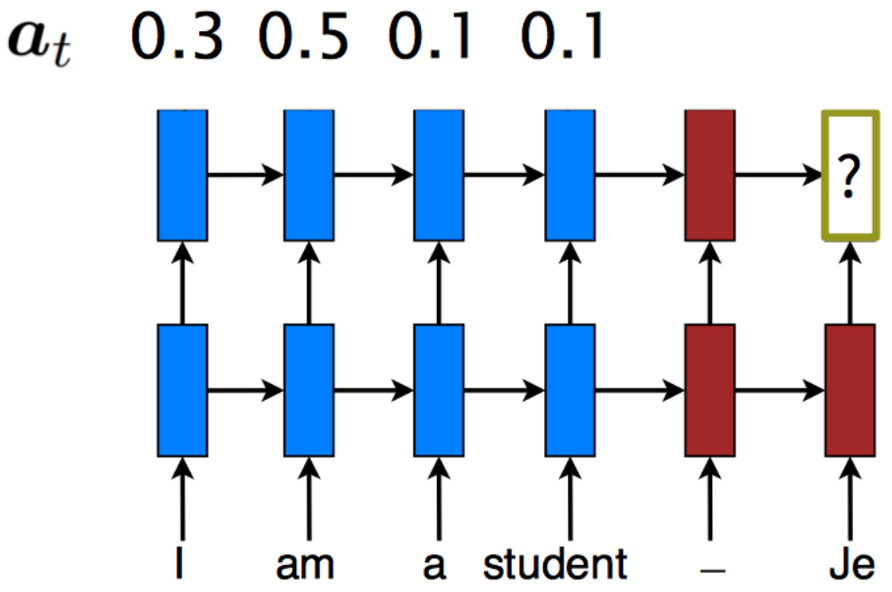
$$\text{score}(\boldsymbol{h}_{t-1}, \bar{\boldsymbol{h}}_s)$$



Seq2Seq with Attention

Seq2Seq Attention 작동 과정

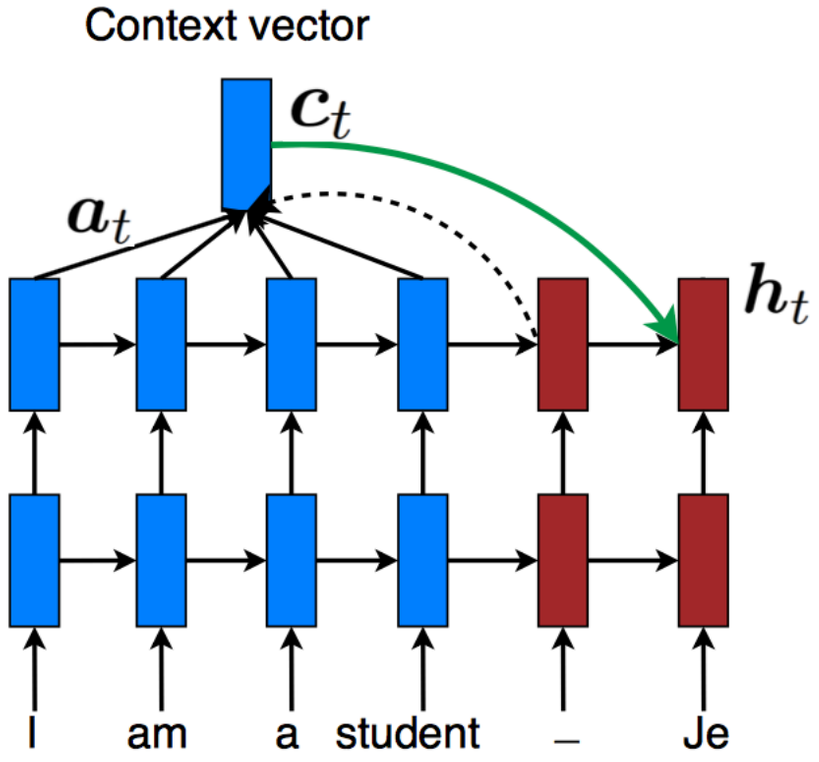
$$\boldsymbol{a}_t(s) = \frac{\exp(\text{score}(s))}{\sum_{s'} \exp(\text{score}(s'))}$$



Seq2Seq with Attention

Seq2Seq Attention 작동 과정

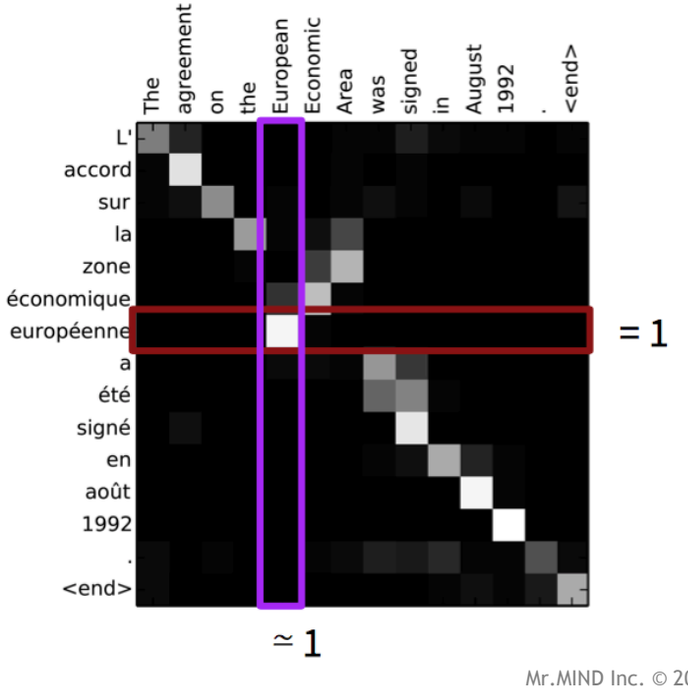
$$c_t = \sum_s a_t(s)\bar{h}_s$$



Seq2Seq with Attention

Seq2Seq with Attention의 효과

- 불필요한 input 정보에 대해서 decode sequence에서 반영할 필요가 없다.
- Decoder가 매 time step에 output을 generate하는 과정에서 encoder에 대한 정보 순서를 유연하게 받아들일 수 있다.
  - ➢ 앞서 설명한 Traditional Machine Translator에서 Alignment의 문제를 어느정도 해결 할 수 있음을 보인다.



References

- Stanford CS224n Lectures
  http://web.stanford.edu/class/cs224n/syllabus.html
- Stanford CS231n Lectures
  http://cs231n.stanford.edu/2016/syllabus.html
- "Attention and Augmented Recurrent Neural Networks" by Colah's Blog
  https://distill.pub/2016/augmented-rnns/
- Deep Learning Summer School, Montreal 2016 - VideoLectures.NET
  http://videolectures.net/deeplearning2016_montreal/
- From Attention to Memory and towards Longer-Term Dependencies by Yoshua Bengio
  http://www.iro.umontreal.ca/~bengioy/talks/RAM_NIPS2015_workshop_12Dec2015.pdf

감사합니다