

## 5장. 딥러닝 – VI

## **6. 종단간 학습 모델**

**6.1 인코더-디코더 망**

**6.2 인코더-디코더 망의 구성**

**6.3 주목 메커니즘**

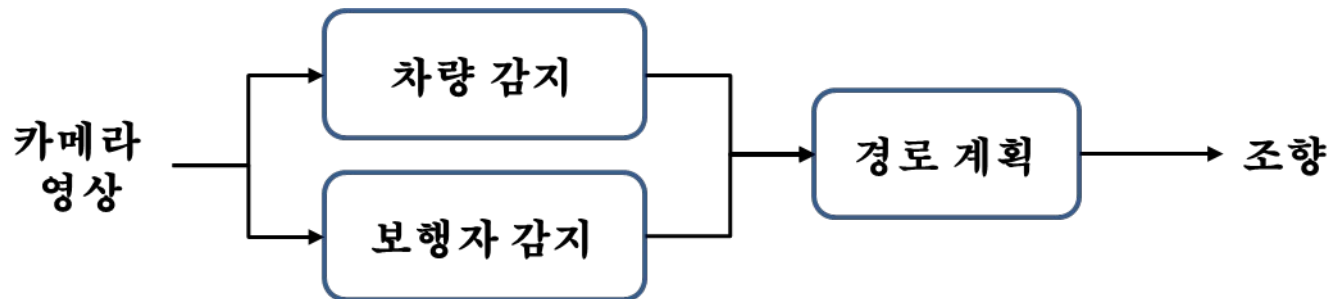
## 6. 종단간 학습 모델

### ❖ 종단간 학습(End-to-End Learning)

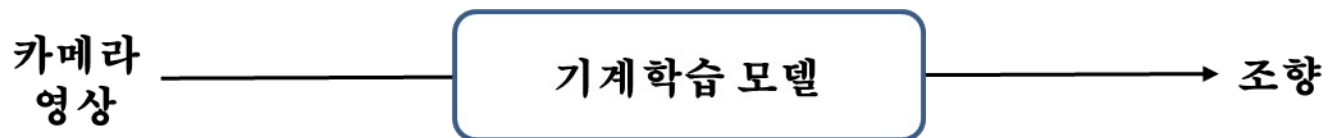
- 입력과 출력 정보로 구성된 학습 데이터를 거의 원본 그대로 사용하여, 개발자의 중간 개입없이 입력으로부터 기대하는 출력을 만들어내는 학습
- 모든 파라미터를 보통 **경사 하강법** 기반의 방법으로 학습

#### ▪ ex. 자율주행 자동차

##### • 전통적인 방법(Conventional method)



##### • 종단간 학습



# 종단간 학습 모델

## ❖ 종단간 학습 모델의 적용

### ▪ 기계 번역

- 어떤 언어의 문장을 다른 언어의 문장으로 변환

### ▪ 영상 주석 달기

- 주어진 사진을 설명하는 문장 생성

### ▪ 동영상 묘사하기

- 주어진 동영상을 설명하는 문장 생성

### ▪ 음성 인식

- 음성을 문장으로 변환

# 6.1 인코더-디코더 망

## ❖ 인코더-디코더 망 (Encoder-Decoder Network)

- 입력 데이터의 구조와 출력 데이터의 구조 간의 대응 관계를 찾는 문제에 적용
- 인코더와 디코더로 구성
  - 인코더
    - 구조적인 정보가 포함된 입력을 일정한 차원의 특징 벡터로 압축하여 표현
  - 디코더
    - 특징 벡터를 출력 공간의 데이터로 변환

# 인코더-디코더 망

## ❖ 인코더-디코더 망의 형태

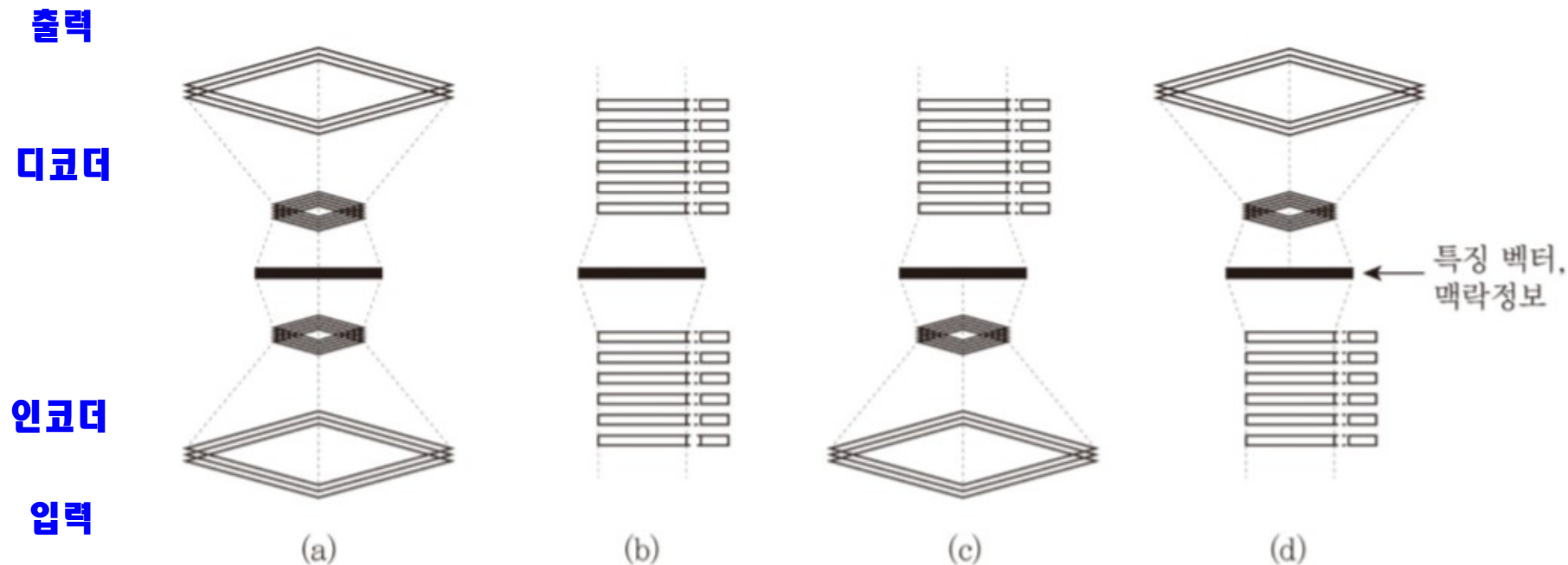


그림 5.72 인코더-디코더 망

(a) 오토인코더 (b) 서열대서열 변환 (c) 영상 주석달기 (d) 영상 합성 [출처: <http://gabgoh.github.io/ThoughtVectors/>]

## 6.2 인코더-디코더 망의 구조

### ❖ 단순 인코더-디코더 망(Simple Encoder-Decoder Network)

#### ▪ 인코더 $f_{enc}$

- 입력 데이터  $x$ 를 읽어들이어 일정한 차원의 실수 벡터  $c$ 을 생성
- 벡터  $c$  : 입력 데이터에 대한 요약으로 맥락 정보

$$c = f_{enc}(x)$$

- RNN, CNN 등으로 구현

#### ▪ 디코더 $f_{dec}$

- 입력에 대한 맥락정보  $c$ 를 사용하여 출력  $y$ 를 생성
- 조건부 확률  $p(y|x)$  계산

$$p(Y|x) = f_{dec}(x, c)$$

- RNN, RBM 등으로 구현

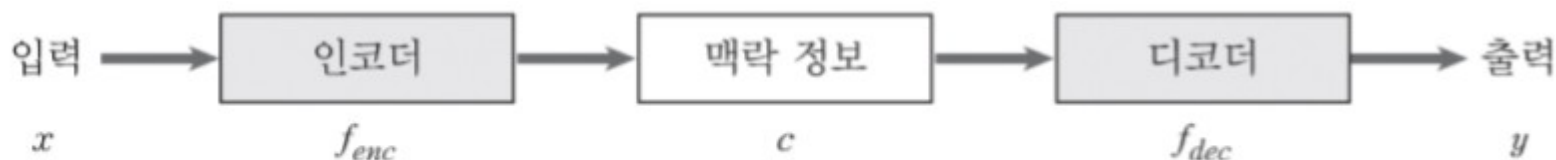


그림 5.73 인코더-디코더 망.

# 인코더-디코더 망의 구조

## ❖ 기계번역 인코더-디코더 망

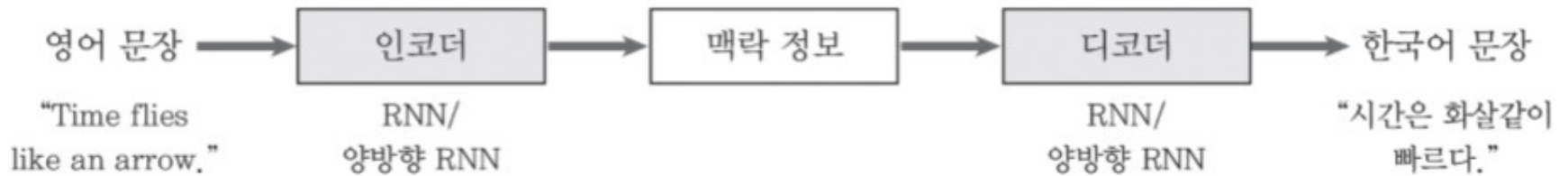
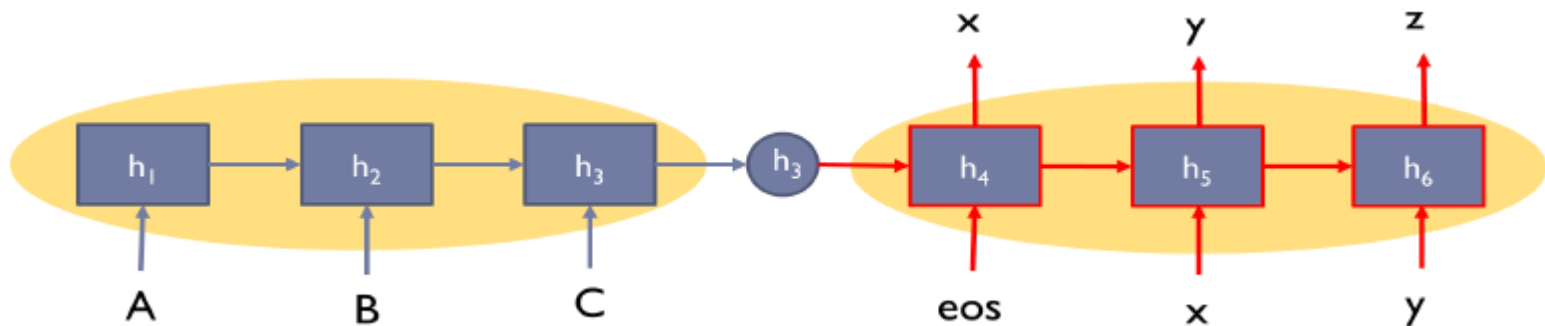


그림 5.74 기계 번역 인코더-디코더 망.

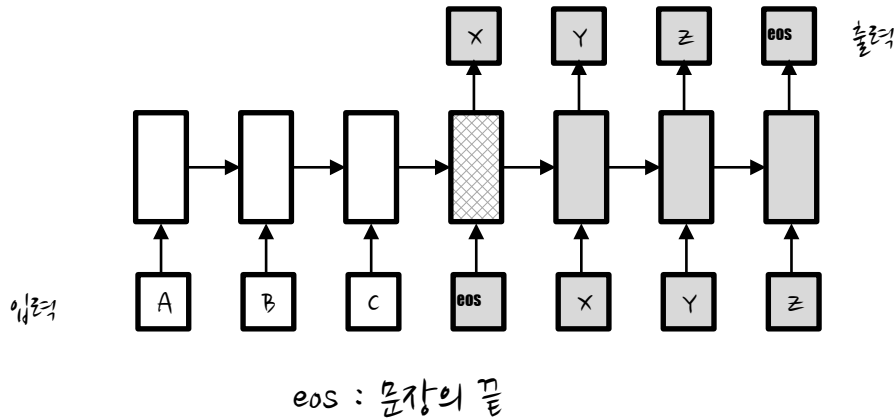
- 어떤 언어의 문장 하나가 수치 벡터로 변환된 다음, 해당 수치 벡터로부터 다른 언어의 같은 의미의 문장 생성



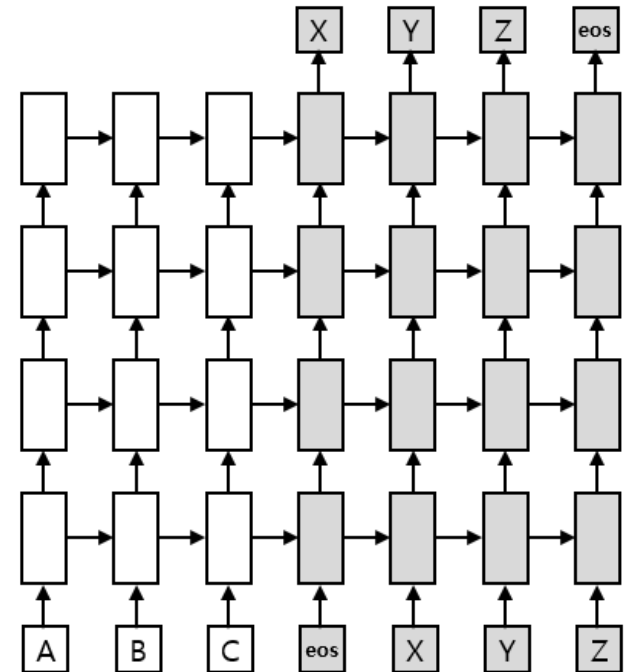


# 인코더-디코더 망의 구조

## ❖ 단순한 RNN 기반의 기계 번역 망의 형태



## ❖ 딥 재귀신경망 기반 기계 번역 망의 형태



# 인코더-디코더 망의 구조

## ❖ 영상 주석 달기(Image Captioning)

- 인코더를 통해 입력 영상의 맥락 정보를 추출하고, 맥락 정보로부터 설명 문장 생성

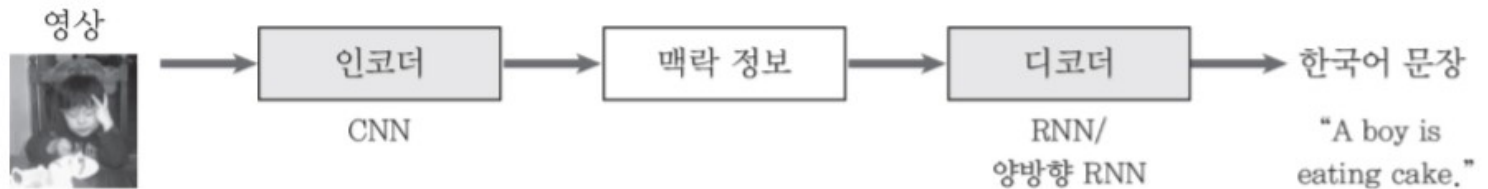
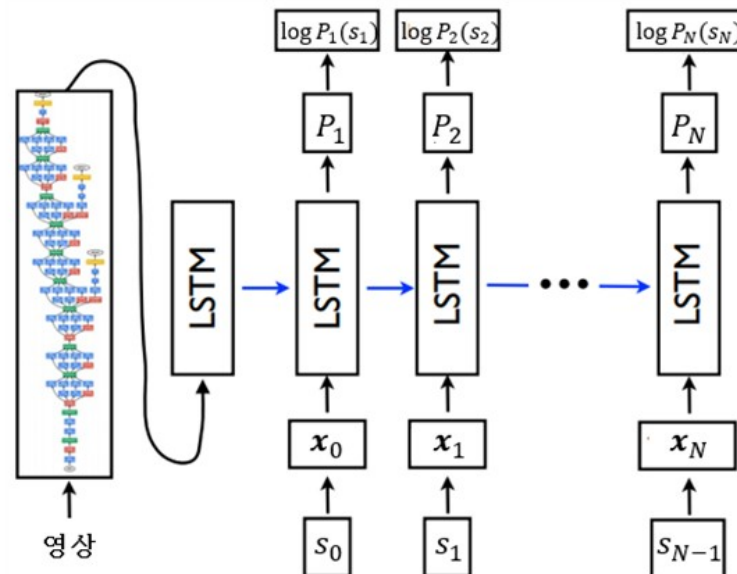


그림 5.75 영상 주석달기 인코더-디코더 망



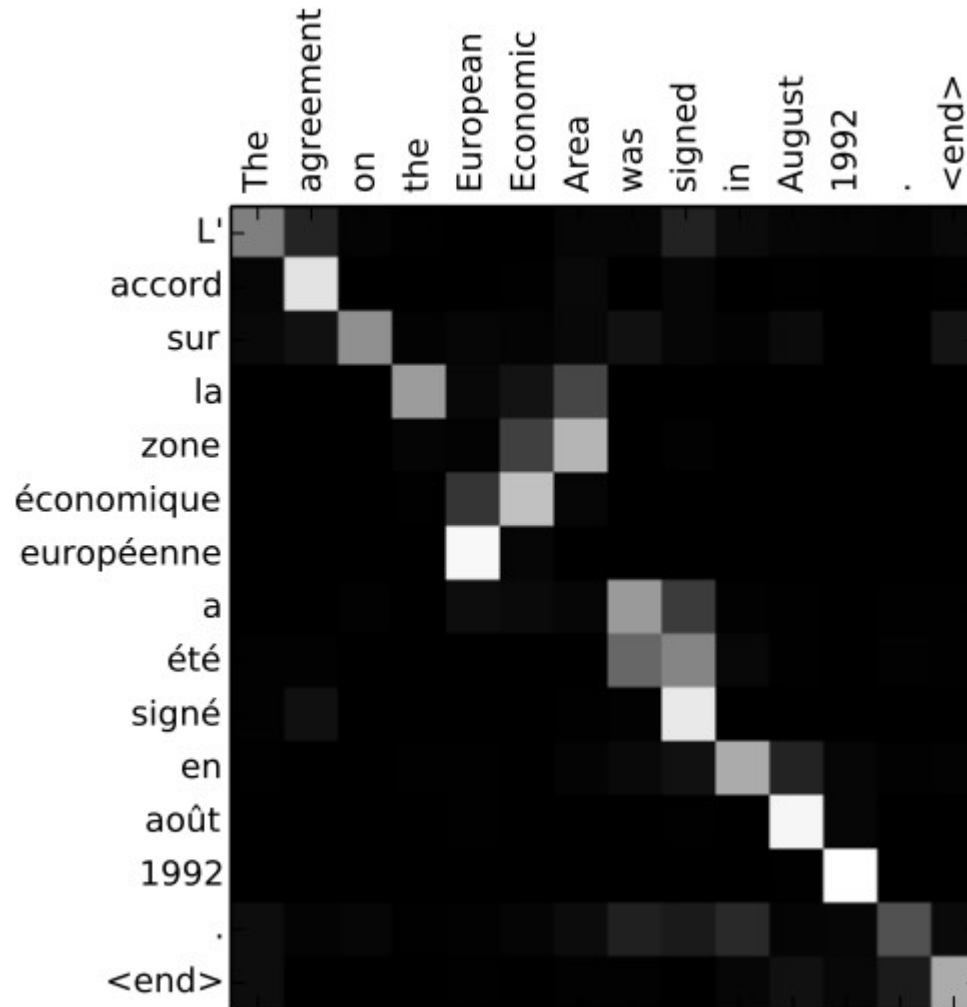
## 6.3 주 목 메 커 니 즘

### ❖ 단순 인코더-디코더 망의 제약

- 입력의 길이가 길어지면 일정한 크기의 벡터로는 입력의 모든 정보를 표현하기 곤란
  - 재귀 신경 망의 은닉층 노드 개수를 증가
  - 각 과거 시점의 은닉층 상태 정보를 디코더에서 함께 사용
  - 특정 시점을 출력을 결정할 때, 인코더의 모든 시점에 대한 은닉층 상태 정보를 필요로 하지는 않음
- 특정 정보에 대한 주목 필요



# 주목 메커니즘



# 주목 메커니즘

## ❖ 예. 영어-프랑스어 번역에서 관련된 단어

- 번역시 모든 단어의 정보를 고려할 필요가 없음

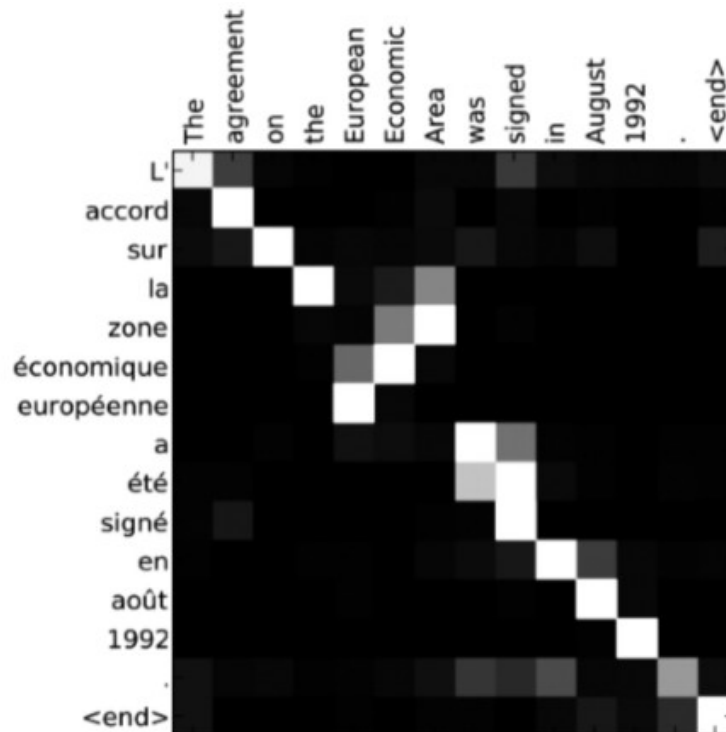


그림 5.76 기계번역에서 영어 단어와 프랑스어 단어의 관련도 [출처 Bahdanau, Cho, Bengio, 2014]

밝을수록 관련도가 큰 것을 의미한다.

# 주목 메커니즘

## ❖ 주목 모델

- 인코더의 시간적, 공간적, 또는 시공간적인 상태들을 결합하여 주목해야 할 부분에 대한 맥락 정보를 만드는 역할
- 인코더와 디코더의 사이에 위치

## ▪ 선택적 주목



- 인코더가 입력 전체에 대한 하나의 맥락정보를 만들어 디코더에 전달하는 대신에, 디코더가 출력을 산출하는 때 시점마다 맥락정보를 계산하여 사용

# 주목 메커니즘

## ❖ 주목 메커니즘을 포함한 인코더-디코더 망

### ▪ 인코더

- 요약 정보 생성

### ▪ 주목 모델

- 해당 순간의 요약 정보를 결합에서 맥락 정보 생성

### ▪ 디코더

- 맥락 정보를 사용하여 출력 생성

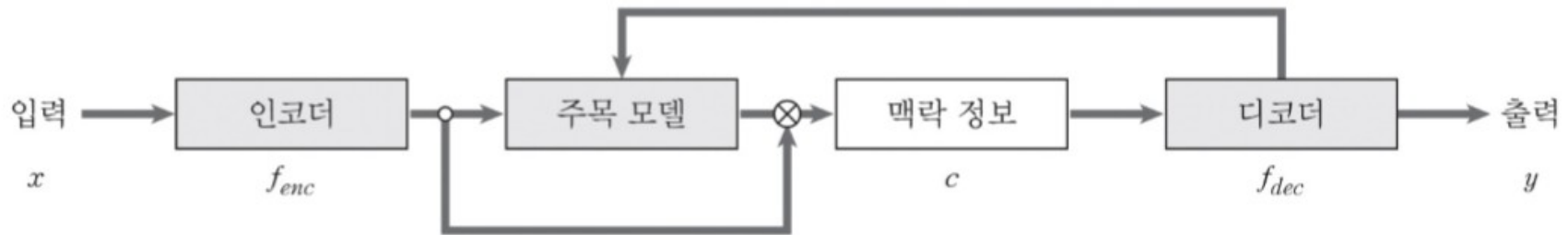


그림 5.78 주목 메커니즘을 포함한 인코더-디코더 망.

# 주목 메커니즘

## ✧ 주목 메커니즘을 포함한 기계번역 인코더-디코더 망

### ▪ 인코더

- 양방향 RNN 등을 사용하여 특징 벡터 서열  $\{h_1, h_2, \dots, h_T\}$  생성

### ▪ 주목 모델

- 시점  $t$ 에 노드의 출력  $h_i$ 에 대한 가중치  $\alpha_i^t$  계산
- 가중치를 사용하여 노드의 출력 결정

$$c_t = \sum_{i=1}^T \alpha_i^t h_i$$

- MLP와 softmax 사용

$$e_i^t = v^T \tanh(W_\alpha s_{t-1} + U_\alpha h_i)$$

$$\alpha_i^t = \frac{\exp(e_i^t)}{\sum_{k=1}^T \exp(e_k^t)}$$

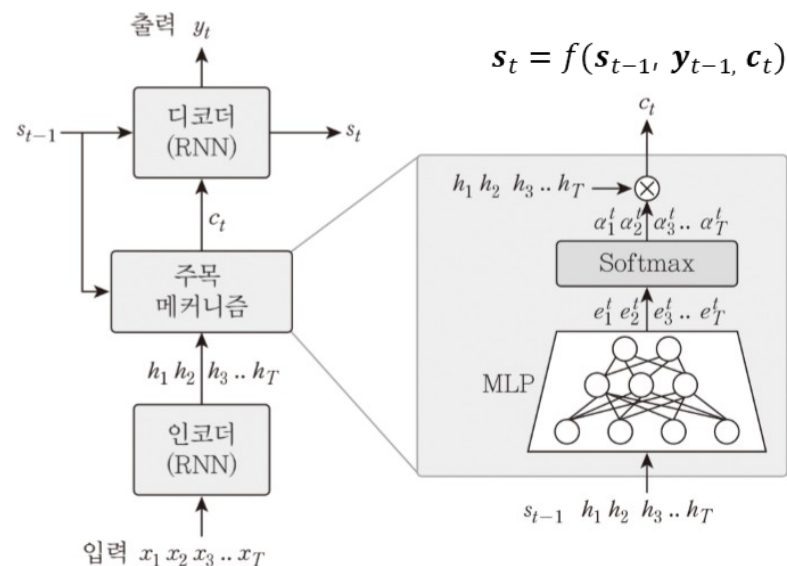


그림 5.79 서열대서열 변환을 하는 주목 메커니즘을 포함한 인코더-디코더 망.



# 주목 메커니즘

## ❖ 영상 주석달기를 위한 인코더-디코더 망

- 입력 영상을 격자 형태로 구분하고 디코더가 격자들을 선택적으로 주목하도록 하는 방법을 사용

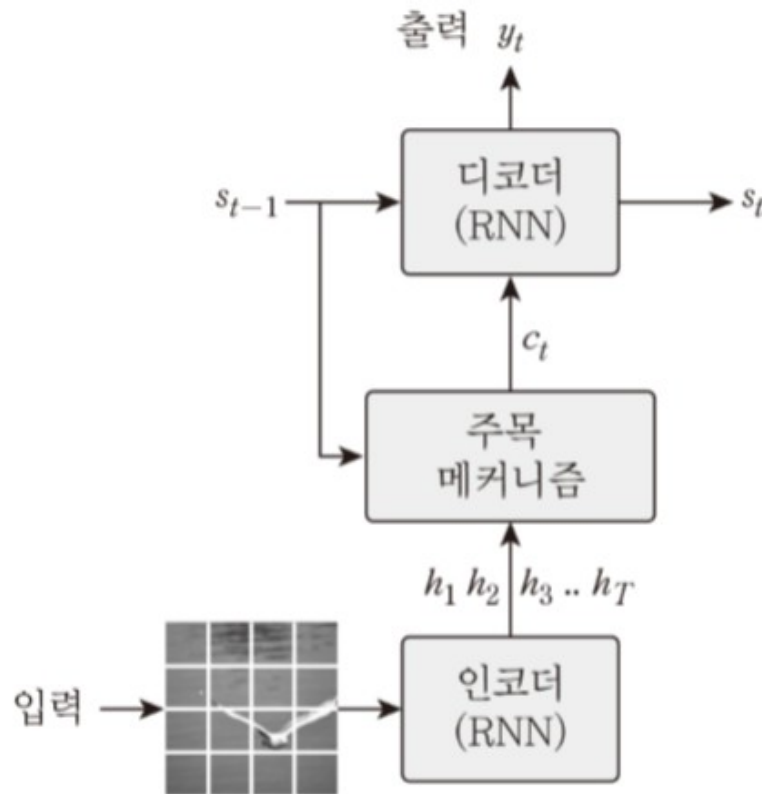


그림 5.80 영상 주석달기를 하는 주목 메커니즘을 포함한 인코더-디코더 망

# 주목 메커니즘

❖ Ex. 시간에 따른 주목

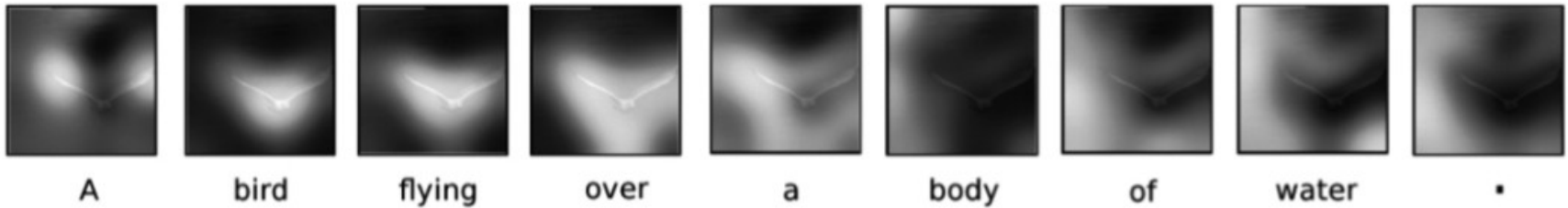


그림 5.81 영상 주석달기를 할 때 디코더가 각 단어 생성시 주목하는 영역 [출처 : Xu, Bengio 등, 2016]



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.