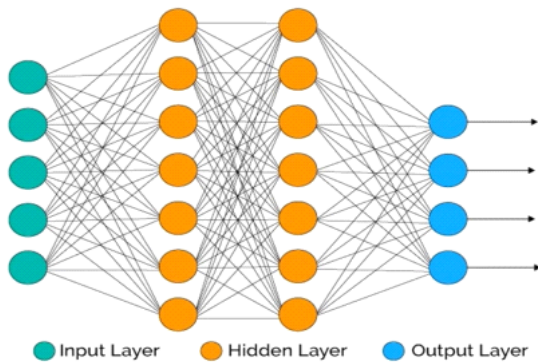


# Attention in NN Basics

Saturday, October 23, 2021 11:43 AM

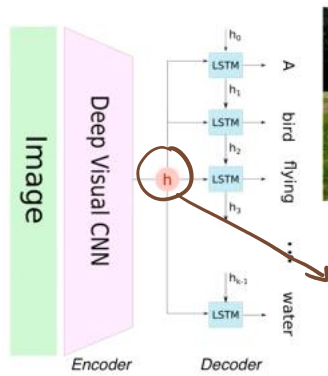
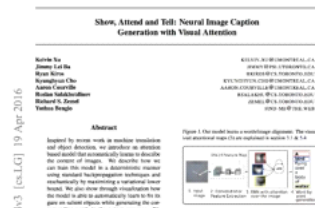
## Attention



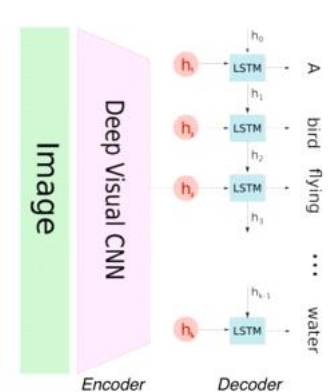
## Attention



Girl throwing a frisbee

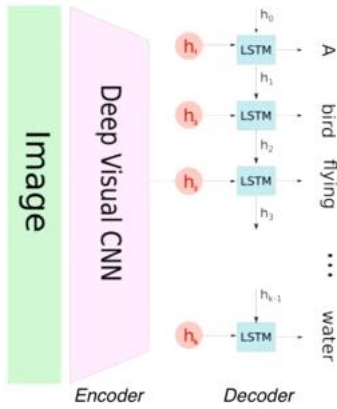


LSTM, look at entire image via this layer to caption even a single word - which is inefficient method  
Ideally it should be very specific



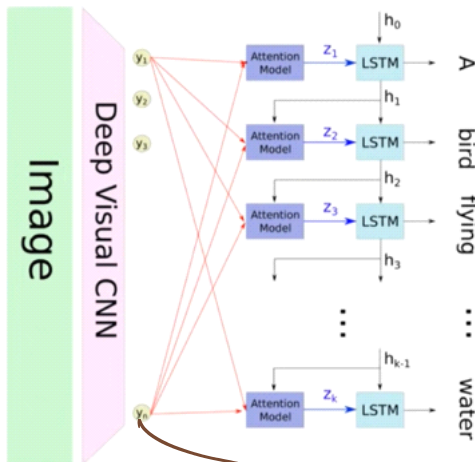
high level Attention Model  
→ each word represents a different vector independent





A Girl throws a Frisbee in the park.

Q-how does it exactly decide the regions to consider??



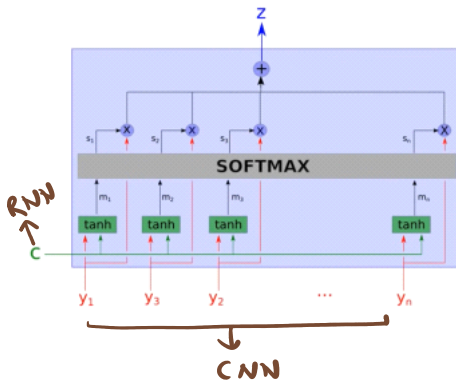
$$E(X) = \sum_n p(X = X_n) X_n$$

Ans- by using Attention unit considers all sub regions and Context for all inputs & output a weighted Arithmetic mean of these region

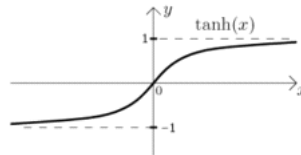
This mean is the Product of Prob. & actual values  
The Prob come from the context

$$m_i = c y_i \rightarrow \text{CNN} \times \text{RNN}$$

Attention Unit



$$m_i = \tanh(y_i W_{y_i} + C W_c)$$



Alternatively  $m_i = c y_i$  can also be applied but tanh is more fine grained

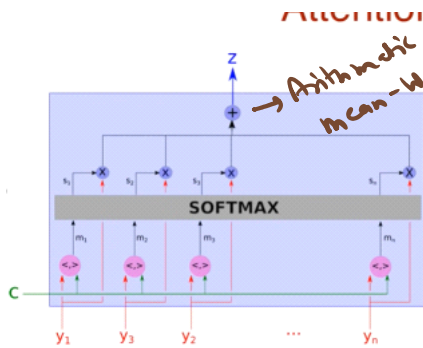
$m_i$  - learnable parameter of Attention unit  
tanh is used to convert very high values very close to 1  
So that they have a small difference & vice versa



Attention Unit

$$m_i = c y_i$$

Arithmetic mean-Weighted avg



$$m_i = cy_i$$

$$s_i = \frac{e^{m_i}}{\sum_n e^{m_n}}$$

$$s_i \in [0, 1]$$

$$\sum_n s_n = 1$$

— Softmax Probability—determine the relevance

$$z = \sum_n s_n y_n$$

— output of a specific region given context

## Types of Attention

1. Soft Attention: different parts, different subregions

$$z = \sum_n s_n y_n$$

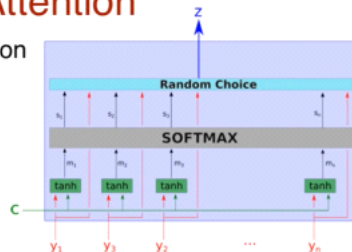
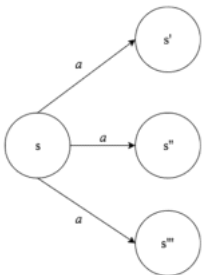
Soft Attention is Deterministic



## Types of Attention

2. Hard Attention: only ONE subregion

Hard Attention is Stochastic



## Types of Attention

2. Hard Attention: only ONE subregion

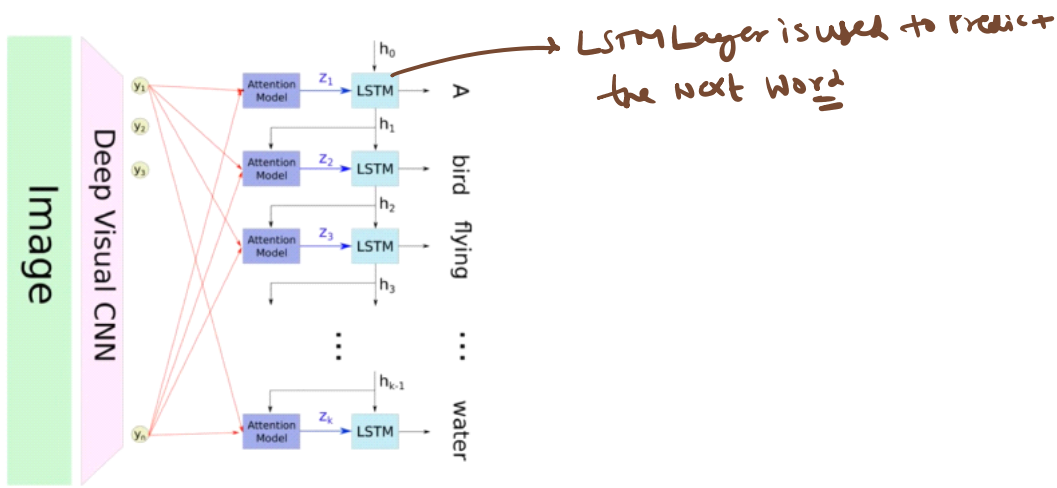
Hard Attention is Stochastic



—The region which has  
high prob.  
has more chances  
to come  
=

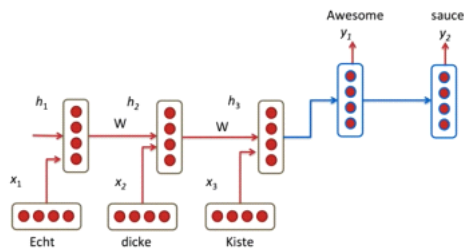


→ LSTM Layer is used to Predict  
the next word



## Applications

### 1. Neural Machine Translation (NMT)



## Applications

### 2. Microsoft's Attention GANs



## Applications

### 3. Teaching Machines to Read & Comprehend

<p>by ent423, ent261 correspondent updated 9:49 pm et, thu march 19, 2015 (ent261) a ent114 was killed in a parachute accident in ent45, ent85, near ent312, a ent119 official told ent261 on wednesday. he was identified thursday as special warfare operator 3rd class ent23, 29, of ent187, ent265. ent23 distinguished himself consistently throughout his career. he was the epitome of the quiet professional in all facets of his life, and he leaves an inspiring legacy of natural tenacity and focused</p> <p>...</p> <p>ent119 identifies deceased sailor as X, who leaves behind a wife</p>	<p>by ent270, ent223 updated 9:35 am et, mon march 2, 2015 (ent223) ent63 went familial for fall at its fashion show in ent237 on sunday, dedicating its collection to "mamma" with nary a pair of "mom jeans" in sight. ent164 and ent21, who are behind the ent196 brand, sent models down the runway in decidedly feminine dresses and skirts adorned with roses, lace and even embroidered doodles by the designers' own nieces and nephews. many of the looks featured saccharine needlework phrases like "i love you,"</p> <p>...</p> <p>X dedicated their fall fashion show to moms</p>
--	--

