Copyright Notice

These slides are distributed under the Creative Commons License.

DeepLearning.AI makes these slides available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite DeepLearning.AI as the source of the slides.

For the rest of the details of the license, see https://creativecommons.org/licenses/by-sa/2.0/legalcode



deeplearning.ai

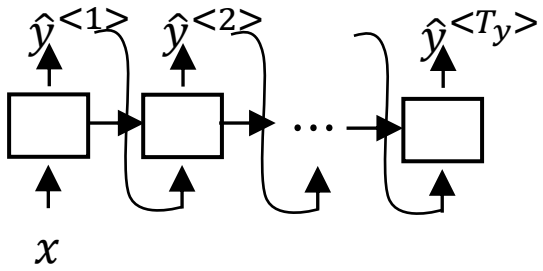
Sequence to sequence models

Transformers Intuition

Transformers Motivation

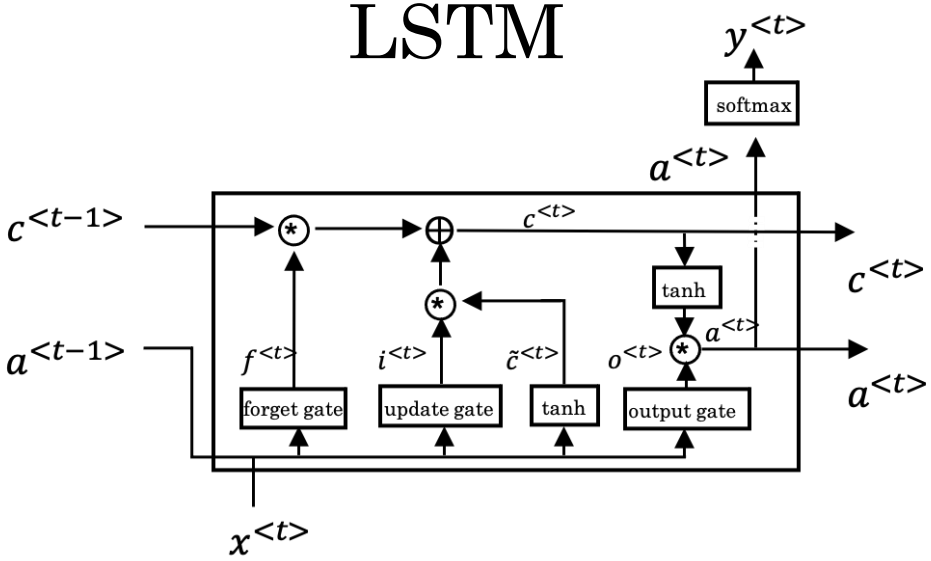
Increased complexity, sequential

RNN



GRU

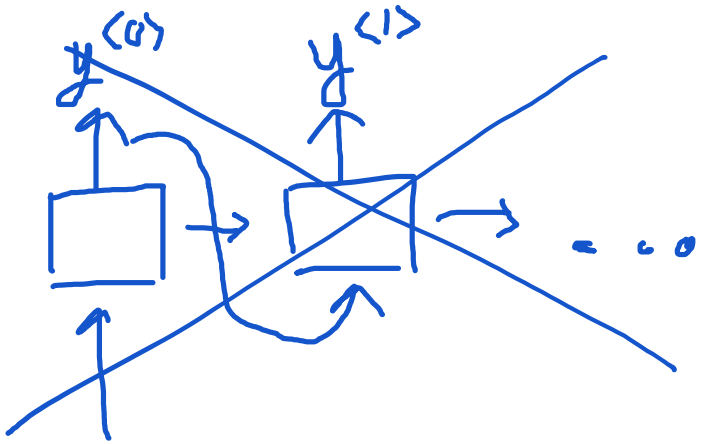
LSTM

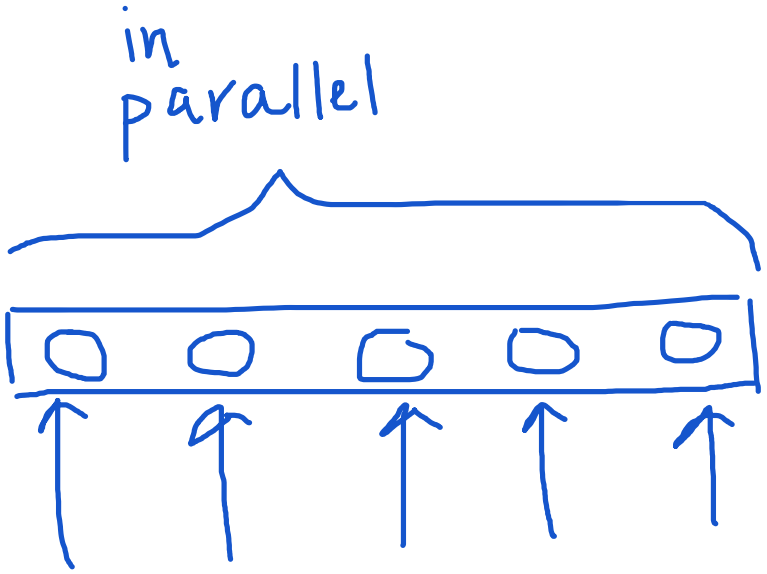


Andrew Ng

Transformers Intuition

- Attention + CNN
  - Self-Attention
  - Multi-Head Attention





[Vaswani et al. 2017, Attention Is All You Need]



deeplearning.ai

Sequence to sequence models

Self-Attention

Self-Attention Intuition

$A(q, K, V)$= attention-based vector representation of a word

calculate for each word

RNN Attention

$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$

Transformers Attention

$$A(q, K, V) = \sum_i \frac{\exp(e^{<q \cdot k^{<i>}>})}{\sum_j \exp(e^{<q \cdot k^{<j>}>})} v^{<i>}$$

$x^{<1>}$	$x^{<2>}$	$x^{<3>}$	$x^{<4>}$	$x^{<5>}$
Jane	visite	l'Afrique	en	septembre

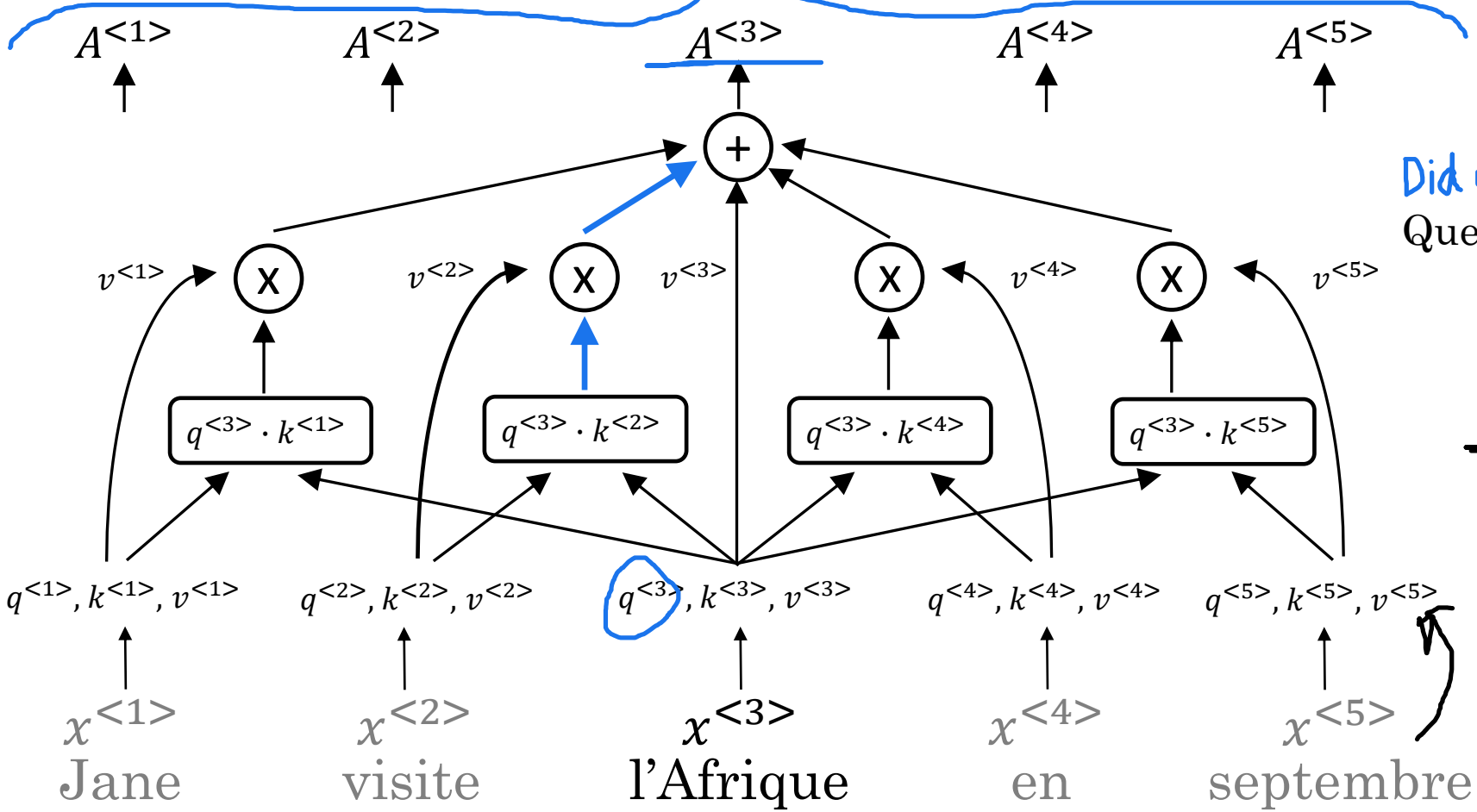
[Vaswani et al. 2017, Attention Is All You Need]

Self-Attention

$$A(q,K,V) = \sum_i \frac{\exp(e^{<q \cdot k^{<i>}>})}{\sum_j \exp(e^{<q \cdot k^{<j>}>})} v^{<i>}$$

Softmax

$$\textbf{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \textbf{softmax}\left(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{V}$$



Did what?

Query (Q)	Key (K)	Value (V)
$q^{<1>}$	$k^{<1>}$	$v^{<1>}$
$q^{<2>}$	$k^{<2>}$	$v^{<2>}$
$q^{<3>}$	$k^{<3>}$	$v^{<3>}$
$q^{<4>}$	$k^{<4>}$	$v^{<4>}$
$q^{<5>}$	$k^{<5>}$	$v^{<5>}$

What's happening there?

person action

Jane visit

W^Q, W^K, W^V

[Vaswani et al. 2017, Attention Is All You Need]

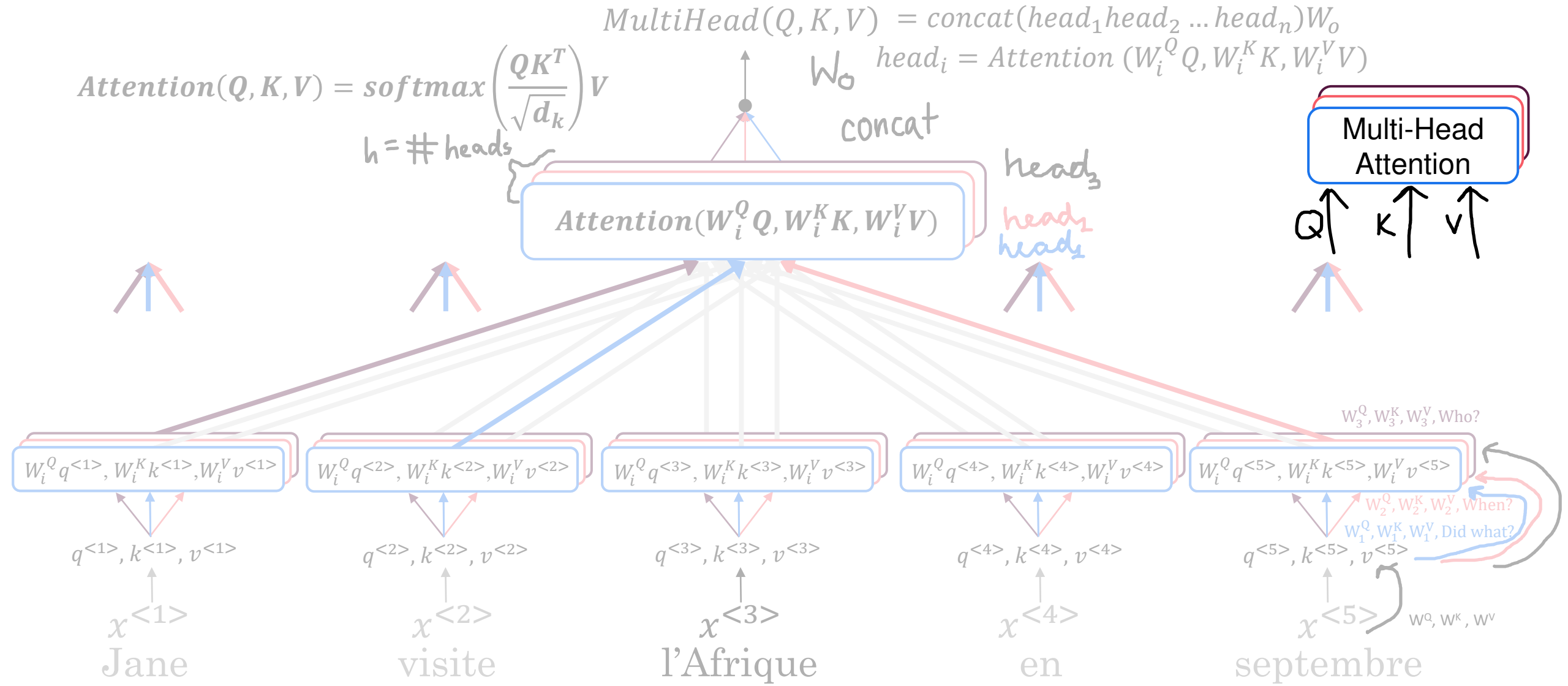


deeplearning.ai

Sequence to sequence models

Multi-Head Attention

Multi-Head Attention



[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng

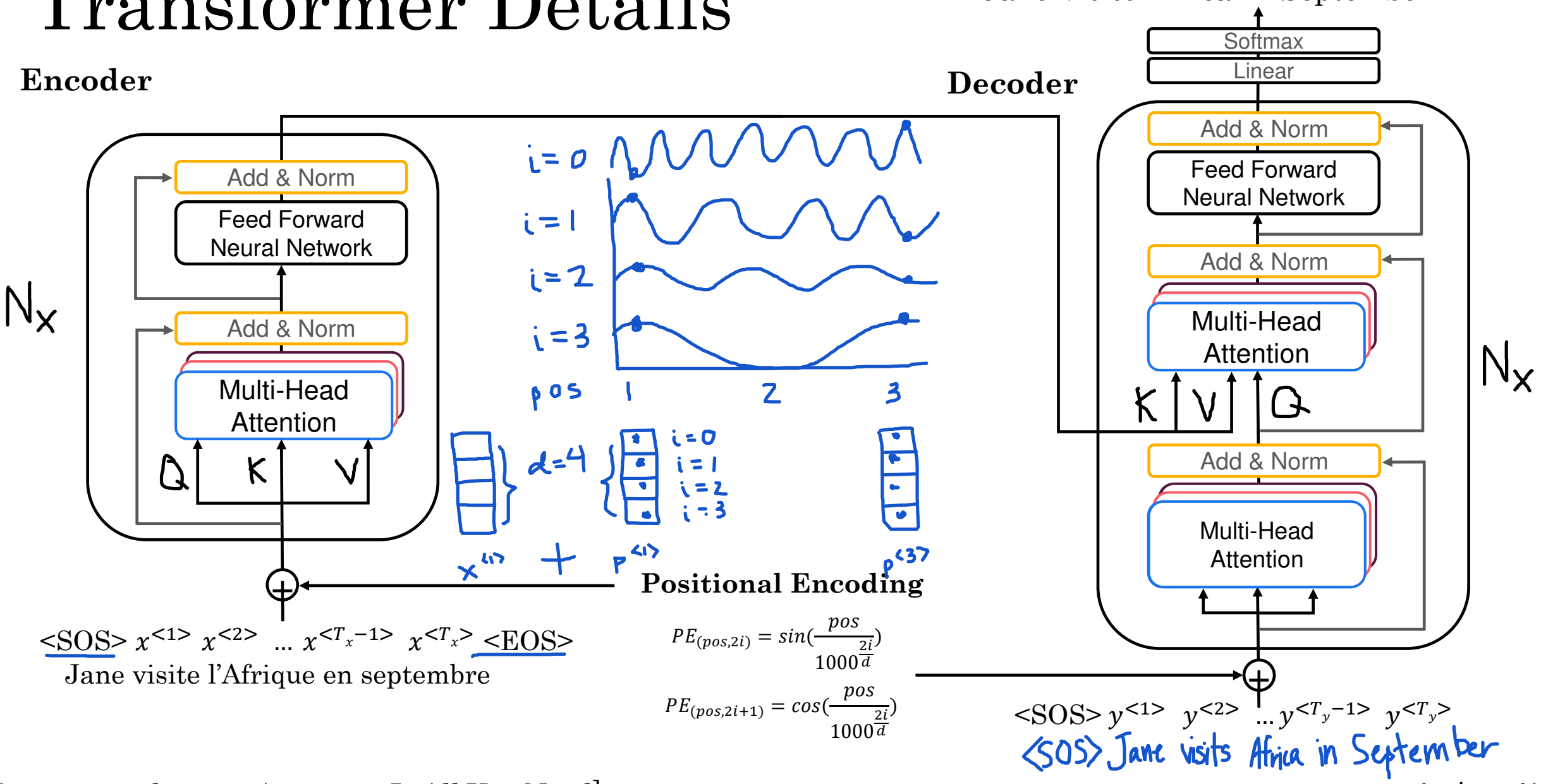


deeplearning.ai

Sequence to sequence models

Transformers

Transformer Details



[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng