ATTENTION

- ## Seq2Seq Models:

Encoders and Decoders

The encoder and decoder do not have to be RNNs; they can be CNNs too!

In the example above, an LSTM is used to generate a sequence of words; LSTMs "remember" by keeping track of the input words that they see and their own hidden state.

In computer vision, we can use this kind of encoder-decoder model to generate words or captions for an input image or even to generate an image from a sequence of input words.

- ## Types of Attention:

Neural Machine Translation by Jointly Learning to Align and Translate

Effective Approaches to Attention-based Neural Machine Translation

- ## Super interesting computer vision applications using attention:

Show, Attend and Tell: Neural Image Caption Generation with Visual Attention [pdf]

Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering [pdf]

Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks [pdf]

Every Moment Counts: Dense Detailed Labeling of Actions in Complex Videos [pdf]

Tips and Tricks for Visual Question Answering: Learnings from the 2017 Challenge [pdf]

Visual Question Answering: A Survey of Methods and Datasets [pdf]

- **Transformer:**

Paper: Attention Is All You Need

Talk: Attention is all you need attentional neural network models – Łukasz Kaiser