

Encoder Decoder Models :-

→ Sequence 2 Sequence was used as we algo of google translate for a while.

Gmail AutoReply, AutoSuggestions also use (d) Seq2Seq.

→ Original seq2seq model paper written by Ilya Sutskever.

Machine Translations :-

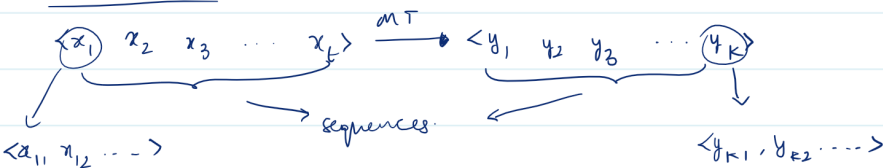
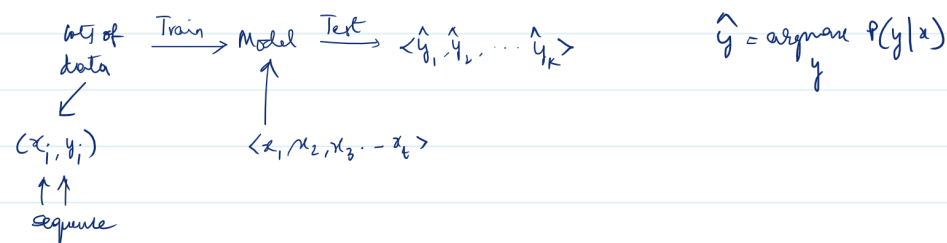


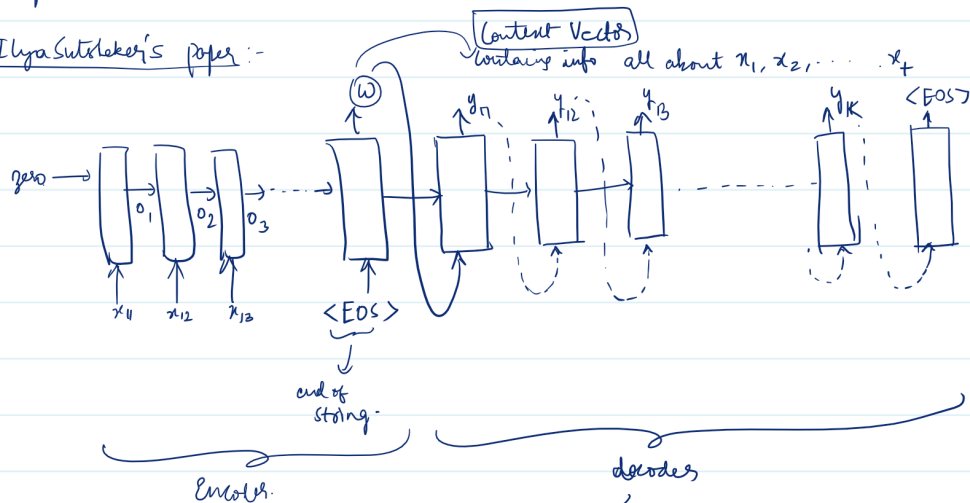
Image Captioning



→ Mathematically, $P(\langle y^1, y^2, y^3, \dots, y^t \rangle | \langle x^1, x^2, x^3, \dots, x^t \rangle)$



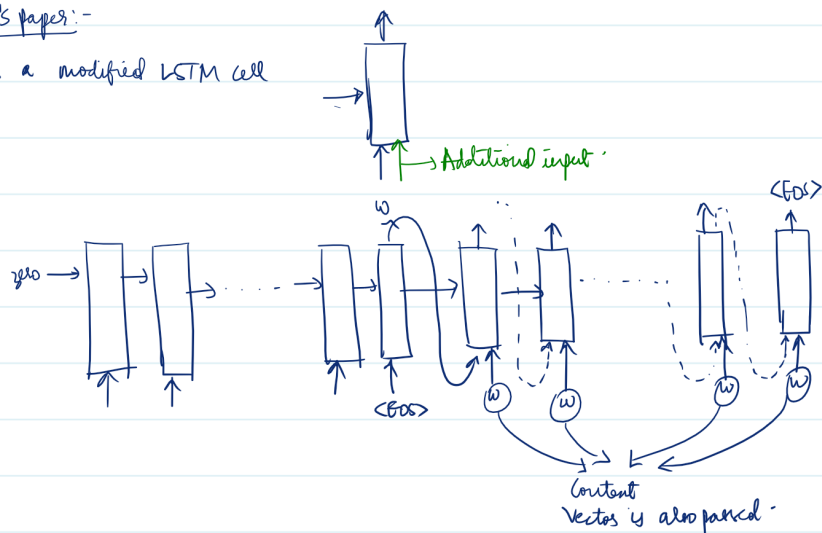
Ilya Sutskever's paper :-



Generally LSTMs are used. Sometimes GRUs too.

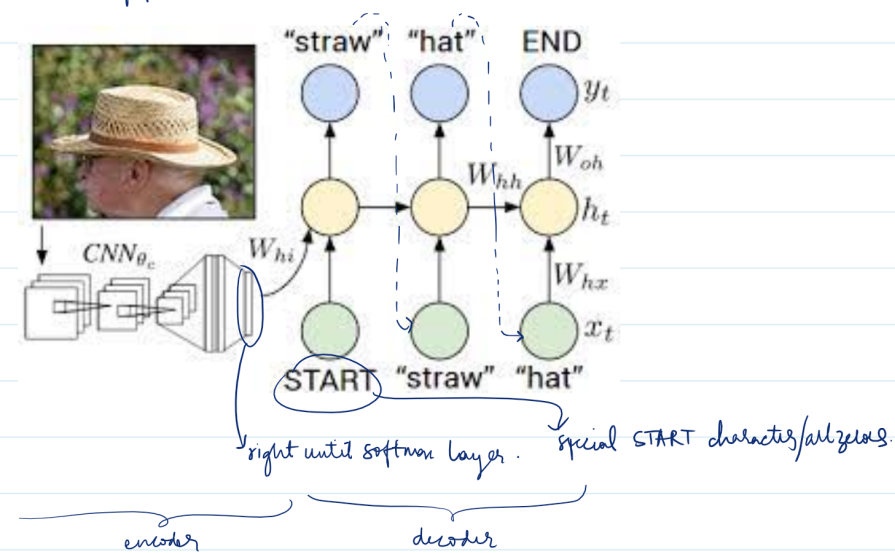
Cho's paper :-

uses a modified LSTM cell



→ Since default LSTM is so well optimized & since this modified LSTM does not show massive improvement, not often used.

Karpathy's paper :-



→ I want food <EOS>

↑ ↑ ↑ ↑ ↑

character level model, if each char is passed one by one after OHE

↑ ↑ ↑

word level model, if each word is passed one by one after OHE/w2v/...

→ Above models do not work well with long sentence. and is not human like.

humans do it by looking at whole input, generate some o/p, go back look at i/p again, then produce o/p.

This is the intuition behind attention models.