

development is happening → course has functions
so don't matter if research
to fast. → too fast

start of the it's a just
find the right architecture & tools
↳ too engineers &
engineering oriented

Coding → lots of them:
How the fuck were gonna
do that if the code is
against us.
Mathematicians are used to solving
problems against us who are the
fuck can fuck with raw code long
against us

- ① Encoder Decoder → research papers
- ② Attention
- ③ Transformer
- ④ BERT

Some of the stuff we know
becomes outdated, so we have
to inform stuff we know well.

Assessment Test:
some people don't know how the
fuck everything works. Tests are essential
to verify. Or just try to solve a
problem using what you know.

- Papers → read after this session
- ① sequence learning with nn - Ilya Sutskever et al
 - ② Learning Phrase Representations using RNN encoders... Kyunghyun Cho. et al

seq2seq → Machine Translation → google now does something like
Gmail auto reply & suggestion
Image - Text → Image Search



Image

Image → <j1 j2 j3 ... jn>

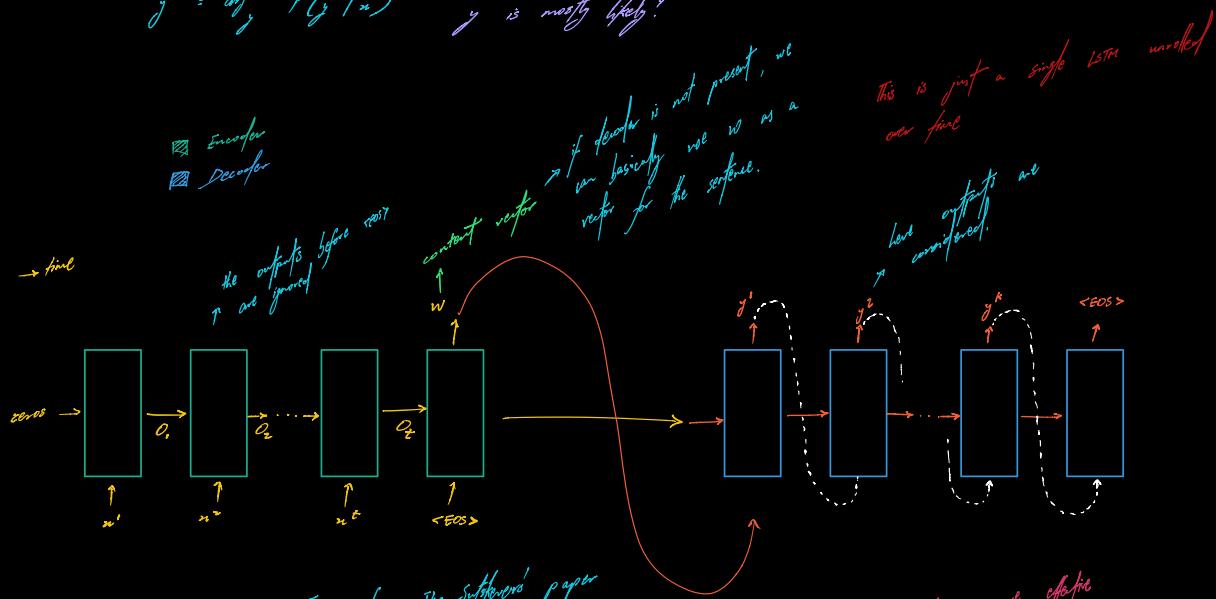
* test
google can now leverage their text search
apps instead of applying with image
changes the unknown into a known problem.

Implementing

$$P(y^1, y^2, \dots, y^T | x^1, x^2, \dots, x^T)$$

(given x , what's the probability
of y ?
The model is trying to model this probability \rightarrow which is very complex
 \therefore lots of data needed to fit it.)

$$\hat{y} = \arg \max P(y|x) \rightarrow \text{given this } x, \text{ which } y \text{ is most likely?}$$



This is from Ilya Sutskever's paper

This paper gives w (content vector) to all decoding final stamps.

LSTM doesn't have a stamp so they changed the LSTM cell architecture.

This change has more performance over Ilya's paper in some problems.

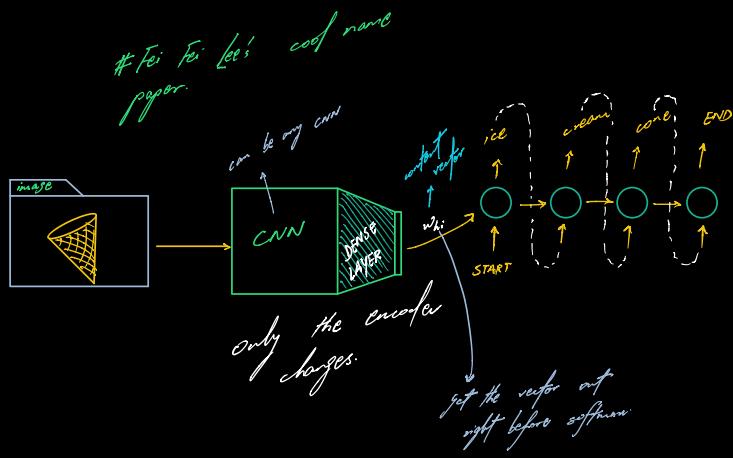
This approach is more effective
but LSTMs are more optimized for
training, if we use

Unless something offers significant
benefit, we choose the simple answer.

So, Ilya wins! Stability wins
over tiny performance
benefit.

This is just a single LSTM unrolled
over time

and



- * Applications
- ① Transfert → google → a net for machine
 - ② Email Auto Reply & smart compose
 - ③ Code Error → Martin Monperrus

1 class Foo { 2 int i; 3 int bar; 4 Foo (int bar){ 5 this.bar = bar; 6 } 7 int decrement(){ 8 return bar-1; 9 } 10 int increment(){ 11 return bar+1; 12 } 13 }	1 class Foo { 2 int i; 3 int bar; 4 Foo (int bar){ 5 } 6 int decrement(){ 7 } 8 int increment(){ 9 } 10 int <START_BUG>{ 11 return bar-1; 12 } 13 }
--	---

Listing 1: Original code Listing 2: abstract buggy context

- ④ Image Copying → Marshall Lamb → super detailed blog

Code

news blog → 10 minute intro to seq2seq

choose a character level model
because it's easier to communicate
their explanations are generally very good.

→ after this we can see word level code.

seq2seq can work with text summarization

Pfeffer with seq2seq

- ① if t is large, context vector cannot capture the essence of entire sentence.
- ② This is not how humans transfert → therefore can do this the human way
↓
we do it in phases & focus on different parts & different phases.