Encoder/Decoder Architecture

Dr. Muhammad Safyan

Encoder-Decoder

- Very good for sequence data.
- Not used today as it is. But foundation for LLM.

- Sequential data is handled by LSTM and GRU
- Now Next challenge is **Seq2eq** Data.
  - Input : sequence
  - Ouput: sequence
  - Example
    - **Nice to meet you**

آپ سے مل کر خوشی ہوئی۔

Why Challenge

- Input is of variable length
- Output is of variable length
- Input length !=output Length
- In RNN/LSTM we learn only to handle variable input length not output.
  So first architecture that can handle variable output length is Encoder/Decoder.

Sequence to Sequence Learning with Neural Networks

2014

Ilya Sutskever
Google
ilyasu@google.com

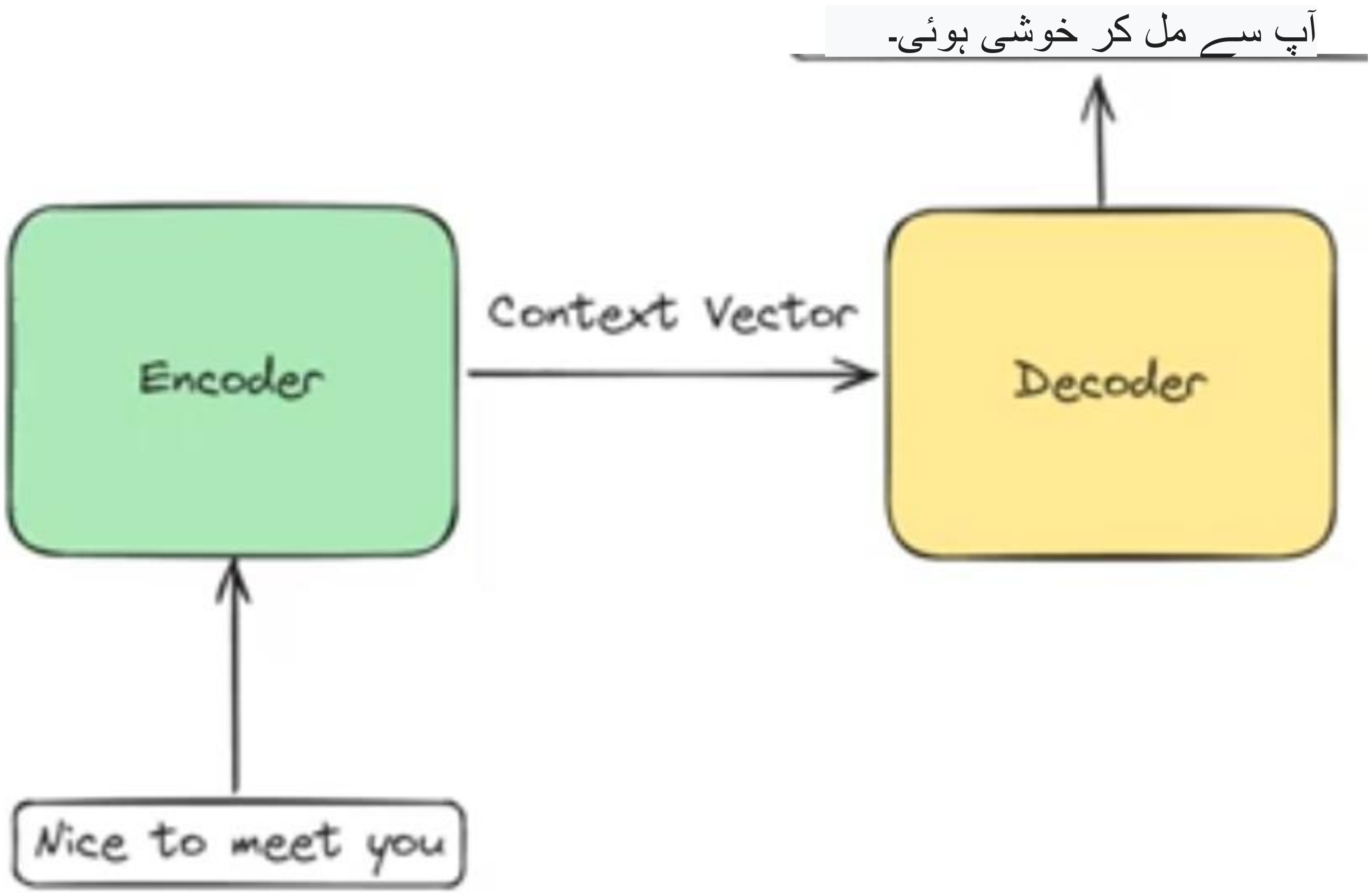
Oriol Vinyals
Google
vinyals@google.com

Quoc V. Le
Google
qvl@google.com

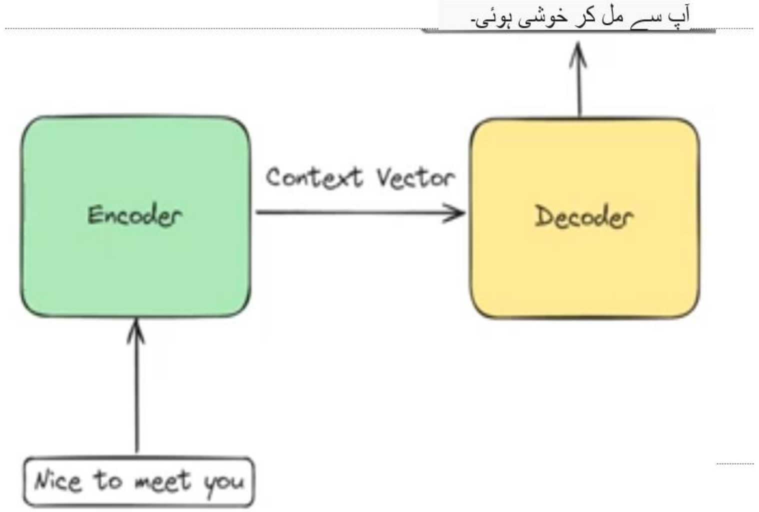
Abstract

Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks. Although DNNs work well whenever large labeled training sets are available, they cannot be used to map sequences to sequences. In this paper, we present a general end-to-end approach to sequence learning that makes minimal assumptions on the sequence structure. Our method uses a multilayered Long Short-Term Memory (LSTM) to map the input sequence to a vector of a fixed dimensionality, and then another deep LSTM to decode the target sequence from the vector. Our main result is that on an English to French translation task from the WMT-14 dataset, the translations produced by the LSTM achieve a BLEU score of 34.8 on the entire test set, where the LSTM's BLEU score was penalized on out-of-vocabulary words. Additionally, the LSTM did not have difficulty on long sentences. For comparison, a phrase-based SMT system achieves a BLEU score of 33.3 on the same dataset. When we used the LSTM to rerank the 1000 hypotheses produced by the aforementioned SMT system, its BLEU score increases to 36.5, which is close to the previous state of the art. The LSTM also learned sensible phrase and sentence representations that are sensitive to word order and are relatively invariant to the active and the passive voice. Fi-

Encoder/Decoder-simple Version



Abstract View

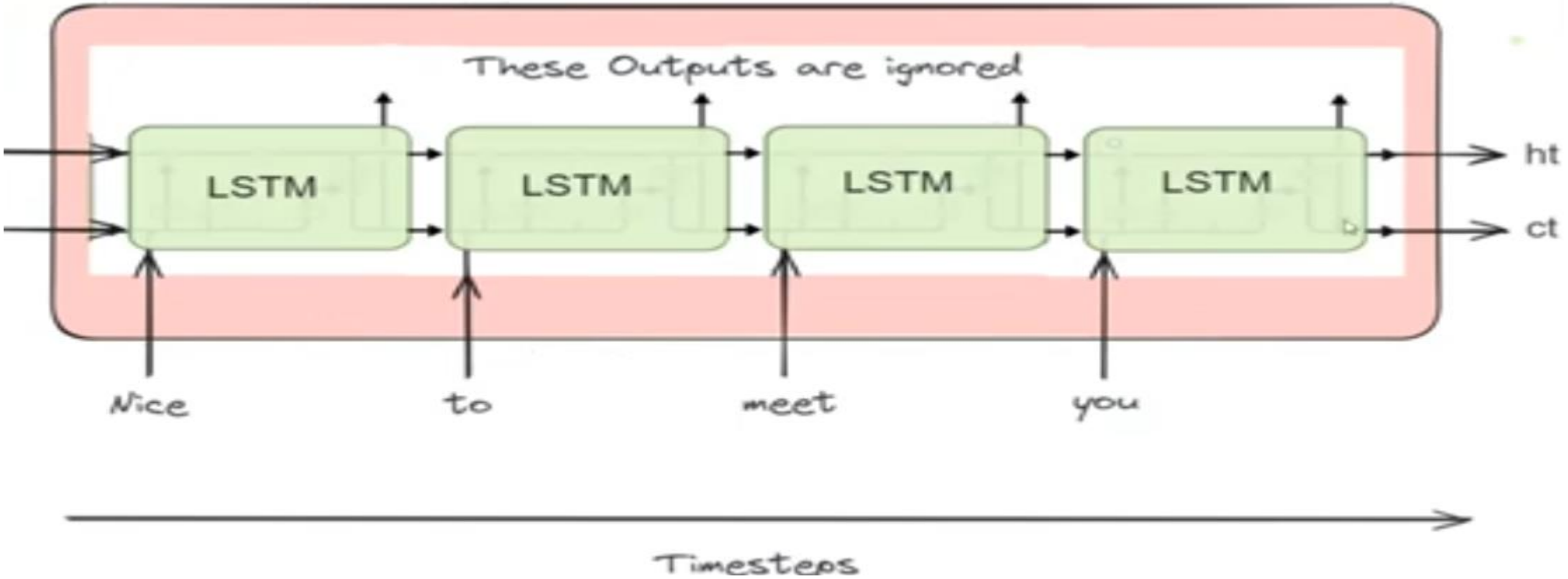


- Two blocks: Encoder -decoder
- Connected with context vector
- Encoder takes the input word by word.
- Encoder try to understand the essence of the sentence and summarize it as context vector(set of numbers).
- Decoder receives the context vector as input and try to understand it and print it in another language.

- Encoder and Decoder have capability to process sequence data.

Encoder

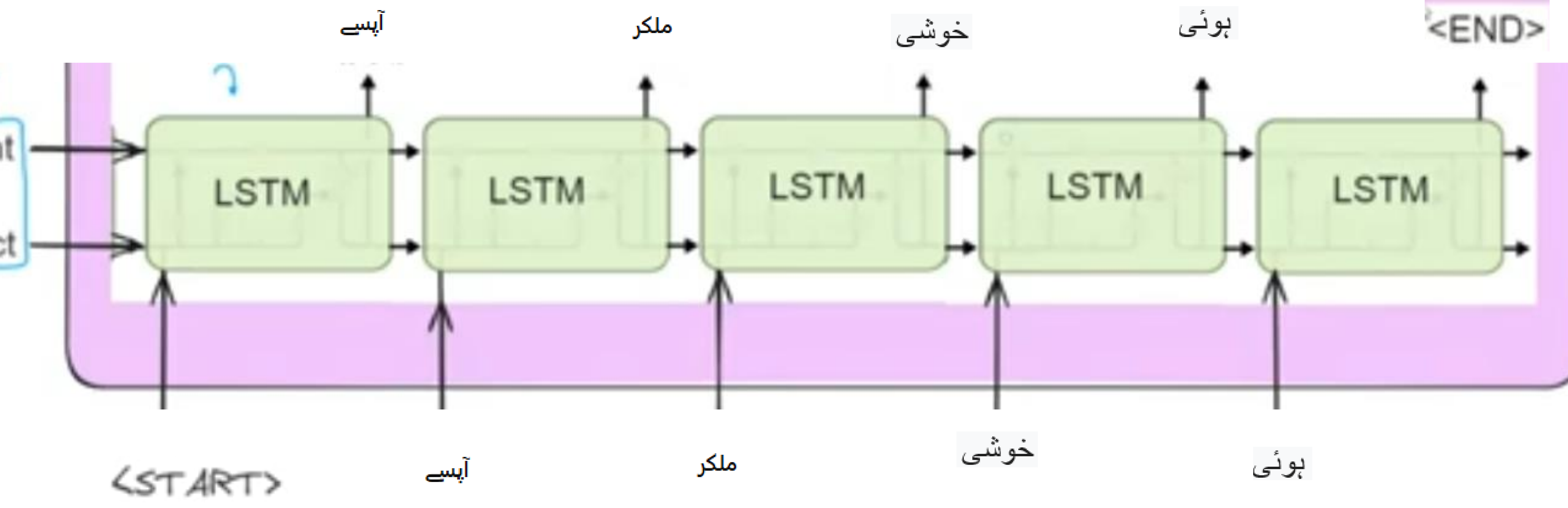
- In Encoder there is LSTM Architecture



Encoder

- In Encoder, we have one LSTM and we unfold it.
- We pass the ht and ct and input the words one by one.
- We ignore the output at each time step
- Get the final Ct and ht.
- Pass the final Ct and ht to the Decoder.

Decoder



آپسے
ملکر
خوشی
ہوئی
<END>
LSTM
LSTM
LSTM
LSTM
LSTM
<START>
آپسے
ملکر
خوشی
ہوئی

Decoder

- At start you give three things
  - Ct, ht, <Start>
- When decoder see <start> symbol. He know start to produce output.
- <End> is to stop the output.
- Dataset : Machine Translation data set
- Called parallel data set

Parallel corpus for English & Urdu langua

Data Card Code (3) Discussion (0) Suggestions (0)

developers working in the field of natural language processing and mac

Dataset (2 files)

english-corpus.txt
462.32 kB

urdu-corpus.txt
930.35 kB

english-corpus.txt (462.32 kB)

This preview is truncated due to the large file size. Cre
download this file to see the full content.

```
is zain your nephew
i wish youd trust me
did he touch you
its part of life
zain isnt ugly
above all be patient
i learned it from him
why am i doing this
i made a bad decision
zain wont care
zain was hesitant
i borrowed zains car
why are you out here
he is just a liar
```

urdu-corpus.txt (930.35 kB)

This preview is truncated due to the large file siz
download this file to see the full content.

```
زین تمہارا بھتیجا ہے۔
کاش تم مجھ پر بھروسہ کرتے
کیا اس نے آپ کو چھوا؟
اس کی زندگی کا حصہ
زین بدصورت نہیں ہے۔
سب سے بڑھ کر صبر کرو
میں نے اسے اس سے سیکھا۔
میں یہ کیوں کر رہا ہوں
میں نے ایک برا فیصلہ کیا
زین پرواہ نہیں کرے گا
```

Bible Dataset with English to Urdu transl

Data Card Code (3) Discussion (0) Suggestions (0)

Detail Compact Column 2 of 2 c

The book of the g...	داود ابن ابرہام کا نسب نامہ...
7936 unique values	**7945** unique values
Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat Judas and his brethren .	ابراہام سے اِضحاق پیدا ہُوا اور اِضحاق سے یعقوب پیدا ہُوا اور یعقوب سے یہوداہ اور اس کے بھائی پیدا ہ...
And Judas begat Phares and Zara of Thamar ; and Phares	اور یہوداہ سے فارص اور زارح تمر سے پیدا ہوئے اور فارص سے حصرون پیدا ہُوا اور

Encoder-Decoder

English	Urdu
Think about it	**سوچ لو**
Come in	اندر آ جاؤ"

Tokenize it

[Think, about, it]   **[سوچ, لو]**

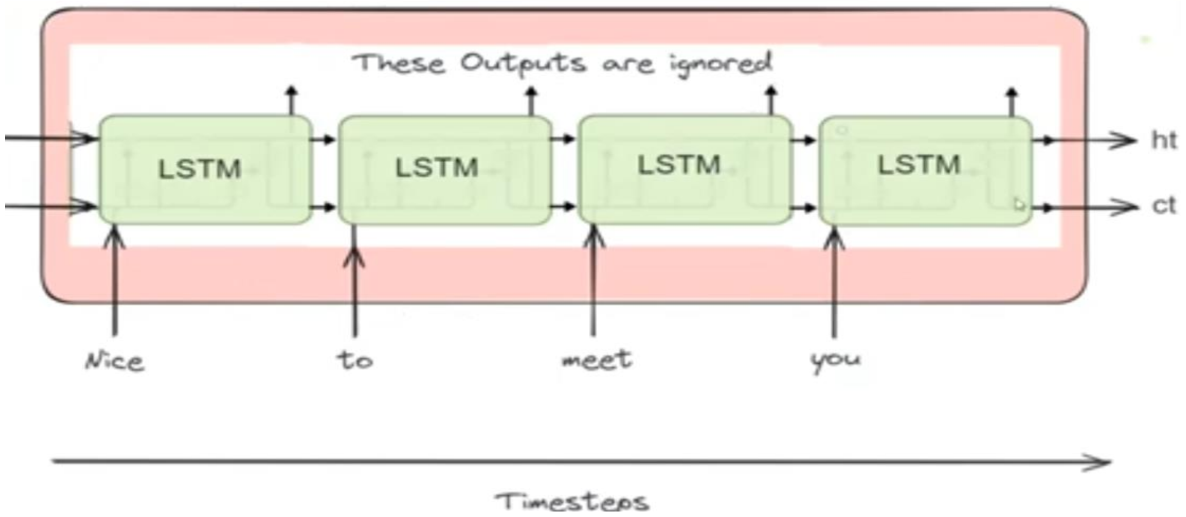
[Come in]   ["اندر آ جاؤ]

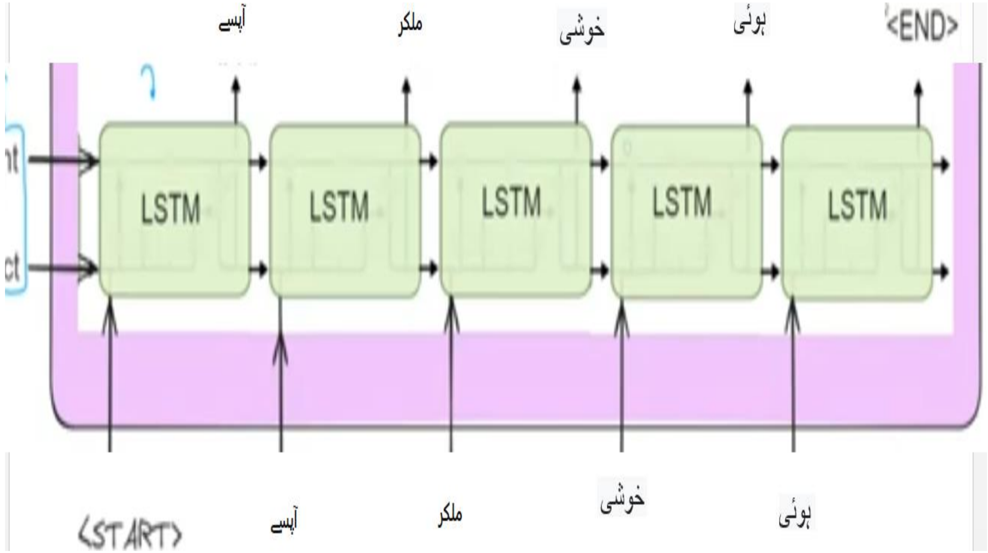
Convert it into one hot-Encoding

English	Urdu
[1,0,0,0,0], [0,1,0,0,0], [0,0,1,0,0], [0,0,0,1,0], [0,0,0,0,1]	<Start> **سوچ لو** اندر آ جاؤ"<end>

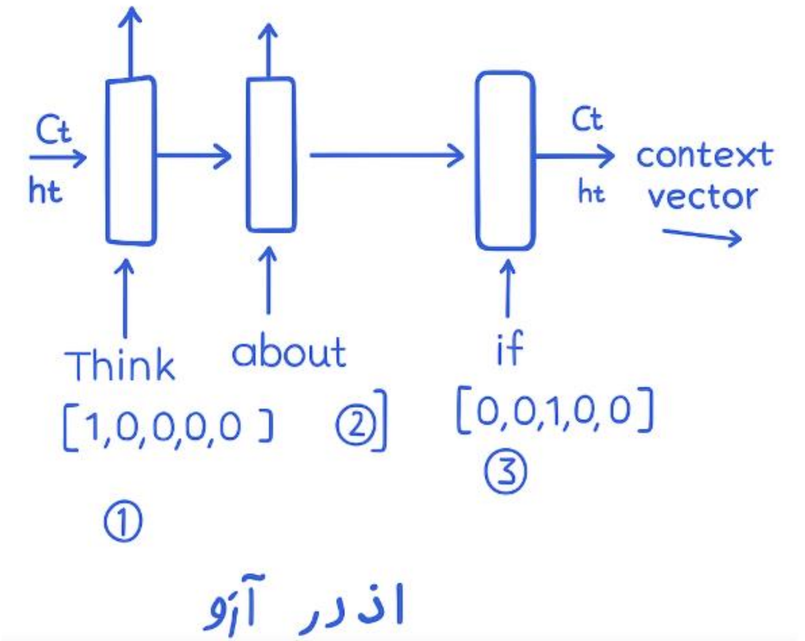
- Row-1
- Will be pass to encoder and decoder

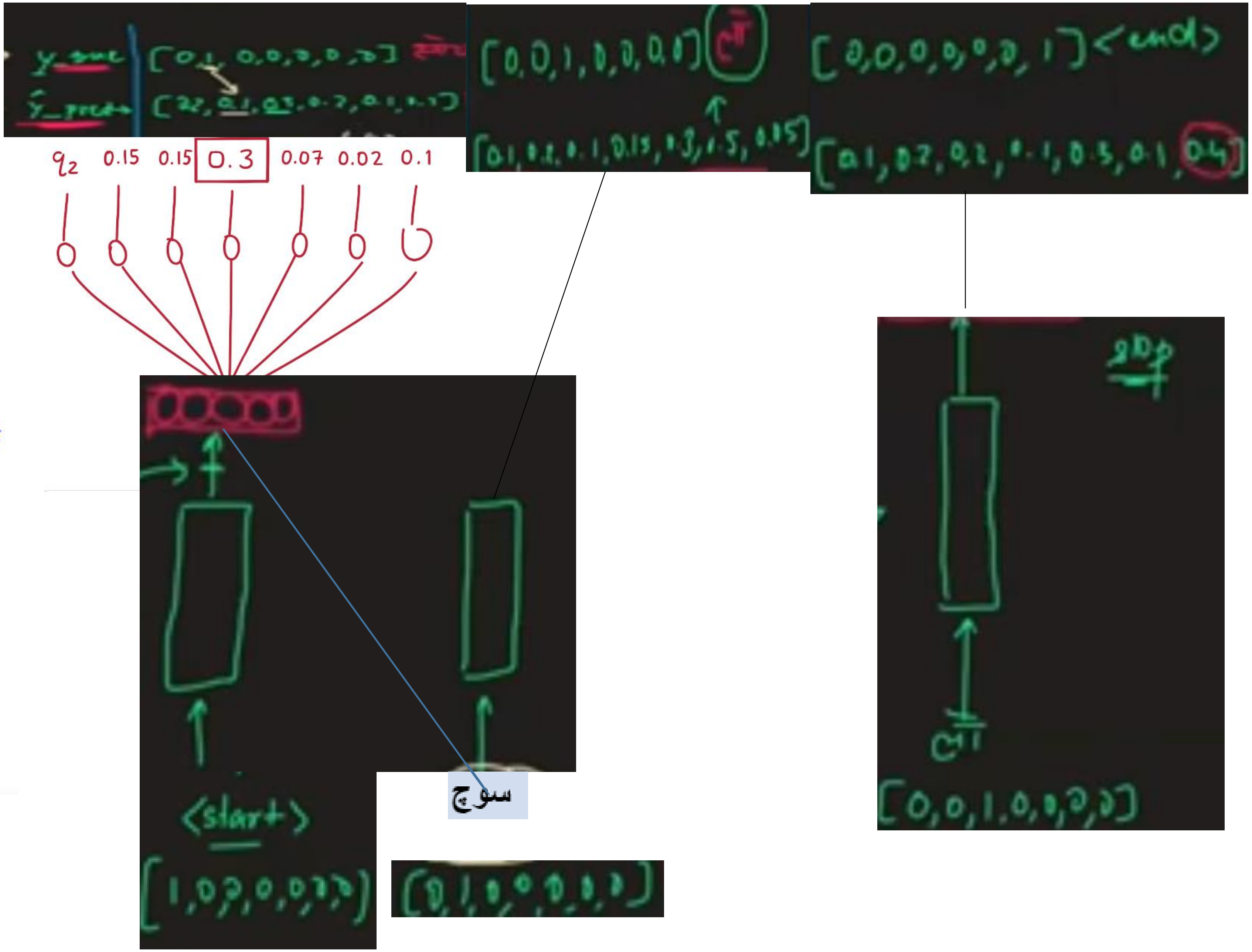
Think about it	سوچ لو



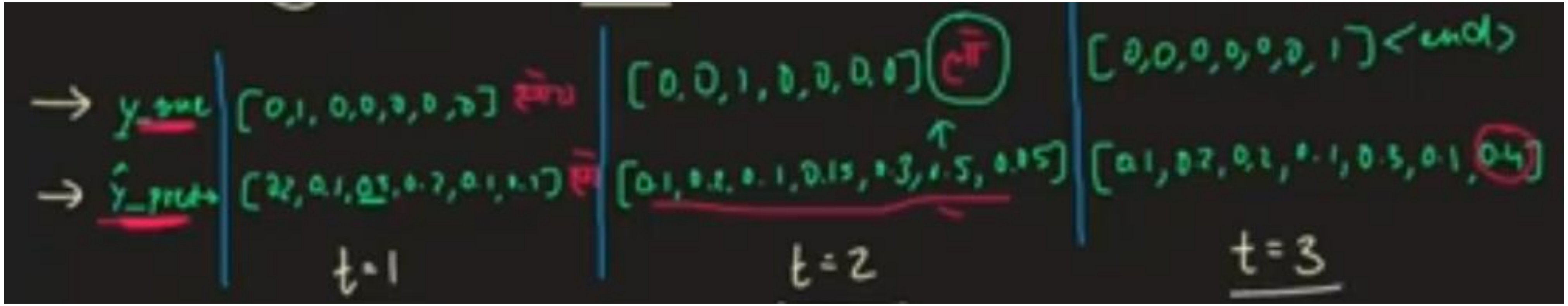


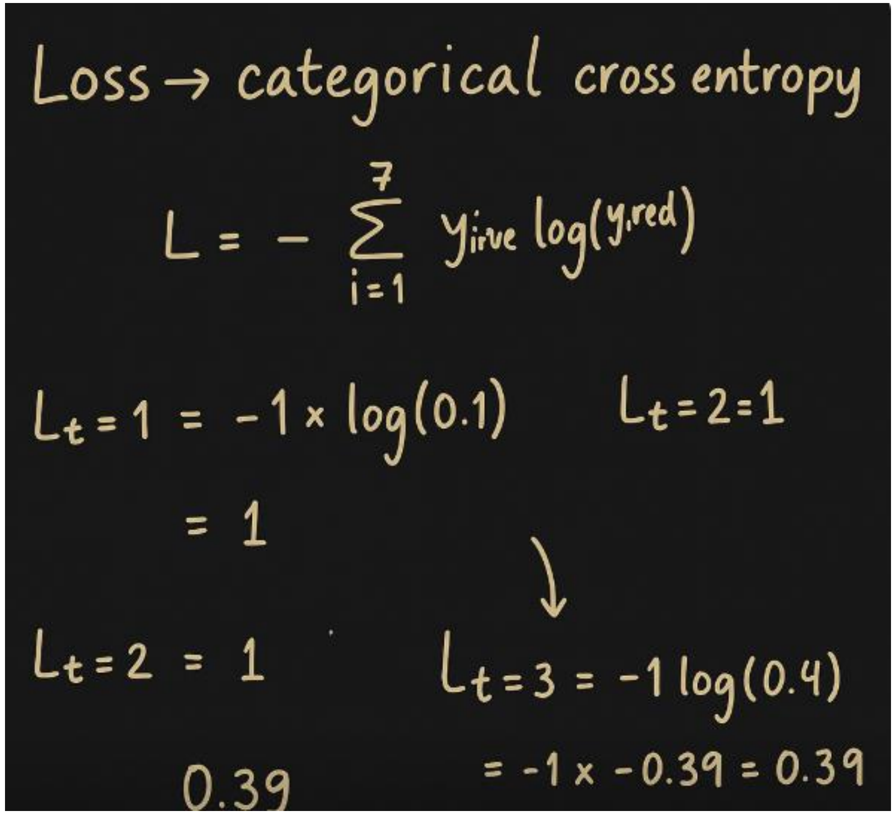
- Weights will be assigned randomly.





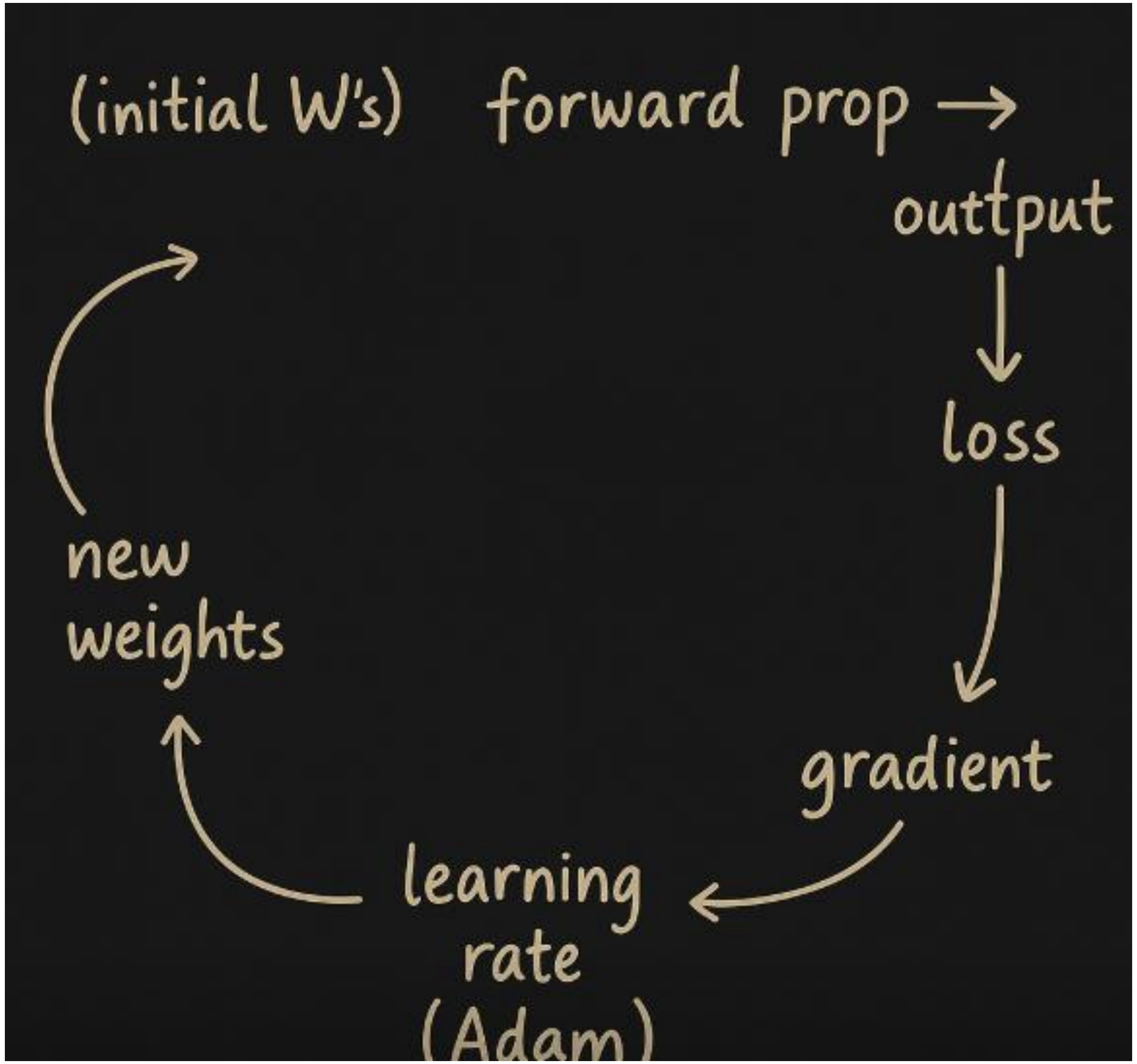
Teacher Forcing

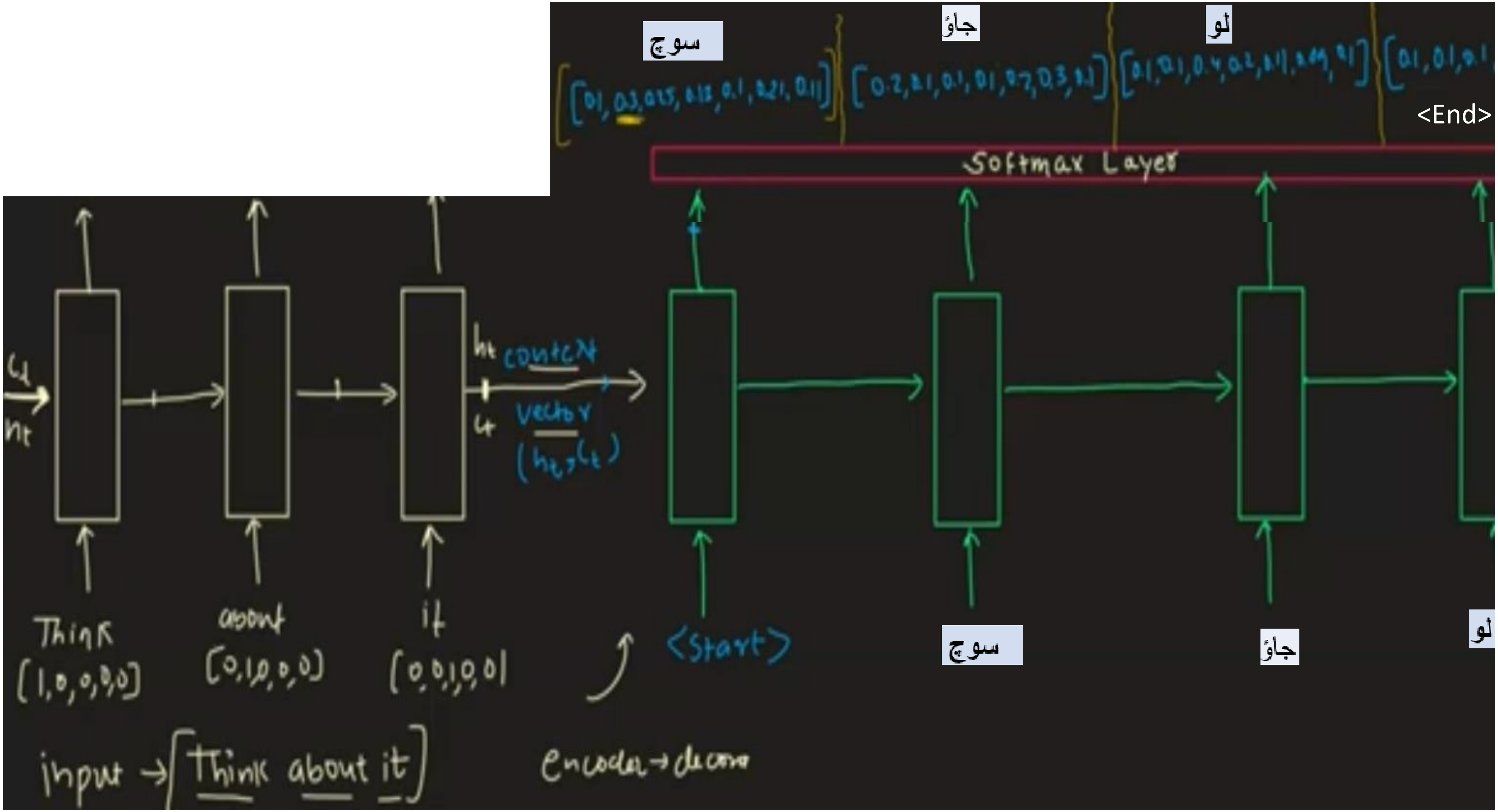




Total Loss= 2.39
Averagge loss=0.7

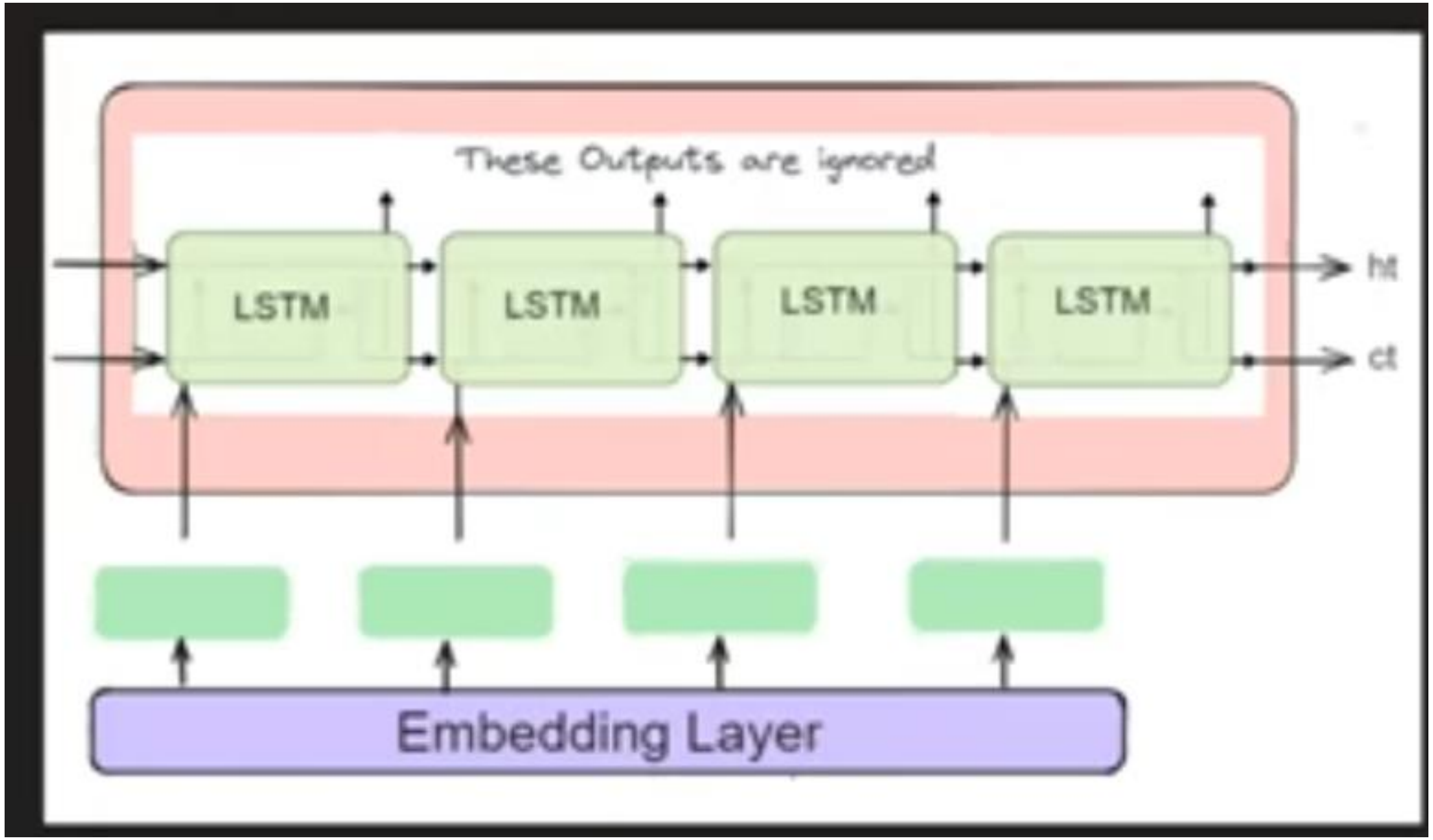
Forward progpgrataion
Backpropagation
    gradient calculation
    update weights


(initial W's)
forward prop
output
loss
gradient
learning rate (Adam)
new weights

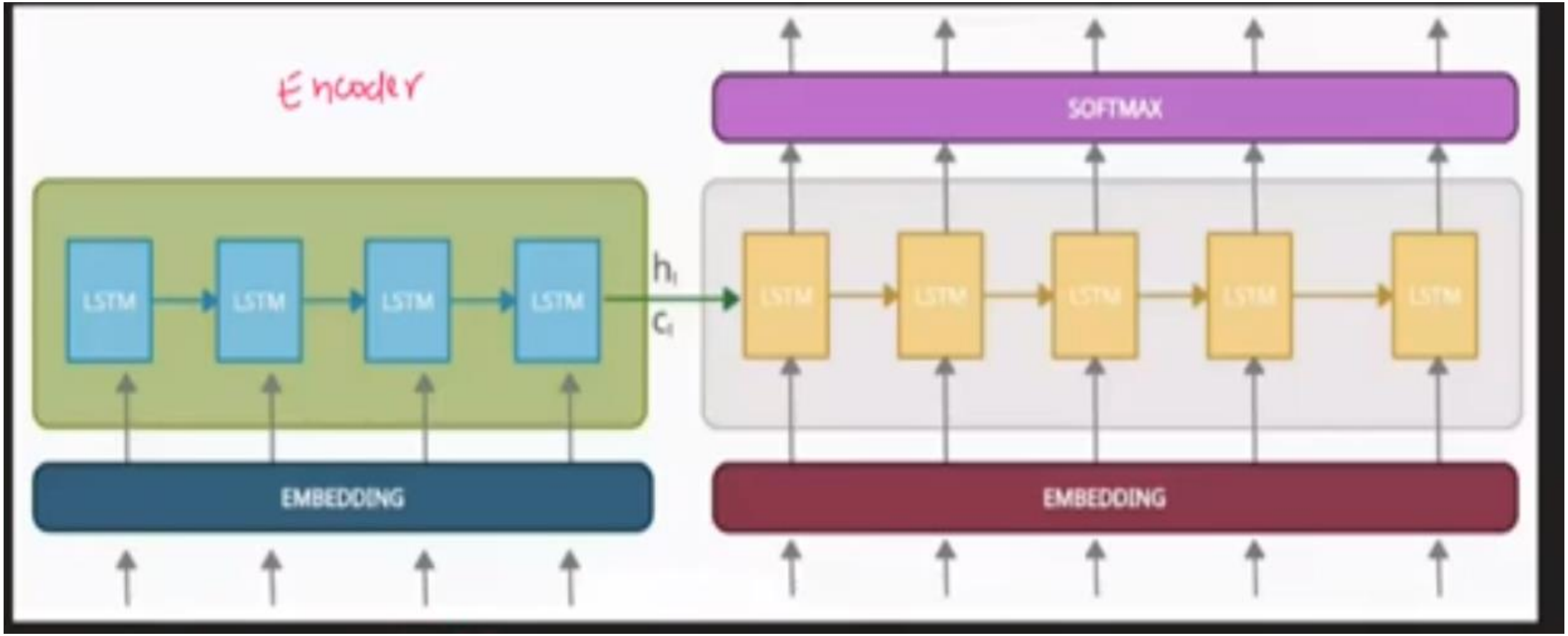


سوچ
جاؤ
لو
<End>
Softmax Layer
context vector
<Start>
Think
about
it
input → Think about it
encoder → decoder

Improvement-Embedding

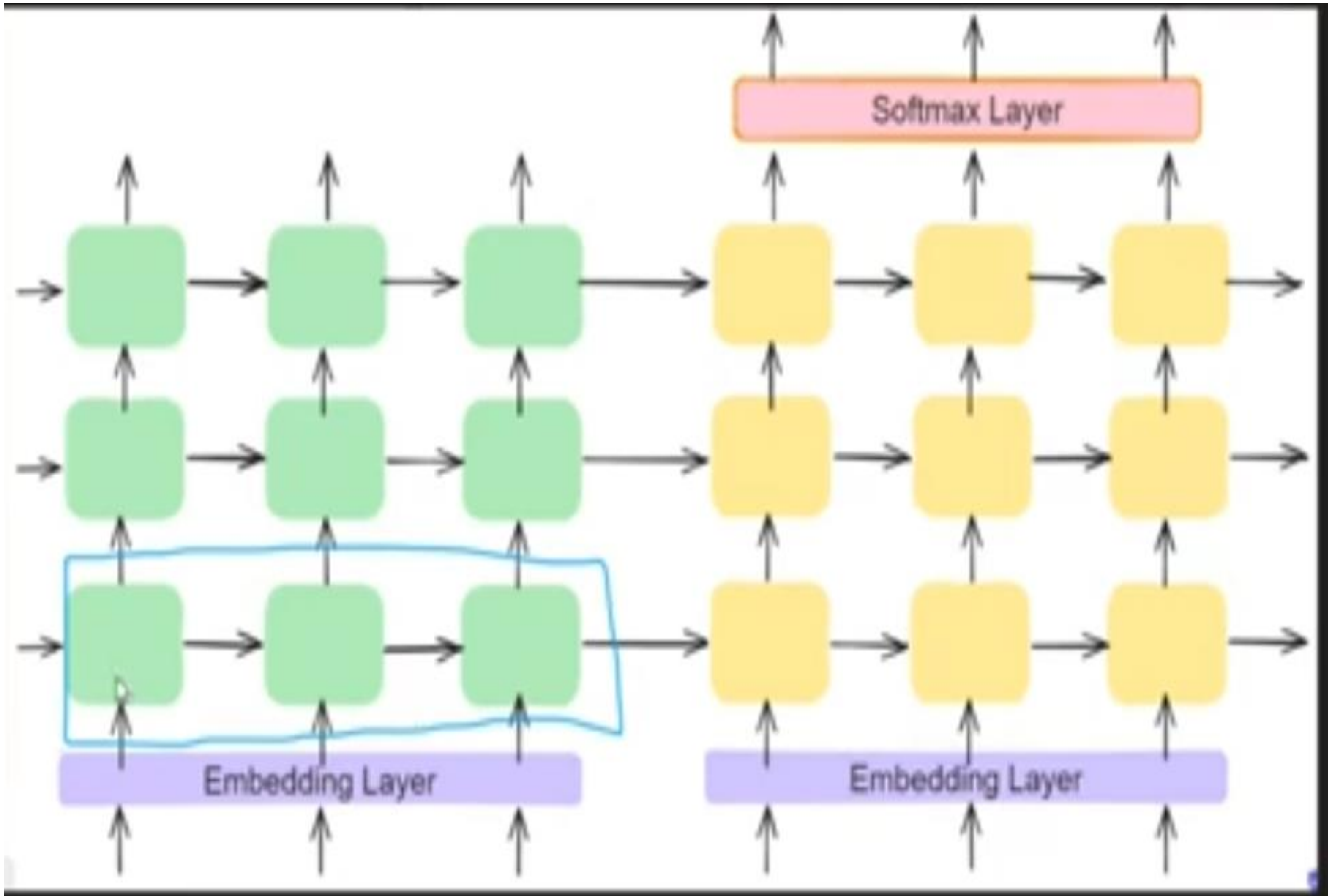


- Use Embedding in Encoding and Decoding Layer.
- Embedding are low dimensions, context
- Eg, W2V, Glove
- Pretrained Embedding's
- Self training



Encoder
SOFTMAX
LSTM
LSTM
LSTM
LSTM
h
c
LSTM
LSTM
LSTM
LSTM
LSTM
EMBEDDING
EMBEDDING

Improvement-Deep LSTM



Handle Long Term Dependency in good way.
Understand Layered hierarchy
Initial layer understand word , intermediate layer understand sentences, and top lay hold the context paragraph
The phone battery is bad but over all phone is good.
As Number of parameters increased ,model capacity also increased , better generalization.

Original paper used 4-layers

Revert the input
It about think