



CpSc 8430: Deep Learning

Homework 2





Copy Right Notice

✿ Most slides in this presentation are adopted from slides of text book and various sources. The Copyright belong to the original authors. Thanks!



HW2

- Video caption generation
 - Sequence-to-sequence model
 - Training Tips





Task Descriptions



Video caption generation

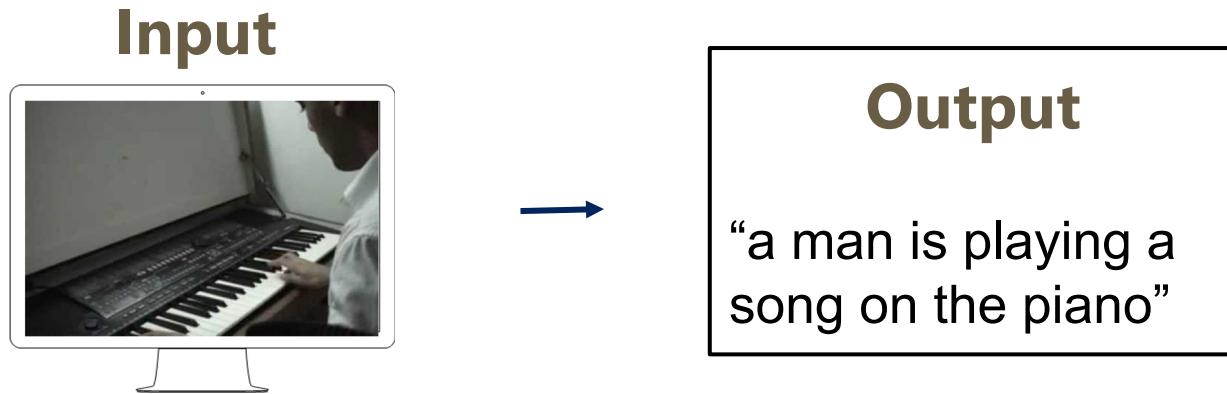
- Introduction
- Sequence-to-sequence model
- Training Tips
 - Attention
 - Schedule Sampling
 - Beamsearch
- How to reach the baseline ?





Introduction

- Video Caption Generation
 - a. Input : A short video
 - b. Output : The corresponding caption that depicts the video



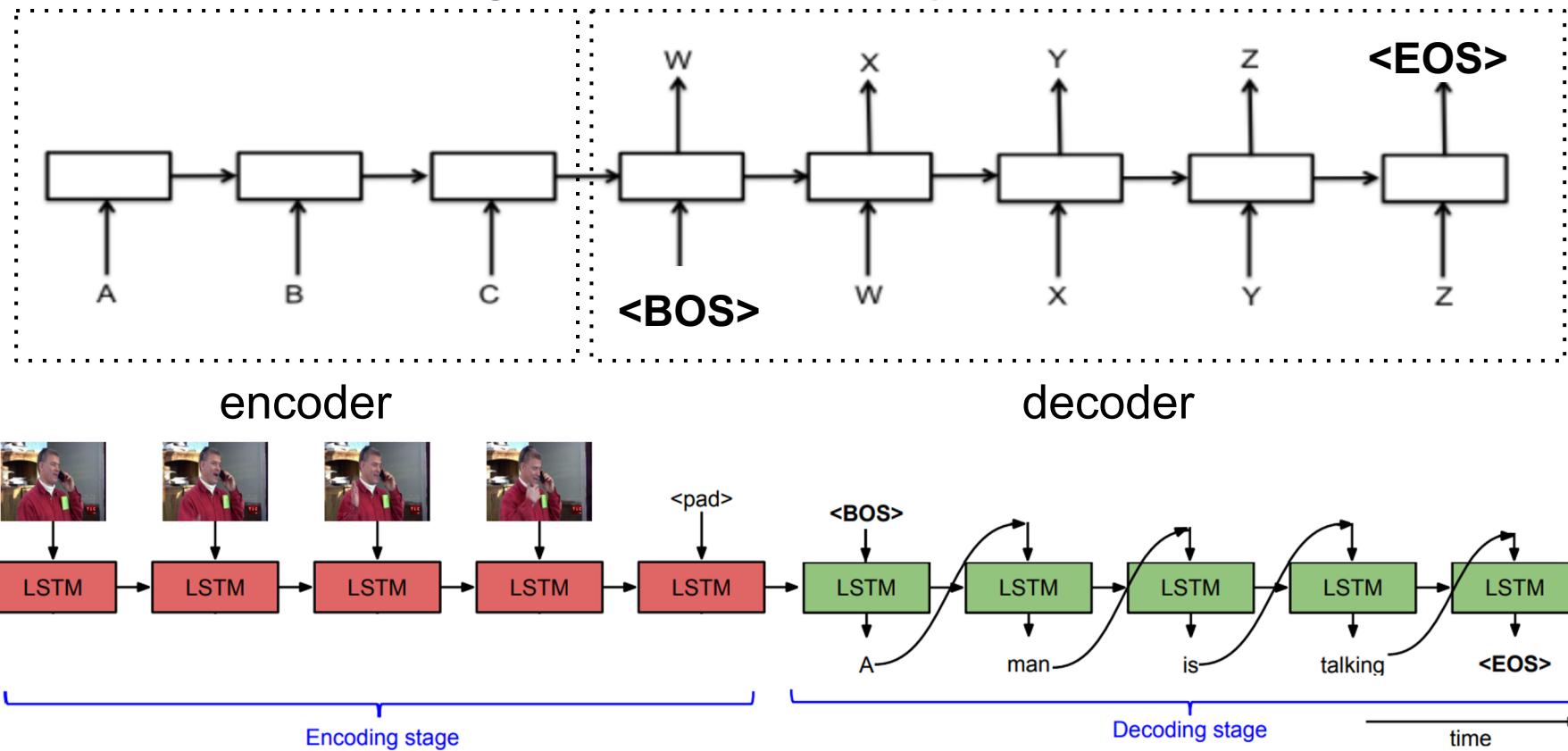
- There are several difficulties including :
 - a. Different attributes of video (object, action)
 - b. Variable length of I/O
- (In this task, video features will be provided)



Sequence-to-sequence

1/5

- **Two recurrent neural networks (RNNs)**
 - an encoder that processes the input
 - a decoder that generates the output



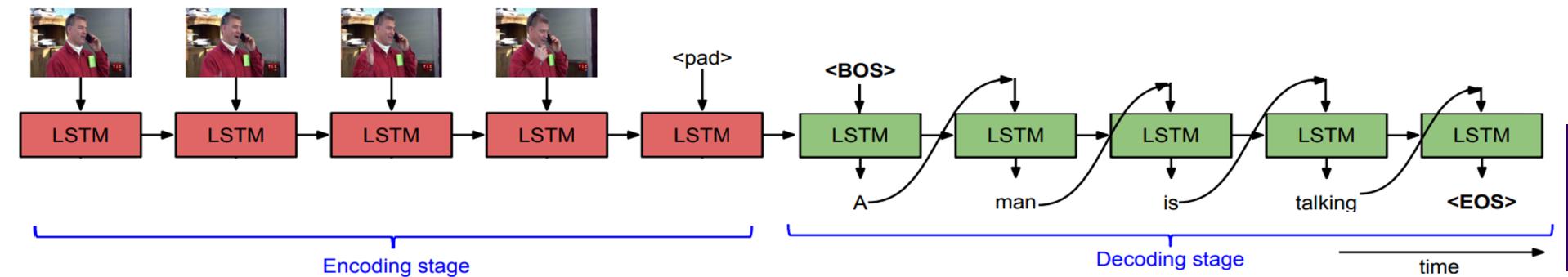


Sequence-to-sequence

2/5

- **Data preprocess :**

- Dictionary - most frequently word or min count
- other tokens : <PAD>, <BOS>, <EOS>, <UNK>
 - <PAD> : Pad the sentence to the same length
 - <BOS> : Begin of sentence, a sign to generate the output sentence.
 - <EOS> : End of sentence, a sign of the end of the output sentence.
 - <UNK> : Use this token when the word isn't in the dictionary or just ignore the unknown word.





Sequence-to-sequence

3/5

- **Text Input :**

One-hot Vector encoding

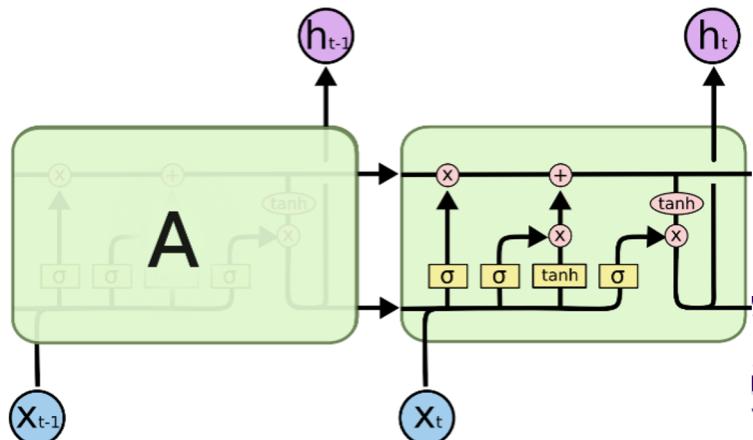
(1-to-N coding, N is the size of the vocabulary in dictionary)

○ e.g.

- neural = [0, 0, 0, ..., 1, 0, 0, ..., 0, 0, 0]
- network = [0, 0, 0, ..., 0, 0, 1, ..., 0, 0, 0]

- **LSTM unit :**

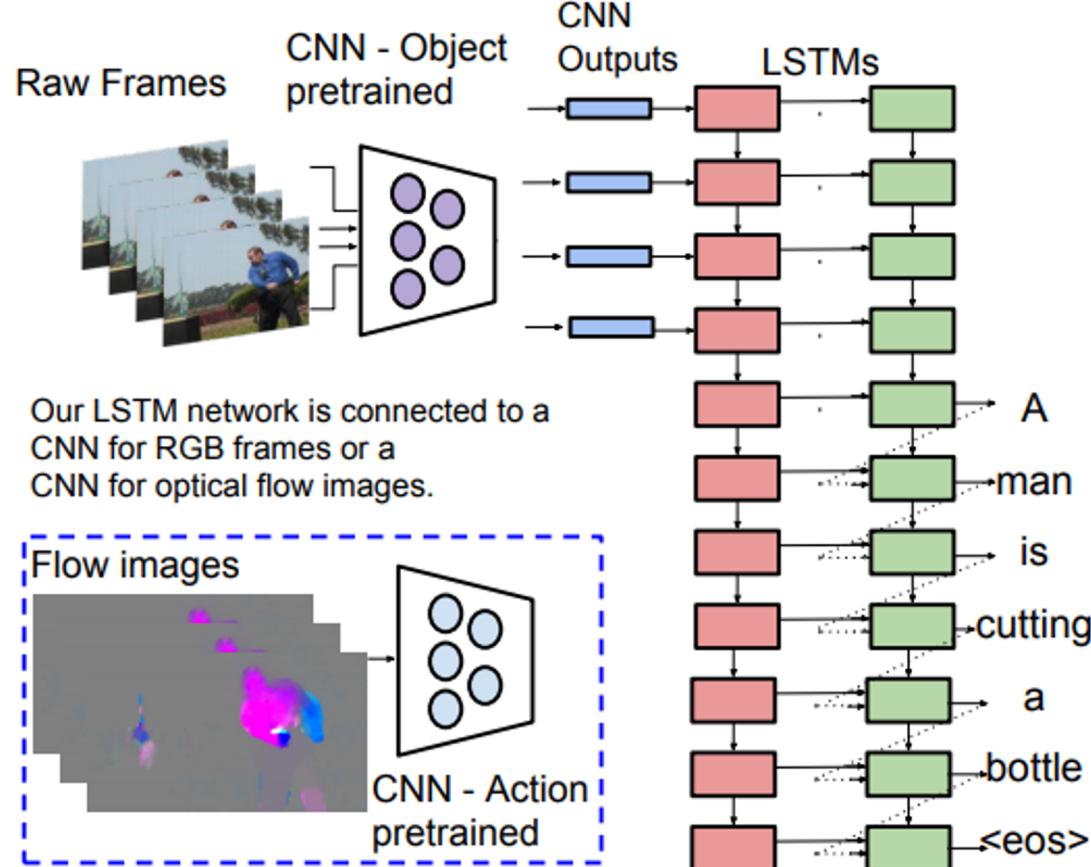
cell output than project to a vocabulary-size vector





Sequence-to-sequence - S2VT 4/5

- Sequence-to-Sequence Based Model: S2VT



Refer to the following paper for detailed info:

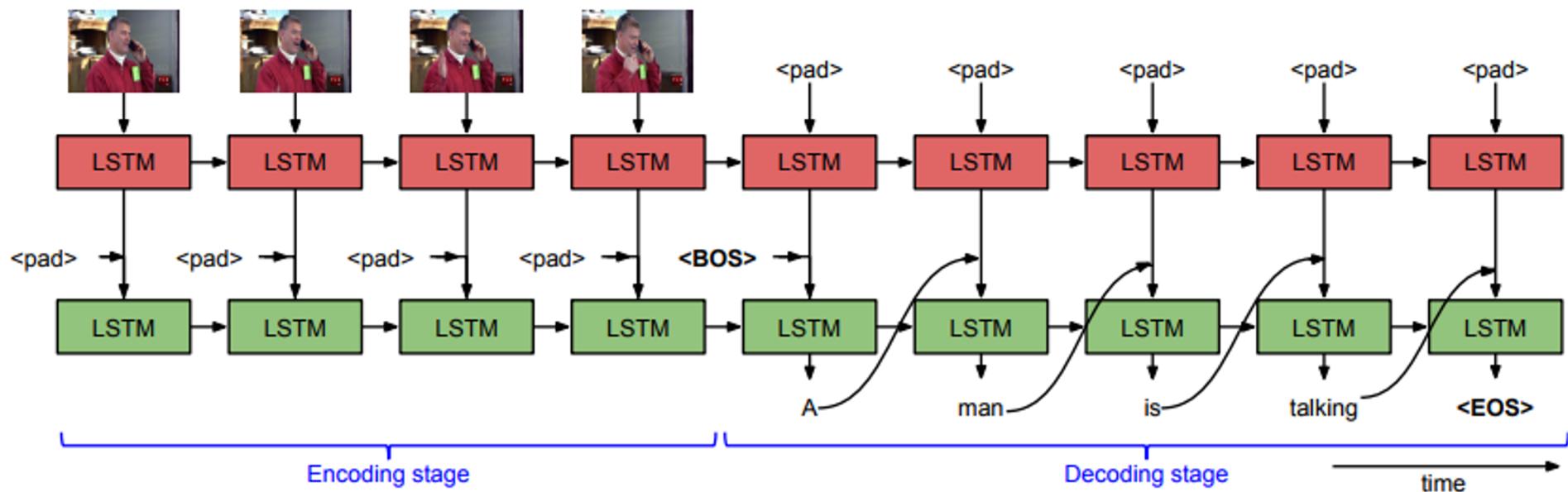
<http://www.cs.utexas.edu/users/ml/papers/venugopalan.iccv15.pdf>



Sequence-to-sequence - S2VT

5/5

- Sequence-to-Sequence Based Model : S2VT
 - Two layer LSTM structure

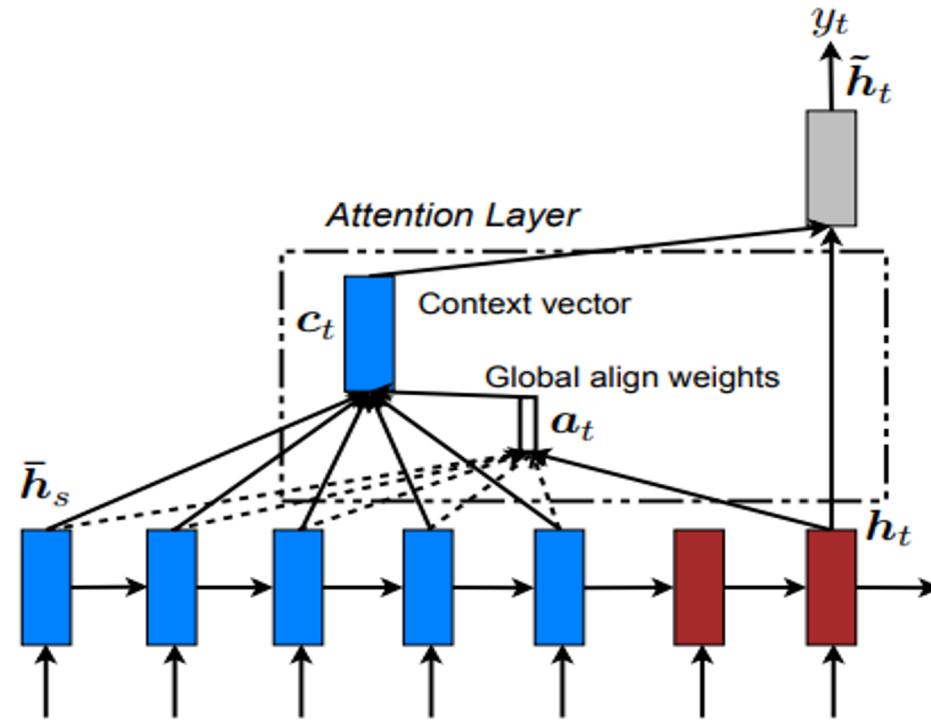




Training Tips - Attention

1/3

- Attention on encoder hidden states :
 - Allow model to peek at different sections of inputs at each decoding time step

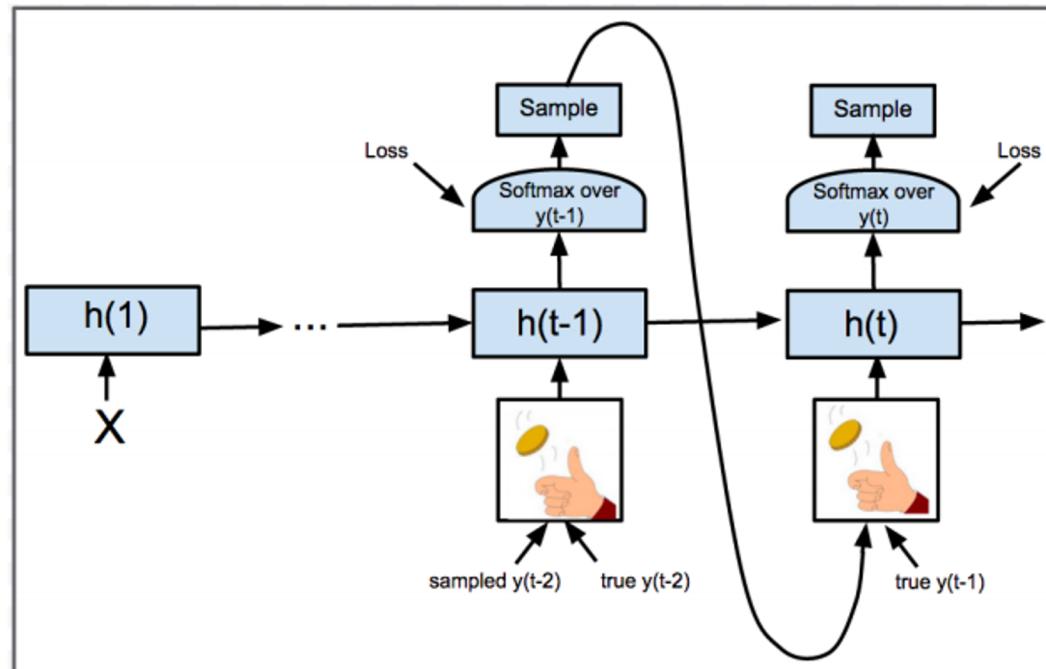




Training Tips - Schedule Sampling

2/3

- Schedule Sampling :
 - To solve “exposure bias” problem,
When training, we feed (groundtruth) or (last time step’s output) as input at odds

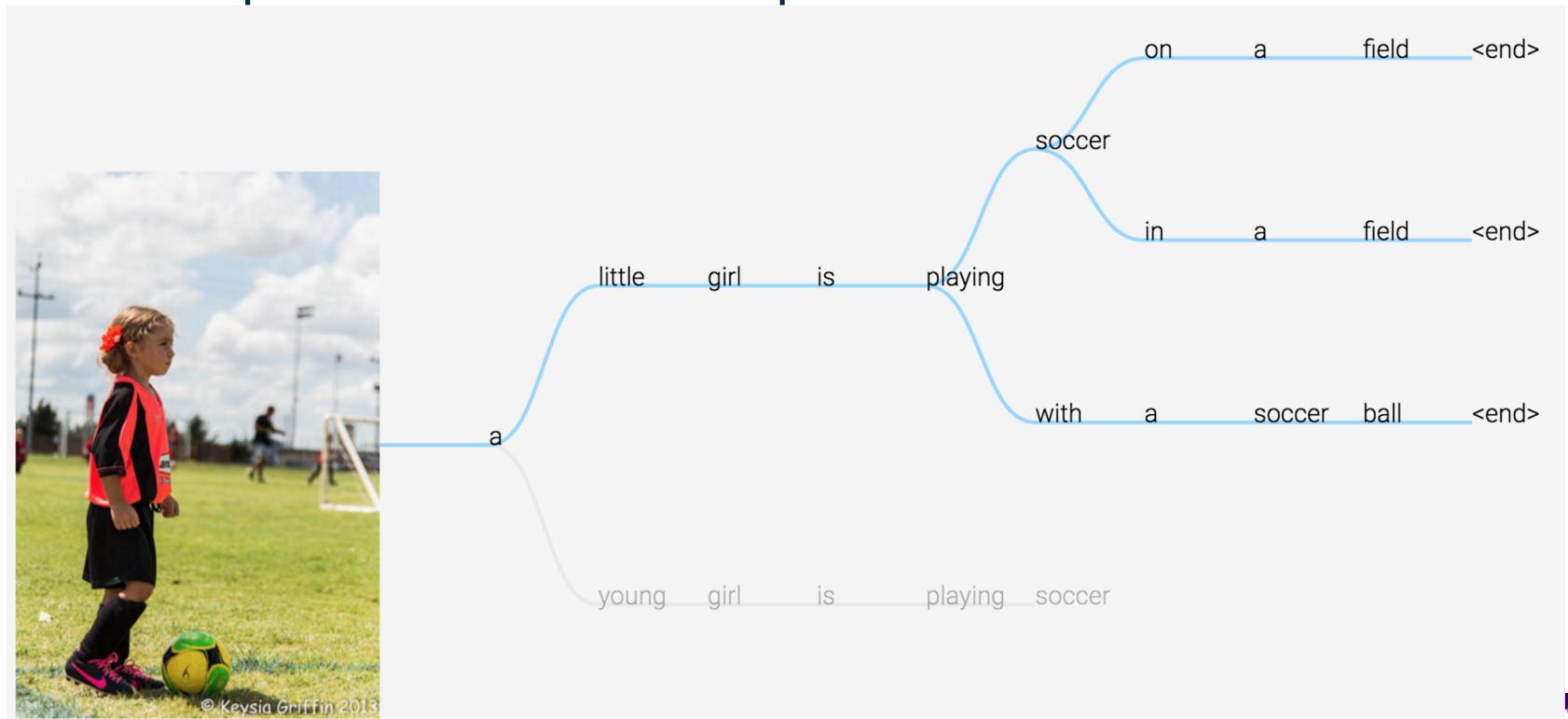


<https://arxiv.org/abs/1506.03099>



Training Tips - Beam search 3/3

- Beam search :
 - keep a fixed number of paths



Demo : <http://dbs.cloudcv.org/captioning>



How to reach the baseline ?

1/2

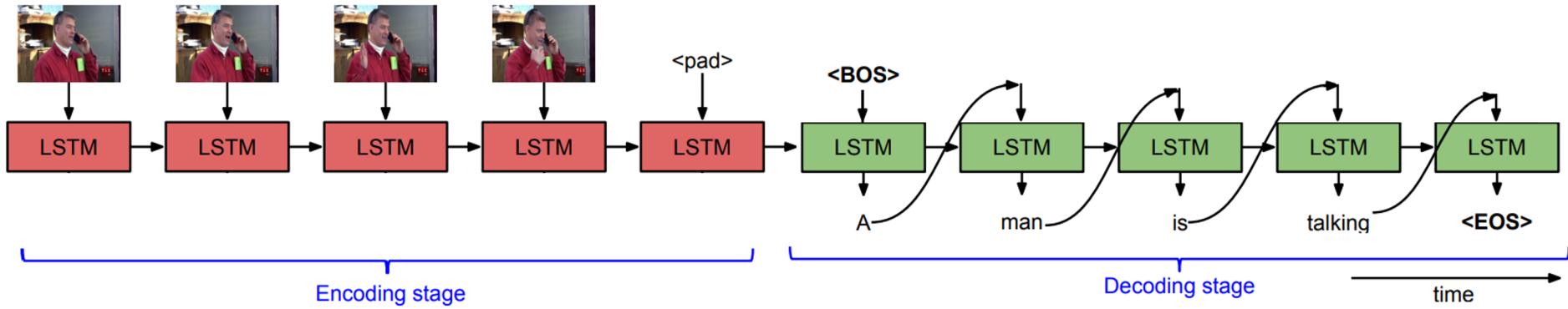
- Evaluation : BLEU@1
 - Precision = correct words / candidate length
 - $$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}$$
where c = candidate length, r = reference length
 - BLEU@1 = BP * Precision
 - e.g. :
Ground Truth : *a man is mowing a lawn*
Prediction : *a man is riding a man on a woman is riding a motorcycle*
BLEU : $1 * 4/13 = 0.308$
 - <https://www.aclweb.org/anthology/P02-1040.pdf>



How to reach the baseline ?

2/2

- Baseline : BLEU@1 = 0.6 (Captions Avg.)
- baseline model :



- Training Epoch = 200
- LSTM dimension = 256
using 960 TX
- Learning rate = 0.001
- vocab size = min count > 3
- AdamOptimizer
- Training time = 72 mins,



Data & format

- Dataset :
 - MSVD
 - 1450 videos for training
 - 100 videos for testing
- Format :
 - [Download](#) MLDS_hw2_1_data.tar.gz
 - Google will remind “Google Drive can't scan this file for viruses”



Submission & Rules

- Deadline: March. 21th 23:59
- Please implement **one seq-to-seq model** (or it's variant) to fulfill the task
- Extra dataset is allowed to use.
- Allow package :
 - python 3
 - **TensorFlow/pytorch ONLY** for CS and ECE student
 - For non-CS/ECE students, Keras is allowed.



Submission & Rules

- For HW2 :
 - Please write shell script "**hw2_seq2seq.sh**" to run your code and follow the script usage below :
 - `./hw2_seq2seq.sh $1 $2`
 - **\$1: the data directory**
 - \$2: test data output filename (format : .txt)
 - Example1 `./hw2_seq2seq.sh testing_data testset_output.txt`
 - Example2 `./hw2_seq2seq.sh ta_review_data tareviewset_output.txt`

Your script should be done within **10 mins** excluding model downloading.



Submission & Rules

- Upload **code** and **report** of HW2 to Github in **different** directory.
- Your github must have directory **hw2/hw2_1/**, and there should be :
 - (1) **report.pdf**
 - (2) **your_seq2seq_model**
 - (3) **hw2_seq2seq.sh**
 - (4) **model_seq2seq.py** (*training code should include*)
 - If your model are too big for github, upload to a cloud space and **write it in your script to download the model.**
 - Please do not upload any dataset to Github (include external dataset).