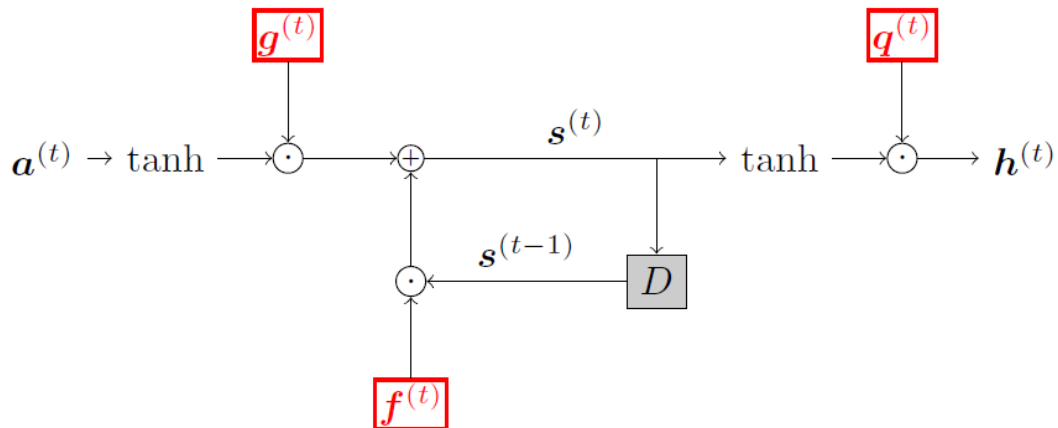


## Lab 2: a LSTM Cell for Image Captioning

### Lab Objective:

In this lab, you have to implement a LSTM cell by yourselves and train an image caption model with your own LSTM cell. After this lab, you will be familiar with LSTM and the basic concept of image captioning.



### Turn in:

Report: 8/22 12:00 pm

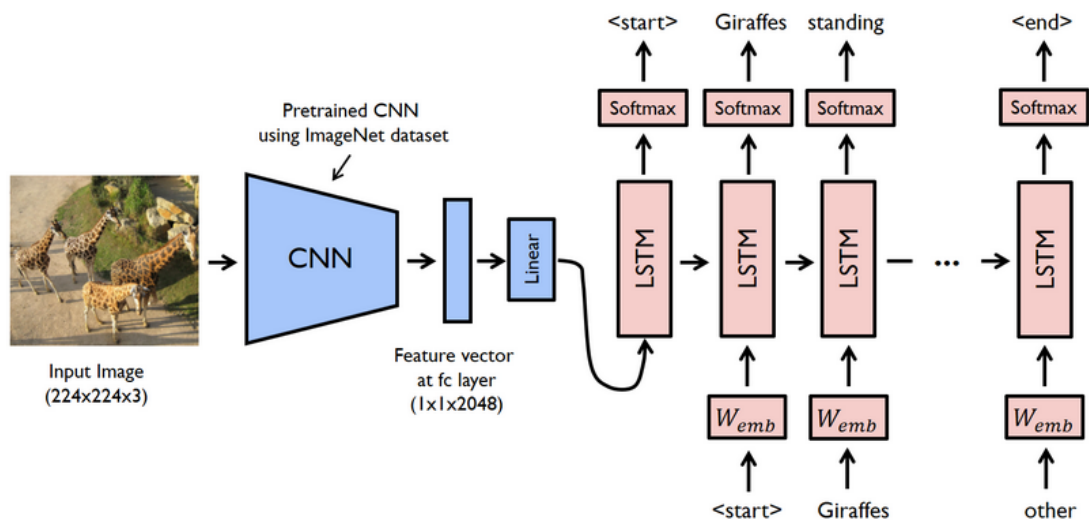
Demo: 8/22 7:00 pm

### Requirements:

1. Implement a LSTM cell
2. Replace the LSTM cell in Pytorch image-caption example with the one you implement.
3. Train an image caption model

### Implementation Details:

- Basic knowledge
  1. Image caption



2. The operation mechanism of RNNs
3. Long short-term memory (LSTM)
  - Memory state:  $s^{(t)}$
  - Input gate:  $g^{(t)} = \sigma(U^g x^{(t)} + W^g h^{(t-1)})$
  - Output gate:  $q^{(t)} = \sigma(U^o x^{(t)} + W^o h^{(t-1)})$
  - Forget gate:  $f^{(t)} = \sigma(U^f x^{(t)} + W^f h^{(t-1)})$
  - New content:  $a^{(t)} = U x^{(t)} + W h^{(t-1)}$
  - Memory update:  $s^{(t)} = f^{(t)} \odot s^{(t-1)} + g^{(t)} \odot \tanh(a^{(t)})$
  - Hidden unit update:  $h^{(t)} = q^{(t)} \odot \tanh(s^{(t)})$
  - Output unit update:  $o^{(t)} = V h^{(t)}$
4. Function `torch.nn.utils.rnn.pack_padded_sequence` (model.py line 42)

#### ● Templates

1. Clone <https://github.com/2018-summer-DL-training-program/Lab-2-Image-Captioning/blob/master/README.md>
2. Change the root of data directory
3. Implement a LSTM cell (DIY\_LSTM.py)
4. Train the model

#### Reference

1. Pytorch nn.LSTM:
  - <https://pytorch.org/docs/stable/nn.html#torch.nn.LSTM>
  - Arguments, inputs, outputs
2. Pytorch pack\_padded\_sequence

- [https://pytorch.org/docs/stable/nn.html#torch.nn.utils.rnn.pack\\_padded\\_sequence](https://pytorch.org/docs/stable/nn.html#torch.nn.utils.rnn.pack_padded_sequence)
3. Pytorch tutorial image captioning
- [https://github.com/yunjey/pytorch-tutorial/tree/master/tutorials/03-advanced/image\\_captioning](https://github.com/yunjey/pytorch-tutorial/tree/master/tutorials/03-advanced/image_captioning)
  - model.py

### Demo

- Given an image, please generate corresponding description



<start> a group of people riding bikes down a street . <end>



<start> a group of giraffes are standing in a field . <end>

- One or two questions related to LSTM

### Report Spec

1. Introduction (5%)
2. Explain how you implement LSTM (45%)
3. Results – generating corresponding descriptions
  - A. example.png (10%)
  - B. ext.png (10%)
4. Discussion (10%)
5. Demo (20%)