# Lab 7: Caption generation with visual attention

## Lab Objective:

In this project, you are going to run a Caption generator by using CNN and RNN language generator to generate a sentence that describes the image.

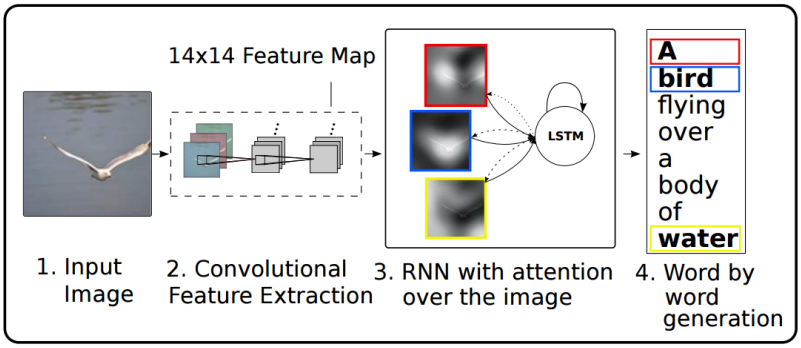
## Turn in:

Report: 4/25(二) 18:00

Demo: 4/25(二) 下課後

## Lab Description:

- Learn how to combine CNN features and RNN language generator.
- Compare two different attention mechanisms.
- The structure is shown below.




## Environment Setup:

- numpy
- matplotlib
- scipy
- scikit-image
- hickle
- Pillow

## Implementation Details:

- Before training, you need to
  - Download training and testing data from coco.
  - Install requirements.
  - Upgrade code for tensorflow 1.0 (model.py, solver.py).
  - Deal with the memory problem.
    - prepro.py main()
    - utils.py load_coco_data(), sample_coco_minibatch()
    - solver.py train()*, test()
  - epoch=10

## Requirements:

1. Implement hard attention.
2. Show the result of two attention mechanisms.

## References:


[1] https://github.com/yunjey/show-attend-and-tell

[2] Xu, K. *et al*. Show, attend and tell: Neural image caption generation with visual attention. In *Proc. International Conference on Learning Representations* http://arxiv.org/abs/1502.03044 (2015).

[3] http://people.ee.duke.edu/~lcarin/Yunchen9.25.2015.pdf

[4] https://www.tensorflow.org/install/migration


## Report Spec: [black: Demo, Gray: No Demo]

1. Introduction (15%, 15%)
2. Experiment setup (15%, 15%)
3. Result (30%, 40%)
4. Discussion (20%, 30%)

Demo (20%) [抽 20 人]