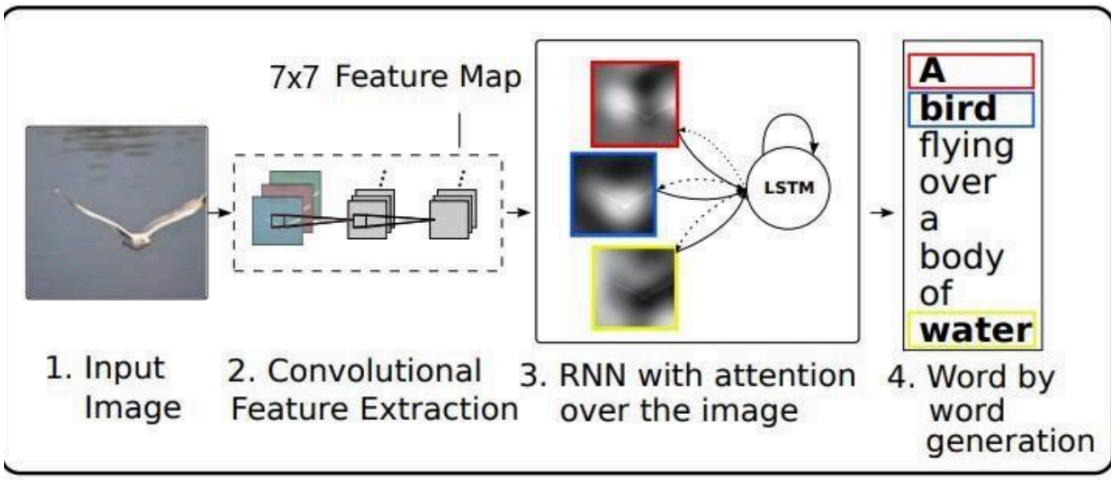
Deep Learning and Practice – Lab3 report

Introduction

Image Caption

將圖片 Input 給電腦後，電腦可以生成一段文字來描述這張圖片，也就是所謂的「看圖說故事」。

使用 Encoder-Decoder 網路結構，先對圖片進行 Encode，將圖片轉換成一個 Feature vector 後，將 Vector 作為 Decoder 的 Input，最後即可產生一段文字。再 Image Caption 中，Encoder 為一個 CNN 網路，而 Decoder 為一個 RNN 網路。



Attention

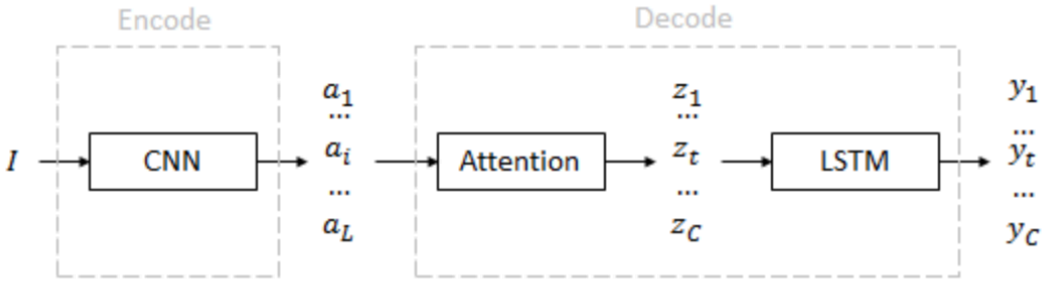
在 Decode 階段時，希望圖片的 Feature 能與 RNN 的時序有所對應，像是表示「鳥」這個物件的 Feature，能在解碼出「鳥」這個文字階段時能有較大的權重，如此能提高 Image caption 的 performance。因此在做 Decode 前，會先計算出該階段下各個 Feature 的權重，也就是 Attention model 所做的事，將此權重與 Feature 做個連結後，再作為 RNN 的 Input。

A giraffe standing in a forest with trees in the background.

A stop sign is on a road with a mountain in the background.

Show attend and tell

在 Image Caption 中結合了 Attention 機制的算法。



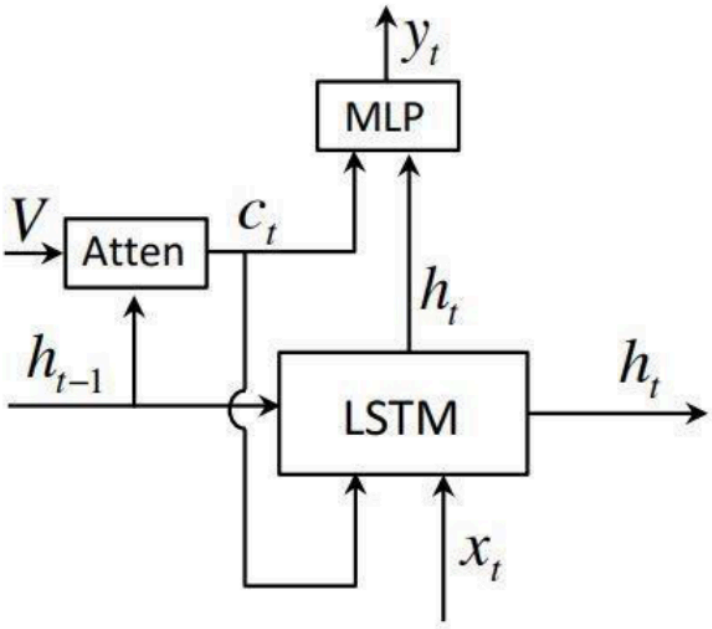
Experiment setup

Model detail

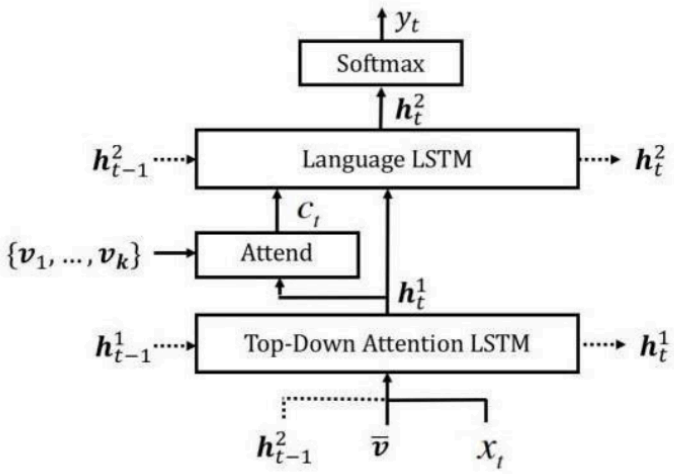
Encoder : CNN - ResNet101

Decoder with Attention Model :

Show attend and tell



Bottom-Up and Top-Down Attention



Parameters

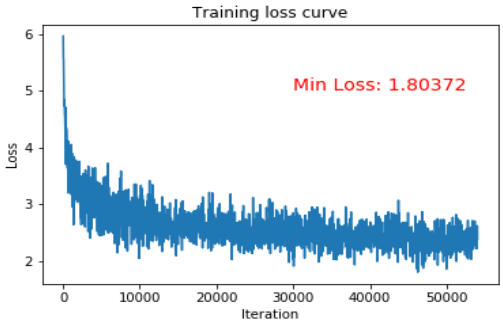
Training parms.

- Epoch : 5
- Batch size : 10
- Input encoding size : 512
- Rnn size : 512
- Att. hid. size : 512
- Fc feat. size : 2048
- Att. feat. size : 2048
- Rnn type : LSTM

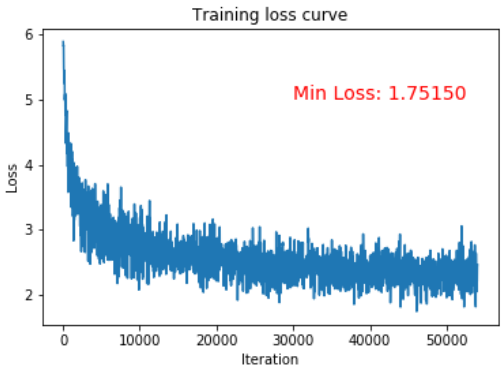
Result

Training loss

Show attend and tell

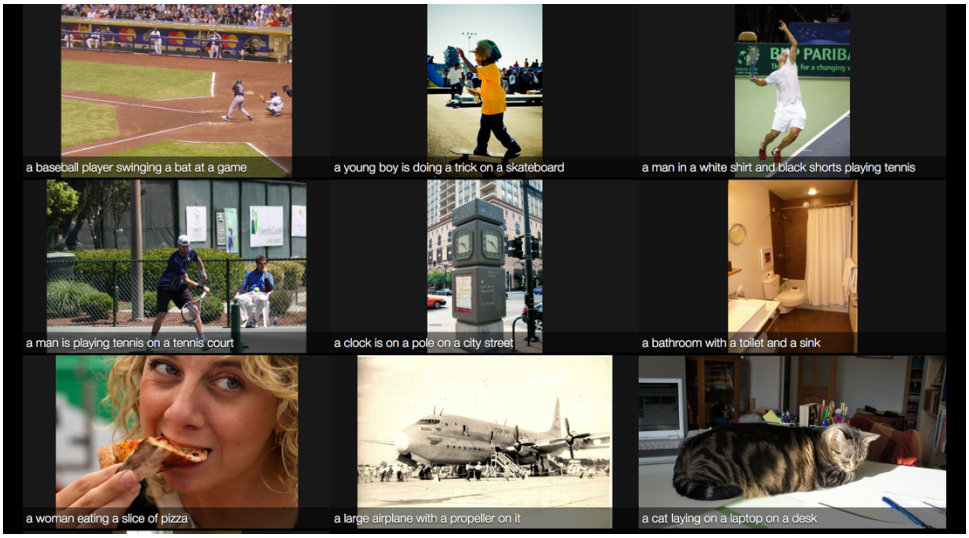


Bottom-Up and Top-Down Attention

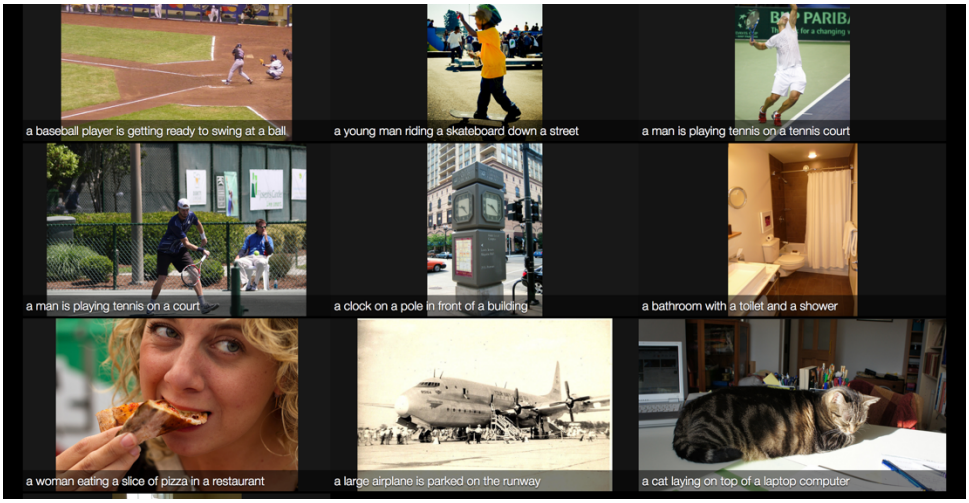


Caption of models

Show attend and tell

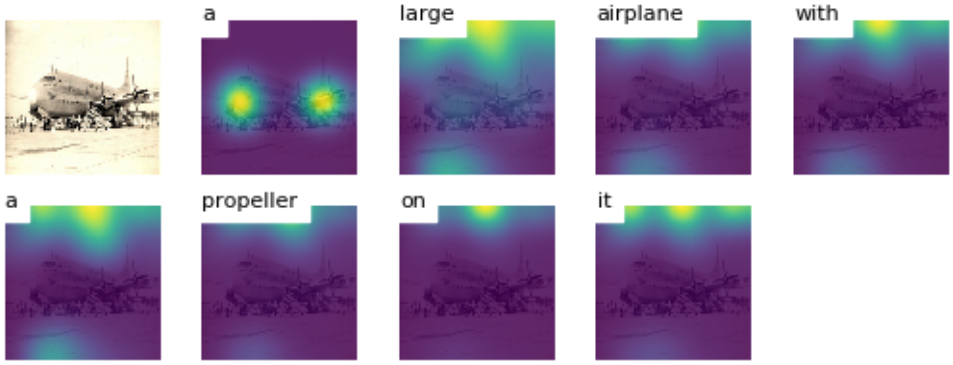


Bottom-Up and Top-Down Attention

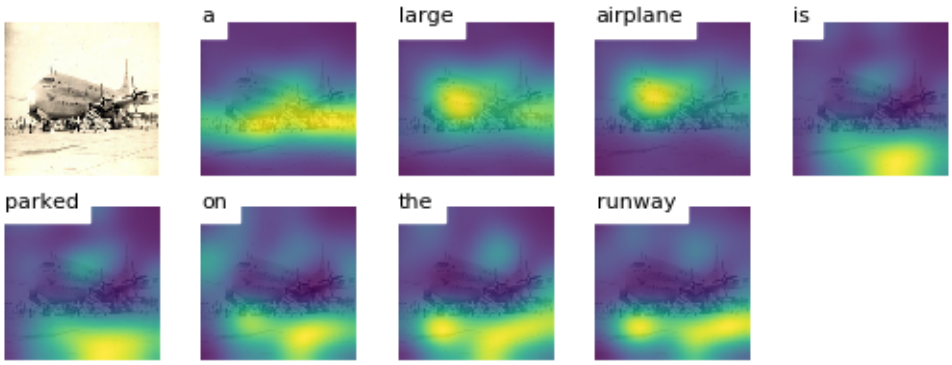


Attention over time

Show attend and tell



Bottom-Up and Top-Down Attention



Discussion

Attention visualization

將 Attention model 計算出來各個 CNN 產生出來 Features 的權重提取出來後，將其 Up-sample 回圖片的 Size，即可與圖片結合進行 Attention visualization。

在我們的 Sample code 中，即提取 ShowAttendTellCore()或 TopDownCore() -> Attetion()中 forward() function 的 **weight**，將每一個時序的 weight(也就是對應到每一個 word)都取出來進行 visualiztion 即可。

Reference

[0] GitHub – ImageCaptioning in pytorch

https://github.com/ruotianluo/ImageCaptioning.pytorch
[1] Show attend and tell (1) :
https://blog.csdn.net/shenxiaolu1984/article/details/51493673
[2] Image Caption :
https://www.cnblogs.com/Determined22/p/6914926.html
[3] Show and tell (1) :
https://zhuanlan.zhihu.com/p/27771046
[4] Bottom-Up and Top-Down Attention (1) :
https://zhuanlan.zhihu.com/p/36151033
[5] Bottom-Up and Top-Down Attention (2) :
https://blog.csdn.net/sinat_26253653/article/details/78436112
[6] GitHub - Attention visualization :
https://github.com/alecwangcq/show-attend-and-tell/blob/master/visualize.ipynb