

隊長：R06922002 姚嘉昇 組員：R06942054 潘仁傑、R06522620 王仁蔚

● 分工表

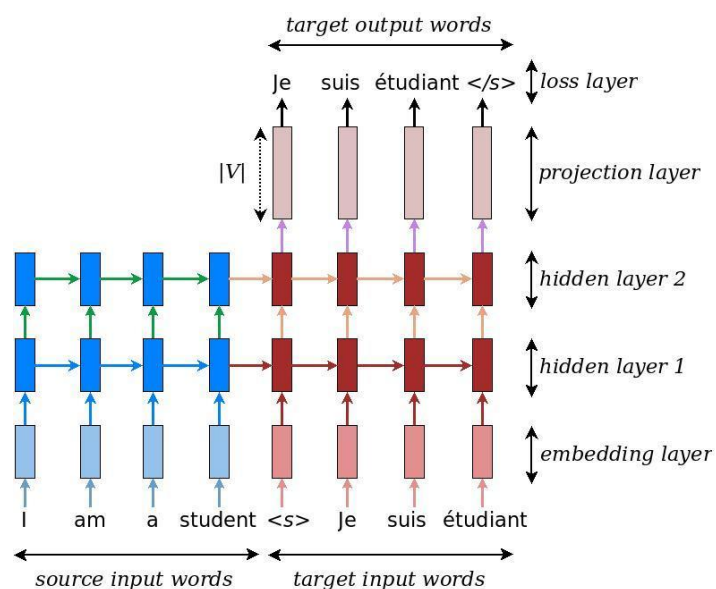
	2-1	2-2	Report 2-1	Report 2-2
姚嘉昇 R06922002		嘗試 pre-train w2v + beam search		
王仁蔚 R06522620		嘗試 attention		
潘仁傑 R06942054		嘗試 scheduling		

● Model description (3%)

參考 Tensorflow 之 Neural Machine Translation (seq2seq) Tutorial 架構圖 (如下圖)，建構這次作業所使用的基礎模型，並將圖中 source input words 改為 source input image。

- ✓ 一層 Embedding layer (size = 1024)
- ✓ 兩層 Dynamic RNN (size = 1024)
- ✓ 最後一層 Projection layer

而 improve model 中分別加入 Luong Attention 及 Bahdanau Attention 機制，從 source input image 中挑選出幾個重要的 frame 丟入 seq2seq model。



- How to improve your performance (3%)

1. Write down the method that makes you outstanding (1%)

我們總共嘗試了四種使用不同 attention 機制的 model :

- ✓ Luong Attention
- ✓ Luong Attention with scale
- ✓ Bahdanau Attention
- ✓ Bahdanau Attention with norm

其中 **Bahdanau Attention with norm** 的 **BLEU@1** 最高，比未加 **attention** 提升 **0.02**

2. Why do you use it (1%)

- 1、不論在 ML 或在 MLDS 都有聽過李老師提到 Attention，但都沒有實際使用過，趁著這次作業，實際 implement Attention 的機制
- 2、Attention 機制能有效增進 seq2seq 輸入長序列圖片之效果
- 3、Attention Mechanism 對輸入的 X 每個部分賦予不同的權重，抽取出更加關鍵及重要的資訊，使模型做出更加準確的判斷
- 4、在 Bahdanau paper 中傳輸路徑是 $h(t-1) \rightarrow a(t) \rightarrow c(t) \rightarrow h(t)$ ，但 Luong 的路徑是 $h(t) \rightarrow a(t) \rightarrow c(t) \rightarrow \tilde{h}(t)$ ，想藉由這次實驗了解兩者結果的差異。

3. Analysis and compare your model without the method. (1%)

沒有使用 attention 的 **BLEU@1** = 0.7010

Bahdanau Attention 的 **BLEU@1** = 0.6965

Bahdanau Attention with norm，**BLEU@1** = 0.7204

可以發現若只單純加上 Bahdanau Attention 反而變差，若將 Attention 初始的 weight 做 normalize 之後提升到 0.7204，根據 <https://arxiv.org/abs/1704.00784> 中 2.4 段有提到 weight normalization 可以降低 energy terms 的影響，也許是因為如此 BLEU 進步，但 Luong Attention with scale 卻沒有好的結果，這也是尚待討論的議題

- Experimental results and settings (1%)

我們使用四種 Attention，分別為 Luong Attention、Luong Attention with scale、Bahdanau Attention、Bahdanau Attention with norm，這四種 Attention model 的架構及參數都一樣，只差在不同的 attention 機制，以下為 model 參數設定：

rnn_size = 1024	num_layers = 2 (RNNlayer 數)	dim_video_feat = 4096
embedding_size = 1024	learning_rate = 0.0001	batch_size = 29
max_gradient_norm = 5	max_encoder_steps = 64	max_decoder_steps = 15
sample_size = 1450	dim_video_frame = 80	

其中 sample_size 指的是每個 epoch 只取 1450 筆 data 做 training。

下表是基於同架構同參數，但替換不同的 attention 機制得到的 BLEU@1 結果。

Method	BLEU@1
Without Attention	0.7010
Luong Attention	0.7054
Luong Attention with scale	0.6977
Bahdanau Attention	0.6965
Bahdanau Attention with norm	0.7204

以下隨機節錄了一些，Without Attention 及 Bahdanau Attention with norm 的 output 結果(上排為 Without Attention，下排為 **Bahdanau Attention with norm**，粗體為較優)：

a woman is cutting a slab of tofu into small (label 中沒有 slab、small)

a woman is cutting some meat (label 中沒有 meat)

a woman is adding some ingredients to a bowl of water (label 中沒有 adding、water)

a woman is adding some meat to a bowl (label 中沒有 adding, water, meat)

a person is adding water to a bowl (label 中沒有 adding 及 water)

a man is adding some ingredients to a bowl (label 中有 ingredients)

a man is slicing a piece of food (label 中並沒有出現 food)

a woman is slicing a vegetable (label 中有 vegetable)

a man is walking with a large in a garden (label 中並沒有出現 walking)

a man is singing (label 中並沒有出現 singing 但句子較短)

可以發現抽出來的五句是 **Bahdanau Attention with norm** 預測較準確。