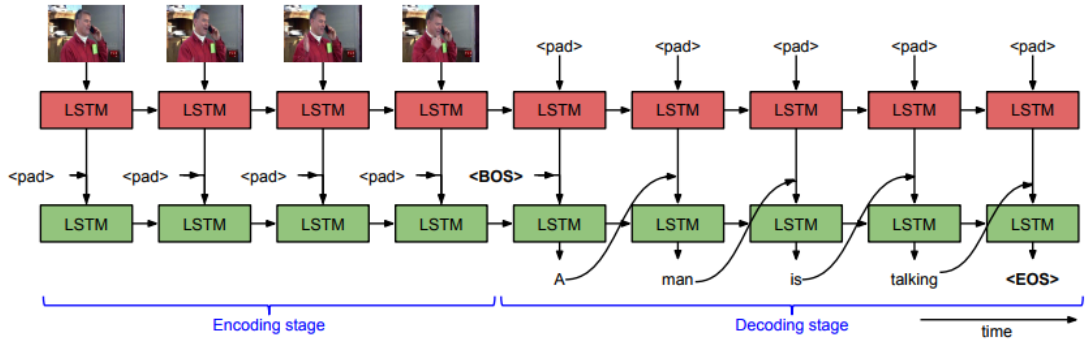
MLDS HW2 report

R04943151 梁可擎

Model Description

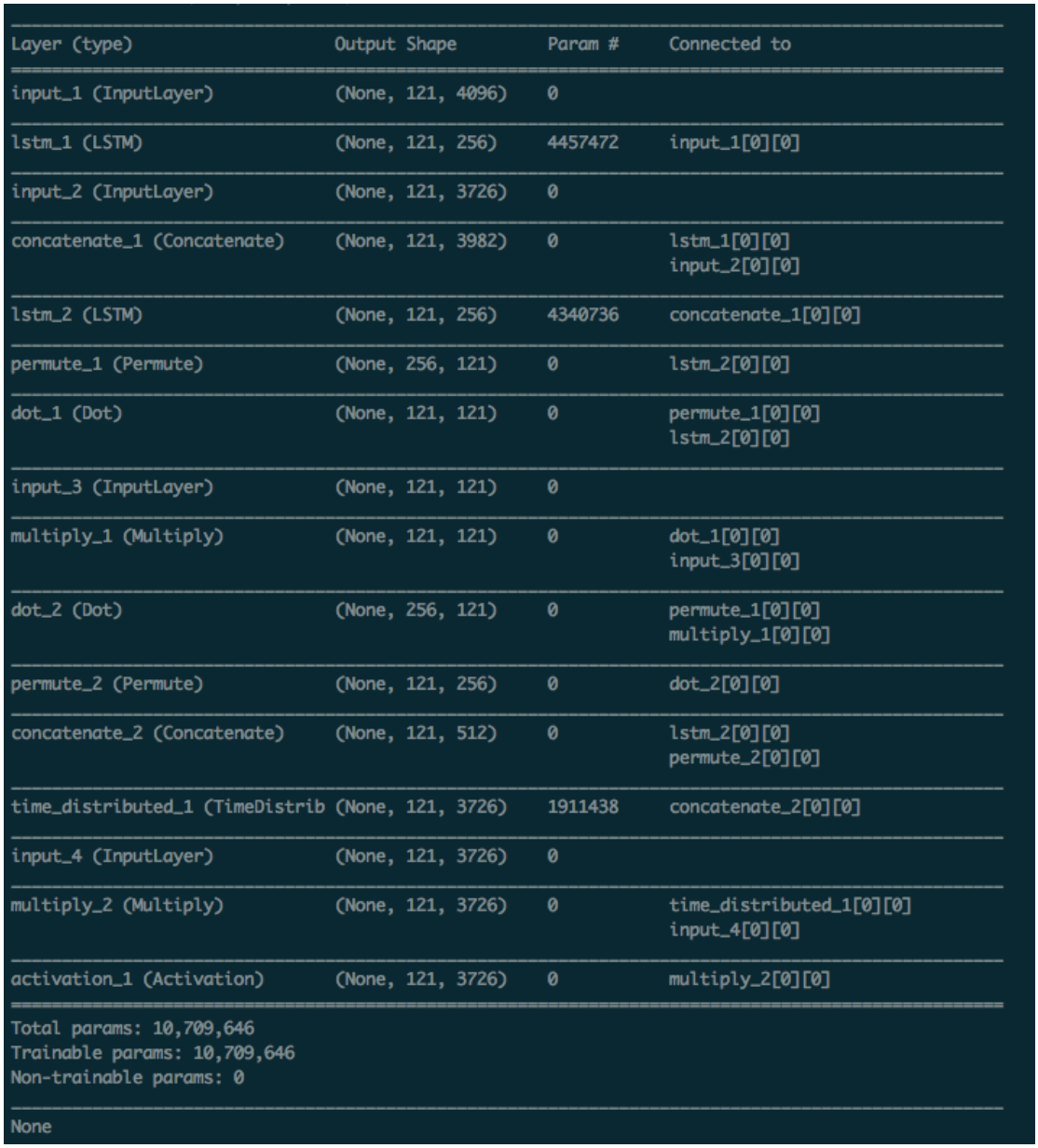
這次的作業模型，依照投影片建議實作 S2VT 的 model，如下圖：



兩層 LSTM 再過一層 Dense。因為一開始選用 Keras，後來發現直接加兩層 LSTM 會有 Encoder 的 loss 算到 categorical_cross_entropy 的問題。因此在過了 Dense 之後做 Softmax 前需要用一個 input tensor 將 encoder time steps 的輸出強制歸零。

Attention Mechanism

使用參考論文（Effective Approaches to Attention-based Neural Machine Translation）裡面描述的 dot 的方法，比較方便在 Keras 上實作。在 LSTM2 之後，將 output (None, 121, hidden_dim) 轉置然後做內積，得到一個 (121, 121) 的 tensor，為了只保留我們要的 weights，會用一個 input mask tensor 將只有 encoder output 內積的地方跟 decoder outputs 去掉。再將 weight 與 LSTM2 outputs 內積後再 concatenate 回原本的 LSTM2 output 得到 (121, 512) 的 tensor 再去做 Dense 和 Activation。如下圖：



Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 121, 4096)	0	
lstm_1 (LSTM)	(None, 121, 256)	4457472	input_1[0][0]
input_2 (InputLayer)	(None, 121, 3726)	0	
concatenate_1 (Concatenate)	(None, 121, 3982)	0	lstm_1[0][0] input_2[0][0]
lstm_2 (LSTM)	(None, 121, 256)	4340736	concatenate_1[0][0]
permute_1 (Permute)	(None, 256, 121)	0	lstm_2[0][0]
dot_1 (Dot)	(None, 121, 121)	0	permute_1[0][0] lstm_2[0][0]
input_3 (InputLayer)	(None, 121, 121)	0	
multiply_1 (Multiply)	(None, 121, 121)	0	dot_1[0][0] input_3[0][0]
dot_2 (Dot)	(None, 256, 121)	0	permute_1[0][0] multiply_1[0][0]
permute_2 (Permute)	(None, 121, 256)	0	dot_2[0][0]
concatenate_2 (Concatenate)	(None, 121, 512)	0	lstm_2[0][0] permute_2[0][0]
time_distributed_1 (TimeDistrib	(None, 121, 3726)	1911438	concatenate_2[0][0]
input_4 (InputLayer)	(None, 121, 3726)	0	
multiply_2 (Multiply)	(None, 121, 3726)	0	time_distributed_1[0][0] input_4[0][0]
activation_1 (Activation)	(None, 121, 3726)	0	multiply_2[0][0]

Total params: 10,709,646
Trainable params: 10,709,646
Non-trainable params: 0

None

How to Improve Performance

在訓練的過程中發現多選用不同的 captions 組合相當有效。即是同一組 captions（一個 batch）不用訓練太多個 epochs，隨機多使用幾組不同的 captions。例如一開始用 10 組 captions 訓練了的 bleu score 剛好過 baseline，使用 20 組之後有提升到 0.27。可能是多看不同的字詞組合會增加學到的東西。0.27 的參數是 hidden_dim = 256、20 iterations，model.fit 的 batch_size = 32、epoch = 5。

Experimental Results

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 121, 4096)	0	
lstm_1 (LSTM)	(None, 121, 128)	2163200	input_1[0][0]
input_2 (InputLayer)	(None, 121, 3726)	0	
concatenate_1 (Concatenate)	(None, 121, 3854)	0	lstm_1[0][0] input_2[0][0]
lstm_2 (LSTM)	(None, 121, 128)	2039296	concatenate_1[0][0]
time_distributed_1 (TimeDistrib	(None, 121, 3726)	480654	lstm_2[0][0]
input_3 (InputLayer)	(None, 121, 3726)	0	
multiply_1 (Multiply)	(None, 121, 3726)	0	time_distributed_1[0][0] input_3[0][0]
activation_1 (Activation)	(None, 121, 3726)	0	multiply_1[0][0]

Total params: 4,683,150
Trainable params: 4,683,150
Non-trainable params: 0

None

原本的 S2VT，hidden_dim = 256 經過 20 iterations 後得到 Average bleu score 0.272592008029。

10 iterations 5 epochs for attention_real.h5 loss: 0.1185

10 iterations 5 epochs for attention_real.h5 with mask = [:,80:,:] loss: 0.0225

100 iterations 1 epoch for attention_epoch100.h5 loss: 0.1497 Average bleu score is 0.26398077121657304