

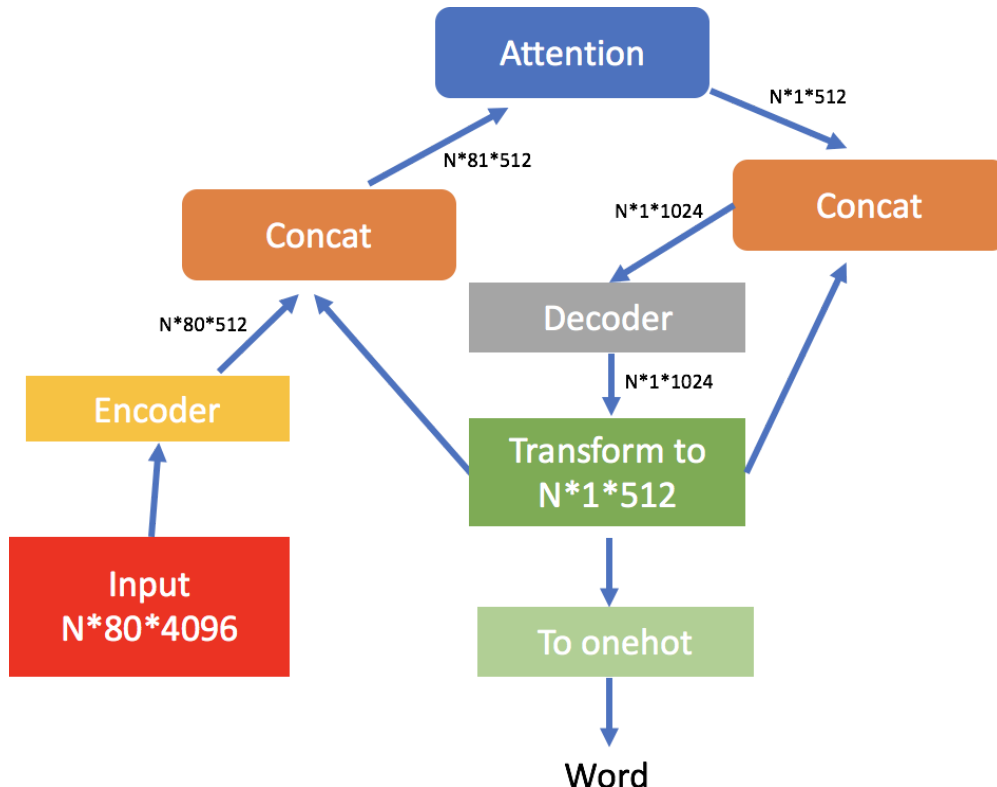
MLDS HW2 Report

許芯瑜、李祐賢、熊展軒

HW2-1 Video Caption Generation:

● Model description: (3%)

- 使用S2VT 架構 (兩個LSTM cells 作為encoder 及 decoder) 。
- Decoder unit 數量為1024 而encoder unit 數量為一半512。
- Decoder output 1024 維會經過一個transform 變成一個512維之輸出。
- Encoder 輸出 ($N*80*512$) 和decoder 之前一個輸出合併(concat)成 ($N*81*512$)之tensor 作為attention之輸入，attention會計算各encoder output 之權重並將各encoder outputs 之weighted sum 輸出給當前之decoder。
- Attention 輸出後會和前一個decoder output 合併(concat) 成1024 維再當作當前decoder 之輸入。
- 產生一個word 知過程如圖：



● Improve Performance: (3%)

- 我們試著用attention 做為提升performance 的手段，但結果並不理想。
- 會選擇使用attention 是因為如果model 可以自己學出應該要在哪個 frame 取得有效的資訊，應該會讓model 更可以解釋，也更相似於人類智慧的概念。
- 比較使用attention 前後之解果發現，bleu 分數差距不大 (attention 0.709, without attention 0.710)，且產生之字句相當相近甚至大多句子

是完全相同的，我們推測應該是沒有對attention 產生之權重作 regularization 導致有做跟沒做差不多。

Attention	Without Attention
<pre> 1 ScdUht-pM6s_53_63.avi,a woman is a a 2 wkgGxsuNVSg_34_41.avi,a man is down down a 3 BtQtrGI0F2Q_15_20.avi,a is is a a 4 k06Ge9ANKM8_5_16.avi,a baby is playing a 5 szf3VDsdDPM_107_114.avi,a man is a a a 6 shPymuahrs_5_12.avi,a dog is a a 7 XOAguUVVwKEA_8_20.avi,a baby is is 8 uffT2BWh3BQ_0_8.avi,a panda panda is 9 5YJaS2Eswg0_22_26.avi,a woman is riding a up 10 lw7pTwpX0K0_38_48.avi,a man is a a 11 UbmZAe5u5FT_132_141.avi,a woman is chopping a 12 xCFCXzDUGjY_5_9.avi,a man is a a 13 He7Ge7SogrK_47_70.avi,a man is a a a a 14 tJHUH9tpgPg_113_118.avi,a woman is pouring a a 15 n016qlw8Q30_2_11.avi,a man is up a 16 RjpbFLOHfPs_8_25.avi,a man are walking down the 17 6JnGBs88sL0_4_10.avi,a man is a a 18 EpMuCrbxE8A_107_115.avi,a man is a a 19 HAJwXjwN9-A_16_24.avi,a men are a a 20 4xVGpDma4LE_23_33.avi,a man is a a 21 k5OKBK2e7xA_19_32.avi,a group is playing with a 22 Jag7oTemldY_12_25.avi,a man is a a are 23 8MVo7fje_oE_125_130.avi,a man is pasta out out 24 bqMmyYlImkI_0_14.avi,a woman is a a 25 jTnrm338_KY_34_42.avi,a man is a a 26 UdCObAQ5OOM_15_30.avi,a man is a a a 27 4PcL6-mjRNk_11_18.avi,a man is a up a 28 3qqEKTpXlNs_1_15.avi,a baby is is a a 29 glriJRGnmc0_211_215.avi,a man is a a 30 q7pOfn8s4zc_263_273.avi,a man is a a a 31 mtrCf667KDK_134_176.avi,a man is a a 32 0lh_UWF9ZP4_62_69.avi,a woman is a into into 33 JntMAcTl0F0_50_70.avi,a man is a are the 34 7NNg0_n-bs8_21_30.avi,a man is playing a 35 IhwPQL9dFYc_124_129.avi,a woman is cutting a 36 Baf3LXFUaGs_28_38.avi,a man is playing a 37 6qlD9irxnhZ0_286_295.avi,a man is playing a 38 RZL9irxnhZ0_34_40.avi,a man is a a 39 WWf0Z6ak3Dg_5_15.avi,a woman is playing a a 40 PeUHy0A1GF0_114_121.avi,a woman is shrimp shrimp shrimp 41 KlteYv1Uv9A_27_33.avi,a man is riding a bike </pre>	<pre> 1 ScdUht-pM6s_53_63.avi,a woman is a a 2 wkgGxsuNVSg_34_41.avi,a man is down down a 3 BtQtrGI0F2Q_15_20.avi,a is is a a 4 k06Ge9ANKM8_5_16.avi,a baby is playing a 5 szf3VDsdDPM_107_114.avi,a man is a a a 6 shPymuahrs_5_12.avi,a dog is a a 7 XOAguUVVwKEA_8_20.avi,a baby is is 8 uffT2BWh3BQ_0_8.avi,a panda panda is 9 5YJaS2Eswg0_22_26.avi,a woman is riding a up 10 lw7pTwpX0K0_38_48.avi,a man is a a 11 UbmZAe5u5FT_132_141.avi,a woman is chopping a 12 xCFCXzDUGjY_5_9.avi,a man is a a 13 He7Ge7SogrK_47_70.avi,a man is a a a a 14 tJHUH9tpgPg_113_118.avi,a woman is pouring a a 15 n016qlw8Q30_2_11.avi,a man is up a 16 RjpbFLOHfPs_8_25.avi,a man are walking down 17 6JnGBs88sL0_4_10.avi,a man is a a 18 EpMuCrbxE8A_107_115.avi,a man is a a 19 HAJwXjwN9-A_16_24.avi,a men are a a 20 4xVGpDma4LE_23_33.avi,a man is a a 21 k5OKBK2e7xA_19_32.avi,a group is playing with a 22 Jag7oTemldY_12_25.avi,a man is a a are 23 8MVo7fje_oE_125_130.avi,a man is pasta out out 24 bqMmyYlImkI_0_14.avi,a woman is a a 25 jTnrm338_KY_34_42.avi,a man is a a 26 UdCObAQ5OOM_15_30.avi,a man is a a a 27 4PcL6-mjRNk_11_18.avi,a man is a up a 28 3qqEKTpXlNs_1_15.avi,a baby is is a a 29 glriJRGnmc0_211_215.avi,a man is a a 30 q7pOfn8s4zc_263_273.avi,a man is a a a 31 mtrCf667KDK_134_176.avi,a man is a a 32 0lh_UWF9ZP4_62_69.avi,a woman is a into into 33 JntMAcTl0F0_50_70.avi,a man is a are the 34 7NNg0_n-bs8_21_30.avi,a man is playing a 35 IhwPQL9dFYc_124_129.avi,a woman is cutting a 36 Baf3LXFUaGs_28_38.avi,a man is playing a 37 6qlD9irxnhZ0_286_295.avi,a man is playing a 38 RZL9irxnhZ0_34_40.avi,a man is a a 39 WWf0Z6ak3Dg_5_15.avi,a woman is playing a a 40 PeUHy0A1GF0_114_121.avi,a woman is shrimp shrimp shrimp 41 KlteYv1Uv9A_27_33.avi,a man is riding a bike </pre>

● Experimental Results and Settings: (1%)

○ 實驗設定：

- Batch size 150
- Train 13 hours
- GPU : NVIDIA 1080 Ti
- Optimizer : Adam
- Training tip : Attention

○ 實驗結果：

- bleu 0.710 without attention, 0.709 with attention
- 會產生很多重複的word，不同的影片都會產生相同的關鍵字 (man, woman, a, playing, is)，和人類產生之句子有很大的落差。

● 分工：

- HW2-1: 熊展軒
- HW2-2: 李祐賢、許芯瑜