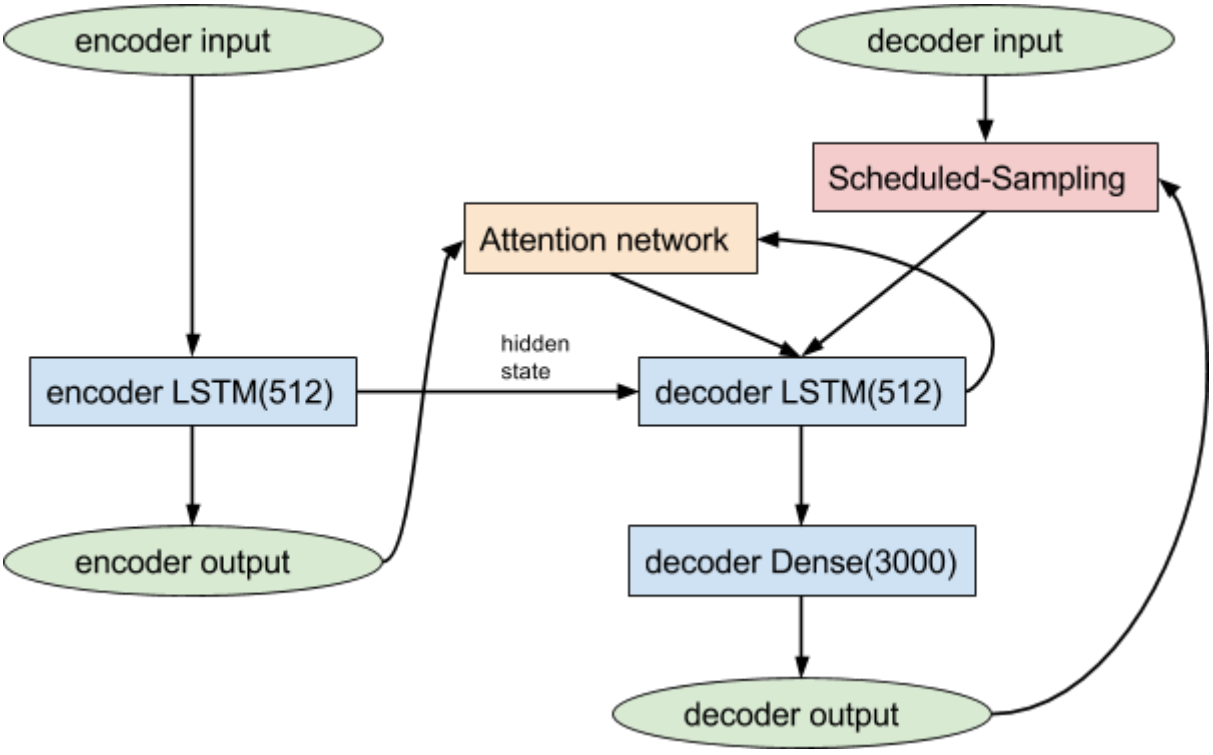
1. Model description (2%) :
    a. seq2seq model (1%) :

主結構 :



encoder input data :
time-step: 80
feature: 4096 dim
decoder input data:
word-dimension: one-hot 3000 word
max-sequence: 50 word
output class:
word-dimension: one-hot 3000 word
max-sequence: 50 word

2. Attention mechanism (2%) :
    a. How do you implement attention mechanism? (1%)
        使用Keras backend及Lambda Layers將每個time-step的decoder state h及encoder outputs分別經過dense transfer後相加，接著經過tanh activation。最後使用一層dense(1, activation=softmax)將每個time-step縮減至一個值，此值即為activation factor，所有time step組成attention vector。
    b. Compare and analyze the results of models with and without attention mechanism. (1%)
        用詞較豐富、較少重疊的'a'，句子結構較完整。
3. How to improve your performance (1%) :
    a. Write down the method that makes you outstanding
        Standardized video feature，Attention mechanism及Scheduled sampling，使用generator產生random training sets。
    b. Describe the model or technique (0.5%)
        Standardized video feature: 將training data video feature取出mean及std，將所有data X 調整至(X - mean)/(2*std)。
        Attention mechanism: 如2.提到的Attention network，接著將attention vector與encoding outputs內積產生contex vector，再將contex vector與decoder input data串接在一起，進入decoder LSTM。
        Scheduled sampling: 使用Lambda Layer即可以調控decoder input data來自ground truth input或是decoder model output，並透過Keras Lambda Callback在每個epoch結尾調整分配係數，一開始採用ground truth的機率為100%，到最後一個epoch採用decoder model output的機率為100%。
    c. Why do you use it (0.5%)
        經過調整與測試，predict出的結果句子較完整，且BLEU score提升。
4. Experimental results and settings (1%) :
    a. latent dim由256調整至512：預測出的句子用詞變得較豐富，且句型結構更完整。
    b. 加入Attention mechanism及Scheduled sampling：減少重複字詞，增加用詞精準度。但BLEU score無法精準反映此現象，有時高分常有斷句，分數較低的結果卻顯得更完整與通順。