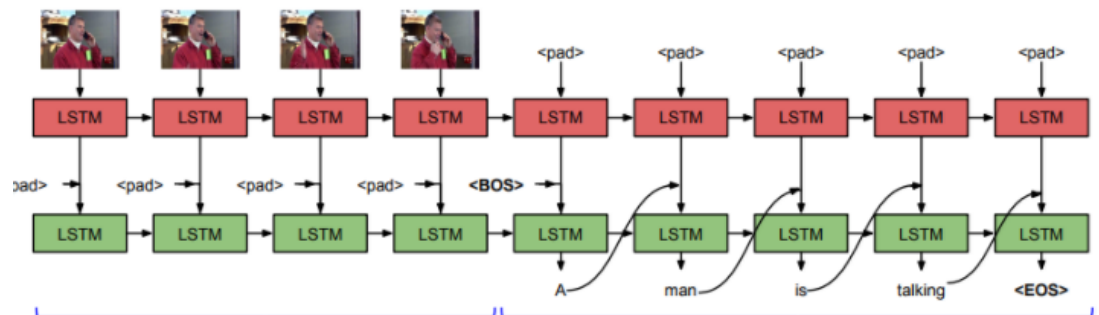


# Report

R05922156

- **Model description (2%)**

我是用助教投影片中的 model – S2VT



大部分都一樣

只有最後加了 dense 層而已

還有一點比較不一樣的地方是綠色那層的 output 我是拉到紅色那層的 input 餵進去

其實我兩種都有試過

但結果好像是從最前面餵會好一點點

- **Attention mechanism(2%)**

我的 attention 也是照投影片裡的

就是額外加一層去 input 所有 encoder 跟當下的 decoder 的 output

然後會吐出目前的 timestep 的最後結果

雖然理論上應該加了 attention 會比較好

但是不知道為甚麼我加了之後變得超爛....

所以我後來就把它拿掉了

- **How to improve your performance (1%)**

一開始還有想過只用單層 LSTM

就是有點像在做翻譯時的 seq2seq 的概念

但單層跑出來蠻爛的

有嘗試過 S2VT 後面多加幾層 dense

有稍微好一點點但 train 的時間大幅上升

所以最後還是只用最基本的 S2VT

- **Experimental results and settings (1%)**

經過了上述的幾次實驗之後就決定用 S2VT 去跑

途中有調整 unit 跟 learning 之類的參數

當我用 adam 但 learning rate 較大時他會很快收斂

可是印出來的東西卻完全不准卻

不知道是不是走到 local optimal 之類的

Unit 方面倒是沒發現太大的差異

最後就用 unit=256 learning rate=0.005 去 train 了

其實大多數時間都花在測試上述 **Model description** 中說的兩種 model 的差異

不過最後還是都沒有過 baseline.....QQ

- **README :**

Code 中主要是用 keras

但因為 keras 有用到 tensorflow 的 backend

所以在自訂義 loss function(加上 mask)那個部分也有用到一點 tensorflow