

# DLCV HW4 LSTM

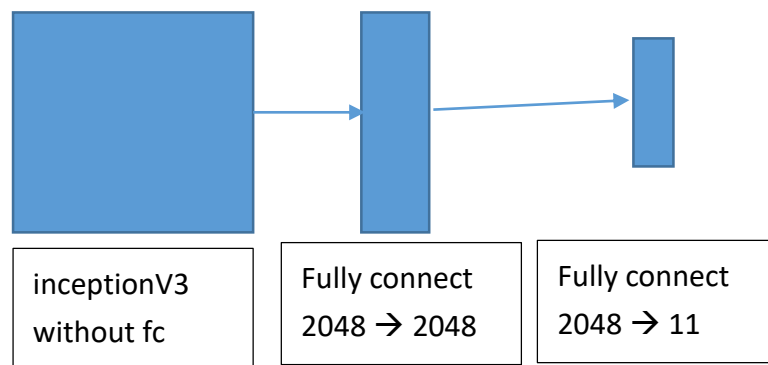
ID : R07922166

Name : 周秉儒

## Problem 1:

第一題只用 CNN 取 feature 然後通過 mean 將 feature 壓成 (1, 2048) 當作 LSTM 的 input.

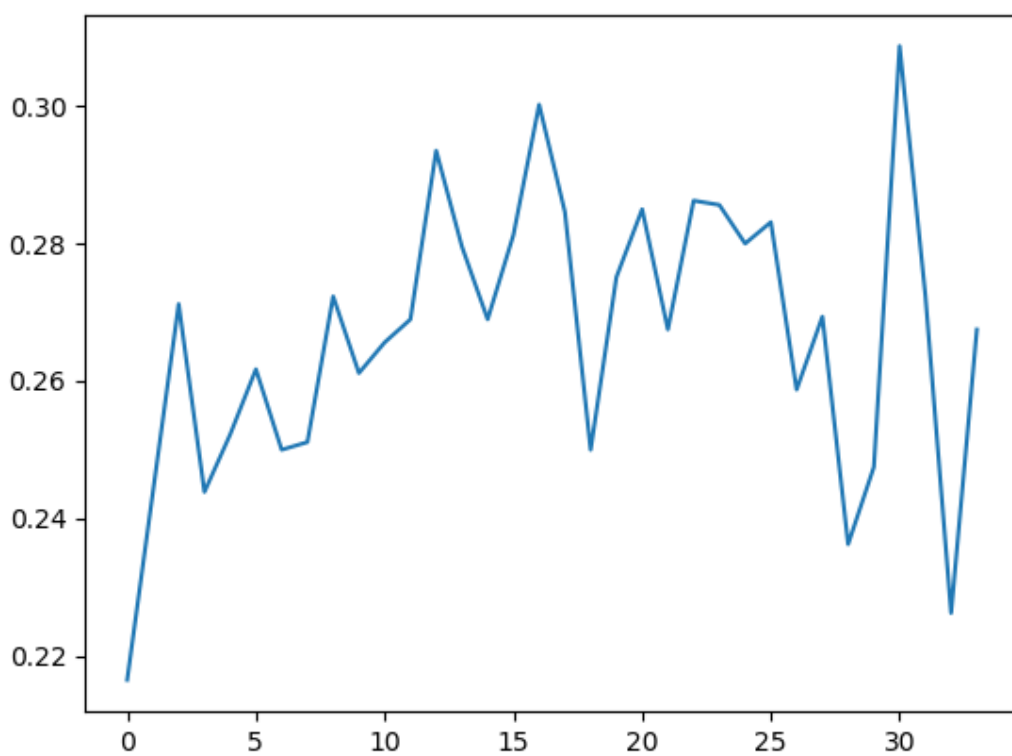
我的 model 架構沒有很複雜但是因為使用的是 inceptionV3 pretrained model, 把最後一個 FC 拿掉, 接上自己的 FC.



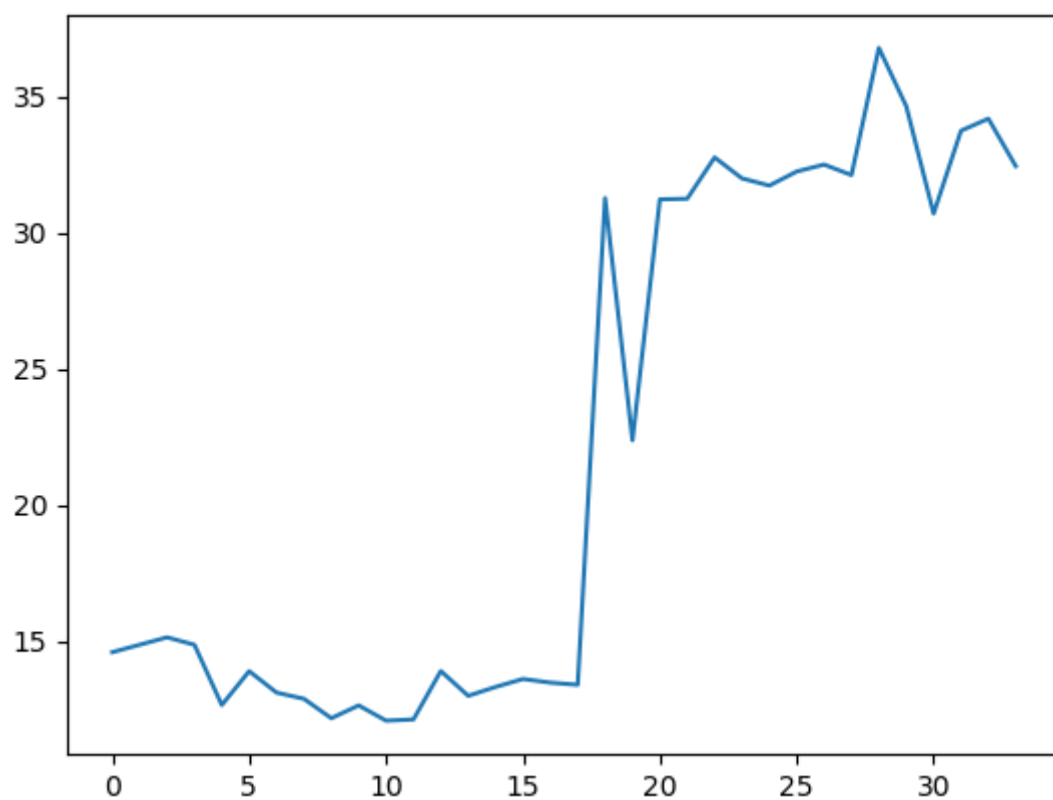
中間有使用 BatchNorm1d, ReLU 等

Loss function 為 Cross Entropy Loss.

以下是我的 learning curve.

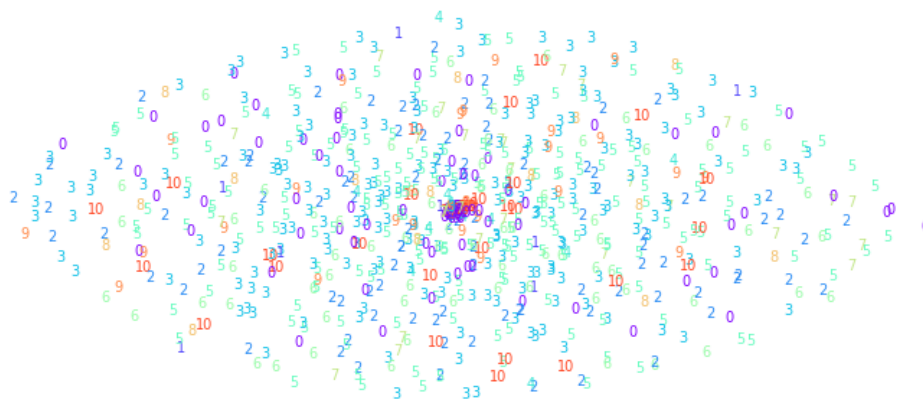


(上圖為準確率)



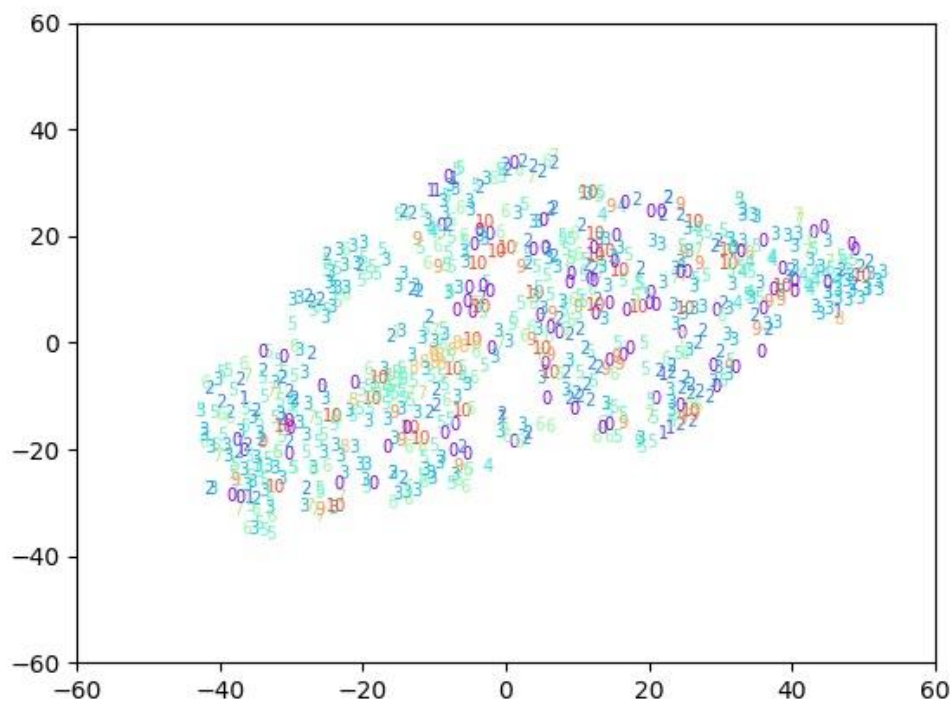
(以上為 loss)

**Loss** 很大是因為我忘記取平均了



將 feature 利用 TSNE 視覺化的結果

結果看起來不是很好，下面為我直接用 inceptionV3 抽出來的 feature 來做 TSNE.



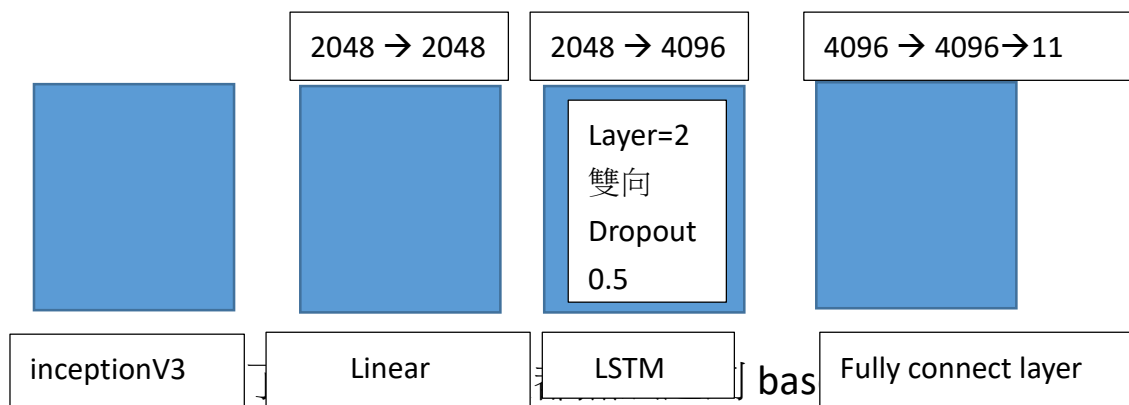
好像也沒有很好，但我怎麼訓練最高都只有到 32%~35%左右。

## Problem 2

第二題是將 CNN 取完的 feature 變成 LSTM 的 input

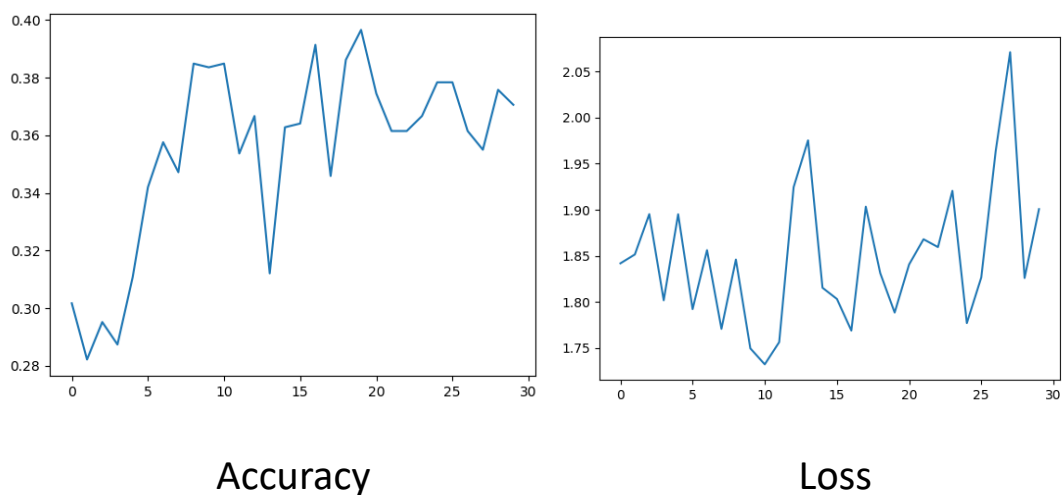
但這題我怎麼 train 都達不到 baseline.

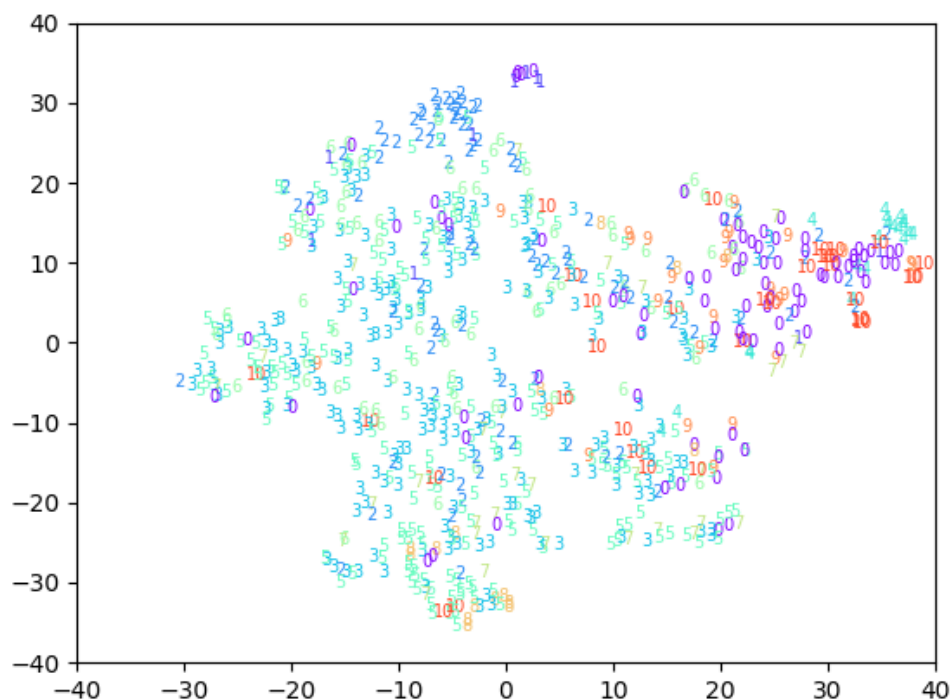
但是 lstm 真的準確率會變高，而且我覺得這次的影片一定要用 Bidirectional 因為 class 3, 5 分別是“get”, “put”完全相反的動作，感覺用 bidirectional 比較合理。



先將抽出來的 feature 通過 FC 降維，在經過 LSTM 然後進行分類，也試過直接將抽出來的 2048 維度的 feature 通過 lstm 進行分類，也試過 Bidirectional, GRU 但都沒到 Baseline...

之前 train 了一百個 epoch 準確率也都在 33%~37%徘徊，導致後面 train 的時候嚴重落後，最後一直在試別的架構，但最好也只有到 40%。



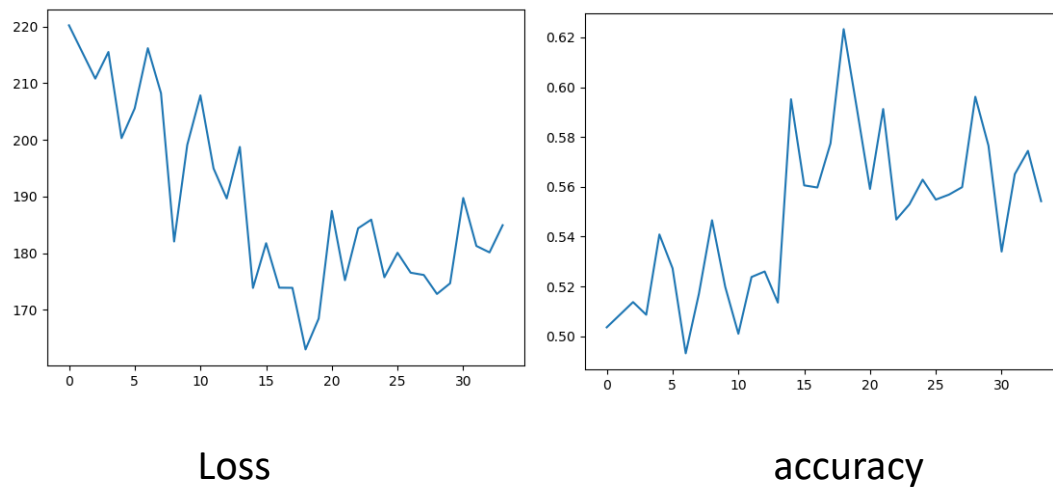


這是第二題 LSTM 的 TSNE 視覺化

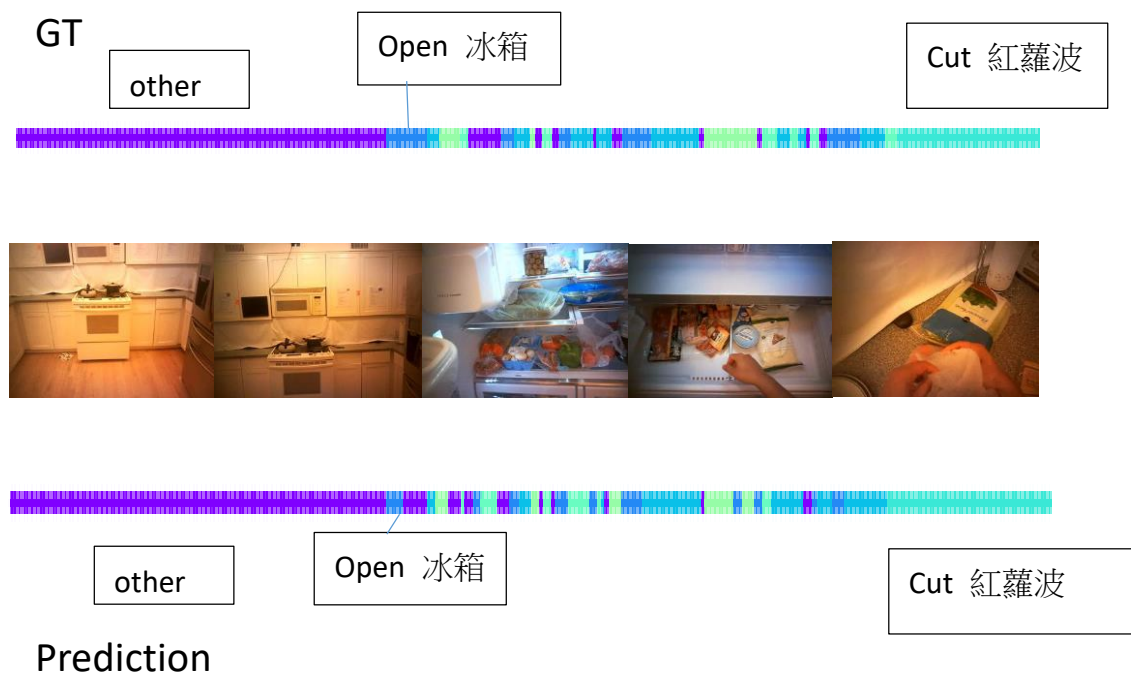
### Problem 3

這題我使用跟 problem 2 一樣的 model(但並沒有 load weight 畢竟沒有過 baseline ), 關於 how to handle long sequence 我使用隨機抽取連續的 50frame(25second).

效果不錯我在 training 的時候, eval 準確率大約在 50%~58%.



因為 500frame 太長了，我分成上下兩段，每段 250frame, 上面試 Ground Truth, 下面是 prediction.



GT

Cut 紅蘿波

Open



Cut 紅蘿波

Open

Prediction