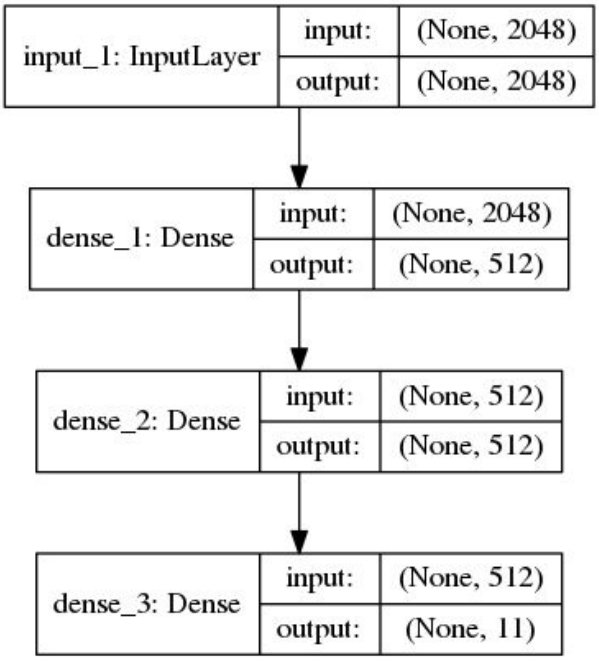
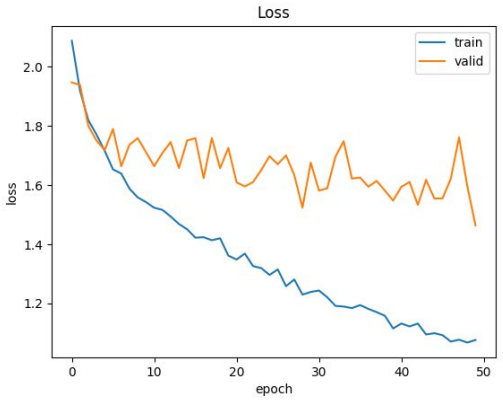
Name: 詹少宏　Dep.:電信碩一　Student ID:R06921017

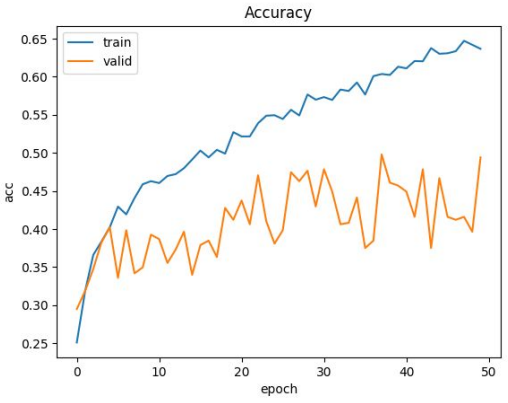
[Problem1]

1. (5%) Describe your strategies of extracting CNN-based video features, training the model and other implementation details.
   Ans: 本次作業使用 Inception v3 做為 CNN的 pre-train model，將影片以 2 fps 採樣後得到的 frame 通過 Inception v3 得到 2048 維 feature 並跟據 frame 數目取平均。之後以簡單的DNN 做動作分類，DNN 架構如下圖 ：



2. (15%) Report your video recognition performance using CNN-based video features and plot the learning curve of your model.
   Ans: 跟據 model.evaluate 得出的結果：loss 為 1.567，accuracy 為0.451。leraning curve 如下圖所示。



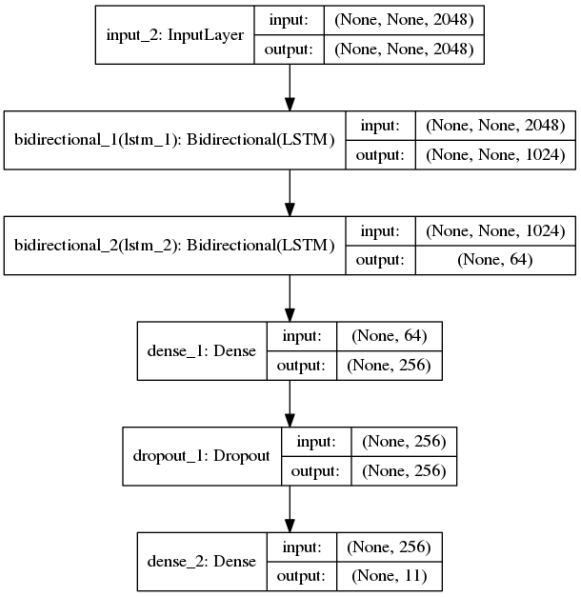


[Problem2]

1. (5%) Describe your RNN models and implementation details for action recognition.

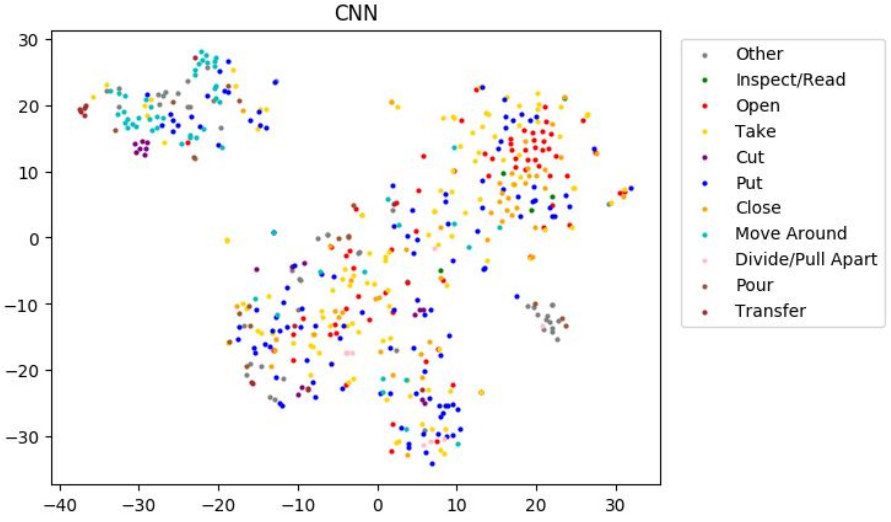
   Ans:

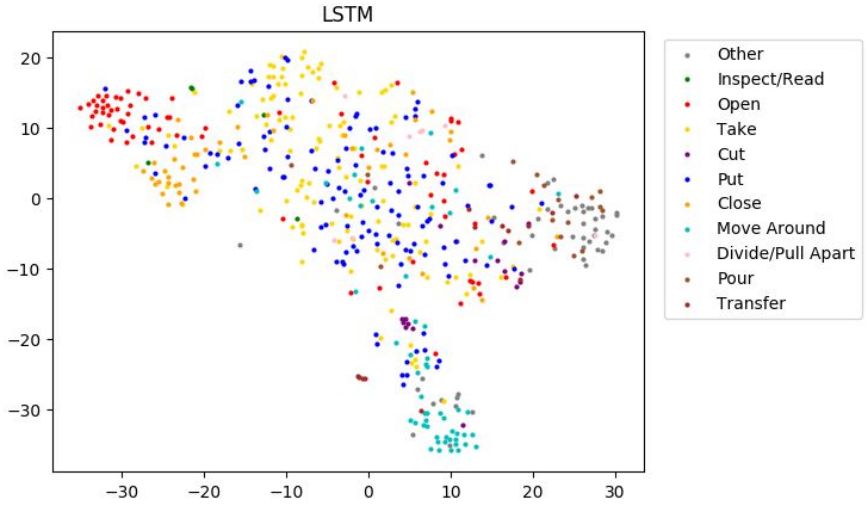
   RNN models: 在這次的model 中，使用兩層 bidirectional LSTM 來提取LSTM的特徵，並接上DNN做 Classifier。



   Implementation details: 將影像採樣出多張圖片後，經過Inception v3 取出feature，並一張張通過RNN model 進一步取出RNN features，最後通過DNN 的 classifier 判斷屬於那一種動作。最後的 validation accuracy 為 0.54。

2. (15%) Visualize CNN-based video features and RNN-based video features to 2D space (with tSNE). You need to generate two separate graphs and color them with respect to different action labels. Do you see any improvement for action recognition? Please explain your observation.





根據上圖的結果，發現CNN的 feature 和RNN的feature 差不多，但就細節而言，RNN 的Other，Close 以及 Move around 都比 CNN來的集中。故 accuracy 較 CNN 平均的結果來高。但也只有上升 9% 而已。

[Problem3]

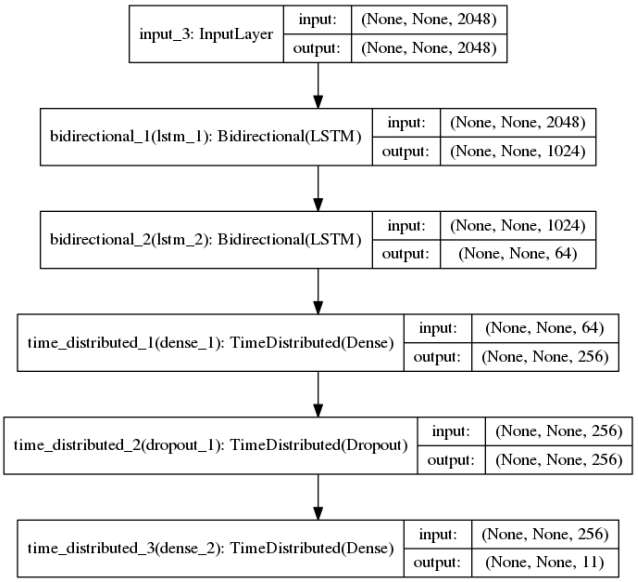
1. (5%) Describe any extension of your RNN models, training tricks, and post-processing techniques you used for temporal action segmentation.

   Ans:

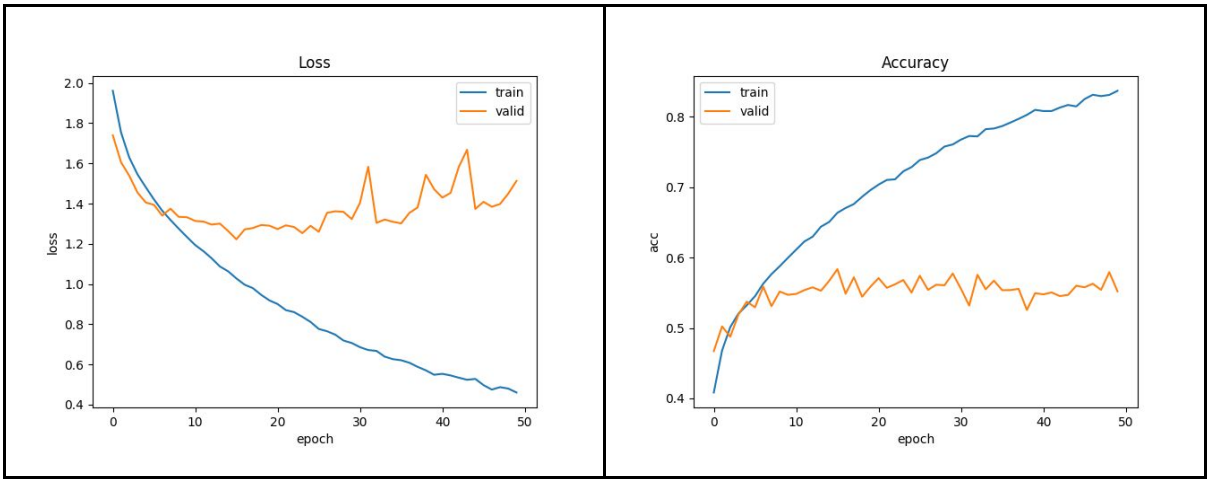
   RNN model: LSTM的部份維持與 Problem2 一樣的架構，至於 DNN 的部份則改用TimeDistributed 來對每個 Frame 做 prediction。架構如下圖所示。

   Training tricks: 以每64 張圖片為一組， normalize 後利用 Inception v3 取出 features 並逐段輸入 model 當中做 training 。

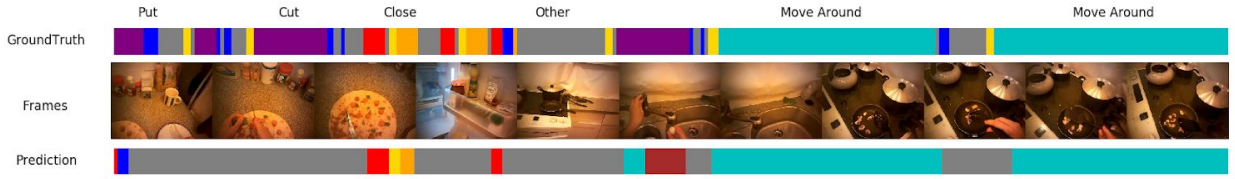
   Post-processing: 觀察資料發現連續為0的情況過多，所以當一組圖片的label 皆為 0 時，將捨棄該組資料。以免 model 之後預測出過多 Other 動作。



2. (10%) Report validation accuracy and plot the learning curve.
   Ans:
   validation accuracy = 0.58
   leraning curve:



3. (10%) Choose one video from the 5 validation videos to visualize the best prediction result in comparison with the ground-truth scores in your report. Please make your figure clear and explain your visualization results. You need to plot at least 300 continuous frames (2.5 mins).
   Ans: 下圖中的 Color bar 顏色與動作的分配與 Problem 相同。



觀察結果中的Move around 以及 Other可知，這次的model 對於時間較長的連續動作有比較準的預測結果，若觀察Close 和 Other 之間，或是Put 和 Cut 之間的結果，則會發現 model 對於快速變化的動作比較遲鈍。 另外，有時候model 也會提早預測出結果，例如 Put 的部份。

[BONUS]