Please use this report template, and upload it in the PDF format. Reports in other format will result in ZERO point. Reports written in either Chinese or English is acceptable. The length of your report should NOT exceed 8 pages.

Name: 郭笛萱 Dep.:電機四 Student ID:B03901009

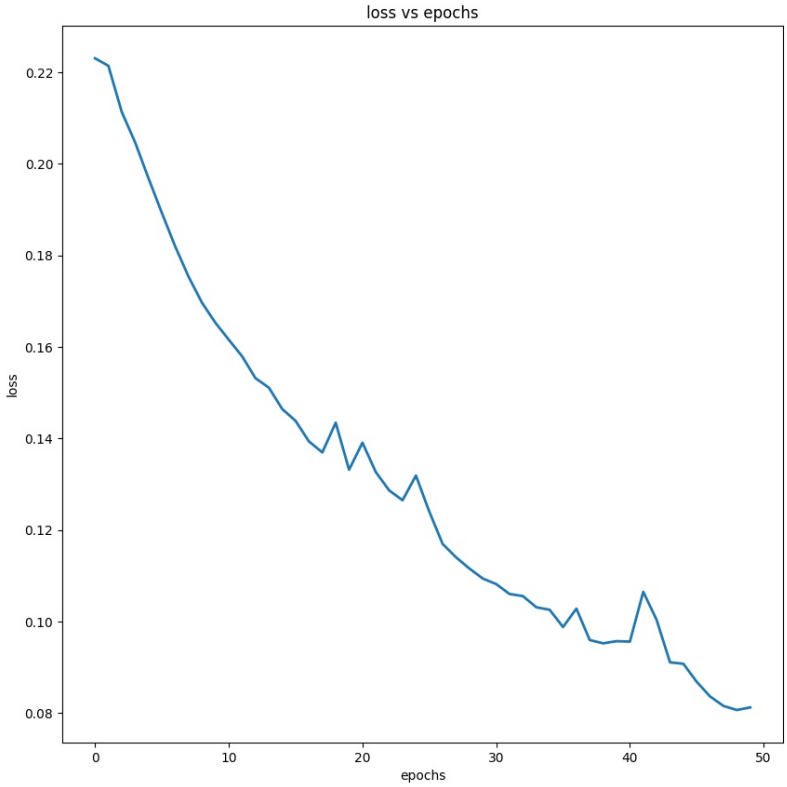
[Problem1]

1. (5%) Describe your strategies of extracting CNN-based video features, training the model and other implementation details.

   首先先固定每個影片的 frame 個數為 16 個，將這 16 個 frame 輸入進 resnet50 來取得每個 frame 的 feature，再來將這 16 個 frame 取平均得到單一一部影片的 feature，最後通過一個 fully connected layer，得到一個 one hot vector，預測出 label。
   Model 選用的 optimizer 為 adam，batch size =32，learning rate = 1e-3。

2. (15%) Report your video recognition performance using CNN-based video features and plot the learning curve of your model.

   此題的 performance 為 0.4119，loss 穩定下降如下圖所示



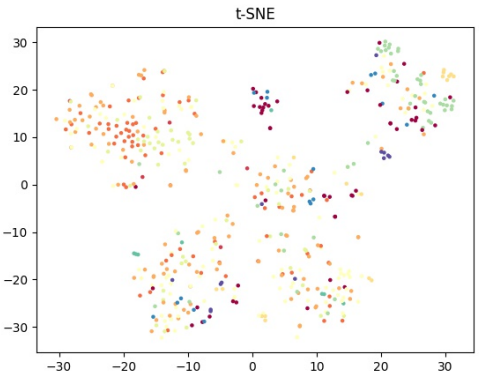
[Problem2]

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

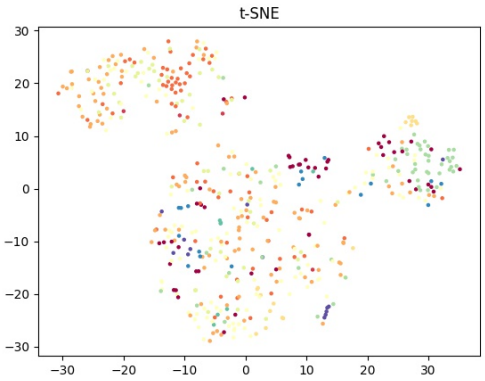
   將第一題所取得的每個 frame 的 feature 依序輸入進 RNN model 中，RNN model 使用的是雙向單層 LSTM，再接上 drop out = 0.3，避免 overfitting，hidden_size = 128，最後接上兩層的 fully connected layer 將 128 維的 hidden vector 轉換為 64 維，64 維轉成 11 維。
   每個影片我固定選取 16 個 frame 的資料，batch size = 32

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.

CNN feature



RNN feature



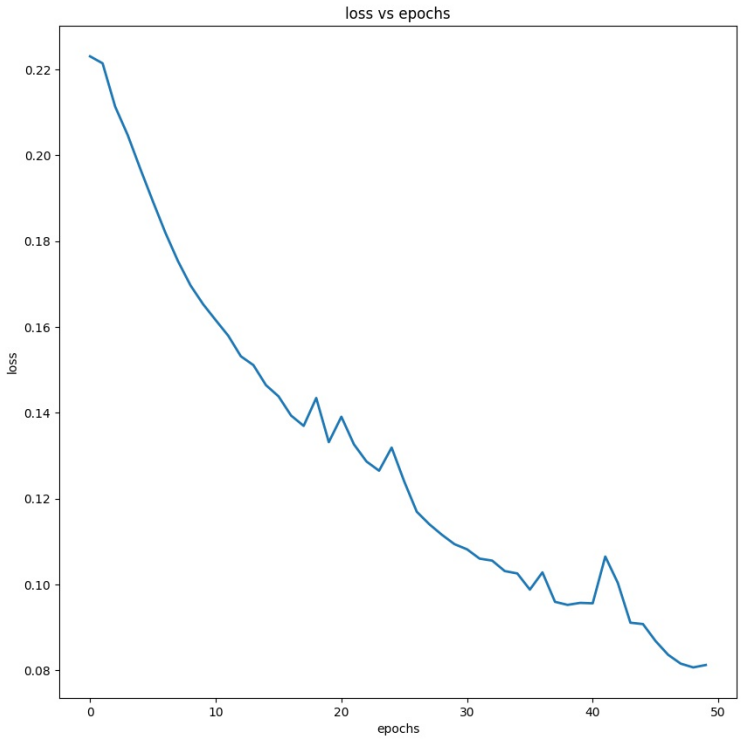
可以看到兩者同類別大致上都沒有很明確的聚集現象，左上角橘色與右上角綠色 label 有比較明顯的聚集現象，但是無法分辨出到底是 CNN 還是 RNN 的 feature 較好。推測兩者分佈會那麼接近的原因是因為我們是拿 CNN feature 來改進 RNN feature 的，所以兩者會有一定的相似程度

[Problem3]

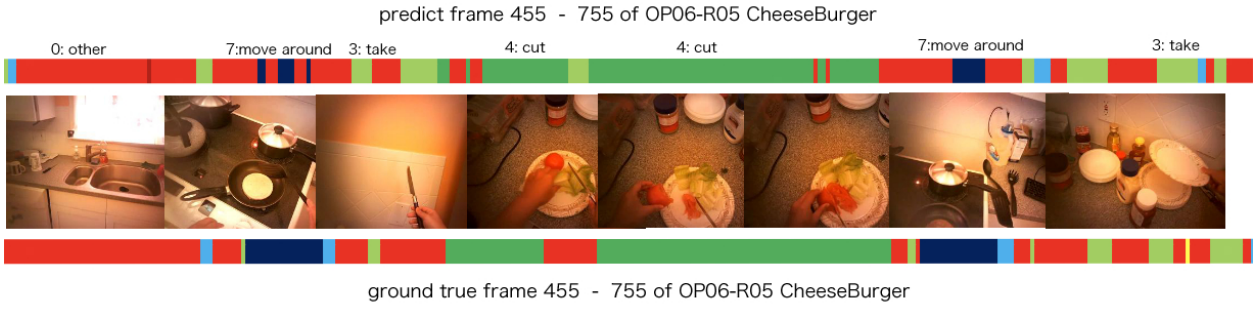
1. (5%) Describe any extension of your RNN models, training tricks, and post-processing techniques you used for temporal action segmentation.

   我固定取每個影片 256 張影像當作 training data，使用與第二題一樣的雙向單層 RNN model 來 train，RNN model 會輸出每個 frame 所 predict 出的 hidden vector，再將這些 hidden vector 接到兩層 fully connected layer，分別為 256 降到 128，128 降到 11，最後接上一層 softmax 來預測出 label。此題沒有使用到 drop out。

2. (10%) Report validation accuracy and plot the learning curve.
   此題的 validation accuracy 為 0.597



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).



[BONUS]