Name:鄭元嘉 Dep.:電機三 Student ID:B04901081

[Problem1]

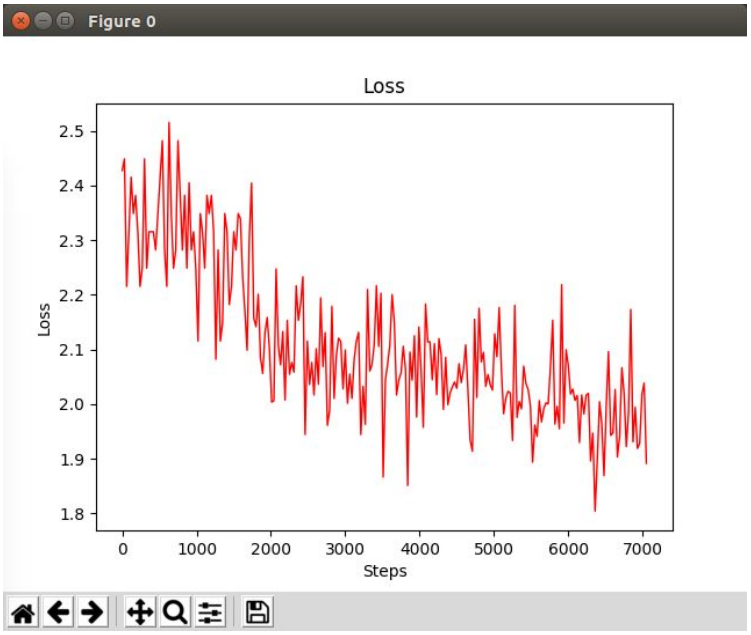
1. (5%) Describe your strategies of extracting CNN-based video features, training the model and other implementation details.
   (1) 採用「Average Pooling」，若該部影片frames大於10則以平均方式取10張出來，若該不影片frames無大於10，則全取。取完n張frames後丟入model得到n張feature map，取平均，再丟入下一層的classifier做分類。
   (2) Model Structure
   主要分為pre-trained vgg16作為extracting features，fully connected layers作為classifier。

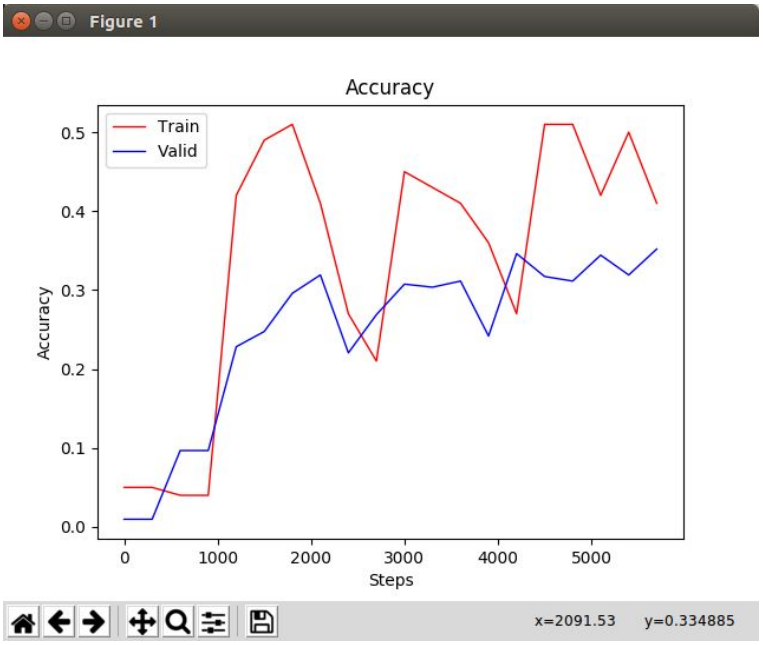
	out_size
Extracting Features	
Pretrained Vgg16	512 * 7 * 10
F.C. Layer	2 ** 10
Classifier	
F.C. Layer / ReLU	2 ** 9
F.C. Layer / Sigmoid	11

   (3) Other Parameters
   Otptimizer : Adam
   Init_Lr : 0.00001
   Batch_size : 1

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

   Final Accuracy on Validation Dataset : 0.369439





[Problem2]

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

(1) Model Structure

使用在Problem1中train好的extracting取出features (transfering)，丟入帶有lstm的model中。

	out_size	Others
Extracting Features		
Pretrained Vgg16	512 * 7 * 10	
F.C. Layer	2 ** 10	
Classifier		
Lstm	1024	layer=1, sigle-direction
F.C. Layer / ReLU	2 ** 10	
F.C. Layer	2 ** 11	
F.C. Layer / Sigmoid	11	

(2) Other Parameters

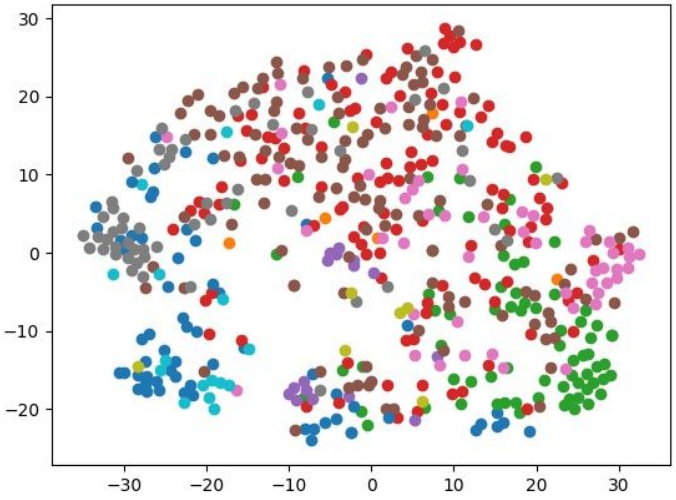
Otptimizer : Adam

Init_Lr : 0.0001

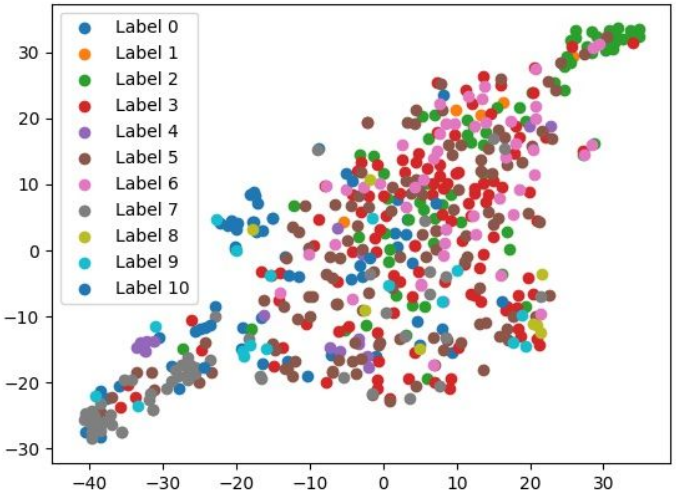
Batch_size : 1

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.

RNN features with tSNE



CNN features with tSNE



整體上RNN features 分佈呈現環狀，CNN features分佈呈現直線狀，就幾何學來說呈現環狀能夠被區分的可能性更大，且由上圖可以很明顯看到RNN features同色的分佈離散程度高，可離性很高，而CNN features相同顏色的點雖然有聚集現象，但因為困在一維方向上，要與其他群聚色點分離的難度比RNN features分佈來得難。

[Problem3]

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

使用在Problem中train好的extracting取出features (transfering)，丟入帶有lstm的model中。

	out_size	Others
Extracting Features	2 ** 10	
F.C. Layer / Sigmoid	2 ** 10	
Classifier		
Lstm	1024	layer=1, sigle-direction
F.C. Layer / ReLU	2 ** 10	
F.C. Layer / ReLU	2 ** 10	
F.C. Layer	11	

架構中在利用pretrained model Extracting Feature後，丟入lstm前，特別加入一層F.C Layer，因為我把pretrained model鎖住，但又怕feature學的不夠好，所以特別加入一層，增加feature彈性，且後面必須加上Sigmoid會比較好，原因是lstm的公式output最後會是經過sigmoid再進到下一步，為了lstm內部運算公式數值scale相近，故加上一Sigmoid。

(2)採用”random sample”的方式，隨機取出連續的50 frames丟入訓練。

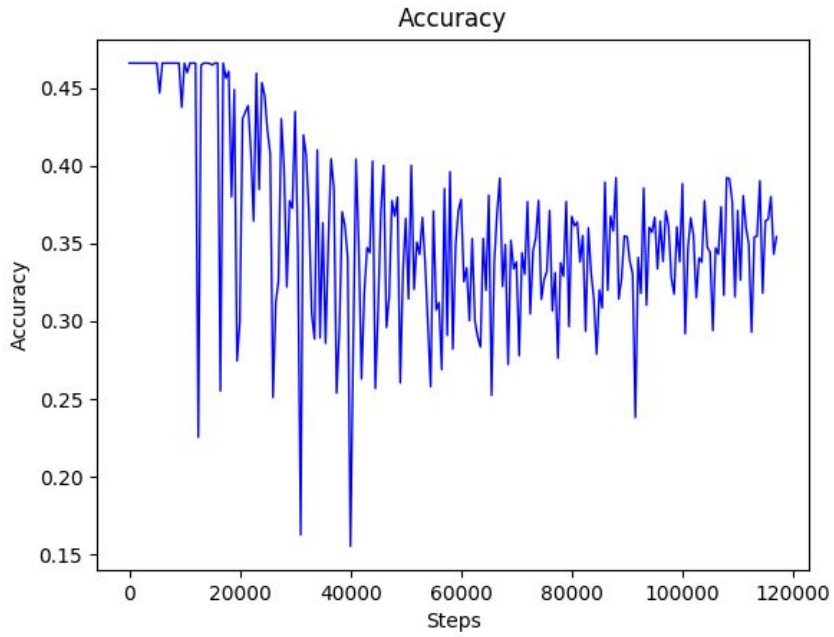
(3) Other Parameters

Otptimizer : Adam

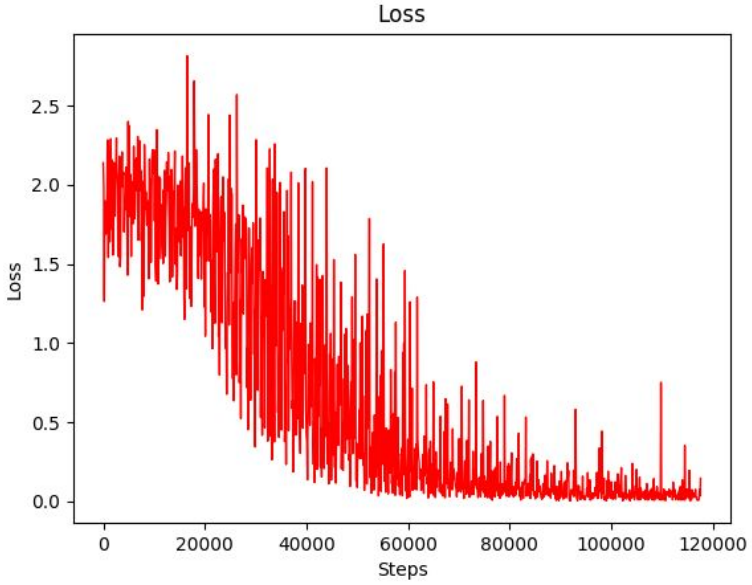
Init_Lr : 0.00001

Batch_size : 1

3. (10%) Report validation accuracy and plot the learning curve.
   Validation Accuracy



   Loss



4. (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).