深度學習於電腦視覺 作業四

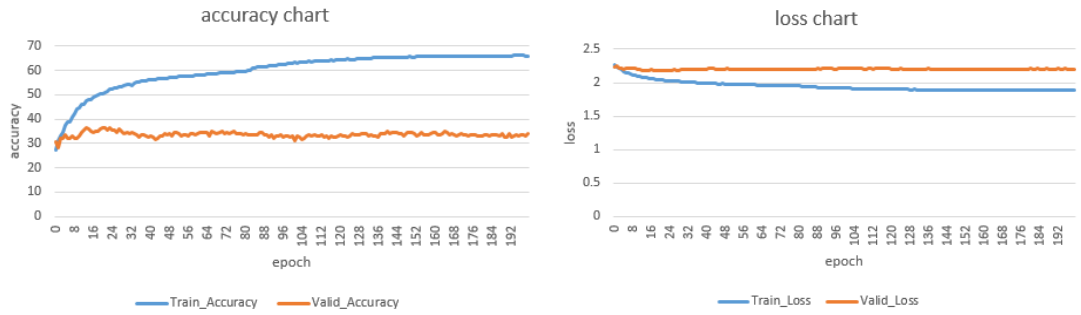
r07922103 資工碩一 李俊賢

Problem 1：

1.

```
FC(
  (fcblock): Sequential(
    (0): Linear(in_features=2000, out_features=11, bias=True)
    (1): Softmax()
  )
)
```

以上是我 Model 的架構，對於一個影片，利用 Pretrained Resnet50 吐出來的眾多 1000 維的前後兩個，接成一個 2000 維的 feature，再利用 fully connect 把該 feature 接到 11 個 class，再過一層 softmax。以下是我畫出來的 Learning Curve，可以看到不管在 train/valid data 上面，loss 都可以有效的下降，accuracy 也可以有效的上升。

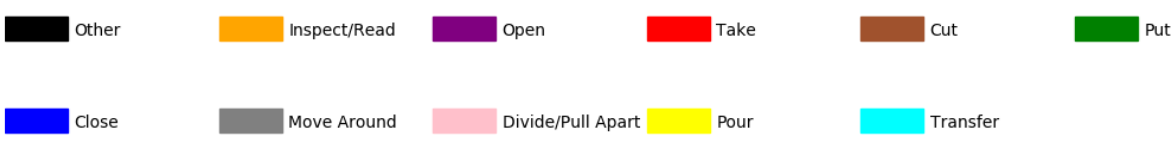


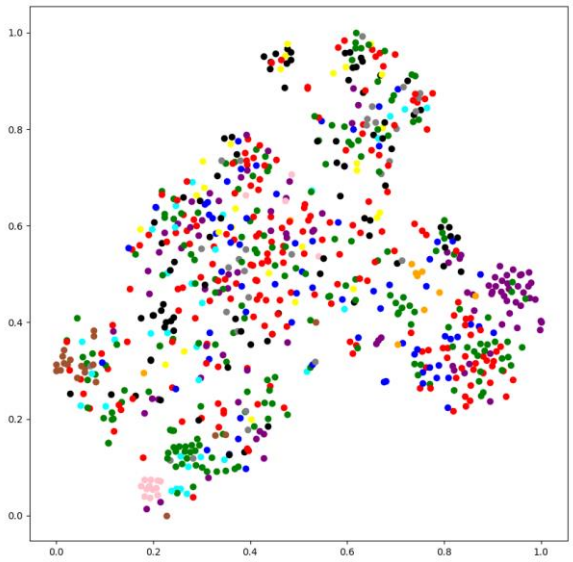
2.

我利用 Pretrained Resnet50 以及以上 Model 所做出來的最高 valid Accuracy 是：36.54%。

3.

以下是我把每一部影片的 2000 維 feature 做 Tsne 畫出來的結果。可以看到分群現象其實並不明顯。



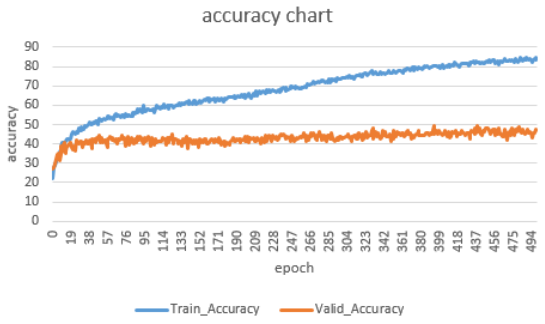


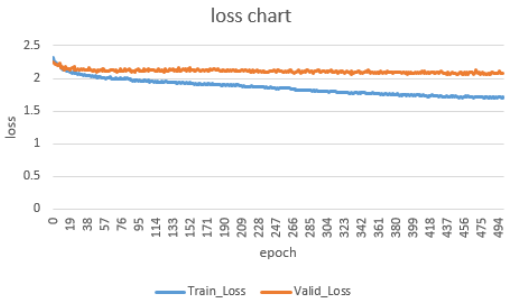
Problem 2：

1.

```
RNN(
  (rnn): LSTM(1000, 128, num_layers=4, batch_first=True, dropout=0.5)
  (out): Sequential(
    (0): Linear(in_features=128, out_features=11, bias=True)
    (1): Softmax()
  )
)
```

以上是我的 RNN Model。主要是利用 LSTM 將 Pretrained Resnet50 的眾多 1000 維 features 餵進去，其中 LSTM 有四個 layers、dropout(0.5)以及 128 維的 output features。我是取最後一個時間點的 128 維 feature 過 fully connected 以及 softmax。而以下是我的 Learning Curve，可以看到 loss 及 accuracy 的表現上都比 Problem 1 還要好許多。



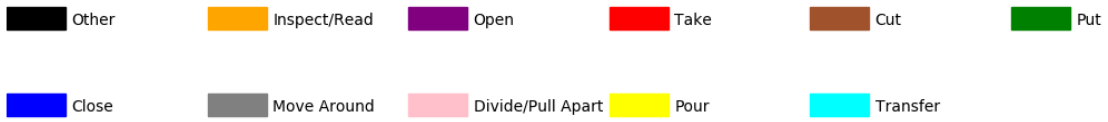


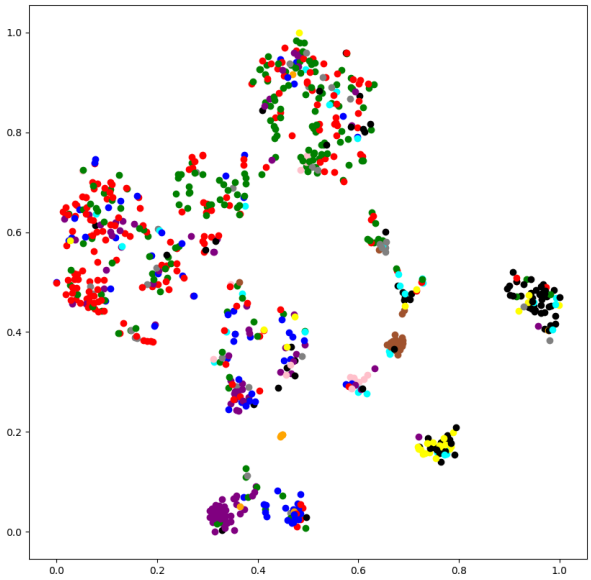
2.

我利用 Pretrained Resnet50 以及以上 Model 所做出來的最高 valid Accuracy 是：49.28%。

3.

以下是我把每一部影片的 128 維 feature 做 Tsne 畫出來的結果。可以看到分群現象其實比 problem 1 好上很多，也比較明顯。其中 Take、Put 的 training data 是數量最多的，所以理所當然也要訓練得比較好，從下面圖可以看到他們兩個的群聚效應蠻好的。而再來 open、close、pour、other、cut 這類動作明顯的影片，也因為 LSTM 加入了時間軸的相關性而有較佳的分群效果。





Problem 3：

1.

```
RNN(
  (rnn): LSTM(2048, 128, num_layers=4, batch_first=True, dropout=0.5)
)
FC(
  (fc): Sequential(
    (0): Linear(in_features=128, out_features=11, bias=True)
  )
)
```

以上是第三題的架構。與第二題不同的是，Pretrained Resnet50 的最後一層被我拿掉了，所以從 Resnet50 出來的眾多 2048 維，會接上 LSTM Model 並依據每個時間點輸出 128 維的 feature，再將這些 feature 過一個 fully connect 來完成 per-frame 的類別預測。其中值得注意的是我並沒有實作 post processing，原因是因為實作之後發現 accuracy 會降低，而助教也說

非必要，所以就不做了。

2.

此題利用上面所敘述的 Model 所預測出來的七部影片平均 valid accuracy 是：60.51%。而這七個 valid 影片的 accuracy 分別是：

(1) OP01-R02-TurkeySandwich：484 / 1012 = 47.82%

(2) OP01-R04-ContinentalBreakfast：596 / 982 = 60.69%

(3) OP01-R07-Pizza：1650 / 2471 = 66.77%

(4) OP03-R04- ContinentalBreakfast：505 / 889 = 56.80%

(5) OP04-R04- ContinentalBreakfast：710 / 1085 = 65.43%

(6) OP05-R04- ContinentalBreakfast：472 / 948 = 49.78%

(7) OP06-R03-BaconAndEggs：983 / 1551 = 63.37%

3.

以下是我實作出來整部 OP01-R07-Pizza 的 Prediction Picture。因為其 accuracy 最高，所以預測圖也相當完整，上排是 Ground Truth，下排是 Predict 的結果。