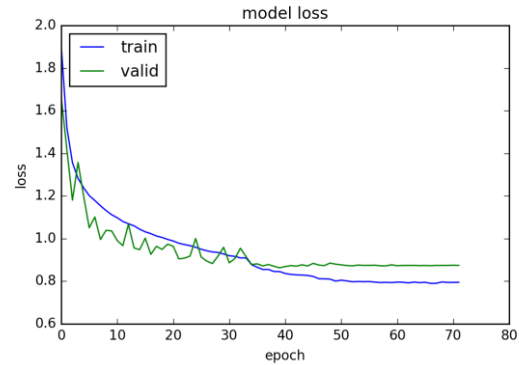
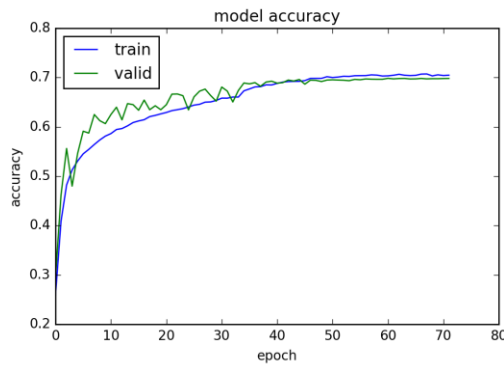
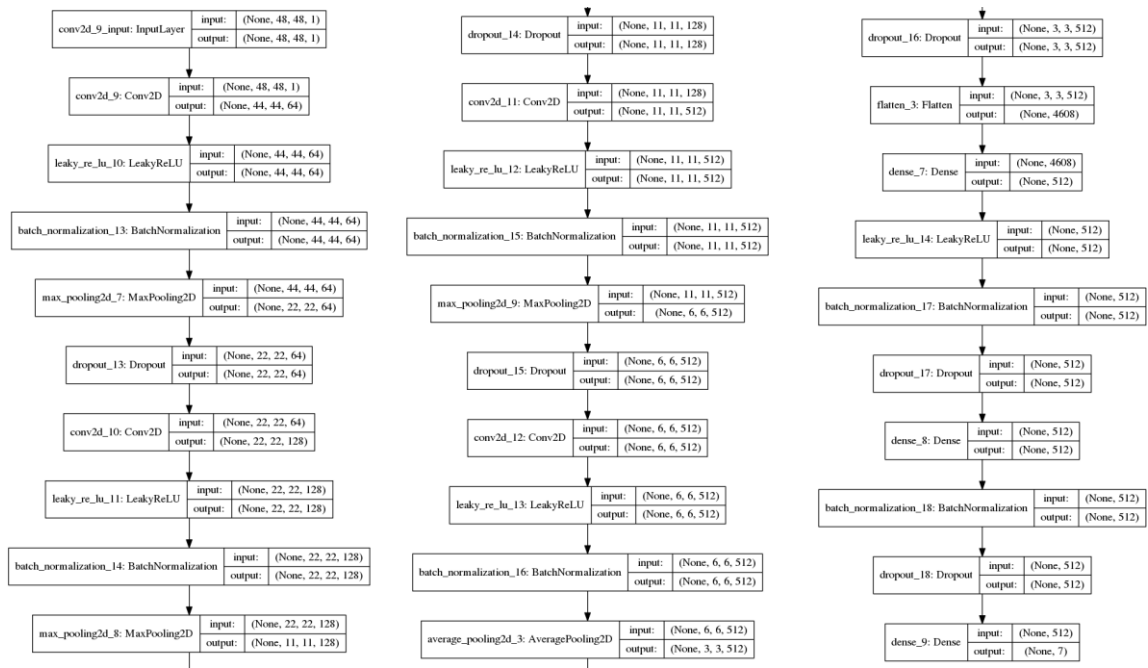


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

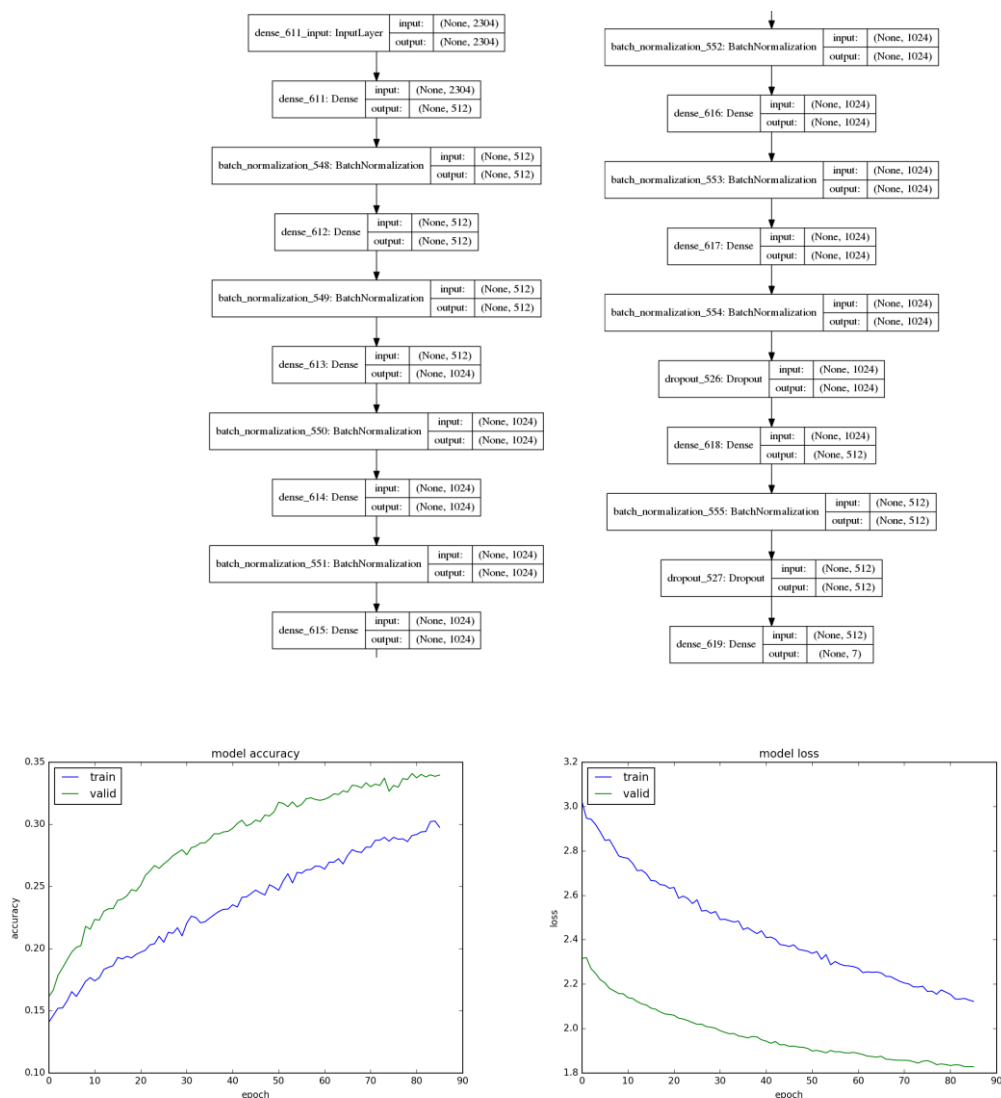


上

圖為 model1 的架構、過程，然而 model 2 為 kaggle 分數最高的 model，訓練手法是使用 ensemble 將四個不同的 model (準確率都差不多)組合起來，而 model 1 是四者中最好的一個，準確率相比 $0.72081/0.70827=1.0178$ (和 best 相差 $1.78\% < 2\%$)。

accuracy	Training	valid	public	private
Model 1	0.703428	0.698363	0.70827	0.70158
Model 2			0.72081	0.71551

2.(1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

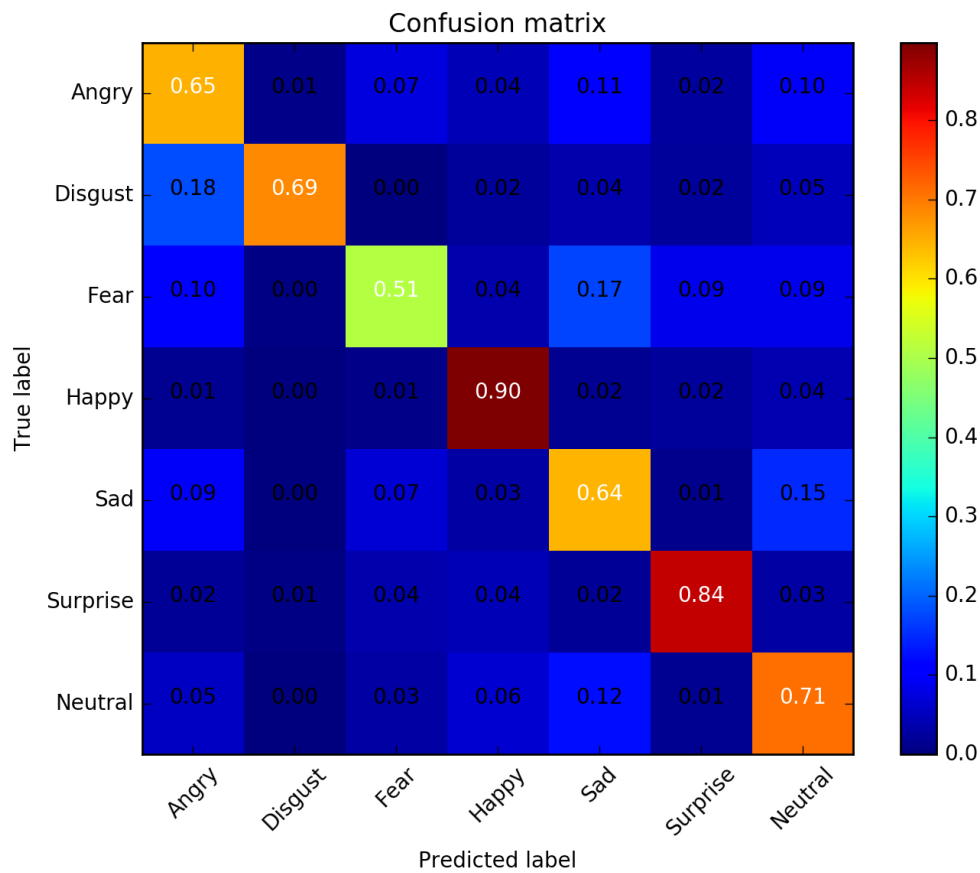


上圖為和 model 1 相近參數量的 DNN 架構、過程。

accuracy	Training	valid
DNN	0.290983	0.340822

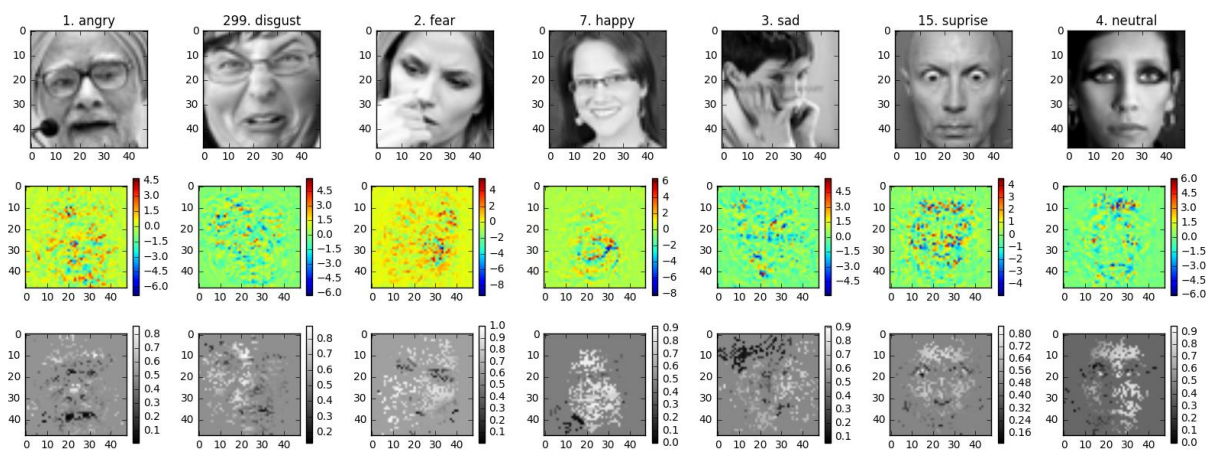
和 CNN 相比更容易發散且 train 不好，太深層學習不出，太胖太低層精準度不高，我猜在同樣的 pattern 出現在不同地方這件事情對於 DNN 是比較難的，需要更多參數量才能解決，比較起 CNN 一個 filter 即可達成這件事情比需要更多參數，故較差。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]



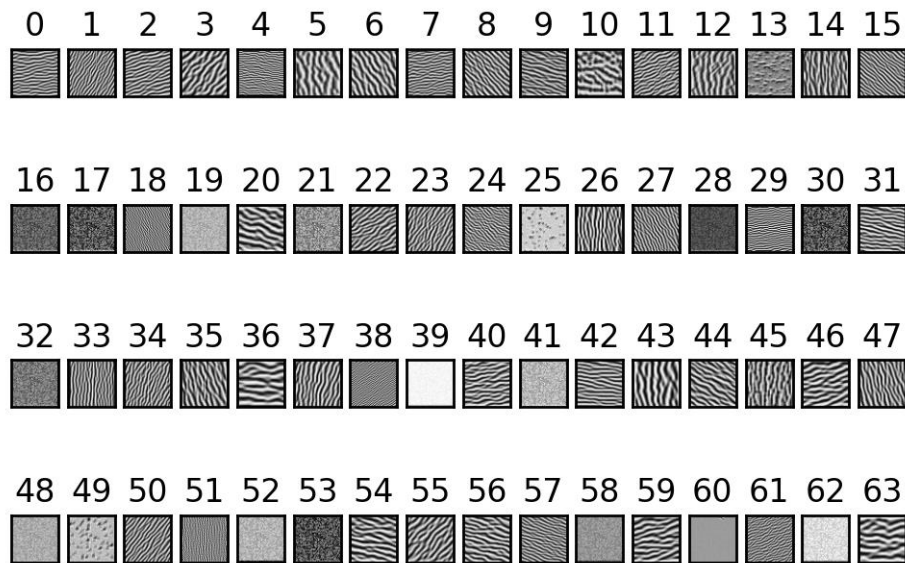
由圖中可看出 Disgust 和 Angry(0.18)是在對角線外發生最高機率，這兩種最易混淆，其次是 Fear 和 Sad(0.17)，再來是 Sad 和 Neutral(0.12)。

4.(1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？



上圖可明顯得知 focus 在五官的部分。

5.(1%) 承(1)(2)，利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate。



觀察第一個 LeakyRelu 得到下面圖形，可見到有很多條紋很深的 filter，我猜測是在臉部表情變化很大(五官特徵變化明顯)的情況下最容易被 activate。