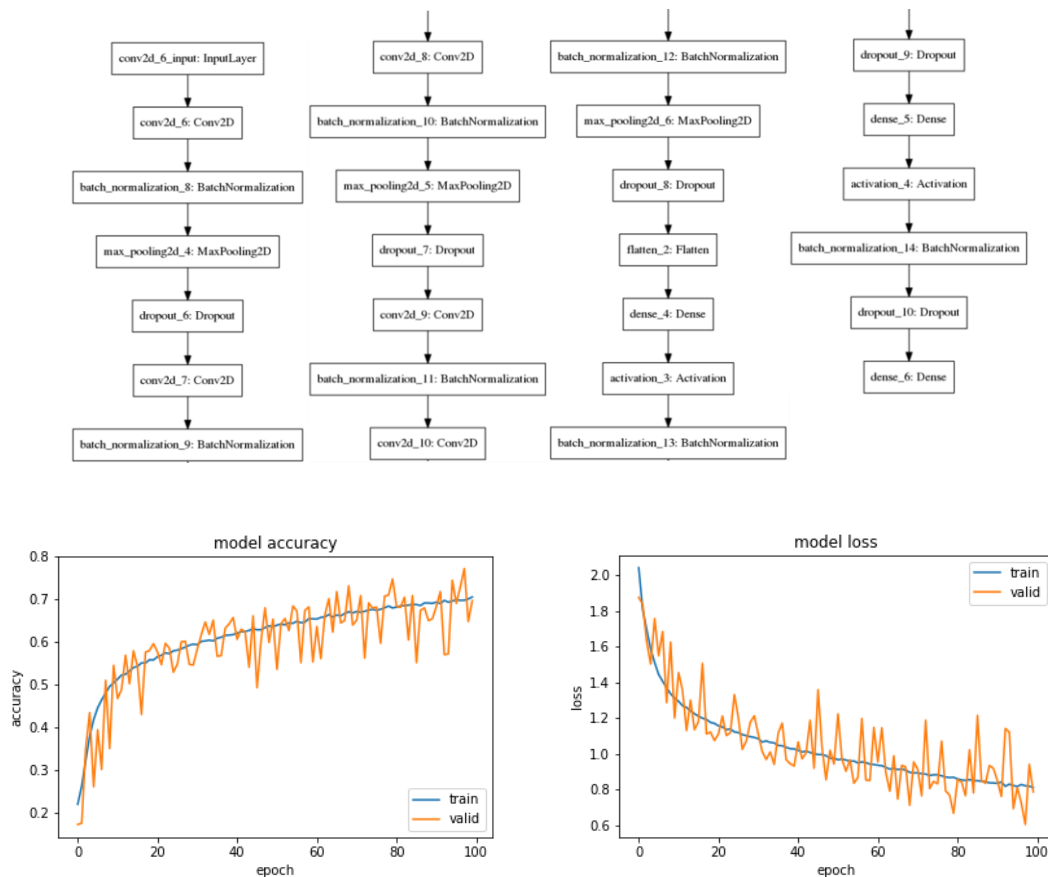


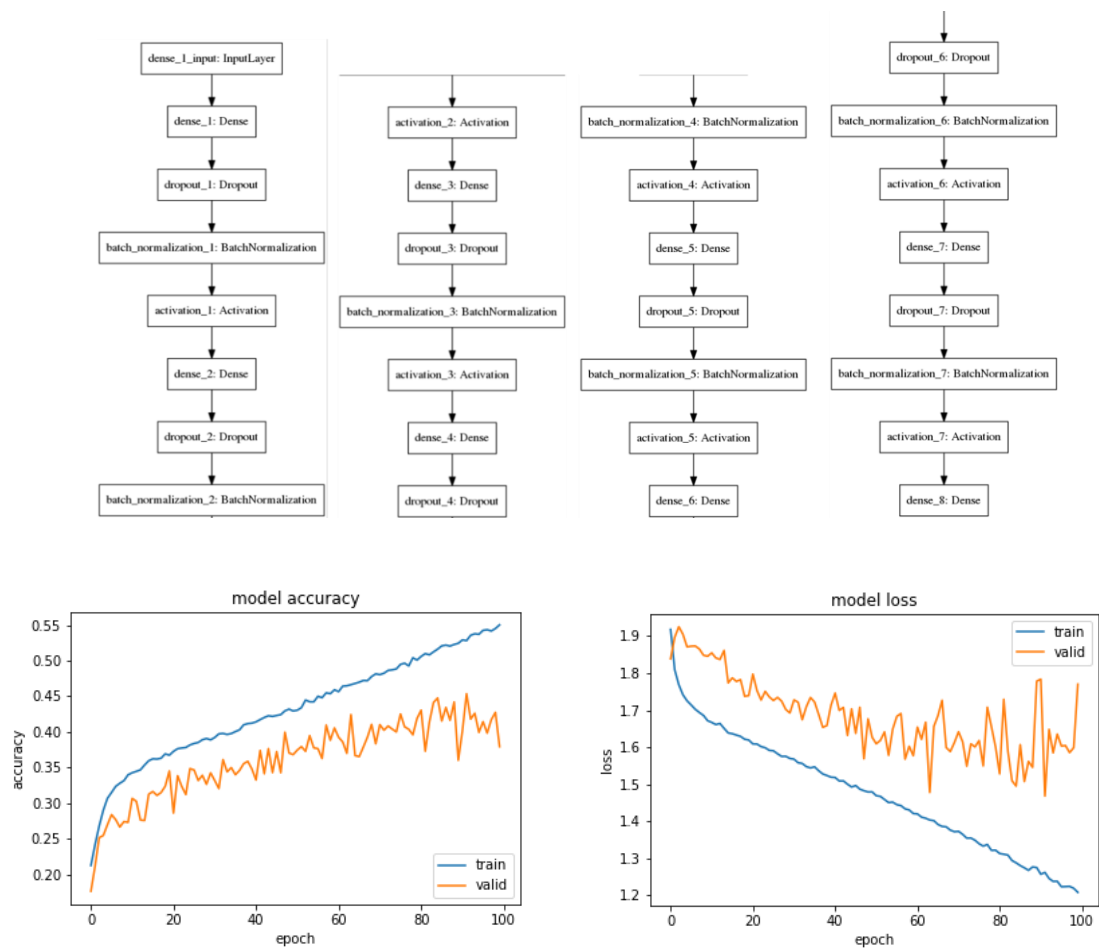
1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？
答：

模型架構圖和結果如下圖。這次實驗有利用 ImageDataGenerator，將圖片進行旋轉、水平、垂直位移、放大、縮小、對稱翻轉，可以使得訓練效果更好，較不容易產生 overfitting 的現象。準確率一般都可以到達 65% 左右，而最好的 kaggle 成績是利用 5 個相似的 CNN models，利用 ensemble 的方式，得到更為準確且穩定的結果。



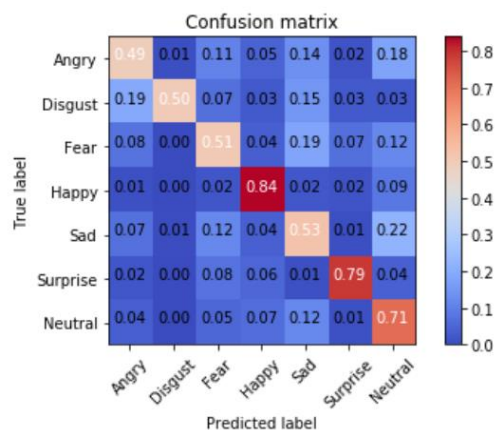
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？
答：

模型架構圖和結果如下圖。DNN 訓練速度上較 CNN 快速很多，然而準確率卻是相較之下只有 30-40%，而且由圖表可以發現 DNN 相當容易造成 overfitting，可能是由於 DNN 不像 CNN 有進行影像擴增，增強對雜訊的抵抗力，而最重要的是無法取得圖片中的影像特徵，只能從 pixel 去分類。



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]
答：

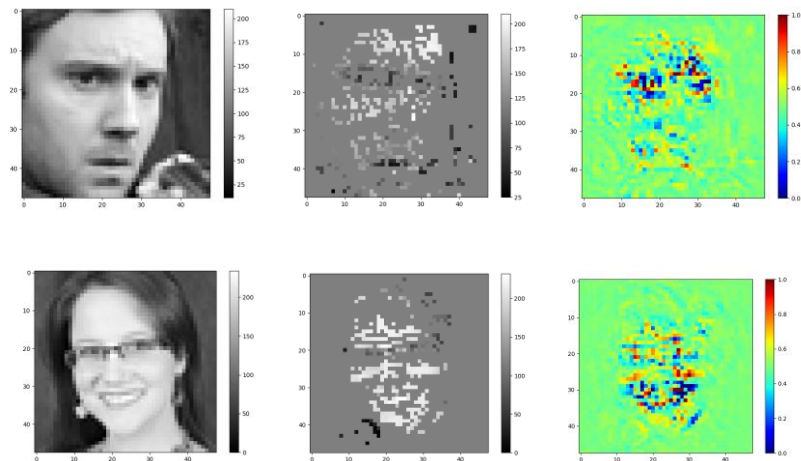
從結果可以發現到 confusion matrix 中，準確率最高的為 happy，因為嘴巴的輪廓特別明顯。而最容易分類錯誤的為 sad 和 neutral，這兩種表情確實連一般人有時都難以分辨，而 fear 和 sad 也是會有同樣的情形，CNN 容易將臉部表情差異不大的分類錯誤。



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？

答：

擷取出 CNN 中的 saliency maps，可以發現 CNN 學習的重點著重於眼睛，而比較特別的是很多微笑的圖片，其特徵會很明顯的在嘴巴周圍，應該是因為嘴巴的輪廓在臉上是最為顯眼的特徵，也呼應上面的 confusion matrix 的結果。



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。

答：

下圖利用 gradient ascent 的方法，將 conv2d_2 的 layers 取出，可以發現最能 activate 的幾乎都是紋理方面的特徵，很類似於 Gabor filter 的感覺。而也有部分的 filter 可以被臉部五官所 activate，可以從圖片中清楚看到人的眼睛、嘴巴，這些較明顯的臉部輪廓。

