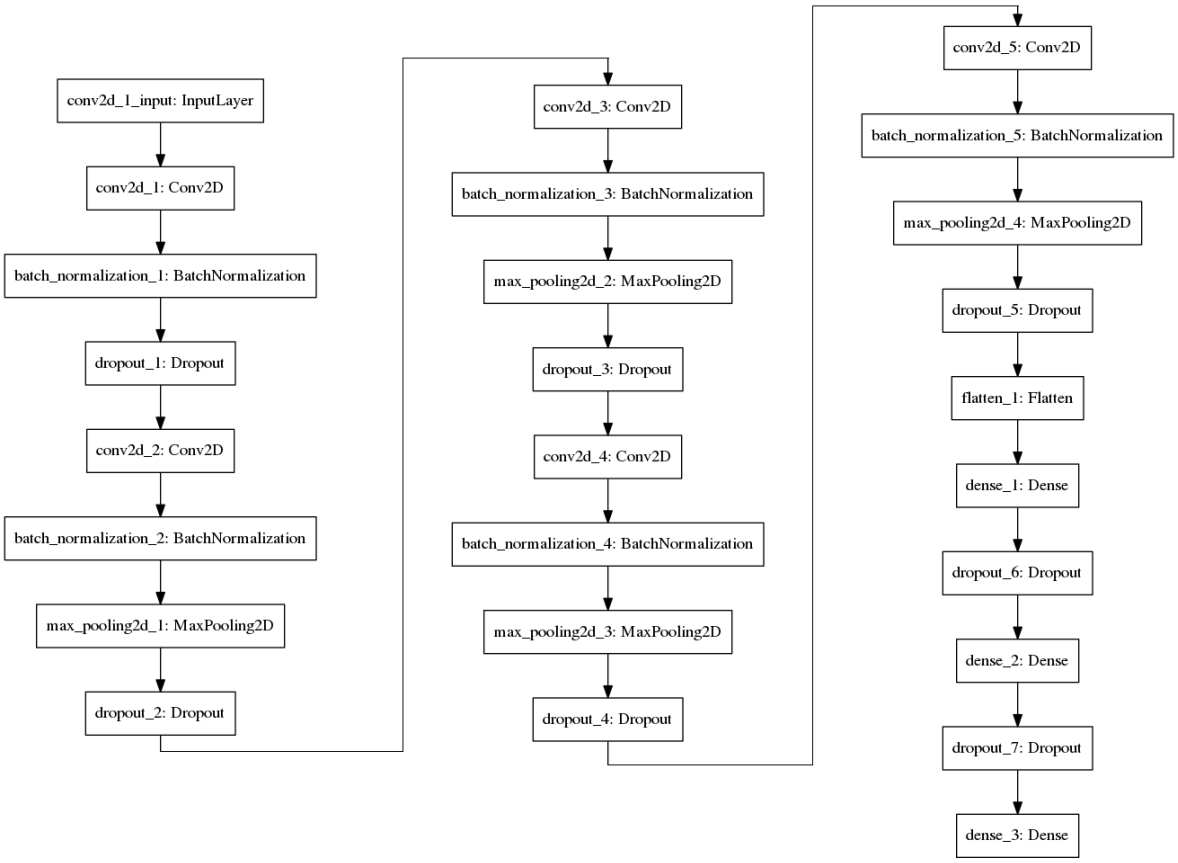
學號：B04902084 系級：資工二 姓名：王建元

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？
模型架構如下圖：

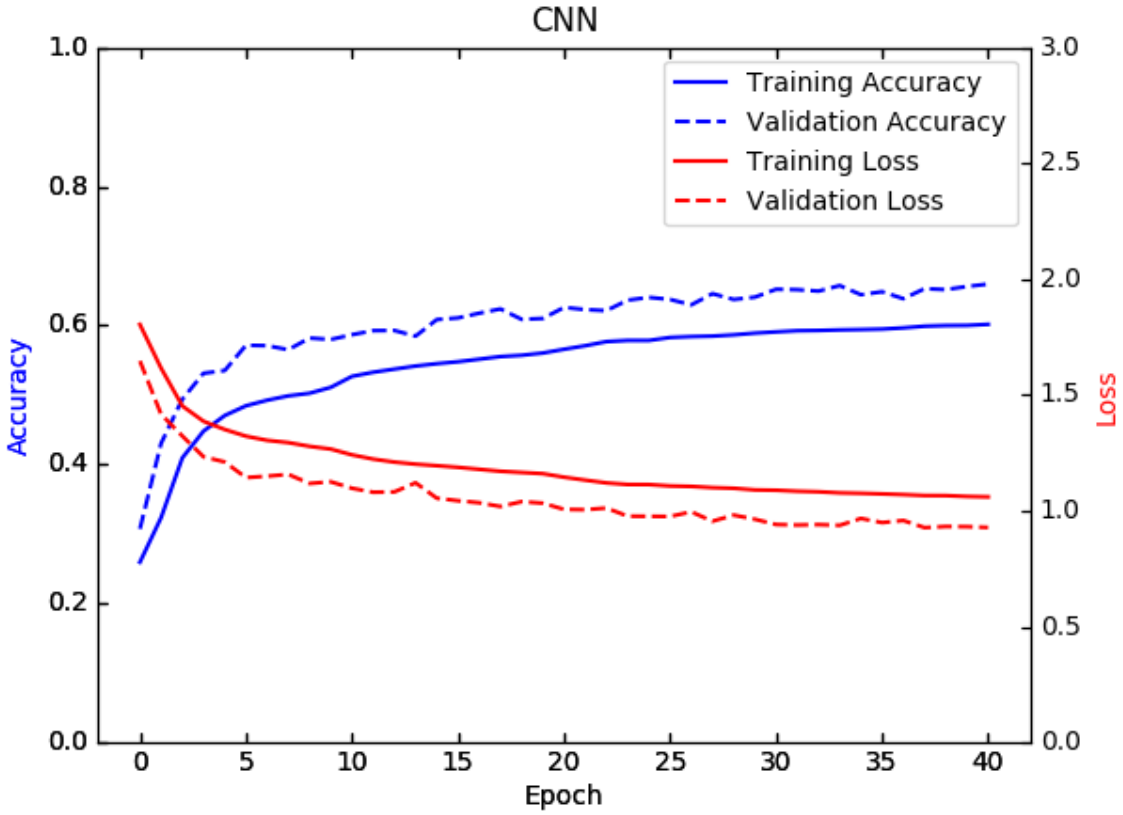


首先是 5 層的 Conv2D, BatchNormalization, MaxPooling2D, Dropout，接著在 flatten 之後經過 3 層 Dense，最後產出結果。

訓練過程：

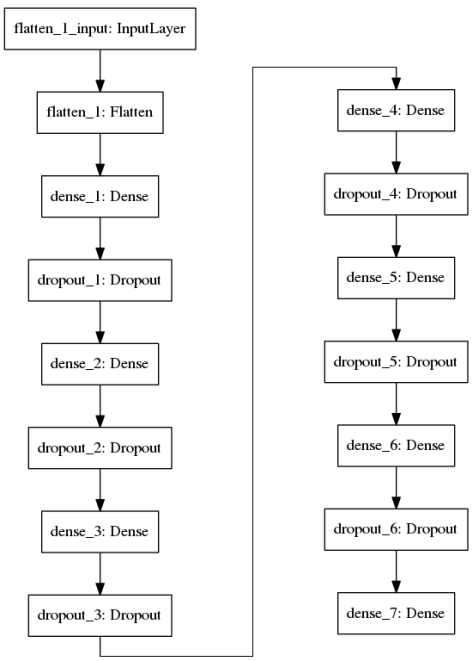
一開始將第 i 個 Conv2D 的 filter 設成 32*i，3 層 Dense 分別為 256, 7, 7，train 50 epoch 便可過 simple baseline。後來加入了 BatchNormalization 和 Dropout，並且將 data 在 train 之前經過鏡像、左右旋轉 10 度，準確率明顯上升，但還過不了 strong baseline。最後是將第 1 層 Dense 的 units 增加到 1024，並且 train 200 個 epoch 才過了 strong baseline。最後總參數量約 540,000（再上去 GPU 跑不動）。

準確率：



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？

模型架構如下圖：

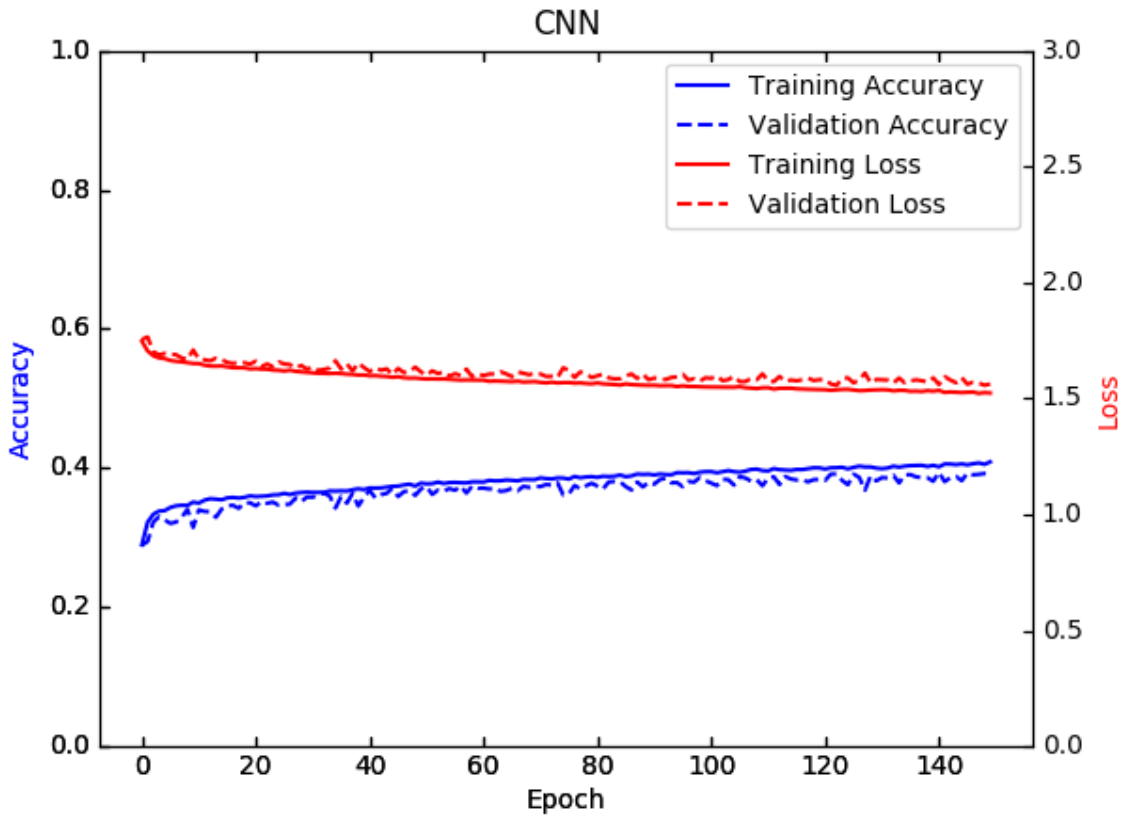


Flatten 完用 7 層 Dense，並且在 Dense 跟 Dense 之間加入 0.1 的 Dropout。

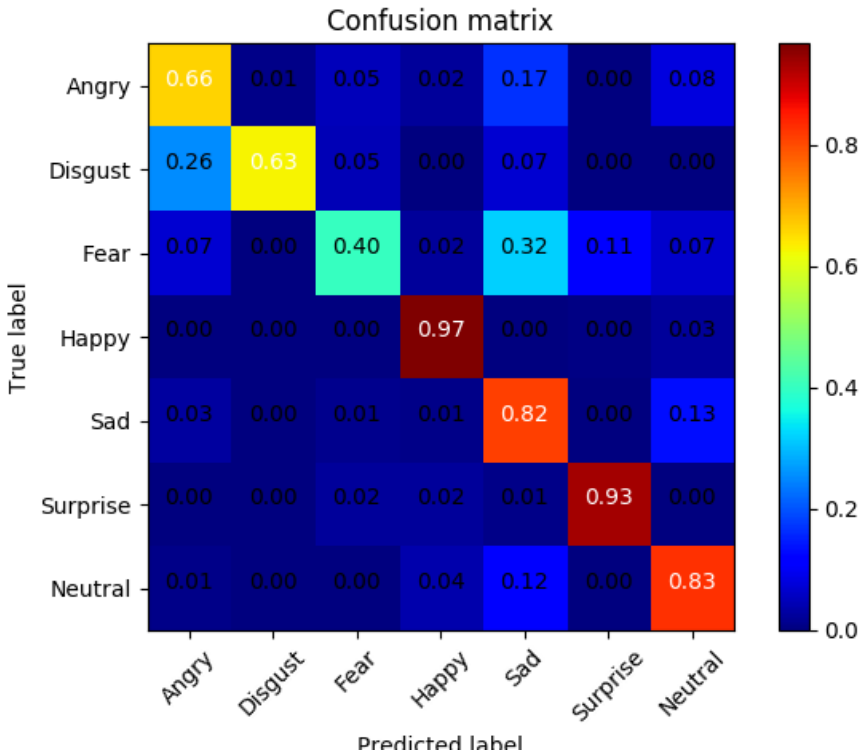
訓練過程：

DNN 在與 CNN 相同參數的狀況下，運算量少很多，又 CNN 較難平行化，因此 DNN 每個 epoch 所花的時間短很多。雖說所花的時間少，但 train 出來的準確率卻也很低。原因是 DNN 在處理資料時看得是絕對位置，而這在圖像辨識上非常不利。相對的 CNN 所看的是相對位置，因此能夠達到更高準確率。也因為這樣，即使 DNN 的參數量達到 CNN 的 10 倍之多，準確率也沒有顯著的提升。

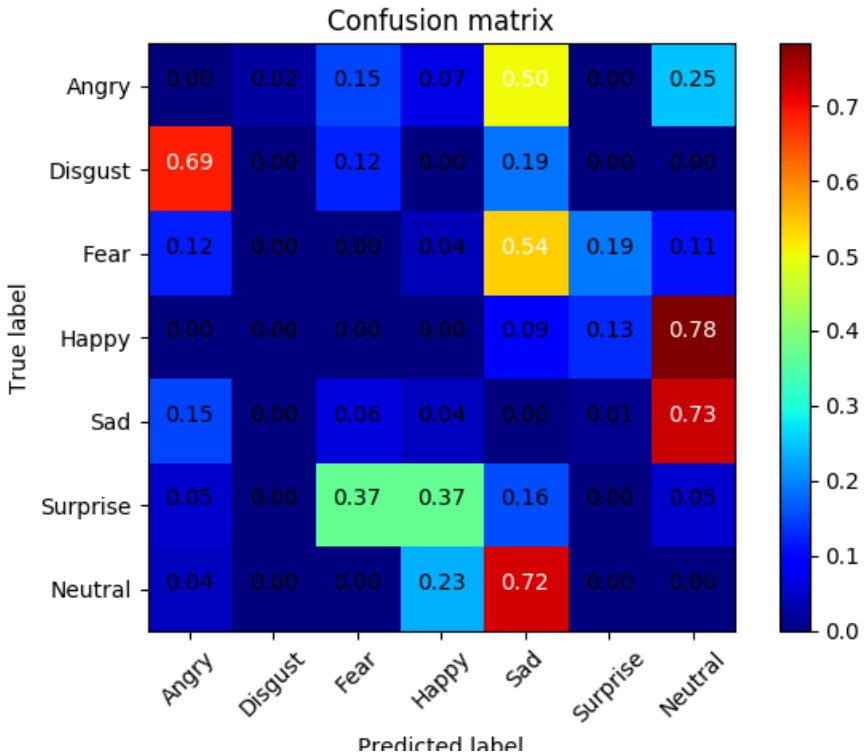
準確率：



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]



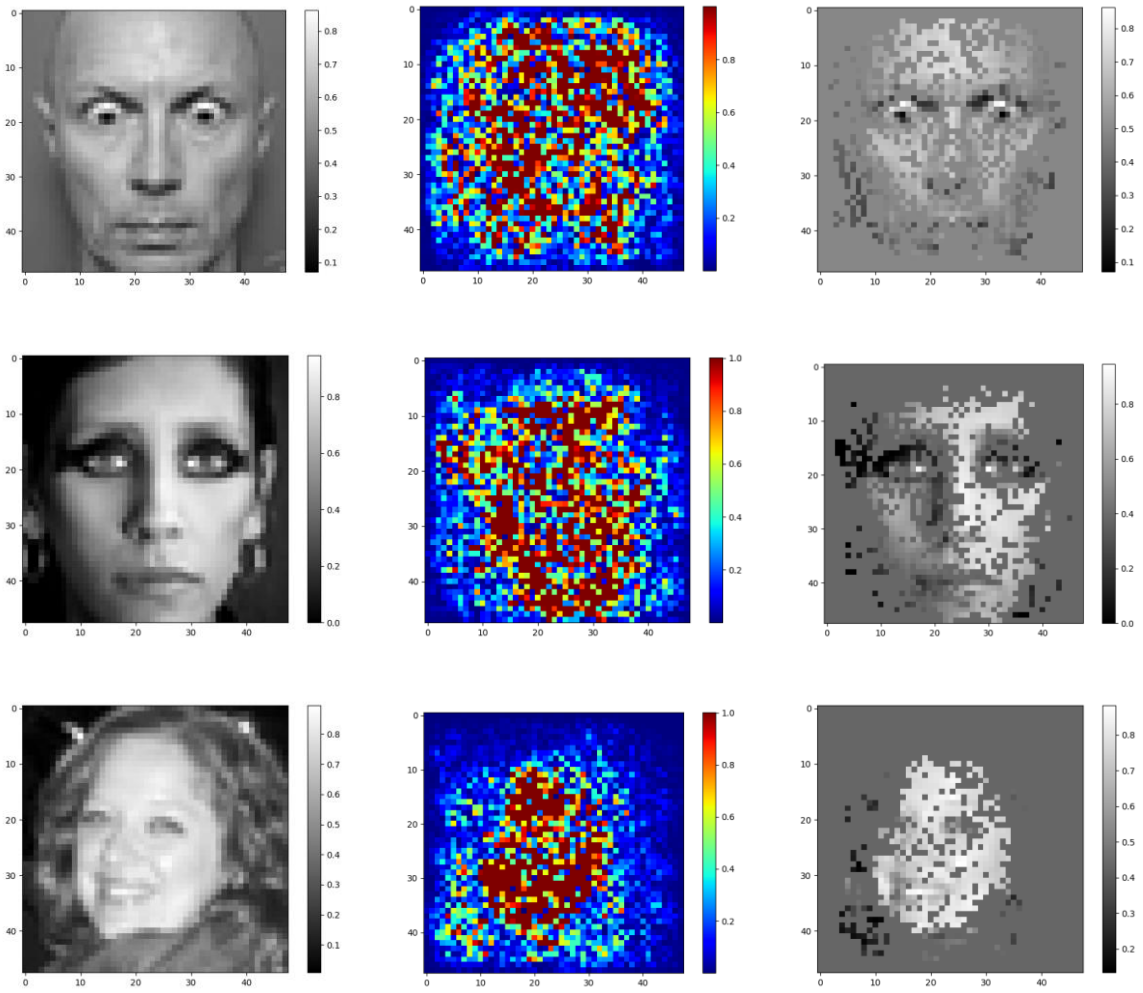
上圖為一正常的 confusion matrix，但為了方便看出哪些 class 易混淆，下圖為去掉正確預測（斜對角線）的 confusion matrix。



結合兩者來看，happy 和 surprised 誤判率十分低，而 happy 在被誤判的情況有很高的機率被判成 neutral。準確率其次的是 sad 和 neutral，他們非常容易被誤

判成彼此。再來 angry 容易被判成 sad 和 neutral，disgust 容易被判成 angry。Fear 是所有裡準確率最低的，而它容易被判成 sad。此外被誤判成 disgust 的機率幾乎是 0，有可能是因為 disgust 的 data 量小導致的，因為 model 會學到「盡量別輸出 disgust」。而許多圖都容易被判成 sad 和 neutral，可能是因為兩者的表情沒有明顯特徵所致。

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？



由上述幾個例子可以看出，DNN 會自動找出人臉，並且聚焦於眼睛、鼻子、嘴巴的部份。

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。