

說明：請各位使用此 **template** 進行 **Report** 撰寫，如果想要用其他排版模式也請註明題號以及題目內容（請勿擅自更改題號），最後上傳至 **github** 前，請務必轉成 **PDF** 檔，並且命名為 **Report.pdf**，否則將不予計分。

-----閱讀完以上文字請刪除-----

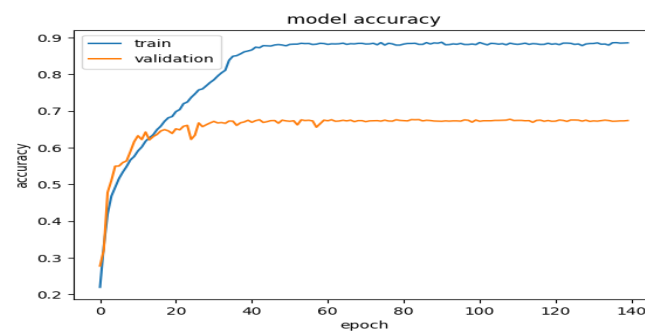
學號：R05942017 系級：電信碩二 姓名：黃梓鳴~

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？

結構:



總共用了 7 層 con 層，每層前面都加個 zeropadding，每個激活函數都用 relu，每層激活層前面都用 batchnormalization，分別在 2,2,3 個 con 層後加 maxpooling，全連接層有四層，output 數量分別為 1024,512,256,128，在 maxpooling 和每層全連階層後都有加 Dropout=0.3，最後一層輸出層的激活函數用 softmax，optimizer 用 adam，loss function 用 categorical_crossentropy，有用 EarlyStopping、ModelCheckpoint、ReduceLROnPlateau 這三個 callbacks 函數。

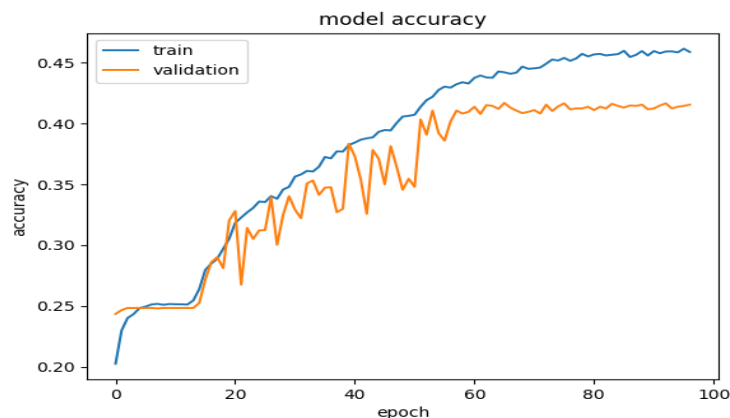


由上圖知在 epochy 在 15 時，val_acc 上升速度開始大幅下降，epoch 到約 40 時，val_acc 就幾乎不再上升，最高約為 67.8，而 train_acc 在 epoch 約 37 時上升速度開始大幅下降，在約 50 開始停止上升

上傳到 kaggle 的 acc 約為 65.1，參數量為兩千三百萬，而這是調升過後的，原本同架構，參數量 3 百萬上船的 acc 為 64.8，兩者相差不大，而訓練時間有大幅差距

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？

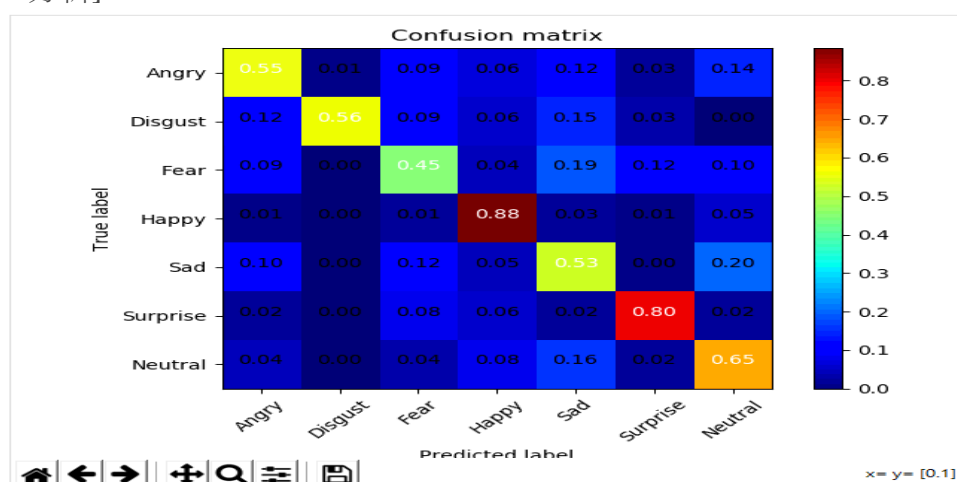
由於層數很長且重複性很高就不放結構圖，有一個 4096,三個 2048,4 個 1024,3 個 512,2 個 256,2 個 128，含輸出層共 16 層 Dense 層。參數和第一題非常接近。



由左圖中觀察到 epoch 到 15 前 acc 都約為 0.25，而 15~55 個 epoch 間 val_acc 有上升的趨勢，但是是上下振盪逐漸上升跟 CNN 的情況明顯有差，val_acc

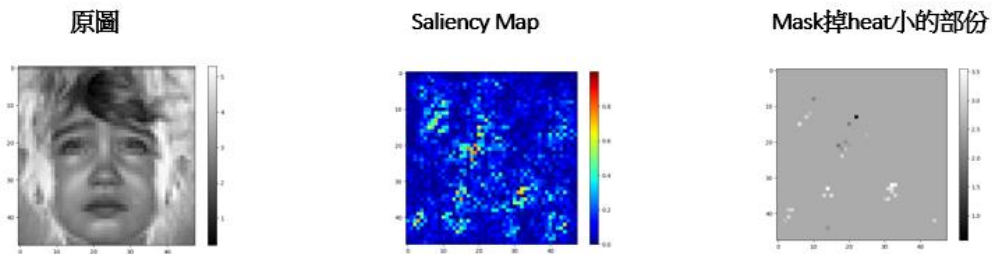
最高也才到 0.4 左右就開始 overfitting，上傳 kaggle 也只得到約 0.4 的準確率，可見 CNN 確實在影像辨識上有顯著的優勢

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]



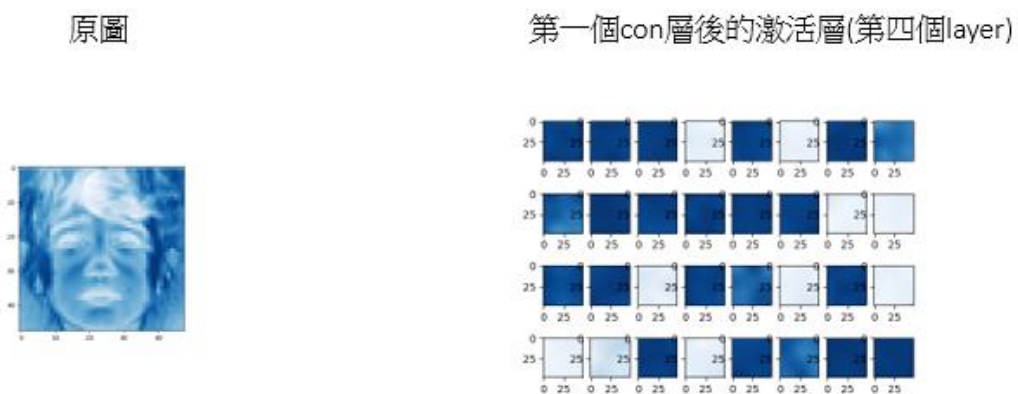
有圖知，表情為 happy 的預測成功率最高(0.88)，fear 的預測成功率最低(0.45)，neutral 和 sad 最容易用混

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？



從 saliency maps 發現機器跟人類觀察到的地方不太一樣，不是關注在眼睛，鼻子五關等，而是圖中特別的一些 pixel

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。



↑上圖中顏色深的為容易被激活的，我原本想說有沒有可能會一個 filter 吃

眼睛，一個 filter 吃嘴巴之類的，但並沒想像中那樣。

第一個conv層後的激活層(第四個layer)output

