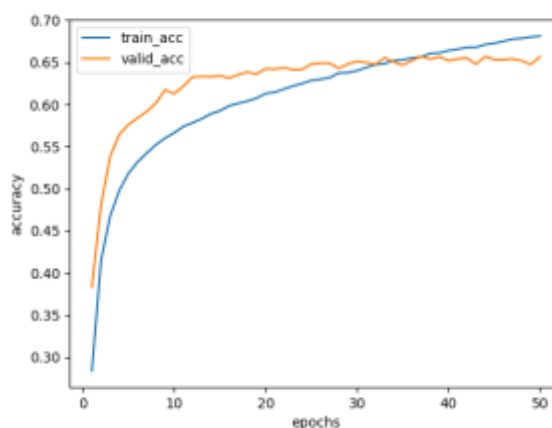


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

我在這個 model 裡做了 7 層 convolution，每一層都 normalize 且 dropout 0.5，並且每兩層做一次 2*2 的 Maxpooling，最後再做兩層 dense。

Data 的部分有使用 keras 的 image generator 將圖片隨機水平反轉、旋轉 5 度、位移 0.1、縮放 0.1。

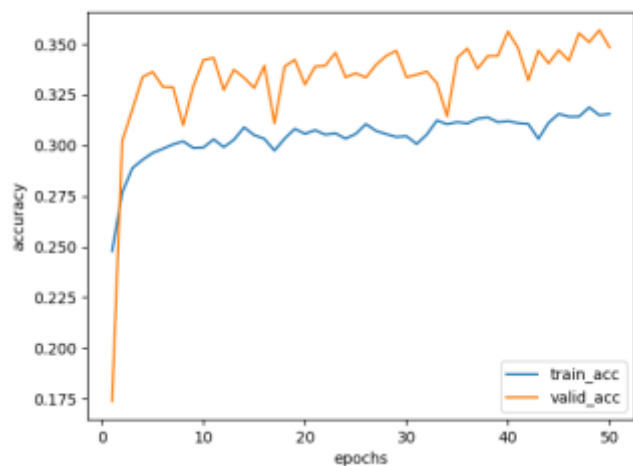
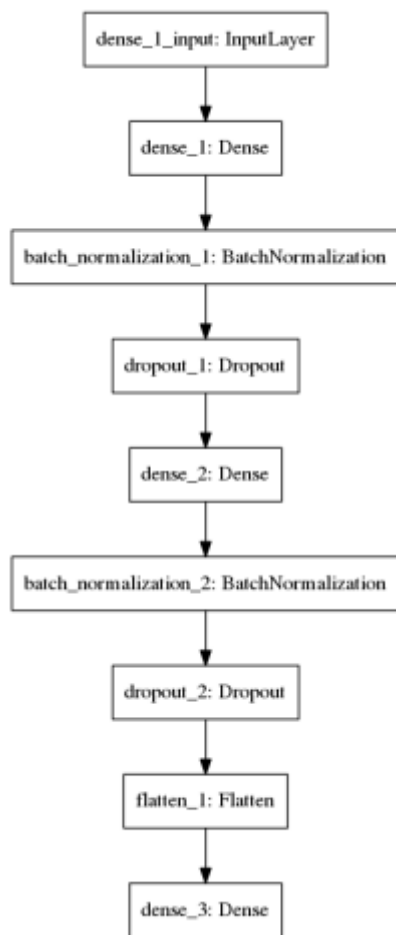
在做了 50 個 epochs 後，valid_acc 超過 0.65 便不再上升，開始 overfitting。



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？

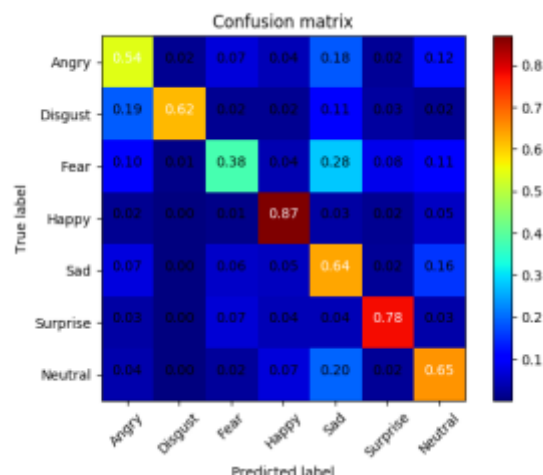
在 dnn model 中，我將前面做的 cnn model 中 convolution 的部分拿掉，並調高每一層的 units 使參數量接近。

做了 50 個 epochs 後，valid_acc 很快的到 0.35 左右便不再上升，此外，在 dnn model 上，training data 似乎也沒辦法 fit。



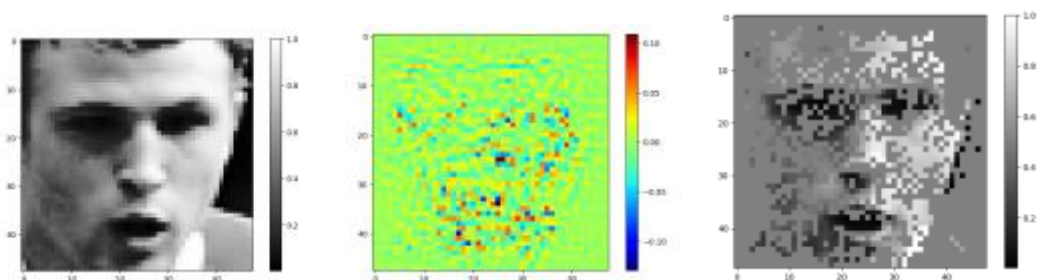
3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]

從矩陣中看出，被辨識錯誤的圖片，大多是被誤辨識為 Sad，如 Fear 和 Neutral 分別有 28%、20%被辨識為 Sad，然而，Happy 和 Surprise 被辨識為 Sad 的比例卻很少，且正確率到高達 70%以上。由此觀察到，似乎負面情緒之間的辨識比較容易被搞混。



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？

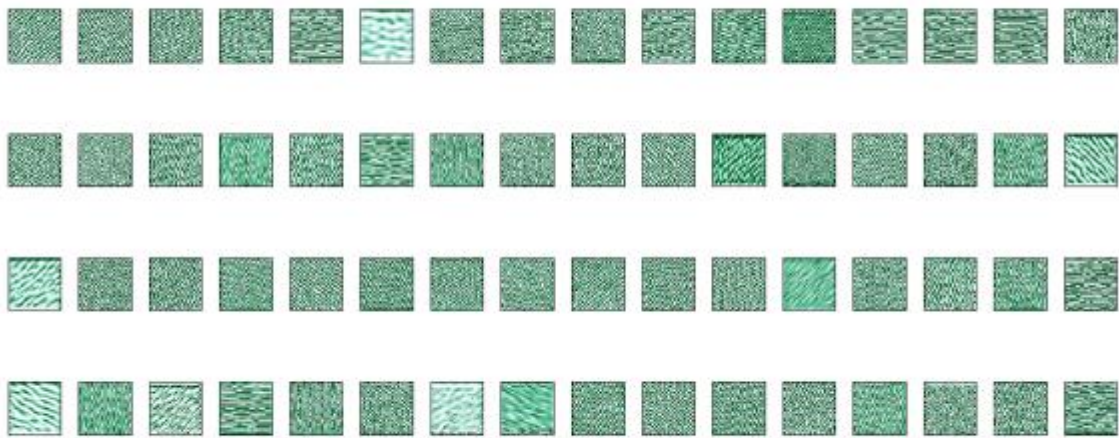
這裡我選擇了編號 7122 的圖片做處理，將 gradient 大於中位數的 pixel 畫出來，發現 gradient 大的 pixel 集中在嘴、鼻、雙眼、跟臉頰。而這張圖中左邊的臉頰比較模糊，因此被 focus 的也較少。



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。

在這個 model 中由 gradient ascent 得到的圖看起來都十分複雜，推測可能是因為 model 層數太多，每一層的 filter 辨識的部份都很細微導致。我也嘗試了其他層的結果，都看不出什麼具有辨識性的圖片。不過從編號 7122 的圖片在該層輸出的結果看出，整體而言圖片灰階的深度跟 gradient ascent 得到的結果大致有相似，顏色越深的 filter 輸出的圖就越深。

Filters of layer conv2d_4 (# Ascent Epoch 150)



Output of layer0 (Given image7122)



[Bonus] (1%) 從 training data 中移除部份 label，實做 semi-supervised learning

這裡使用跟(1)依樣的 model，實作 sel-training。

先將全部 data 中 0.8 的 label 拿掉，只用 0.2 的 data 做第一次 training，
valid_acc 大約落在 0.5 左右，第二次 training，valid_acc 只有在稍微上升一點，到
0.52 左右，但 train_acc 仍持續上升。整體而言，表現仍不如 supervised learning。

