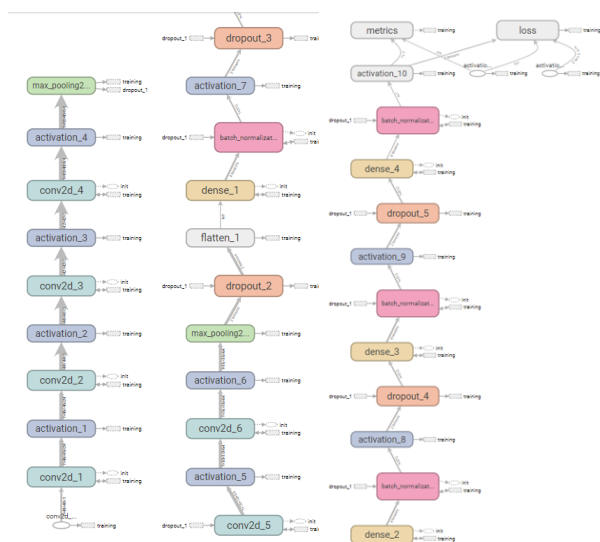


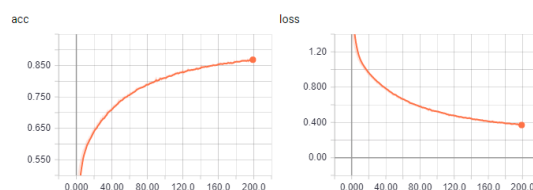
1. (1%) 請說明你實作的 **CNN model**，其模型架構、訓練過程和準確率為何？
(Collaborators:)

答：



訓練過程的 accuracy 和 loss 在 tensorboard 中顯示如右，在 kaggle 上的 public score 為 0.66453。

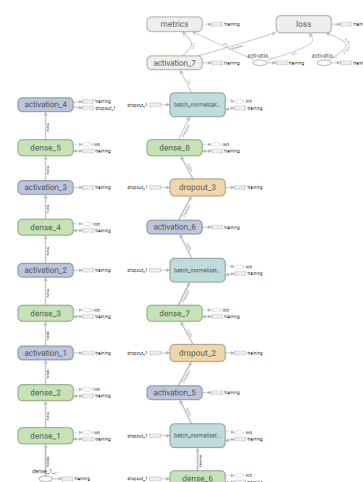
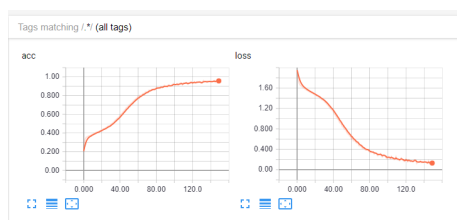
先經由資料前處理產生更多組不同的資料，避免 overfitting。經過 3 層 2D 捲積 + maxpooling + 3 層 2D 捲積 + maxpooling + Flatten + 3 層 FCN 及 batch normalization 來增加模型容忍度，最後用 softmax 產生答案，左圖為 tensorboard 顯示的架構。



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？
(Collaborators:)

答：

我疊了 8 層 FCN 及三層 batch normalization，tensorboard 顯示的架構如右圖。訓練過程中，經由觀察 tensorboard，accuracy 並沒有像 CNN 一樣不斷往上，而在 kaggle 上的 public score 也只有 0.43326。這樣看起來，圖片這種有區域性的 2D 問題，利用 CNN 來訓練會比較理想。

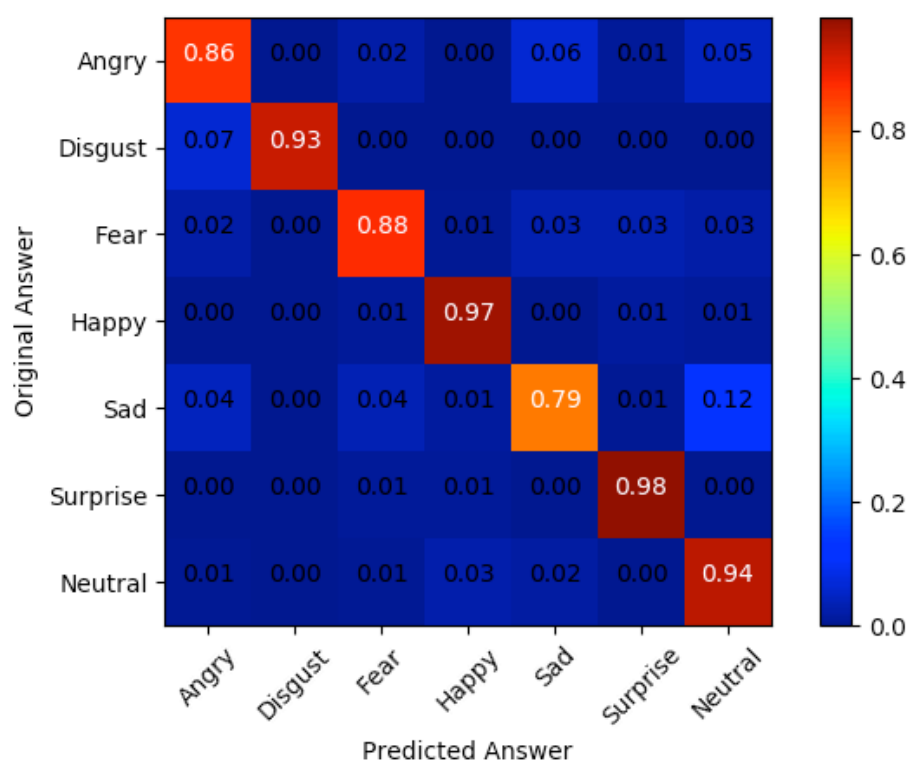


3. (1%) 觀察答錯的圖片中，哪些 **class** 彼此間容易用混？[繪出 **confusion matrix** 分析]

(Collaborators:)

答：

隨機取一群 validation 來繪出 confusion matrix 如下圖。



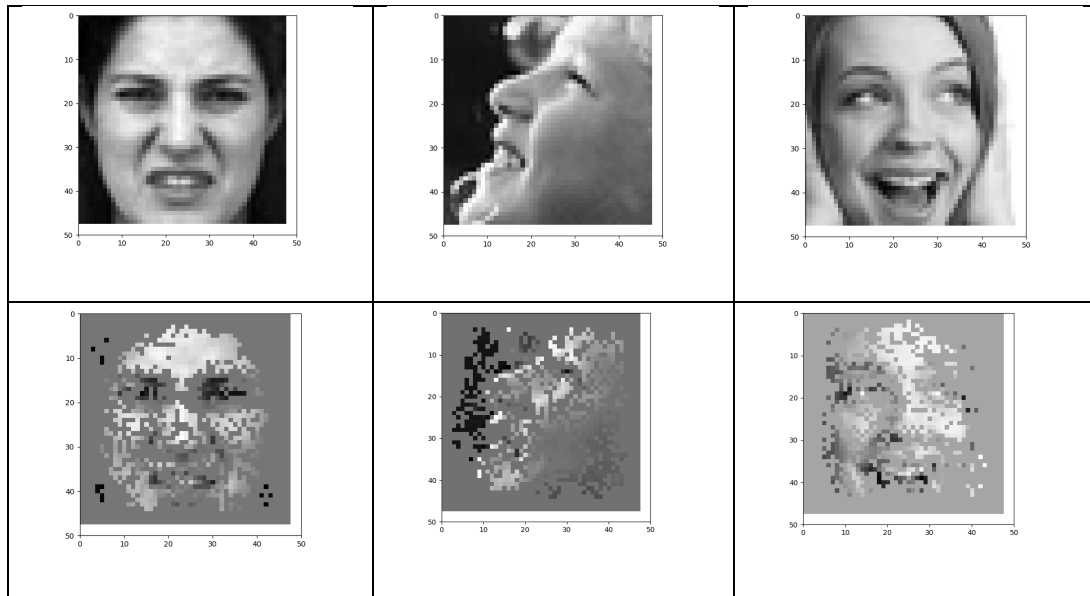
可以發現，除了 sad 較易與 neutral 搞混(0.12)以外，其他的準確率都超過 85%。

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？

(Collaborators:)

答：

隨機挑選 training set 中的三張圖片繪出原圖和 saliency map 如下，來比較。



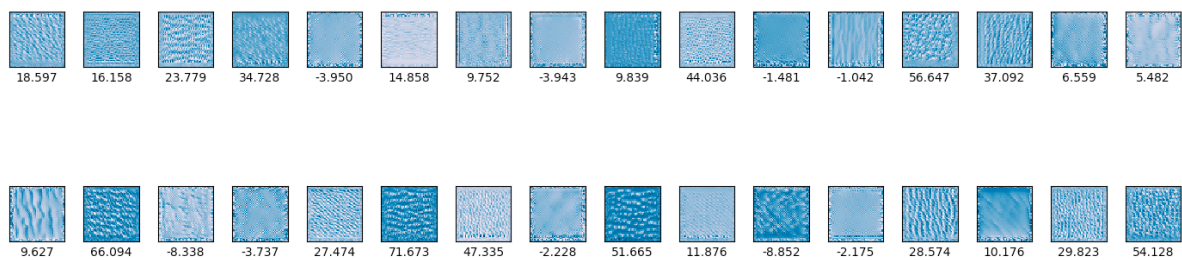
可以看到，經過訓練後，模型會注意在臉部輪廓內的像素，並且會將注意力放在眼睛、鼻子、嘴巴及兩頰的肌肉，就連側臉的圖片也能夠準確抓到重點，這樣的注意力移動和聚焦區域，跟人類在觀察對方表情時的行為一樣。

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**。

(Collaborators:)

答：

我取第四層捲積層來觀察。下圖是最能 activate 該層 filter 的圖片。



發現點狀及霧化最能夠激活這些 filter。