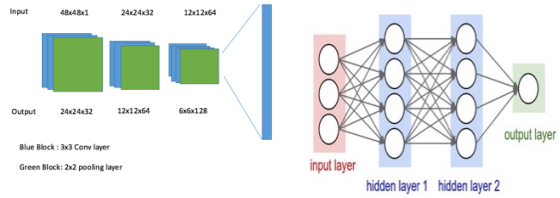
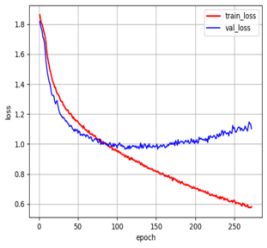
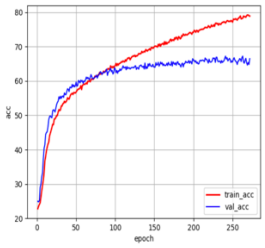
學號：R06921066 系級： 電機碩一 姓名：劉宇閎

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：使用的架構如下圖所示，使用 conv 後加上 max pooling 的數量分別為 2,2,3, 把最終的 feature map 拉成一個 vector 再透過 unit 數分別為 256,256,7 的 fully connected 將特徵一一分開，中使用的最佳化方法為 Adam , learning rate = 3e-4 dropout = 0.5,對影像另外使用 image augmentation 不僅可以增加 train 難度也可以有效抑制 overfitting，此架構及訓練方式在 kaggle 的 public 上得到 0.668 的準確率而 private 也有 0.67762，從圖形也可以看出 loss 在第 100 個 epoch 開始就逐漸上升，如果希望整體 loss 可以更加下降可以考慮加入 decay 使 learning rate 逐漸下降也許可以得到較佳的成果，而因為架構並不算深且在收斂的過程也不算慢，所以我在較少層的 model 中並未加入 batch normalization 來使整個函數收斂更快，另外在 dense 的數量上由於層數不深且輸入影像本身小也使得資訊量亦不大，所以使用的 unit 數也只有使用 256,256,7 如果增加數量有可能導致的問題會是 fully connected 將雜訊也都分入判別時的重要考量資訊中，所以這邊就沒有把ㄙ unit 數加大了。

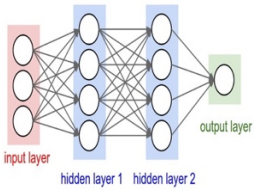


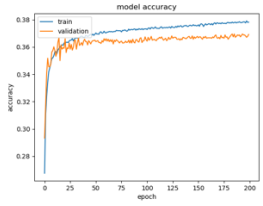


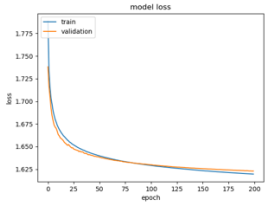


2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼

答：





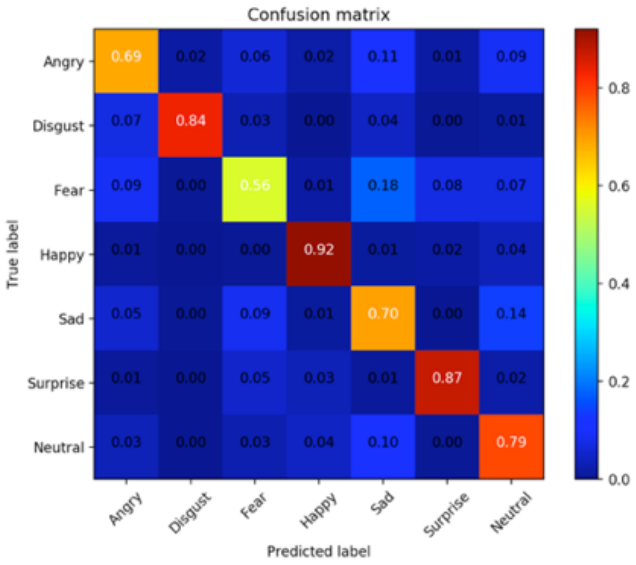


是用共五層的 dense layer 數量分別為 48x48 , 1024, 512, 256, 7 為防止 overfitting 所以賦予 dropout 值 0.3 而這樣的結果丟上 kaggle 後只有 0.41 的準確率，在 CNN 及 DNN 中都對 feature 做了分類(fully connected)但兩者的正確率卻相差慎大，雖然我並沒有將 DNN 做更多的嘗試來使他正確率在更往上升，但稍微經過幾次微調參數及 unit 數後都不見變好所以就以這個 model 來做解釋，在 CNN 中因為 image 透過 kernel 提取出具有空間上特徵的 high level feature 再將她丟入 DNN 中進行分類，所以可以合理的推論出，在表情辨識上 pixel 間的關係是相當重要的，所以 CNN 是因為具有空間上特徵的關係所以可以得到更佳的準確性。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？

答：

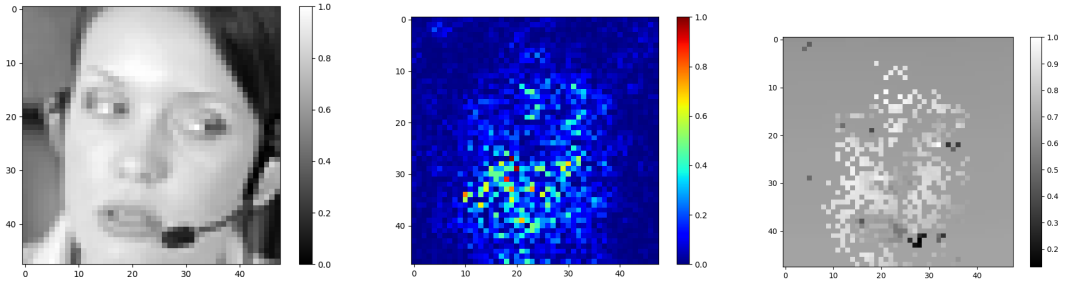
從 confusion matrix 可以看出 fear 和 sad 會容易分辨錯誤，此 dataset 目前人的辨識率大約為 67%，這兩個表情人眼也較難分辨，故此種結果是可以接受的



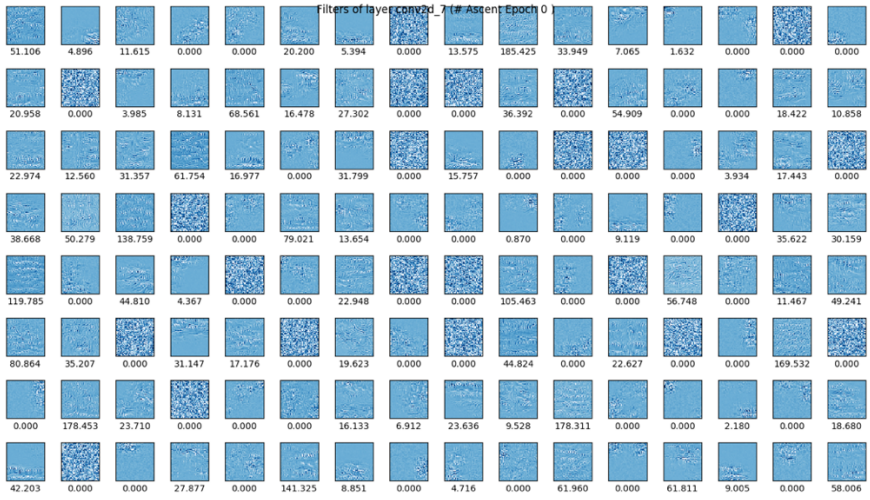
4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？

答：

可以看出使用 CNN 的時候會將重點的部分（五官）提取出來，並將五官與表情做連結，而 kernel 沒有興趣的部分就是被消去的部分，也可以視為判別時的雜訊，雖然值不大，但仍會影響判別結果，所以在一般來說的判別中如果可以做 background normalization 有機會可以在準確度上得到一定的提升，且也不用擔心 train 好的 model 過度 fit 在某種背景上。



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:陳禹齊)
由於 Conv layer 的 kernel 都為 Deep feature(即電腦自行生成)，所以 kernel 判別到的程度也會出現極大的差異。



Output of layer0 (Given image17)