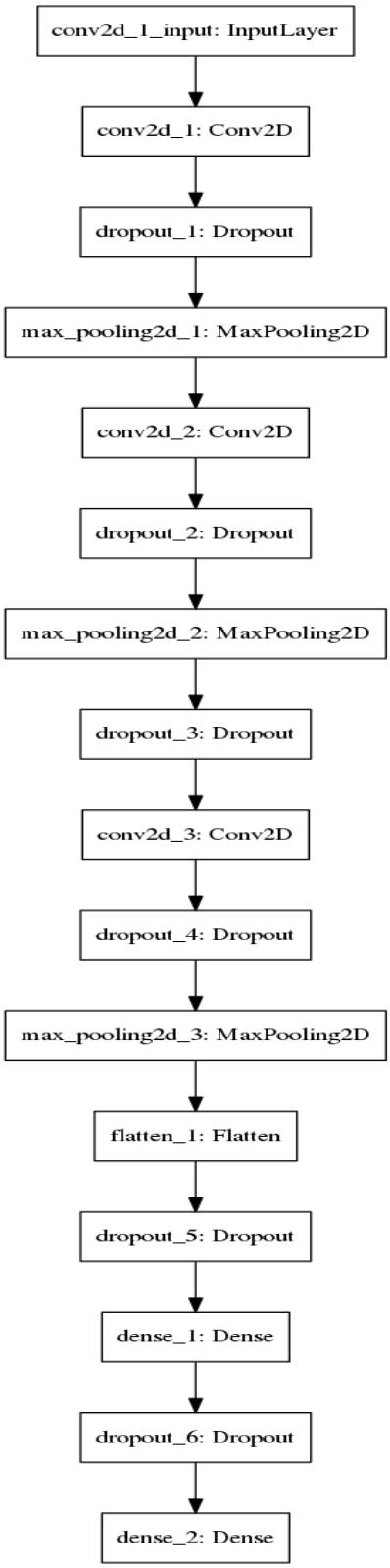
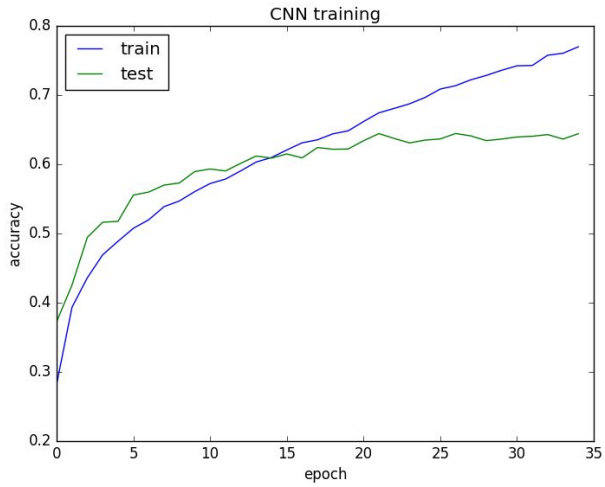
學號：B03901103 系級： 電機三 姓名：陳學平

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： 準確率(在validation set 上)：65.13%





模型架構如左圖所示，conv_layer和max_pool組成的配對一共有三層，使用的activation是 ReLu 。接下來 flatten 後，連接兩層 fully connected layer，使用的activation同樣是 ReLu，最後再連接到 output layer，使用的activation為 softmax。

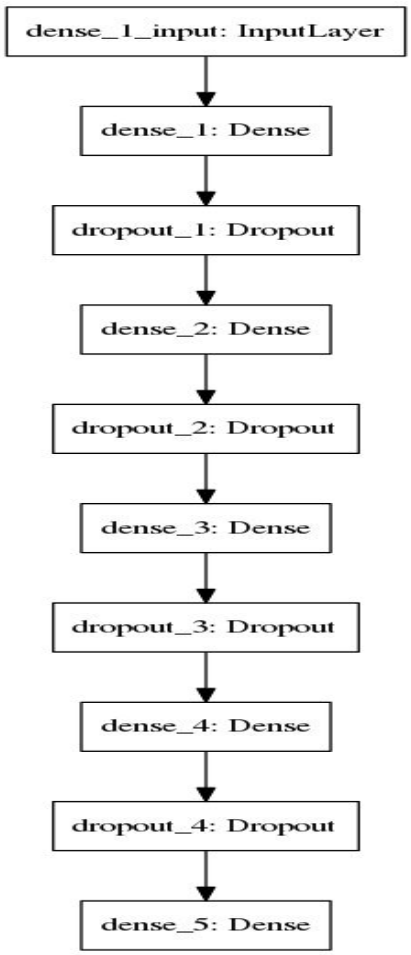
為了避免 overfit 的情形發生，在整個network的架構中有許多地方使用了Dropout，在調整模型的過程中，會發現到適當的使用Dropout能夠使 CNN在validation set上的performance有顯著的進步。

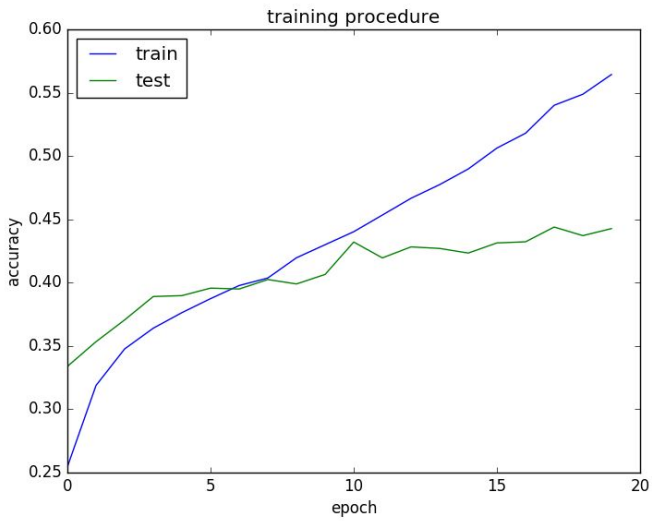
另外，這次唯一對input做的前處理只有正規化，將每張圖片的 pixel 之平均值調整成0，標準差調整為1，這項前處理對於訓練出來的CNN之performance有著比預期中更大的影響。

Model training的過程中，在validation set上，大約25個epoch後，accuracy便趨於平緩，不再上升。

最後值得一提的是，這次的CNN我用來training的optimizer使用的是Adamax，其在validation set上之accuracy大約比 Adam 多了 0.5 %

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： 準確率(在validation set上)：45.20%

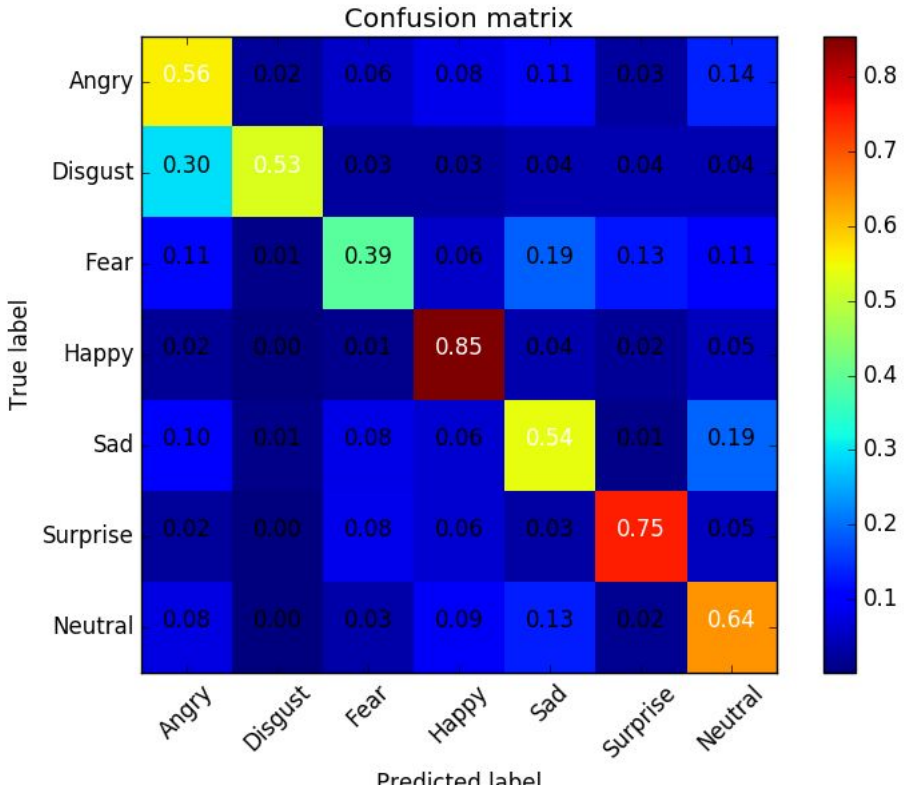




模形架構如左圖所示，簡單的堆疊了4 個 fully connected 的 hidden layer，使用的activation 是 ReLu。和上題相同，連接到output layer後，使用的 activation 是 softmax，並且同樣為了避免overfit，在整個network的架構中多處使用了Dropout，對Input做的前處理也和上題一樣只有正規化。

比較第一題之CNN與本題的DNN，使用的參數數量CNN：9,024,455，DNN：8,920,064，兩者之參數數量十分相近，但performance卻是CNN(65.13%) 遠優於 DNN(45.20%)。另外，在每個epoch上traning花的時間，CNN(35s)大約是DNN(4s)的8、9倍。訓練所花時間的差距十分容易解釋，因為CNN的參數(filter)是shared的，故在參數數量相同的情況下，CNN要做的計算量是遠大於DNN的。至於performance的差距，這則是因為CNN的架構本身就是特別適合用來處理 image 的問題，故我們可以推測CNN的bias比DNN的bias來得小。

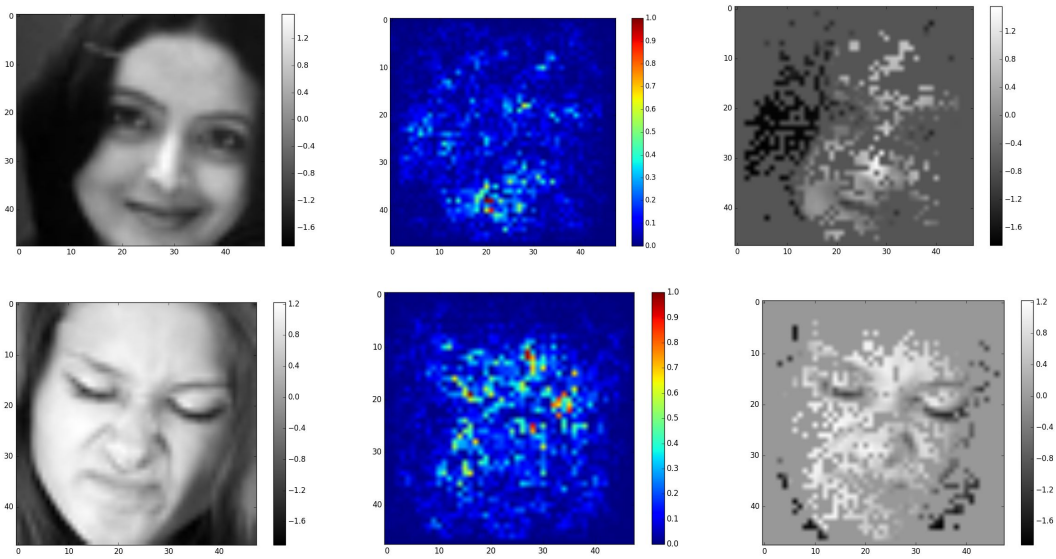
3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]
答：



從 confusion matrix可以看出，Disgust有30%的機會被誤判成Angry，另外 Fear有19%的機會被誤判成Sad，Sad有19%的機會被誤判成Neutral。這樣的結果還蠻直觀的，因為Disgust、Angry、Fear、Sad都屬於負面的情感，因此這些表情被誤判的機率自然也比較大。

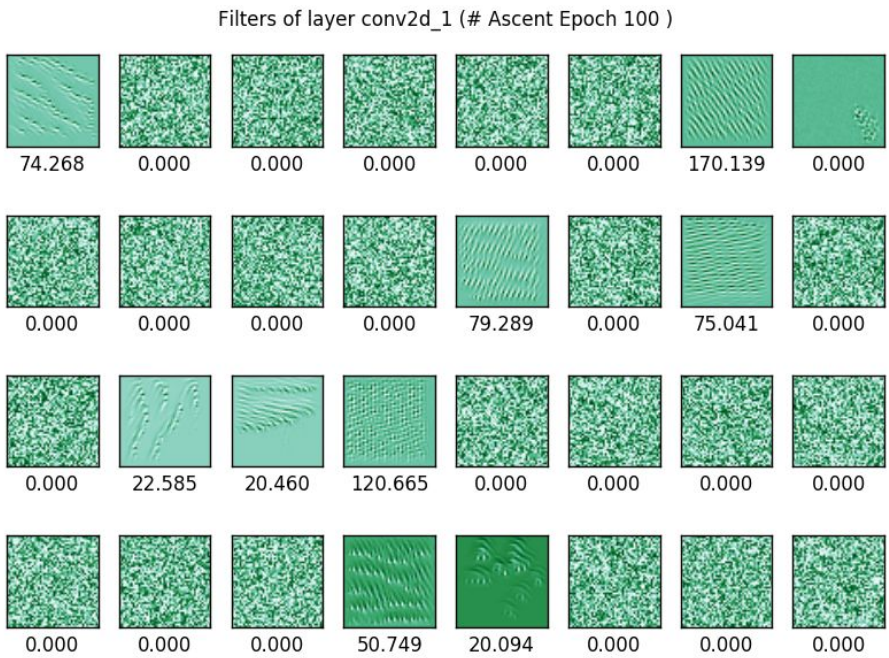
其中，觀察confusion matrix的對角線，可以發現Fear被判斷正確的機率非常小，只有39%，而Happy則有高達85%的機會被判斷正確。這與我們一般人判斷他人情緒的狀況相似，要判斷一個人是否高興是相對容易的。

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：在下一頁的圖中，第一列為高興的表情、第二列為厭惡的表情

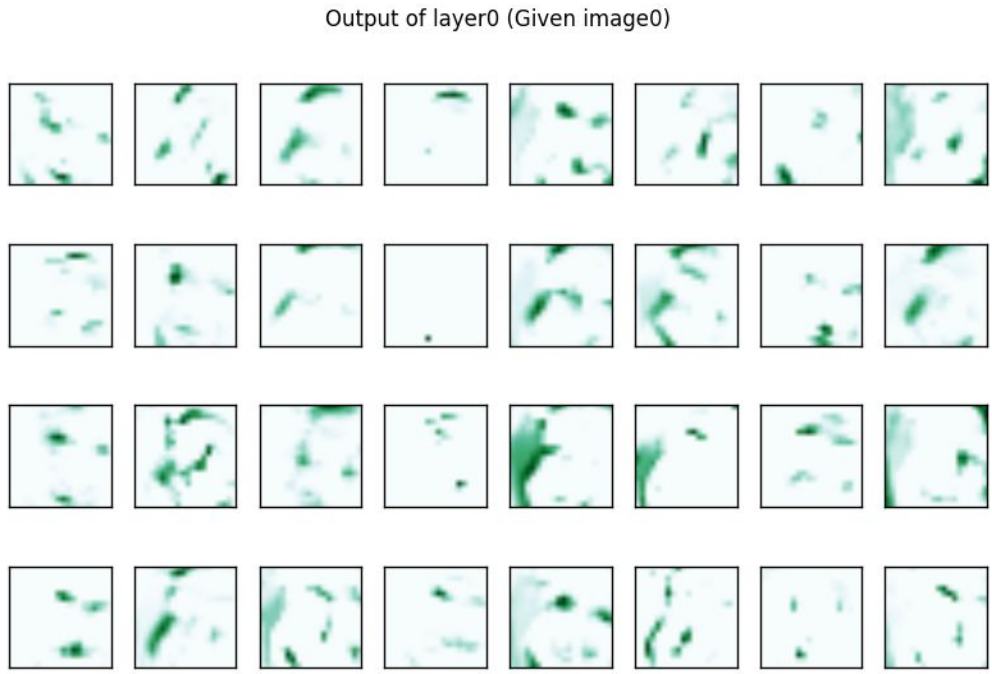


由saliency maps可知，在處理高興的表情時，模型是focus在嘴巴的部份，這與一般人高興時會嘴角上揚的概念相符合；另外，在處理厭惡的表情時，模型主要是focus在眼睛的周遭，這也與大部份的人會藉由眼神來表示厭惡的現象相符。

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。
答：本題之挑選第一層conv2d中的32個filter來分析



本題使用之Input圖片與第四題之高興表情相同：



比較上下兩圖，稍微容易辨識的特徵有兩個：其中一個可以發現第四列的第五個filter是在選取眼睛的特徵，而對應的gradient acent形成的條紋有類似眼睛的輪廓。另外第一列的第七個filter雖然比較不明顯，但是勉強能看出是在辨識嘴形，對應的gradient acent形成之條紋也與嘴巴形成的曲線相似。

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

先將training data 切成 training set (23000筆資料)和 validation set(5708筆資料)，再移除training set 中一半的label，使用 self training 來實作 semi-supervised learning。

未使用self training(也就是只有11500筆training data)：val accuracy = 58.53%；使用self training：val accuracy = 59.95%， 由以上結果可以看出，利用 unlabled data確實可以提升performance。

[Bonus] (1%) 在Problem 5 中，提供了3個 hint，可以嘗試實作及觀察 (但也可以不限於 hint 所提到的方向，也可以自己去研究更多關於 CNN 細節的資料)，並說明你做了些什麼？ [完成1個: +0.4%, 完成2個: +0.7%, 完成3個: +1%]