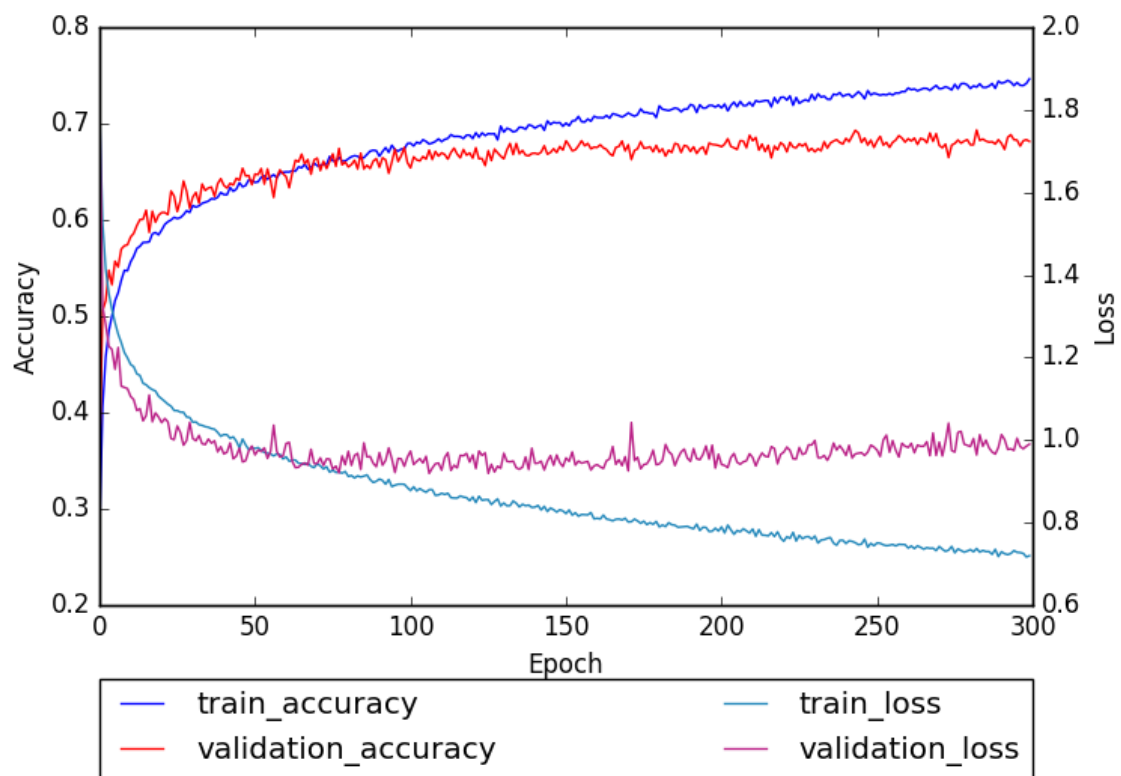


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

答：

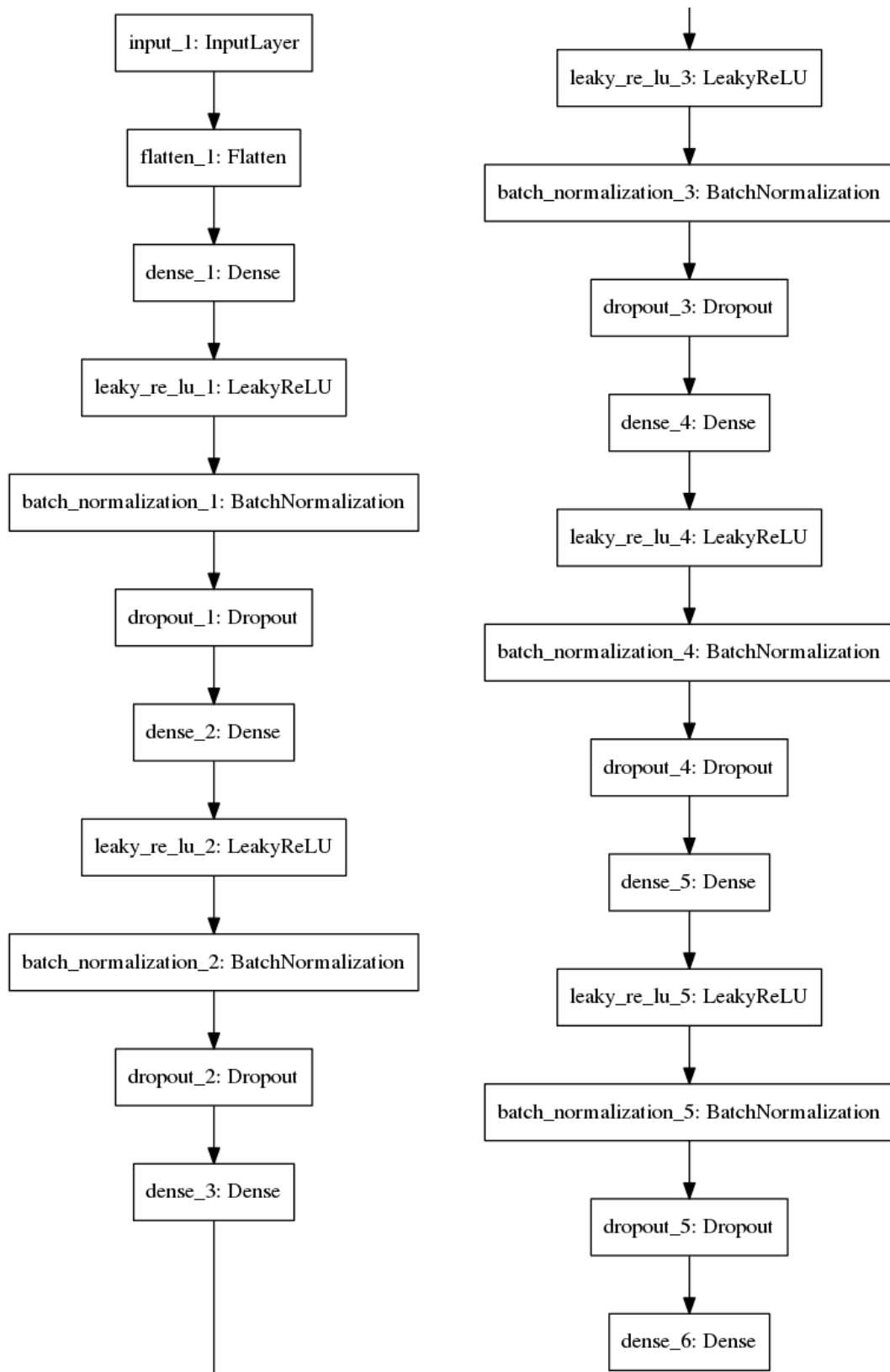




五組 layer，每組裡面有兩個 Conv2D 和一個 Maxpooling，最後一組用 GlobalPooling 削去位置資訊，同時有用正規化和 Dropout 在訓練時使用 keras 的 ImageGenerator 對原始輸入變形，就可以訓練出不錯的結果。前處理做了拉高對比和銳利化。有測試過 gamma correction, edge detection, CIE La\*b\* color space(在人類視覺接近線性的色彩空間), 但都沒有獲得比較好的結果。

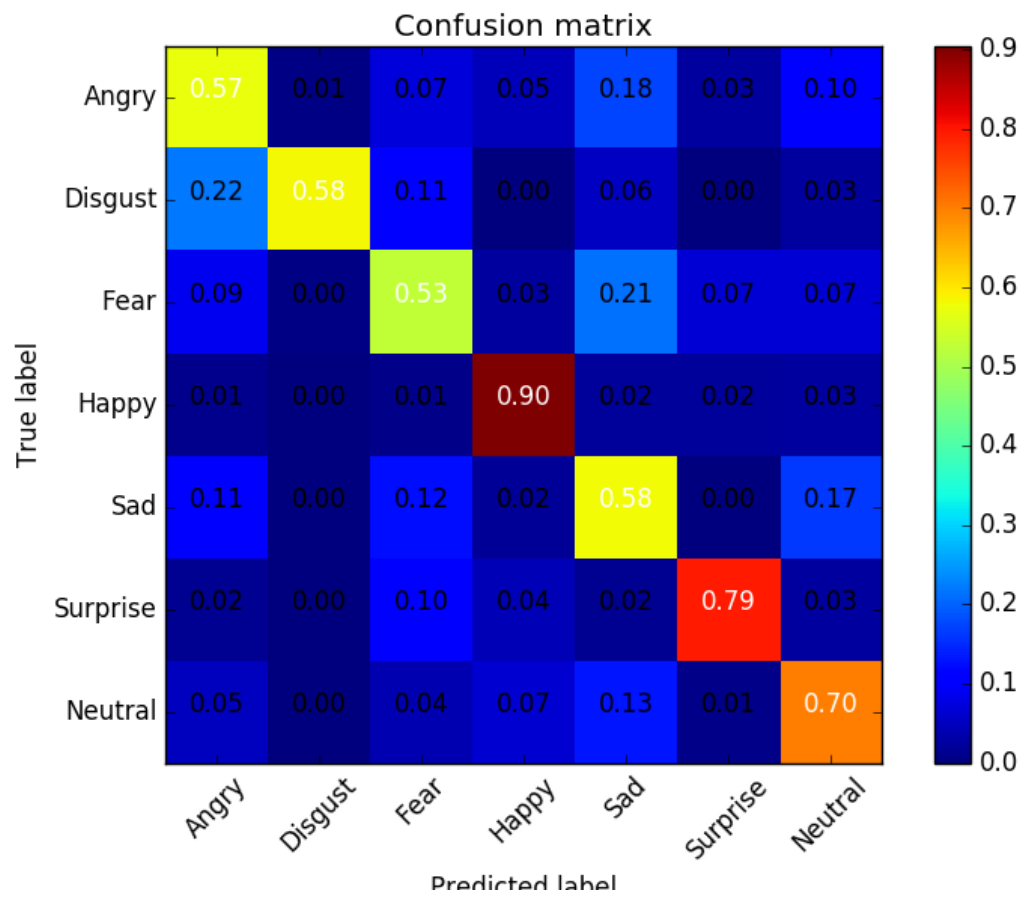
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

答：

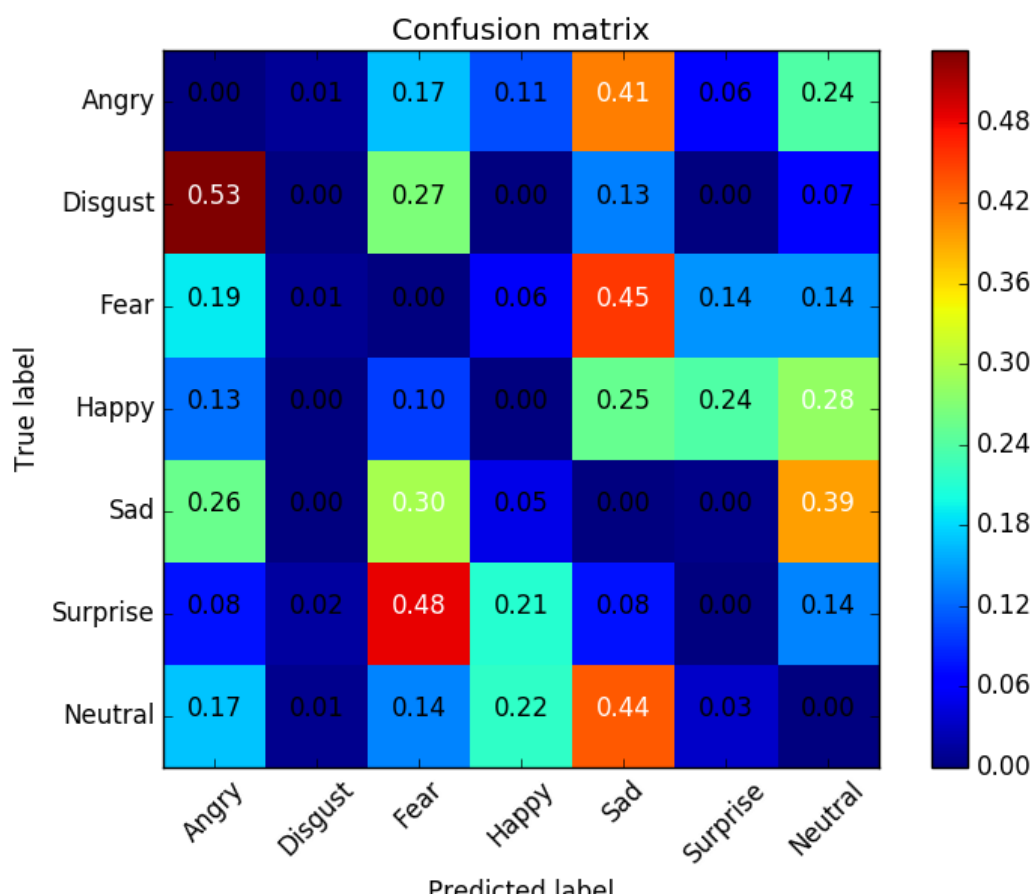


儘管參數已經略多於 CNN，但是訓練效果還是很差，畢竟圖片平移應該不影響結果，但是 DNN 會是看絕對位置來計算，所以效果很差也不太意外。在剛開始嘗試的時候有加入 FFT 之後訓練到 0.54 左右，但是 model 遺失了。

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



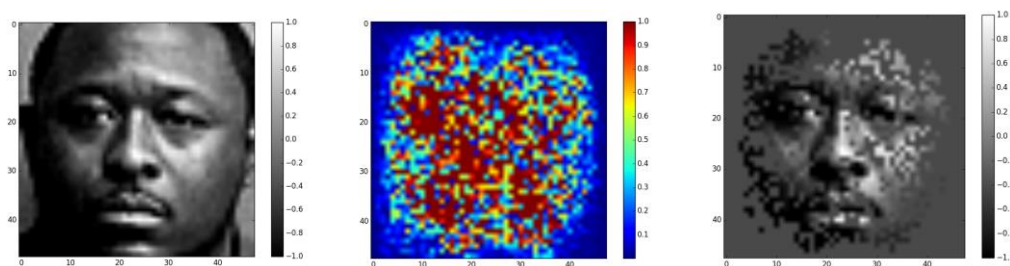
輸出的 Confusion matrix.

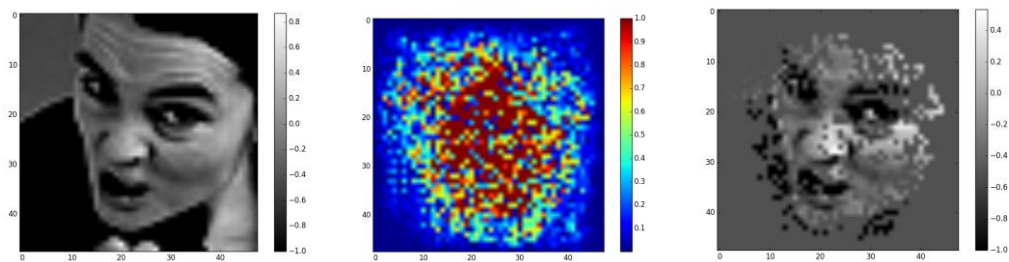


去除正確判定的 confusion matrix.

可以明顯看出 Disgust 弄錯的量很少，但是這其實是因為 Disgust 的 Sample 數量在整個 dataset 裡面很少，Model 可能學到傾向不輸出 Disgust，Happy 的正確率很高是因為他有明顯的特徵，這部分會在最後的分析提到。

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：

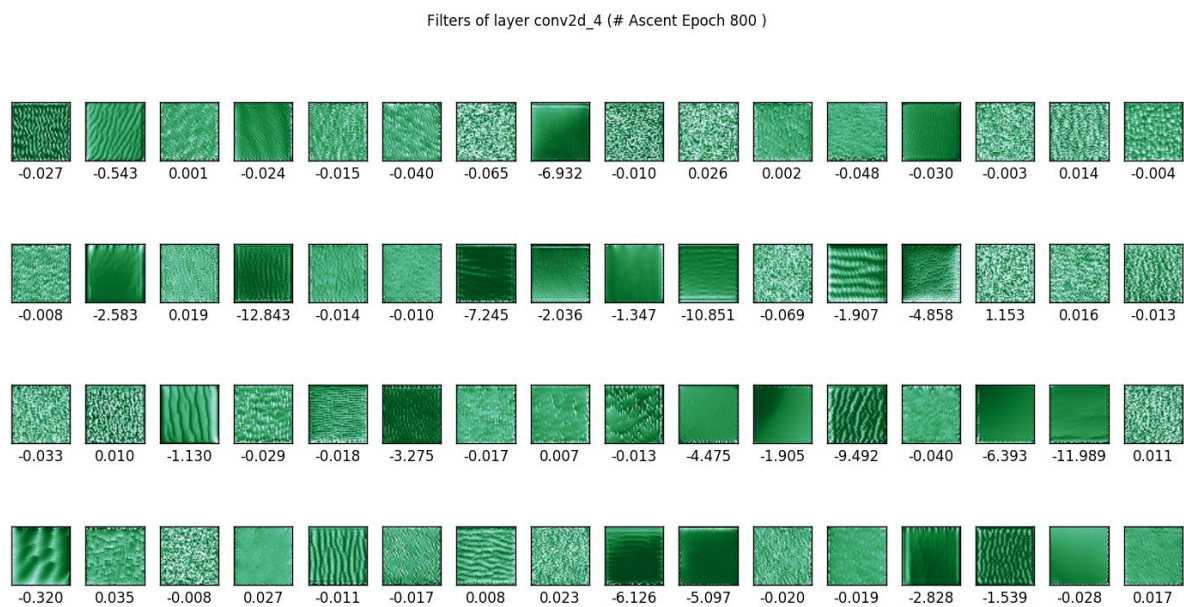




可以看出來 Model 可以把表情的部分從圖片中分離開來

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：



Gradient Ascent from 雜訊

Filters of layer conv2d\_4 ( # Ascent Epoch 800 )



Gradient Ascent from 正確預測而且信心很高(e.g 輸出 0,0,0,0,0,0,1)

Filters of layer conv2d\_4 ( # Ascent Epoch 800 )



Gradient Ascent from 錯誤預測而且信心很低(e.g 輸出[0.15,]\*7)

優化器使用 adam 進行優化,我選的那一層很神奇的只能用 adam 優化,試過 rmsprop, adagrad, vanilla 都跑不起來,結果上來看從雜訊開始訓練的大多都還在紋路的階段,不像助教的範例有類似五官的圖形,而從照片開始訓練的雖然有臉,但是兩個極端的正確和錯誤預測的輸入卻都得到相似的效果,看來這個視覺化的方式在我的模型並不是很適用.

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

無實作

[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%]

除了前五個 problem，我另外分析了預測信心最高且正確預測的圖片(e.g 預測值像是 [0,0,0,0,0,1])。可以明顯看出，預測信心最高的的圖片，大部分都是 Happy 且有很明顯牙齒特徵，從 Saliency mask 中也可以看出他有抓到牙齒。

同時我也分析了預測信心最低的 150 張圖片，很多都有多種特徵在裡面，雖然人工看可能很容易判別，但是抓特徵可能就無法確定，裡面也有兩張非真實照片的卡通圖混在裡面 XD 第一張卡通圖的表情在 Saliency mask 中完全壞掉了

圖片附在下一頁

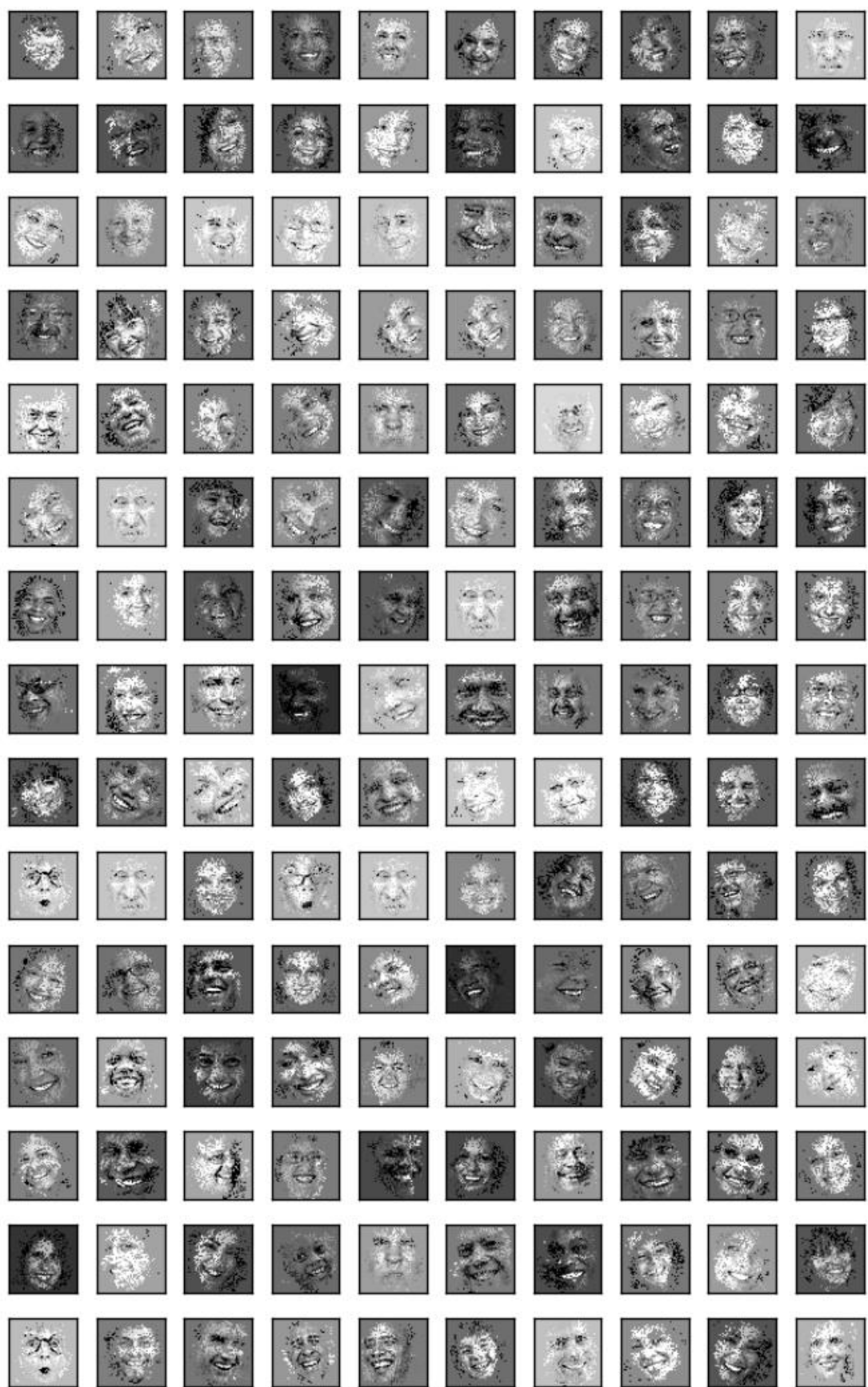




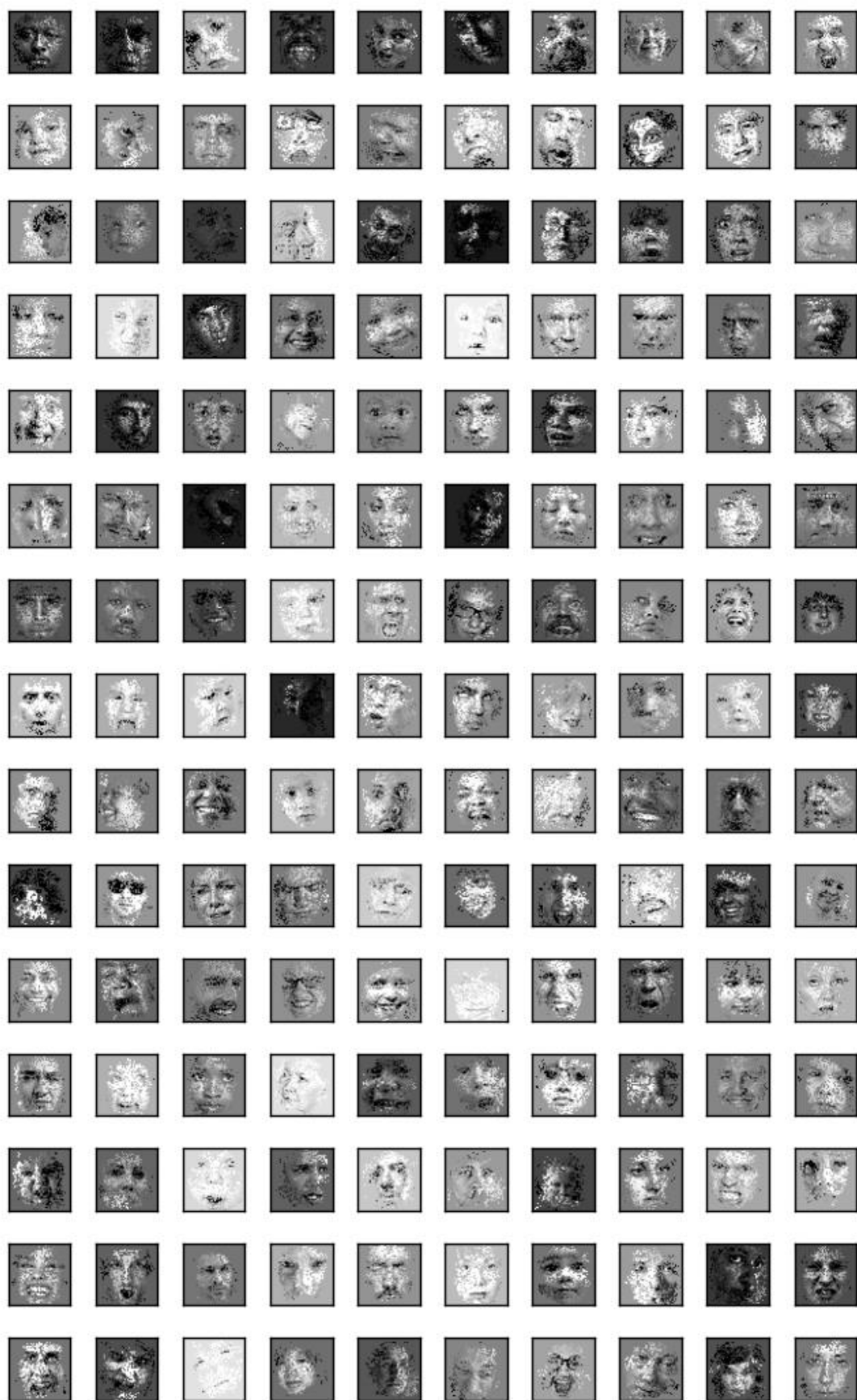
預測信心最高的照片(e.g 預測值像是[0,0,0,0,0,0,1])



預測信心最低的 150 張照片 (e.g. 預測值像是  $[0.15, ] * 7$ )



預測信心最高的照片的 Saliency mask



預測信心最低的 Saliency mask