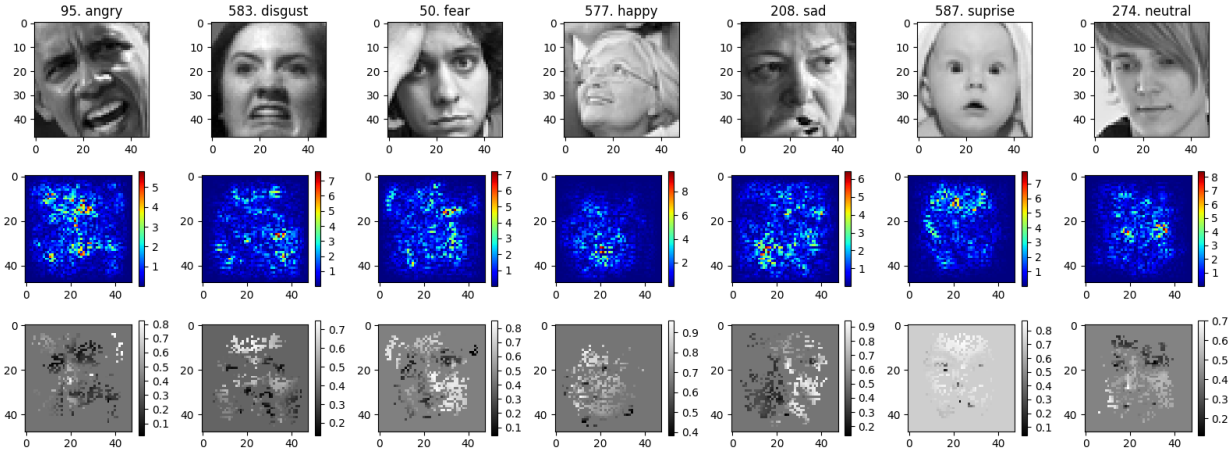
學號：R07942115 系級： 電信碩一 姓名：謝硯澤

1. (2%) 從作業三可以發現，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: )

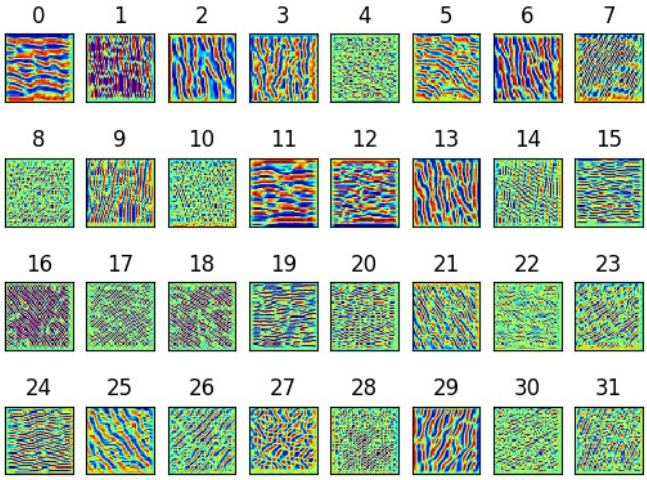
   答：




觀察發現在做 classification 時，主要是 focus 在圖片的臉部上，特別是眼睛、鼻子、嘴巴、眉間。而臉部以外的背景圖案（如圖片的四個邊角）幾乎沒有任何決定 classification 的能力。

2. (3%) 承(1) 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 與觀察 filter 的 output。(Collaborators: )

   答：

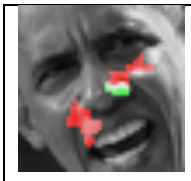
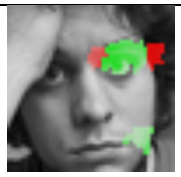


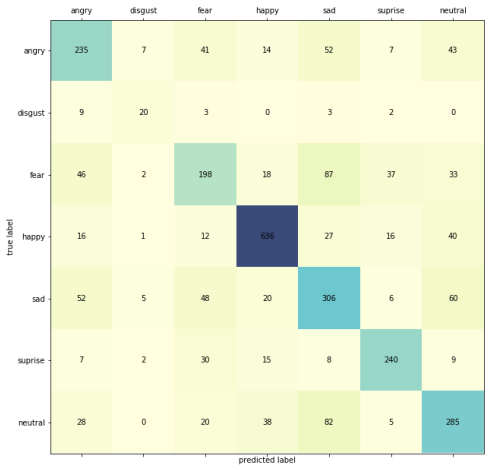

可以發現在第一層中的 filter，最容易被一些簡單的紋理，如直線段、斜線段所 activate。而很多 filter 的結果其實也長蠻像的，但是可能旋轉角度不太一樣。

第一層的 conv_layer 感覺是抓取一些輪廓特徵而已，而這些特徵好像都是由簡單的線條所構成，跟上述的圖片蠻吻合的，都是抓取比較粗糙的特徵。

3. **(3%)** 請使用 **Lime** 套件分析你的模型對於各種表情的判斷方式，並解釋為何你的模型在某些 **label** 表現得特別好 **(**可以搭配作業三的 **Confusion Matrix)**。

答：

| | | | | | | |
|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |
| angry | disgust | fear | happy | sad | surprise | neutral |

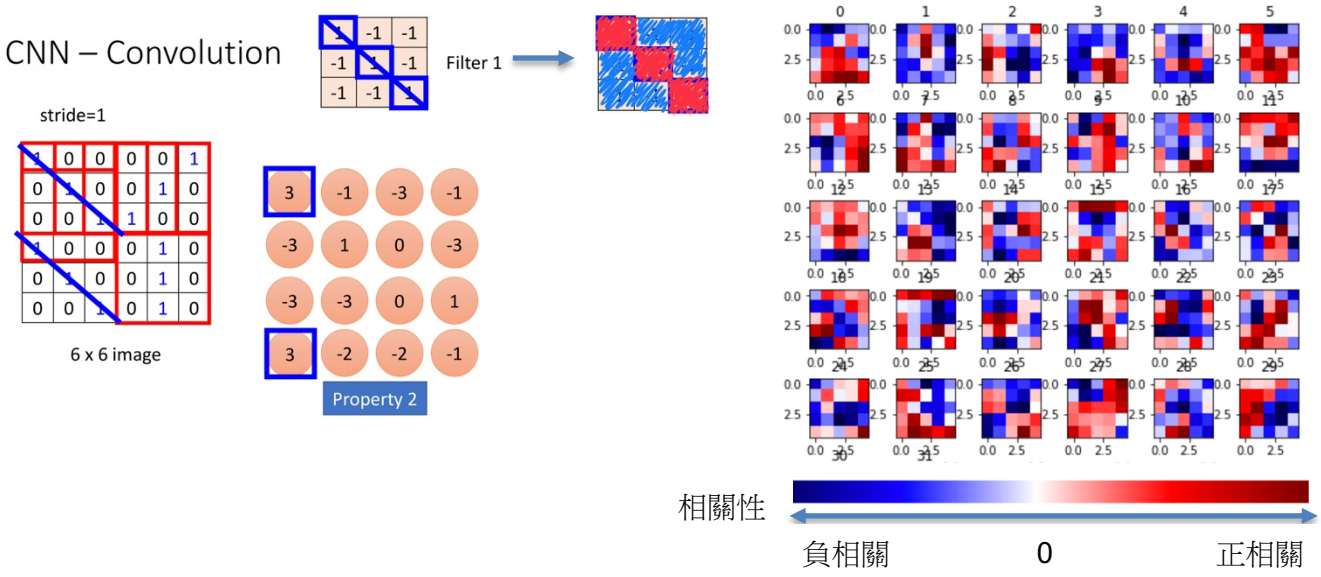



首先，我挑選了 validation set 中，model 分類正確的照片來做 Lime。用人類比較直觀的想法會認為說，去判斷一個人的情緒，應該會從他的眼睛、眉毛、鼻子、嘴巴這些地方去判斷，而 Lime 的結果好像也是如此！？上述的七個類別，綠色和紅色色塊分別代表正相關和負相關，這些色塊大多是落在五官上，只有少部分的類別會跑到其他地方。這些容易跑到其他地方的類別，在 confusion matrix 中好像都是辨識較低的類別。其中，我覺得 angry, disgust, fear, sad 我覺得辨識率都不是很好，在作業三中也有提到特別容易搞混。而會有這樣的問題，我覺得可能是 data 量較少，如 disgust。或是這些情緒的表情差異並不是很大，我自己也有拿 validation 的圖片來做個測試，發現其實人類去看某些類別其實錯誤率比機器還高，我選了二十張照片只答對了 12 張，比 model 還弱，所以也有可能 model 會學到一些不一定是五官上的特徵？而在 happy 和 surprise 的表現最好，可能的原因是正面情緒和負面情緒的表情本來差異就蠻大的，如果今天再新增更多類別如興奮、尷尬笑？可能快樂的辨識率就會下降了吧。

4. (2%) [自由發揮] 請同學自行搜尋或參考上課曾提及的內容，實作任一種方式來觀察 CNN 模型的訓練，並說明你的實作方法及呈現 visualization 的結果。

答：

我畫出某一層 conv_layer 所有 filter 中的參數，理由是如果 filter 中的參數越大或越小代表了高度正相關及高度負相關，預期可以看到一些如上課投影片中的 pattern。但視覺化（右下圖）後的結果好像並不明顯。




最後我嘗試取絕對值，只在意參數是否具有高度相關性，並把參數值超過某個門檻值給予紅色，其餘為藍色。如下圖：