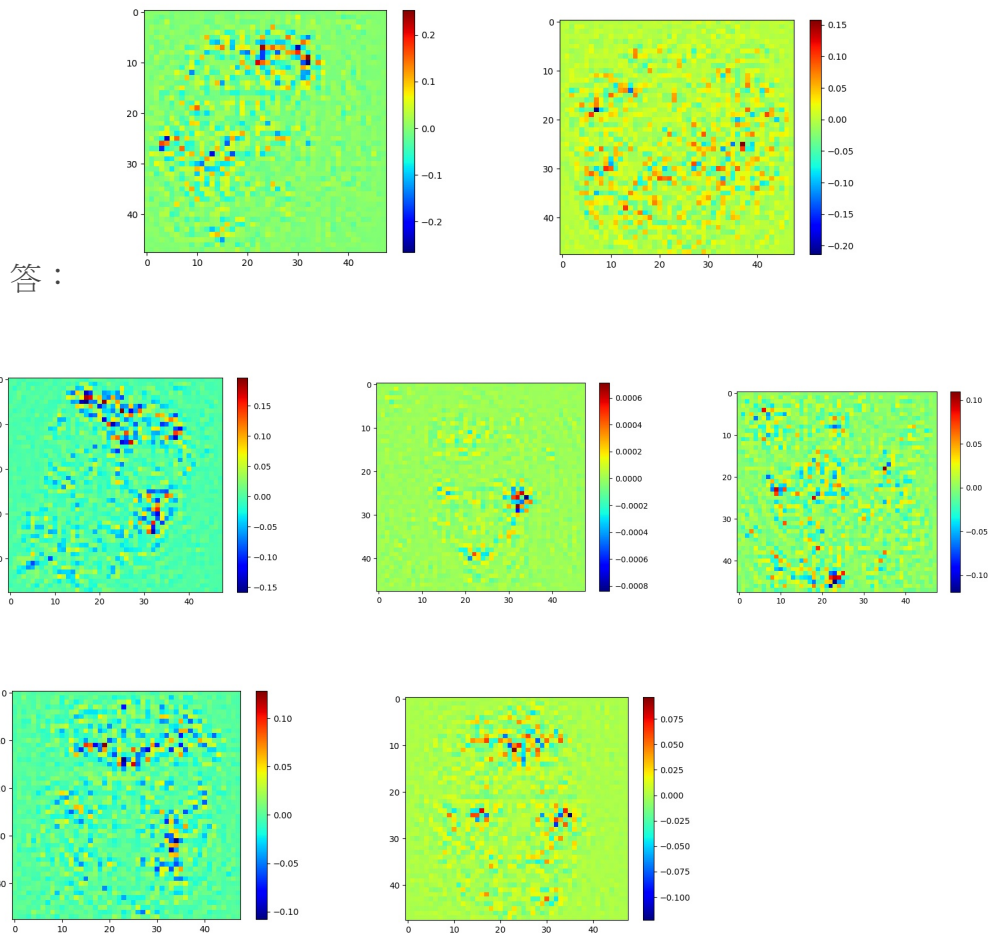


學號：B05902018 系級：資工三 姓名：張凱程

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: 無)

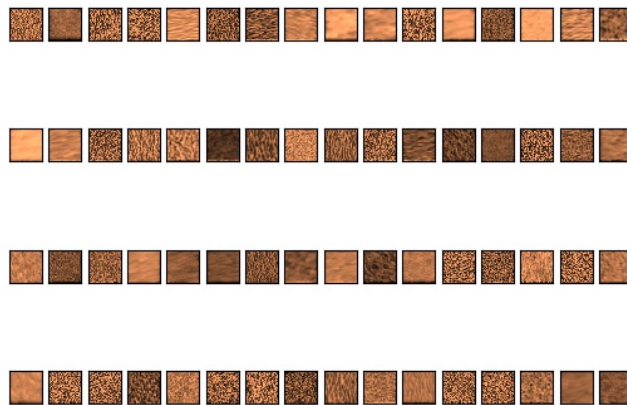


圖依序為憤怒、厭惡、害怕、高興、難過、驚訝跟中立，每張都是該 label 的第一張圖，圖中可知主要都是觀察臉部的輪廓，例如額頭靠近眼睛的部分以及臉頰，這兩個比較明顯。

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: 無)

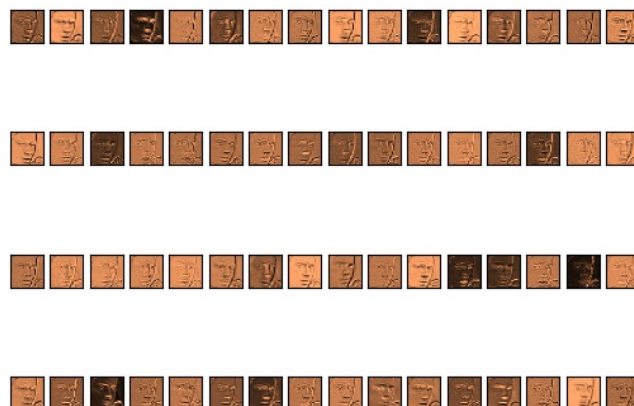
答：以第二層 Conv2D 為例，最容易被 activate 的圖片如下，部分還是 noise，但主要以單向的紋路為主，且紋路清晰度不一。

Filters of layer conv2d\_2



顏色比較深的對應到的 filter activation 比較像 noise，而會讓 output 的特定方向的輪廓更深的通常是紋路，還有 output 輪廓跟原圖較相近的，其 filter activation 較像 Gaussian noise (如 1-2 跟 3-11)。

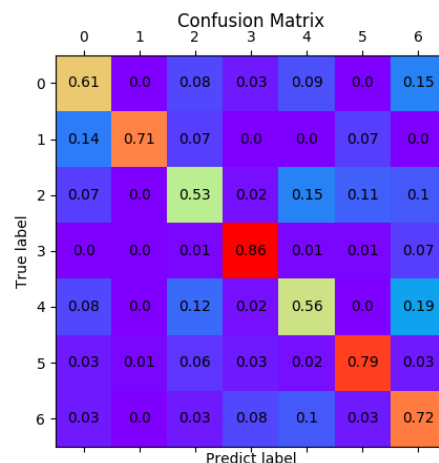
Output of layer conv2d\_2 (Given image 0)



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）。



可以看出這七個例子中，高興跟驚訝是相對明顯的，所以高興跟驚訝的相關區域跟臉部表情相似，至於憤怒、厭惡的表情相似度較高，有可能有誤判或分析不夠對位（但我的 model 貌似把憤怒誤判為中立的機率大一些），而害怕跟難過也是相互容易影響的部分，所以表現也沒那麼好。底下為對應的 Confusion Matrix。



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：使用方法：grad-cam

作法：將 feature map 對特定層的不同 filter 做梯度，然後除掉梯度的「方均根加 epsilon」，接著就以該梯度作為權重跟 filter 做內積就能獲得 CAM，把它加在原圖上做比對。

圖：（左側為針對 conv2d\_2、右側則是 conv2d\_4，以第一張照片實作）

