

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？

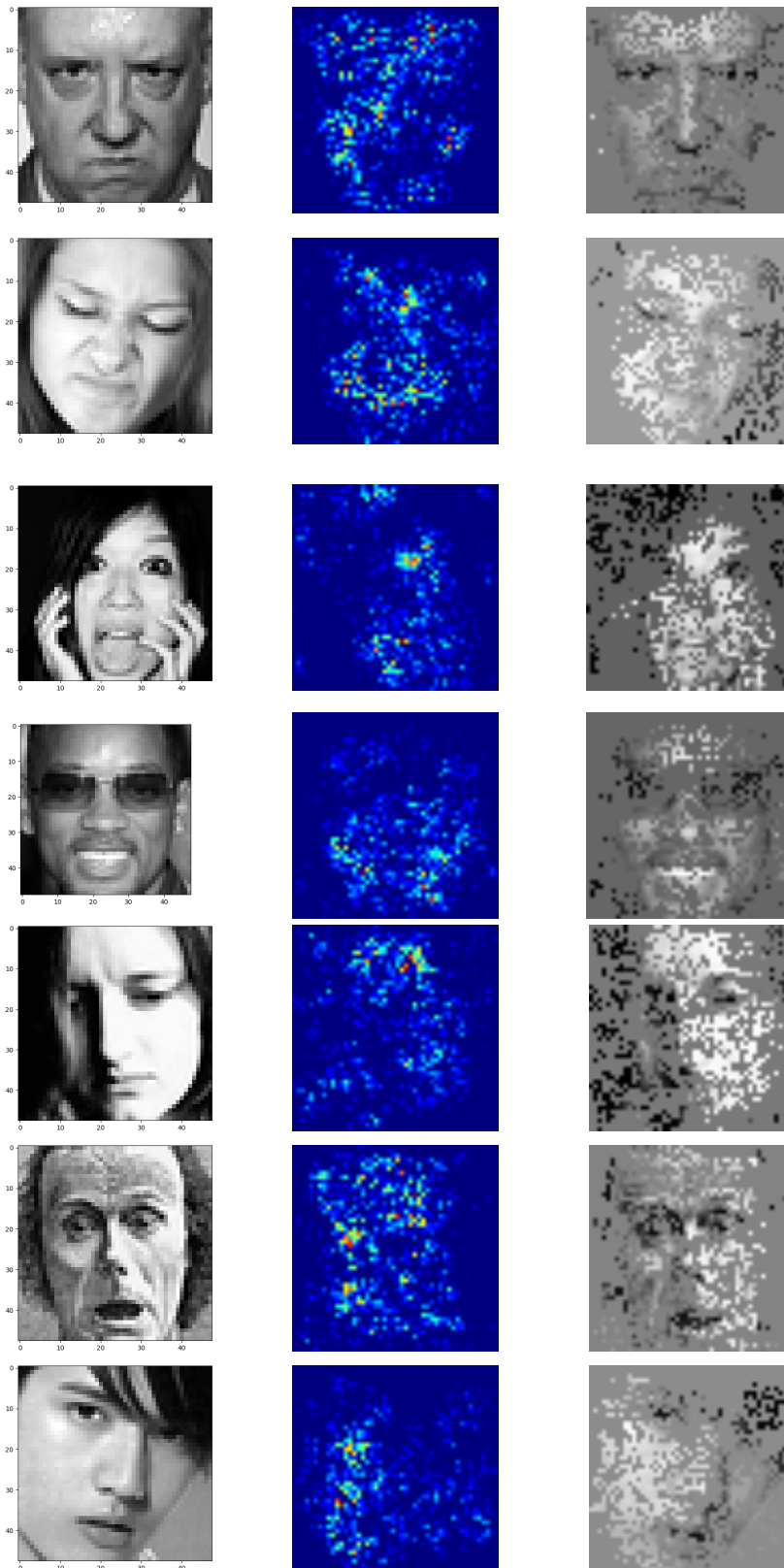
從左到右分別是原圖、saliency map、原圖 mask 掉 gradient 小的點

從上到下分別是生氣、噁心、恐懼、開心、難過、驚訝、神經質

可以看出模型主要是聚焦在眼睛、鼻子、嘴巴、臉頰，不同類別似乎關注的點不太一樣。

例如生氣時眼睛兇狠，噁心時眼睛微咪對模型來說都是很重要的資訊

或者是開心時嘴巴就很重要（對比於難過那張圖，嘴巴沒什麼特別的特徵）

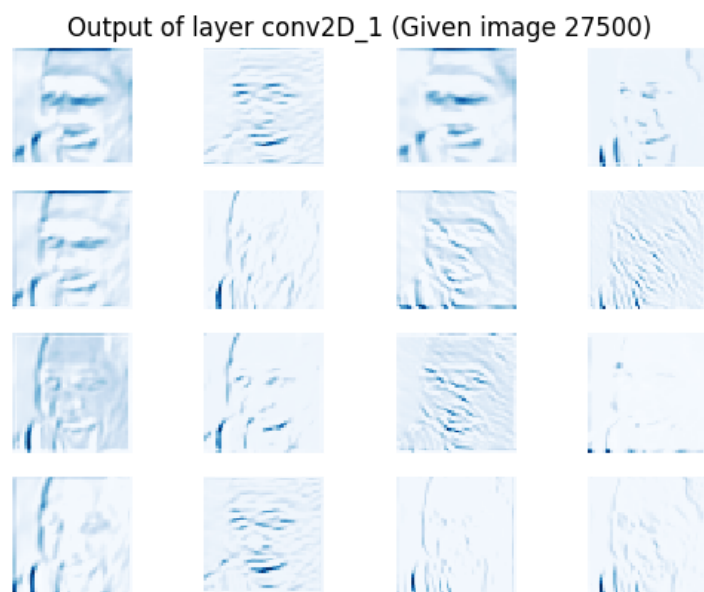
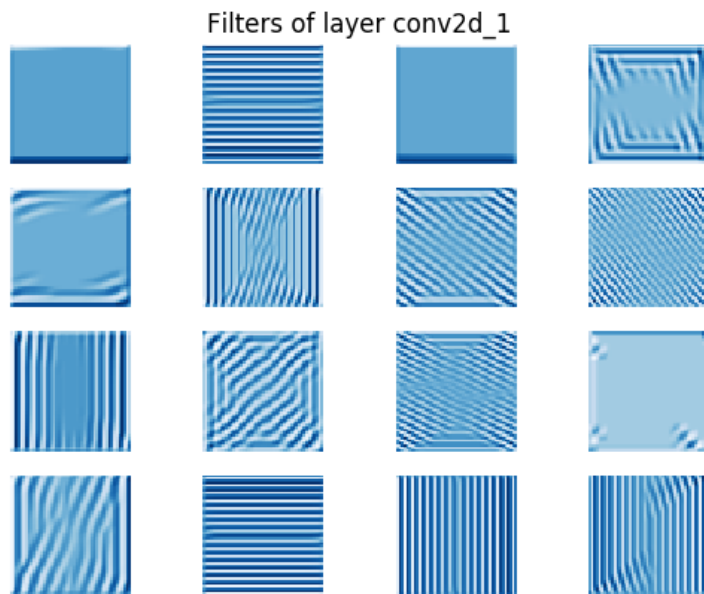


2. (3%) 承(1) 利用上課所提到的 gradient ascent 方法，觀察特定層的filter最容易被哪種圖片 activate 與觀察 filter 的 output。

(Collaborators: b06705058 劉品桢)

Reference:<https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html>)

答：



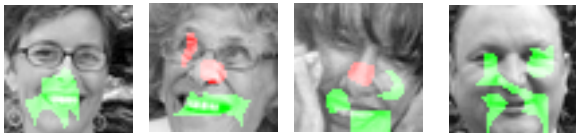
我發現蠻多 filter 是以素色或是某種規律的質地呈現，表示第一層 filters 抓的是一些規律的東西，很符合一開始設計 CNN 的想法

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)。

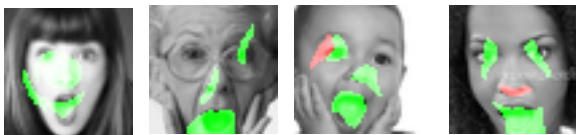
答：



由上次作業分析，label 為開心表現得特別好 (正確率92%)，驚訝的表現也很好（正確率91%），模型認為這兩個 label 的關鍵在於嘴巴。這很合理，如果我們只看嘴巴的話，我自己只能把開心的圖片分出來。嘴巴可以很明顯（相較於眼睛、鼻子、臉頰）的區分它們的類別，因此準確率會比較高。



(標籤為開心，的確嘴巴是很重要的特徵)



(標籤為驚訝，模型認為嘴巴開開的特徵很重要)

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：

我是用上課所教的 local explanation 的其中一個方法，就是選擇一個大小適中的正方形方塊將圖片遮住，並計算其預測的準確率，若準確率下降許多，代表很可能是重要的資訊。我將方塊設為 5\*5 並完全搜遍整個圖，將損失的準確率累加起來並作圖。方塊的顏色選取蠻重要，有些圖片整體偏暗，那麼應該要用白方塊遮住，相反的要用品黑方塊遮住，因此我乾脆每張圖都用兩種顏色去跑並把結果疊加（越亮代表可能越重要）。



有幾張圖是嘴巴地方是白色，其餘都是黑色，代表這圖片對模型來說嘴巴是非常重要的，例如墨鏡男 (標籤為開心)。也蠻多圖片是幾乎整張臉都有一些資訊，這也蠻合理，畢竟一個表情可以從不同部位來綜合判斷。

我個人認為這種方法與 lime 不同地方在於，lime 是事先分割完圖片再去解釋，但交界點的地方到底重不重要就比較難表現出來。而這方法是慢慢移動方塊，因此是連續性的。與 saliency map 不同是，saliency map 關注的只有一個點，而這方法是一個區塊。