

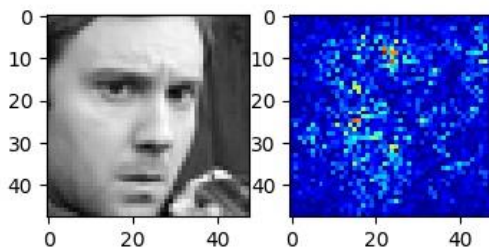
作業四 Report

學號：B06901020 系級：電機二 姓名：張恆瑞

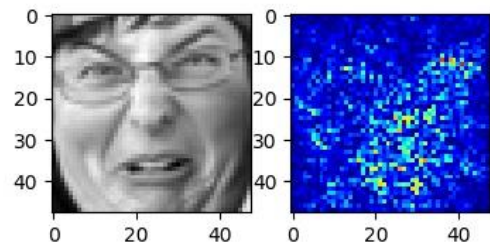
1. (2%) 從作業三可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？

答：經由觀察 saliency maps 發現，CNN 在做分類時，主要 focus 在人臉的部分，尤其是眼睛、眉毛和嘴巴。

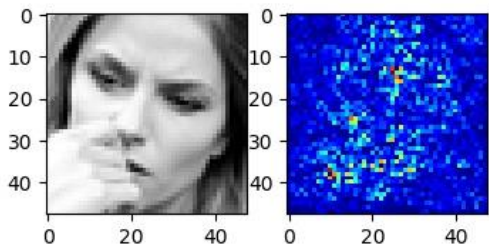
Angry



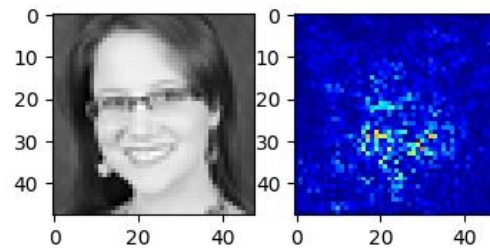
Hate



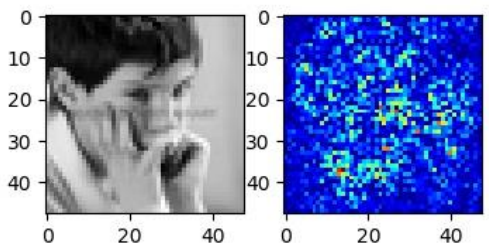
Fear



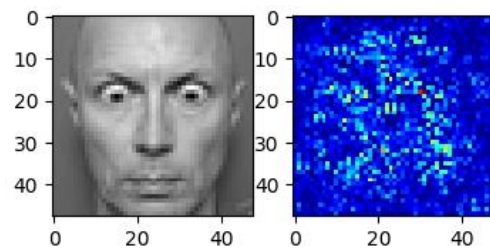
Joy



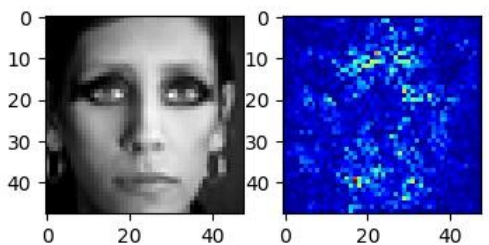
Sad



Surprise

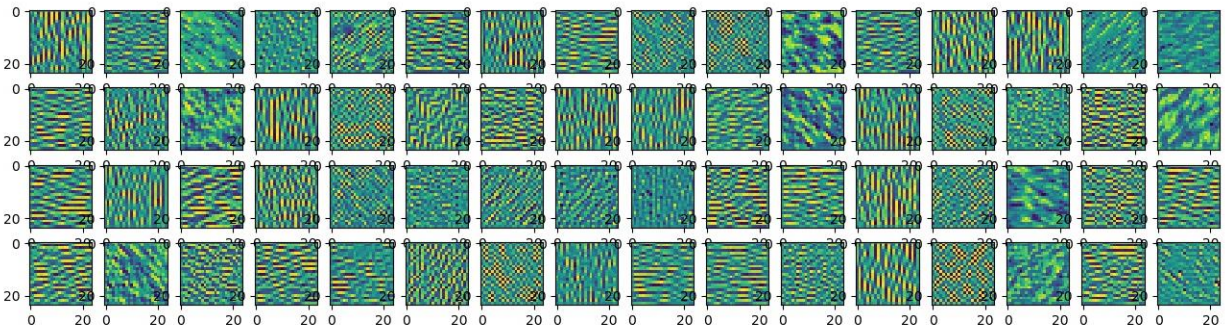


Neutral

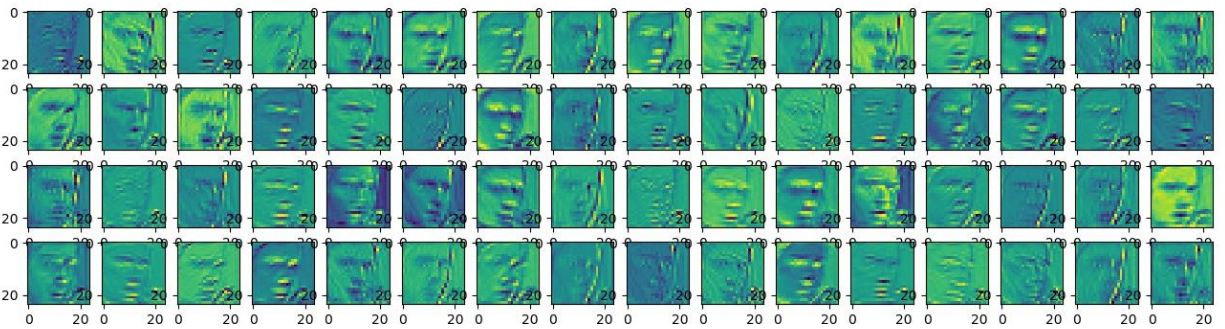


2. (3%) 承(1) 利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate 與觀察 filter 的 output。

答：第一張圖是利用 gradient ascent 觀察第一層 convolution layer 的結果。大致上看得出來每張 filter 的功能都不太一樣，像是處理橫向條紋、縱向條紋、斜向條紋等等。

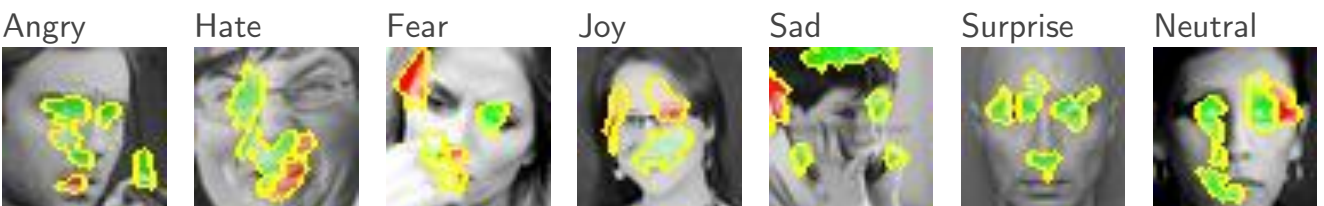


第二張圖是將一張照片塞進第一層 convolution layer，然後觀察它的輸出。可以看出每個 filter 都針對照片的不同地方增強或減弱，好讓下一層繼續處理。



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：由圖片中可以觀察到綠色區塊主要集中在眼睛鼻子嘴巴部分，這和我們人類的辨識方法是一樣的，因為人們在變換表情時變化最大的就是眼睛到嘴巴的 T 型部分。此外，angry 以及 hate 的綠色及紅色區塊分佈很相似，同時對照 confusion matrix 可以發現 hate 的照片有大約 24% 被分類為 angry，因此利用 saliency map 可以驗證原本 model 在辨識這兩類的時候可能誤判的原因。



(註：我發現轉成 pdf 檔之後圖片不知為何爛掉了，不過顏色區塊還是看得出來是跟原本一樣的)

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：抱歉助教最近我有點忙，沒能研究這一題 QQ