

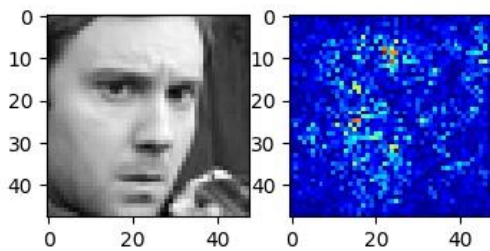
## 作業四 Report

學號: B06901020 系級: 電機二 姓名: 張恆瑞

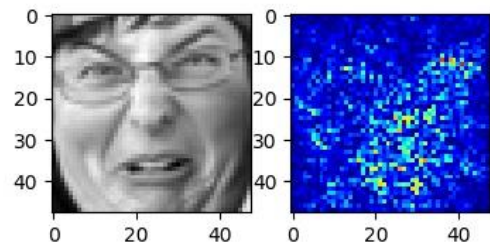
1. (2%) 從作業三可以發現, 使用 CNN 的確有些好處, 試繪出其 saliency maps, 觀察模型在做 classification 時, 是 focus 在圖片的哪些部份?

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。

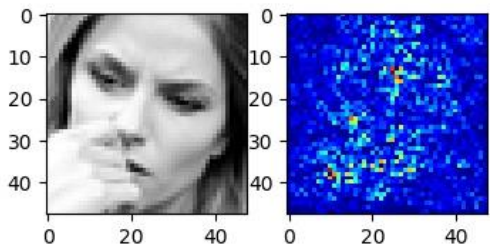
Angry



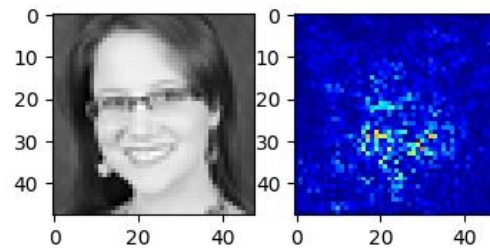
Hate



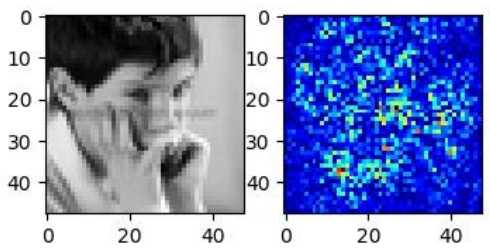
Fear



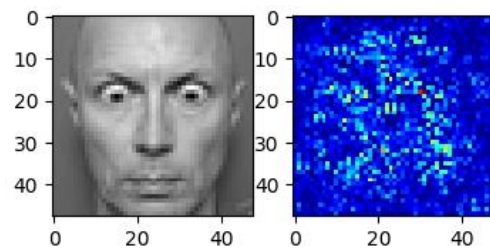
Joy



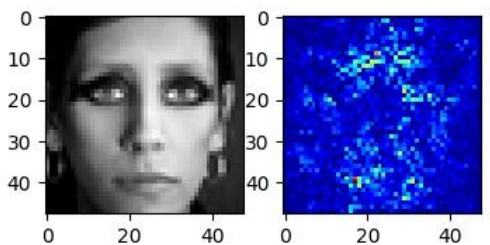
Sad



Surprise

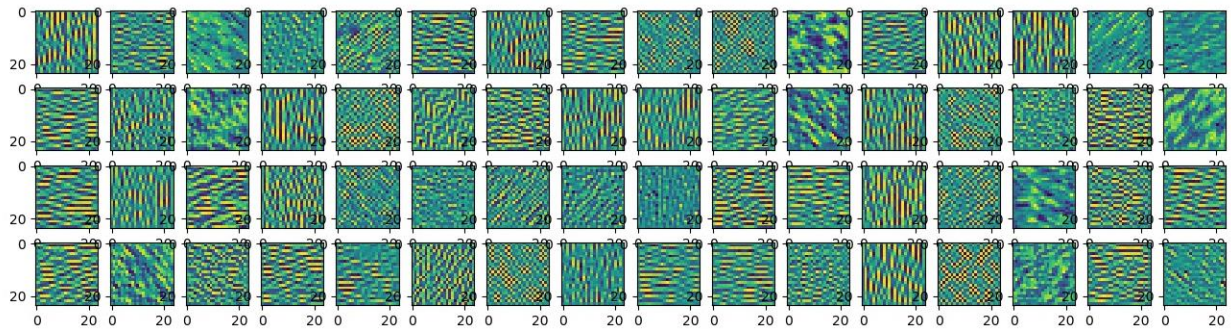


Neutral

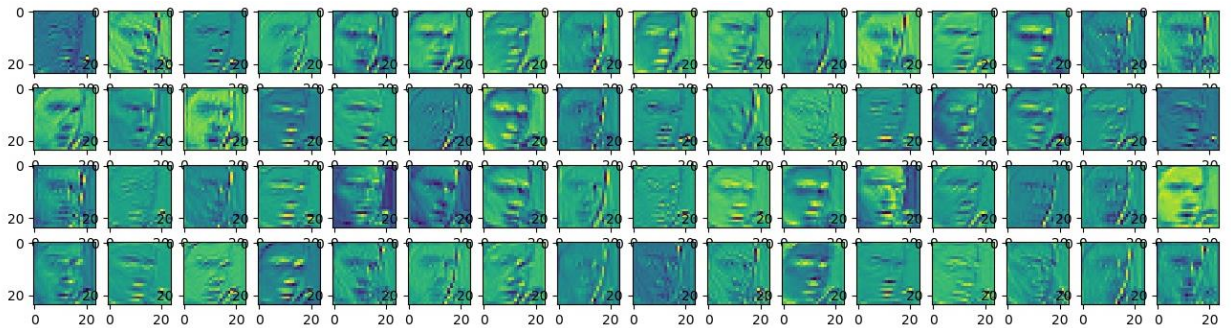


2. (3%) 承(1) 利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate 與觀察 filter 的 output。

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、斜向條紋等不同花紋，可能是去分析並強化人臉的一些線條，像是眼睛、嘴巴等。



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。接著，*angry* 以及 *hate* 的綠色及紅色區塊分布很相似，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。

此外，從 confusion matrix 中發現，辨別 *surprise* 以及 *joy* 時準確率可以到 80% 以上，從 saliency map 看得出 *joy* 辨識集中在正在笑的嘴巴，而 *surprise* 則集中在瞪大的眼睛上，可見機器領悟到了人臉在表現喜悅與驚訝的表情應為如何。



(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)

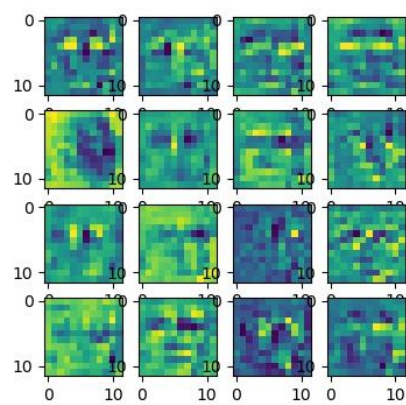
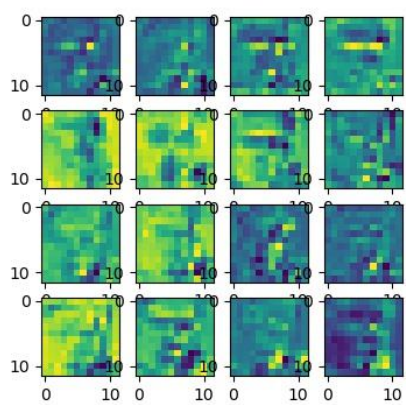


4. (2%) [自由發揮] 請同學自行搜尋或參考上課曾提及的內容，實作任一種方式來觀察 CNN 模型的訓練，並說明你的實作方法及呈現 visualization 的結果。

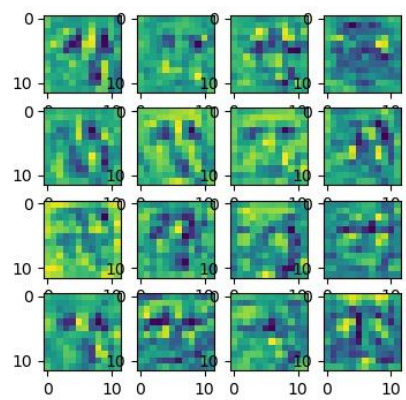
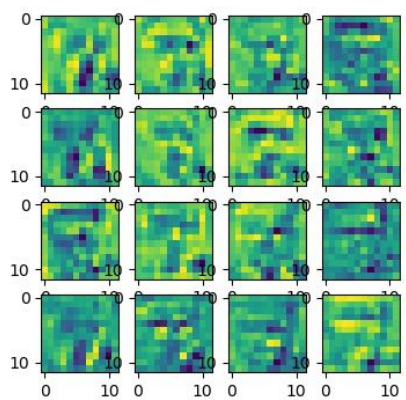
答：我想到一種方法是將照片丟入 CNN，然後觀察每一層輸出結果的變化及不同類別間的差異。到比較後面的 layer 圖片的解析度愈來愈低、愈來愈糊，可以看出每個 filter 的亮點都會集中在某處，因此可以推論，每層的 filter 會漸漸將圖像的一部分凸顯、強調，最後以一個較集中的亮點來傳達給後面的 layers。

Layer	Angry	Neutral
Conv2D 1		
Conv2D 2		
Maxpooling 1		

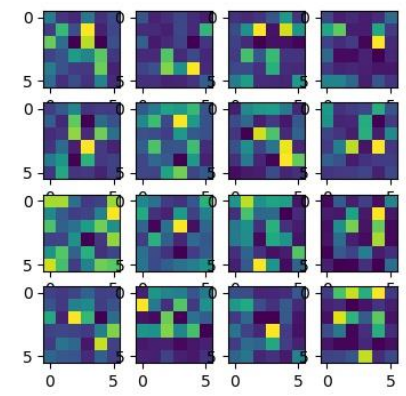
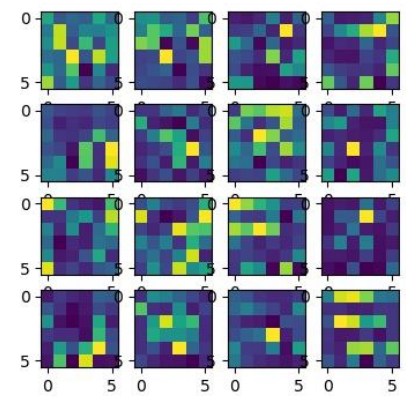
Conv2D 3



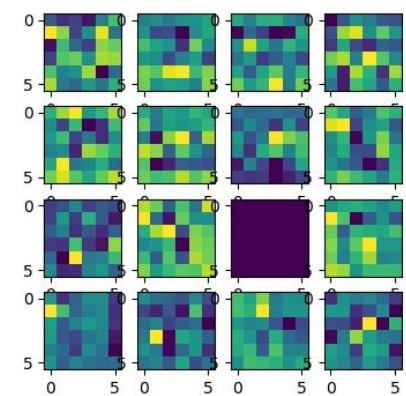
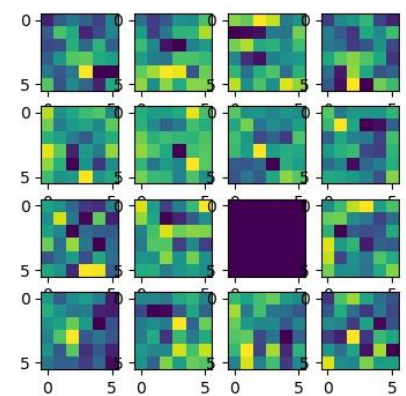
Conv2D 4



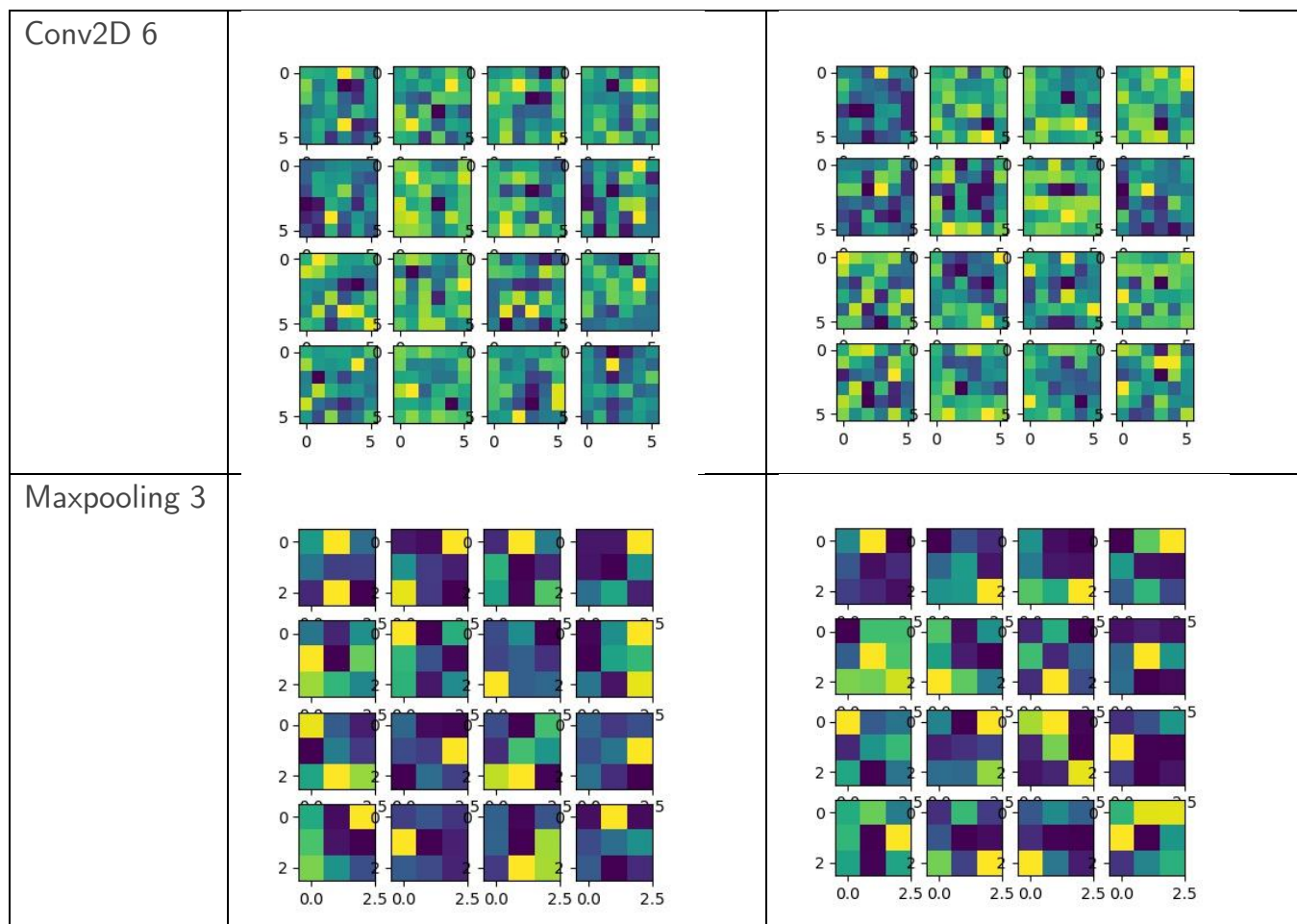
Maxpooling 2



Conv2D 5

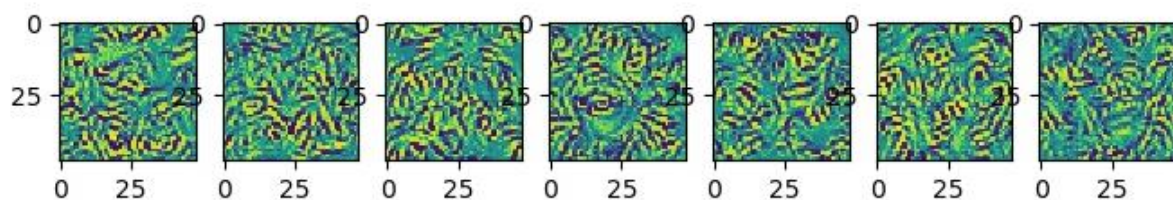






另外，可以看出某些 filter 輸出是 全部深色的，不因輸入相片類別不同而改變，有點小任性，應該可以被解讀成是分組報告時的躺分仔，但說不定他的存在還是有點意義，所以還是不要全然否定人家好了。

此外我想到另一個方法，也就是利用 gradient ascent 的方法，試試看哪樣的圖像可以 將某個類別的分數特別高：(由左至右為類別 0、1、2、3、4、5、6)



然後就得到了一堆 印象派義大利麵，偏醜。那些義大利麵應該是在表現人的眼睛、眉毛、嘴巴等的輪廓，在慢慢感受圖案裡傳達的訊息後，一些人臉從照片中浮現...

看來我的 model 運作應該是正常的，這樣的視覺化 model 看起來還不賴。