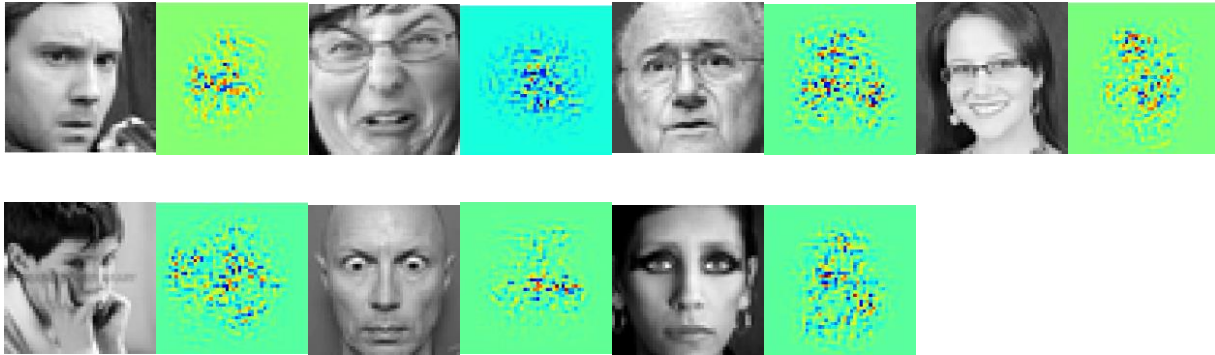


學號：B05901025 系級：電機三 姓名：王鈺能

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？

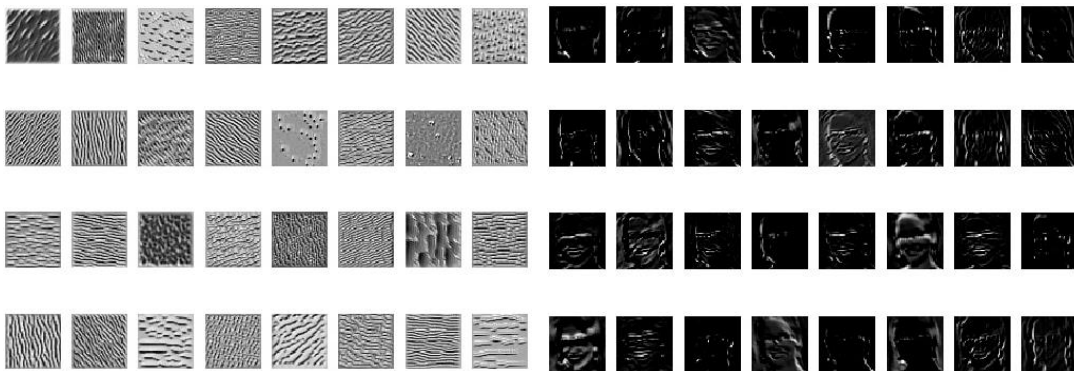
(Collaborators:) 答：



左至右、上至下依序為生氣、厭惡、恐懼、高興、難過、驚訝、中立。從 saliency map 分析，模型在分類時主要 focus 在眉毛上部沿著鼻翼兩側延伸至嘴角部位，成一個「大」字形(去掉上方突出)，而辨識方式似乎是觀察顏色較深的細條，所以皺紋、人中、鼻翼兩側都有比較強的反應，也因此左下圖的手指早成的線條似乎有誤導到模型的現象。

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：



觀察第二層 convolution layer(前 32 個 filter)，gradient ascent 繪成的圖片顯示各 filter 主要會偵測不同角度、寬度、長度、深度的線條，像最左下角的 filter，會偵測與旁邊對比很大的線條，所以從眉毛、鼻翼兩側或交界處理論上就會讓其 activated，對應到該 filter 的 output(右圖最左下)，很明顯的眉毛、臉的邊界 output 特別大，與 gradient ascent 結果相符。

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：



生氣：眉毛、鼻翼部分，該處若皺起來(線條明顯)代表生氣

厭惡：鼻、嘴周圍的紋路，向外擴可能是厭惡時產生的皺紋

恐懼：鼻、嘴周圍緊縮

高興：眉、鼻、嘴周圍紋路，向上或非緊縮的線條可能表示高興

難過：似乎沒有特別辨識哪個部位，對線條較多處有反應

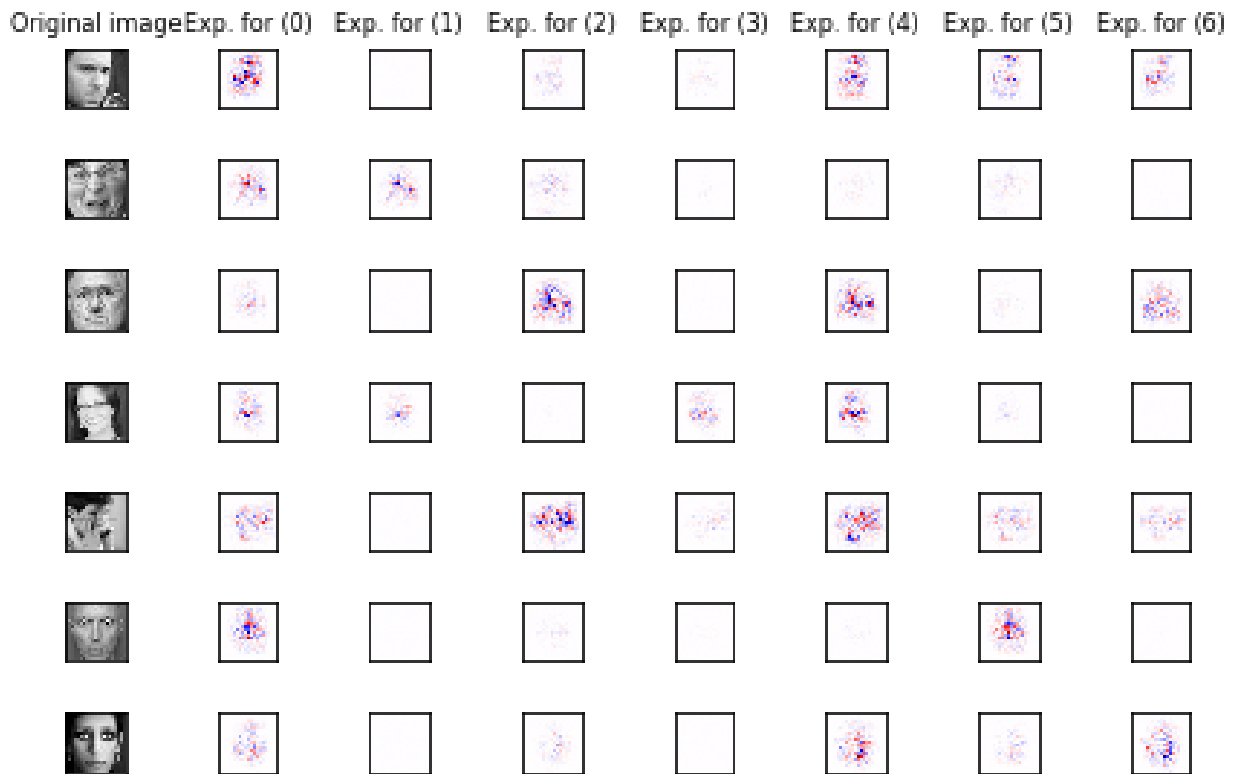
驚訝：鼻翼擴張(鼻翼周圍紋路擴張)、嘴型擴張(嘴周圍紋路擴張)

中立：眉、鼻、嘴周圍紋路，平、沒有彎曲的線條可能表示中立

我的模型從 **confusion matrix** 中可以看出高興是表現最好的，由 **lime** 的解釋就是其他情緒大部分都是偵測某部分集中的線條(因為偏向負面情緒)，但快樂反而是要舒緩的線條才会有反應，和其他情緒完全相反，因此只要線條是不集中然後紋路有向上彎曲，偵測其他情緒的 **filter** 很不活躍、偵測快樂的很活躍，**output** 的機率差被拉大，就很容易被辨識出來。

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 integrated gradient visualize 我的模型，參考 <https://arxiv.org/abs/1611.02639>，此方法可以解決 gradient saturation 造成的問題，用 <https://github.com/hiranumn/IntegratedGradients> 的套件，得到上圖的結果，結果與 saliency map 類似，在眉毛、鼻翼、嘴周圍是偵測的部位，另外也與 confusion matrix 一致，可以看出其實每個情緒在難過部分都有被偵測到，因此很容易與其混淆。