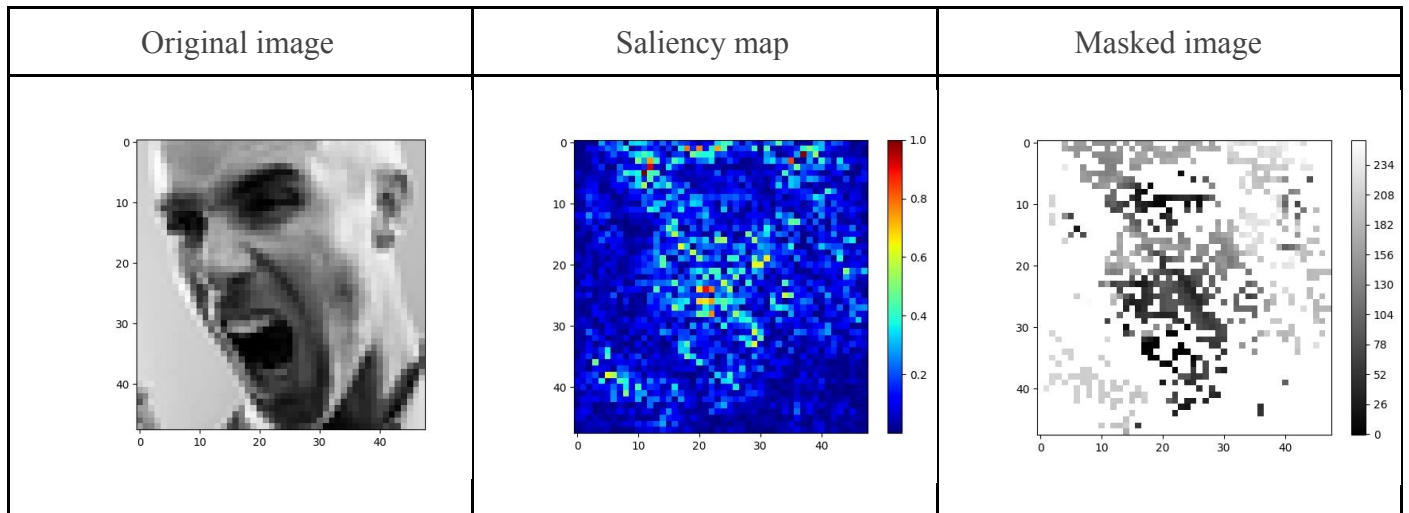


學號：R07922134 系級：資工碩一 姓名：陳紘豪

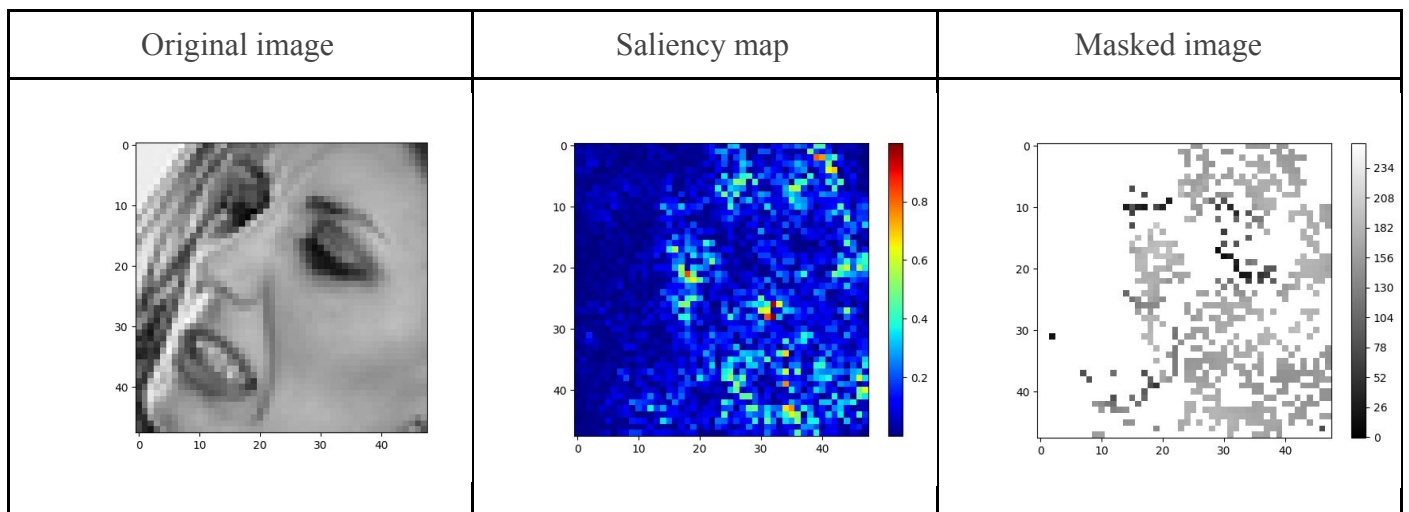
1. (2%) 從作業三可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？  
(Collaborators: )

答：

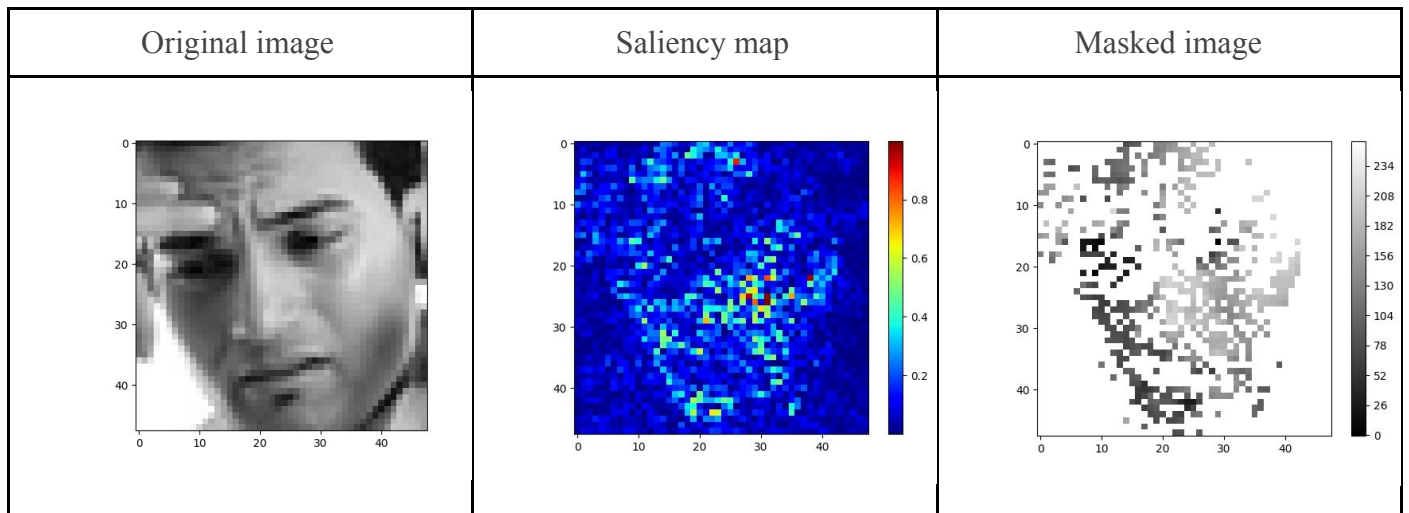
**Class 0: angry**



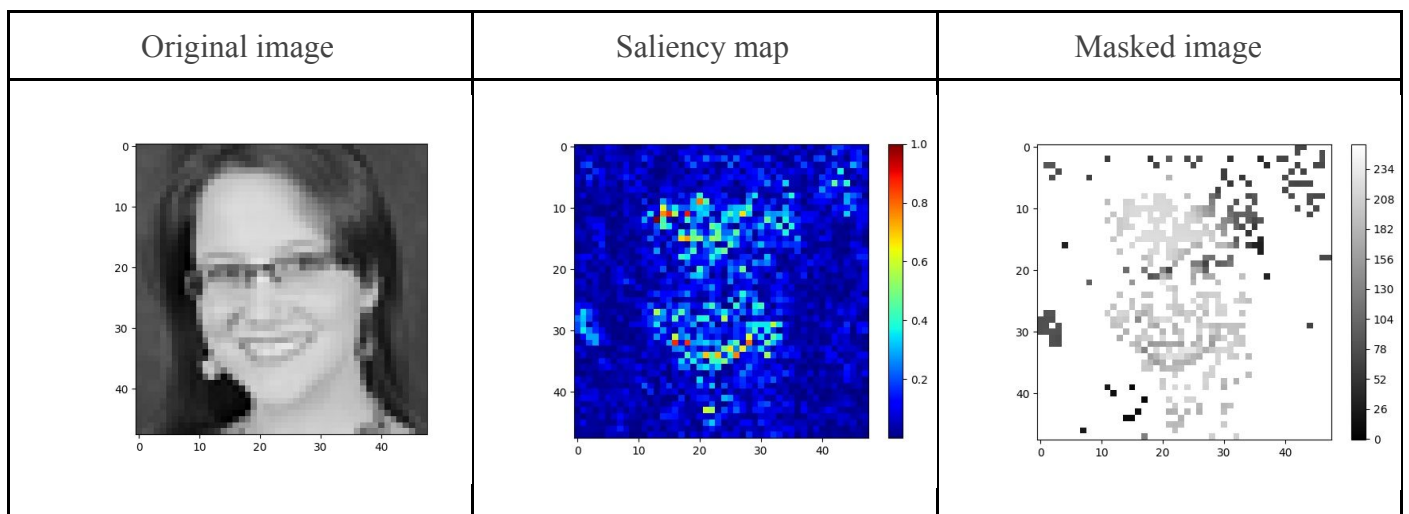
**Class 1: disgust**



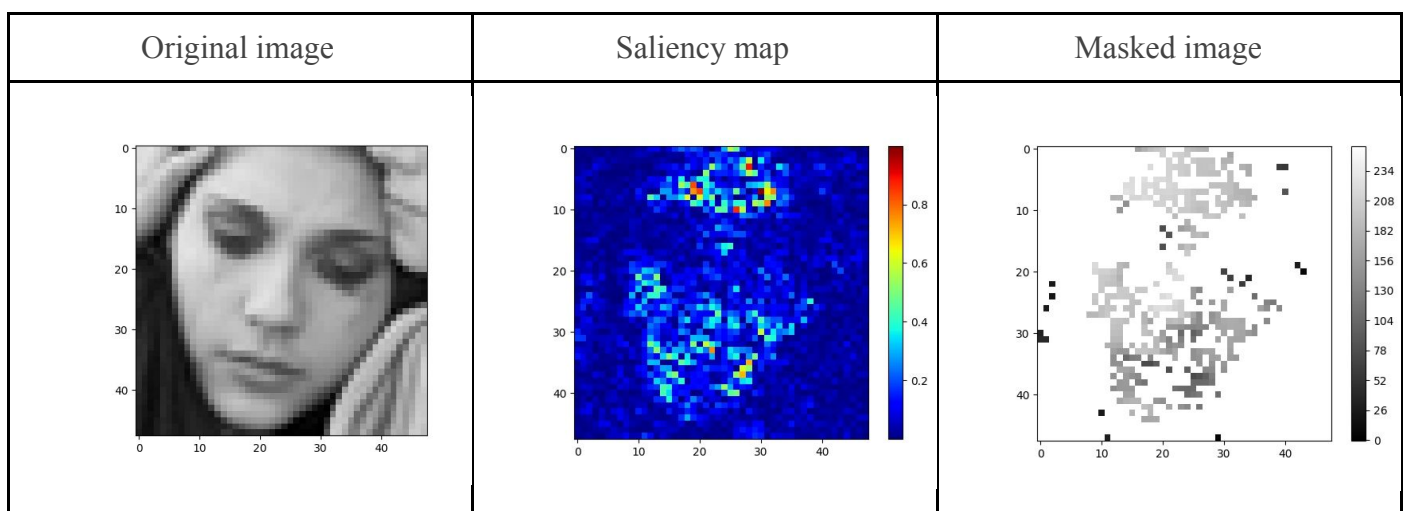
### Class 2: fear



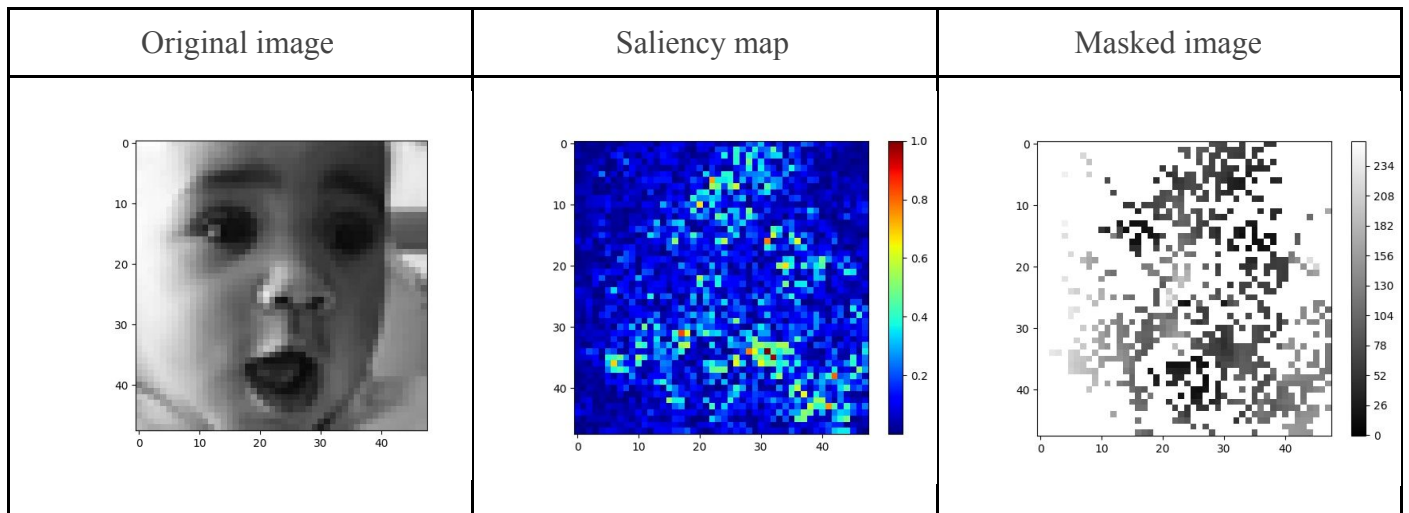
### Class 3: happy



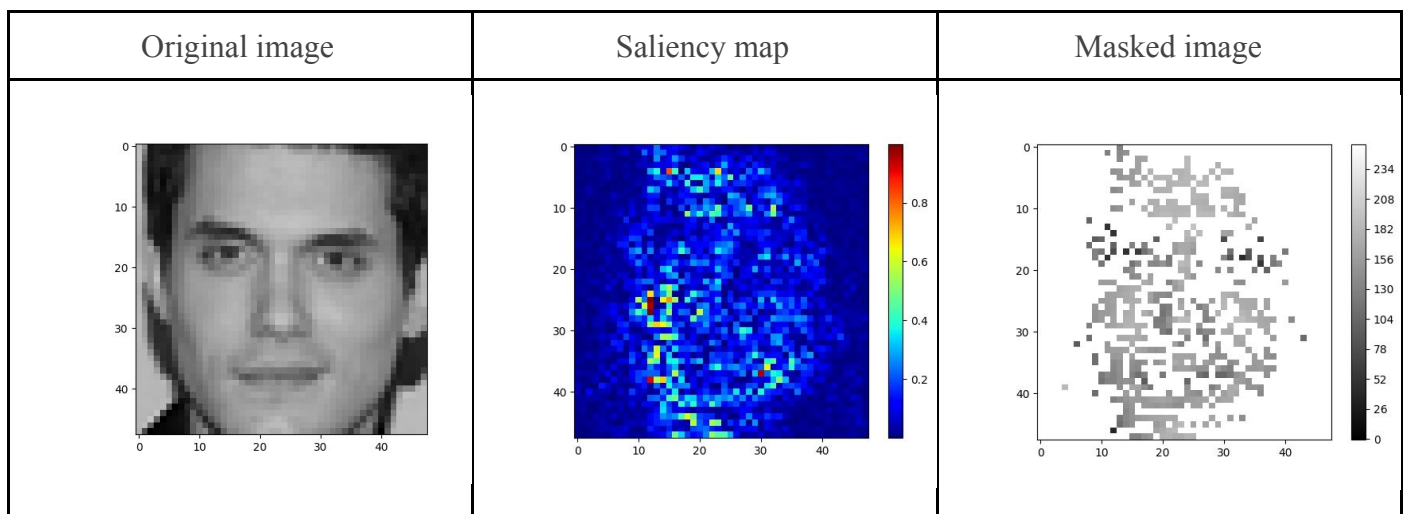
### Class 4: sad



### Class 5: surprise



### Class 6: neutral



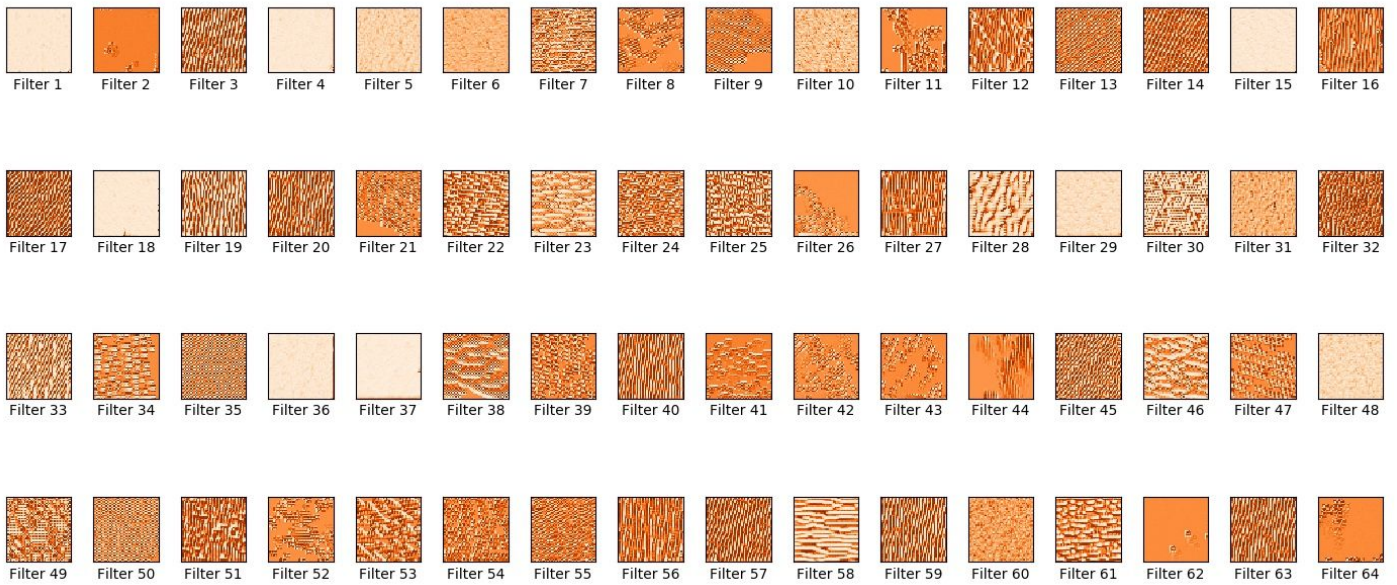
由觀察發現，CNN比較著重在五官藉此來辨識是什麼表情，尤其是嘴巴這個部位，像是「happy」中整個嘴巴包括牙齒都在 saliency map 有相對較大的 gradient 值，可以發現模型來辨認是否 happy 很大一部分就是去觀察嘴巴以及有無露出牙齒，而在「neutral」中也是嘴巴附近有較大的 gradient 值，所以比較沒有起伏的嘴唇就可能被辨識成 neutral。

- (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：所觀察的 filter 是來自於 model 中第二層 Convolution layer



Fig.2-1  
Filters of 2nd Conv2d in block1 (# of ascent epoch: 100)



由觀察可知，此層中的 64 個 filters 比較著重在各種不同的紋路上，有些方向是上下、左右甚至是斜對角的，這些 filters 主要是在找出輸入圖片中各種不同方向的邊（edge）。

Fig. 2-2  
Outputs of 2nd Conv2d in block1 (Given image 6987)

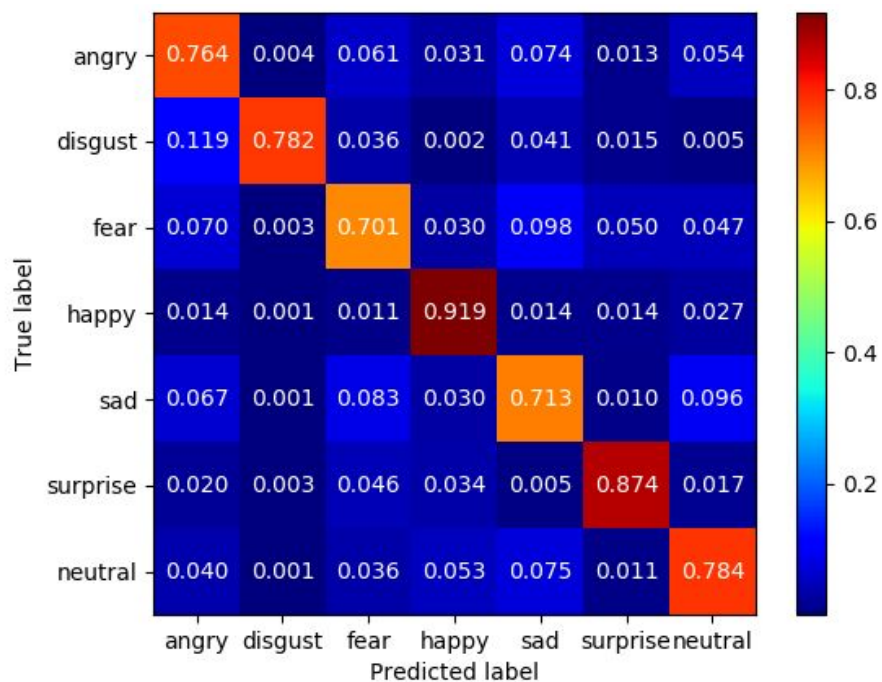
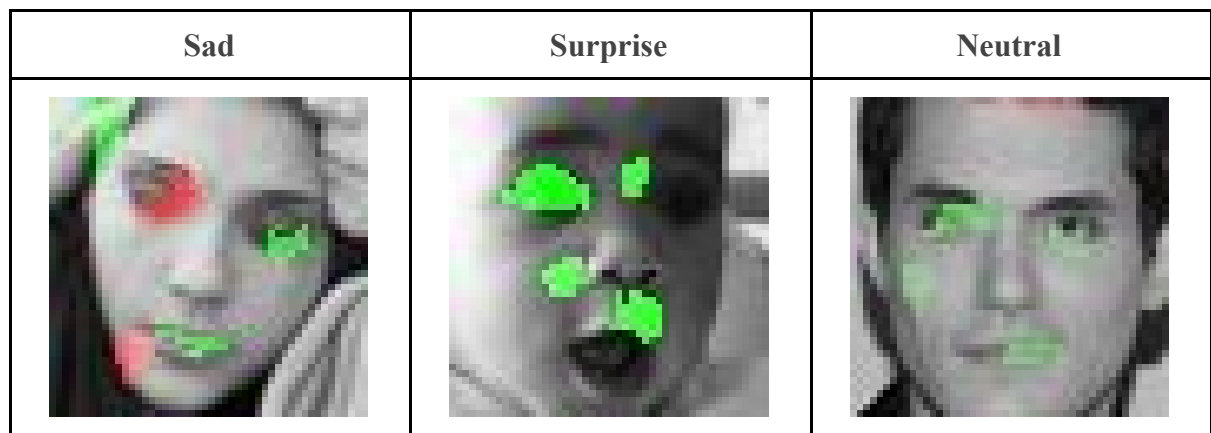
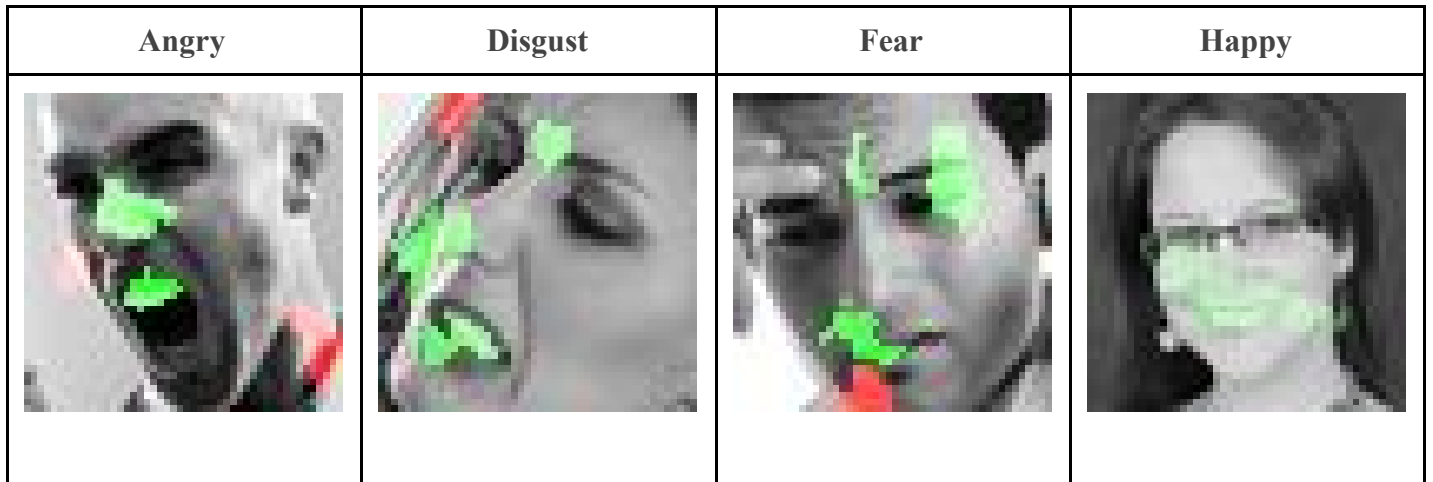


把 image[6987] 丟入 model 後並把第二層 Convolution layer 的 64 個 filters output 出來如上，可以觀察到某個 filter 所 output 的圖片和可以最 activate 該 filter 的 image，其紋路基本上有很大的相似度（例如 Fig 2-1 和 Fig 2-2

的 filter 40) 。

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：（pdf放大看的話有些顏色可能會比較淡，像是 neutral 額頭上的紅色）



辨識結果較好的 label 為「happy」、「surprise」，因為他們皆沒有出現紅色的 negative region，果不其然在 confusion matrix 上這兩個 label 也是辨識結果相對較高的 label，而可以觀察到「fear」、「sad」這兩個 label 的圖片在嘴巴附近都有 negative region，也就是該區域對辨認該 label 有負面的效果，很有可能是這兩個 label 的嘴附近都會較為下垂，所以可能辨識錯誤，在 confusion matrix 上「sad」和「fear」之間也具有較大的 misclassification rate。

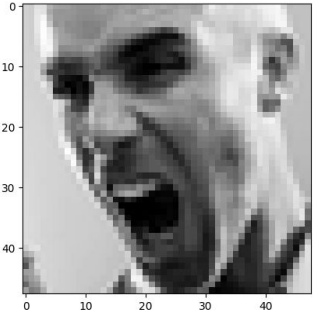
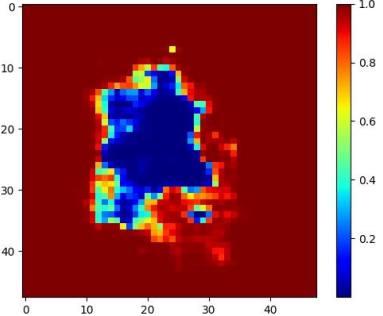
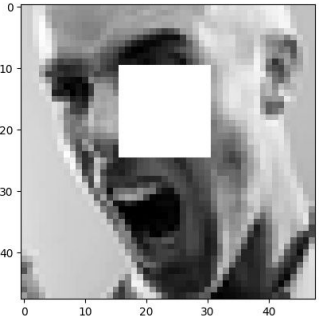
在七個圖片上，positive region 大多數都在眼睛和嘴巴的區域，可以得知 CNN model 幾乎都是靠這兩個五官來判斷屬於哪個類別，尤其是「happy」的類別只有在嘴巴出現，因此 model 可能只要看到露出牙齒的嘴巴就直接判定為「happy」。

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：

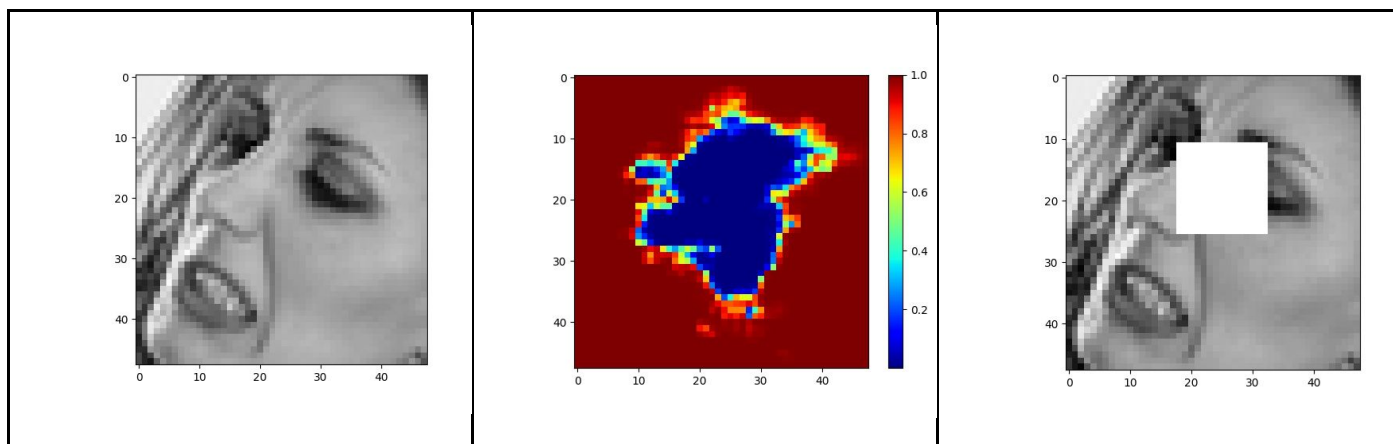
我想要找出圖片中哪個區域拿掉會對於辨識結果影響最大，所以我在原圖上罩上一個 15x15 的 white box，去看當這個 box 罩著每個區域時其對應的預測機率的變化，heatmap 上 (x,y) 位置上的機率值代表 white box 的中心罩在原圖 (x,y) 上後丟入 model 所 output 的預測機率。

#### Class 0: angry

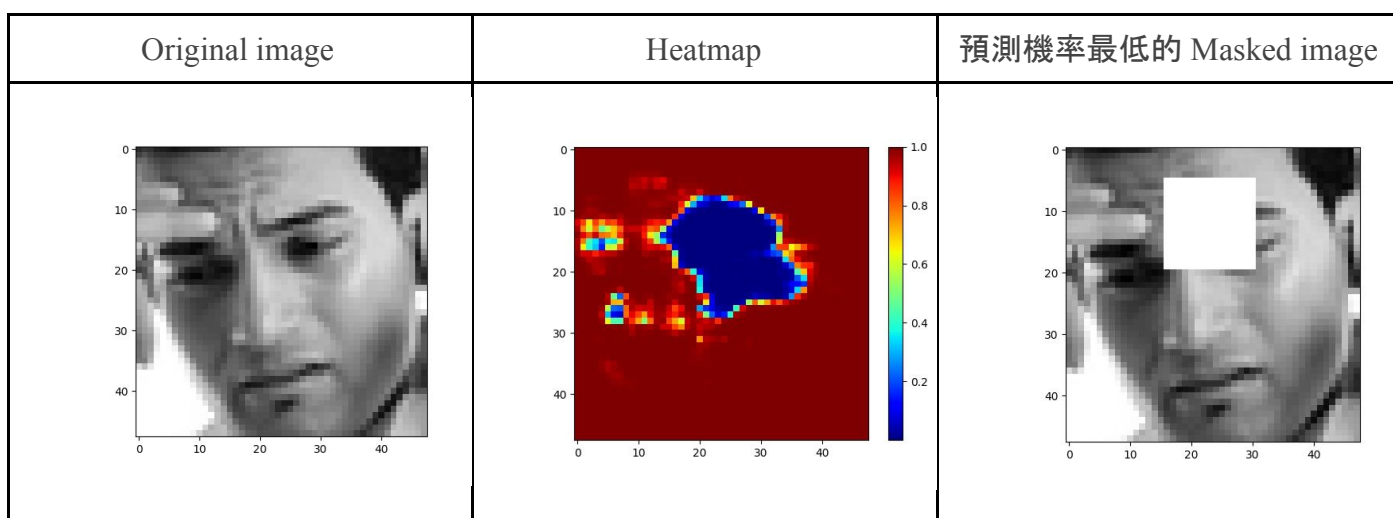
Original image	Heatmap	預測機率最低的 Masked image
		

#### Class 1: disgust

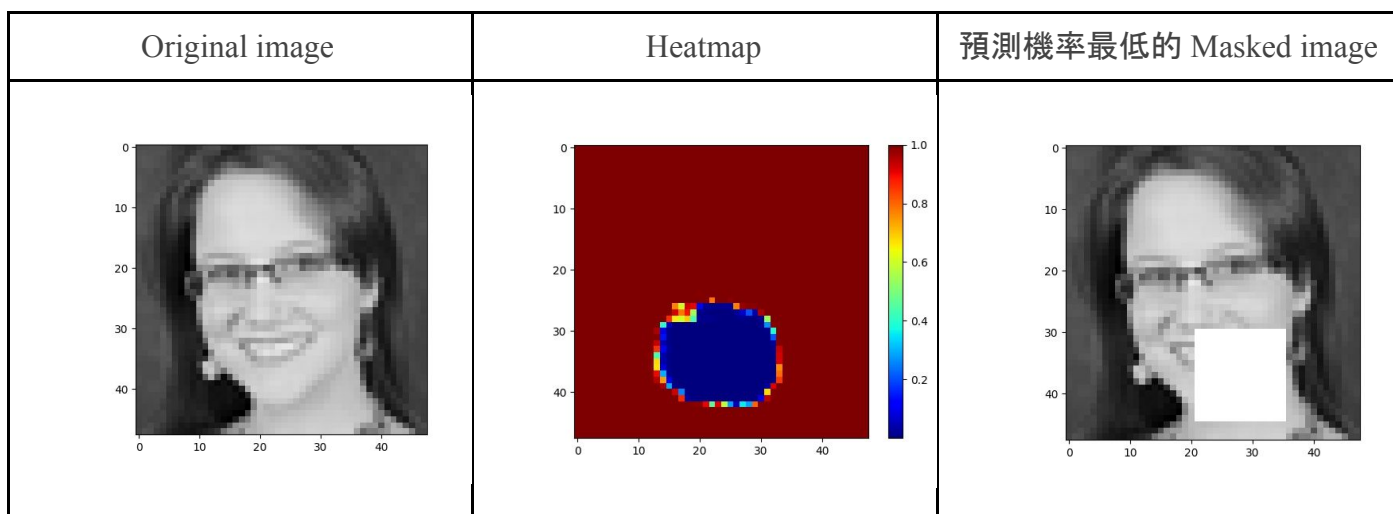
Original image	Heatmap	預測機率最低的 Masked image
----------------	---------	----------------------



**Class 2: fear**



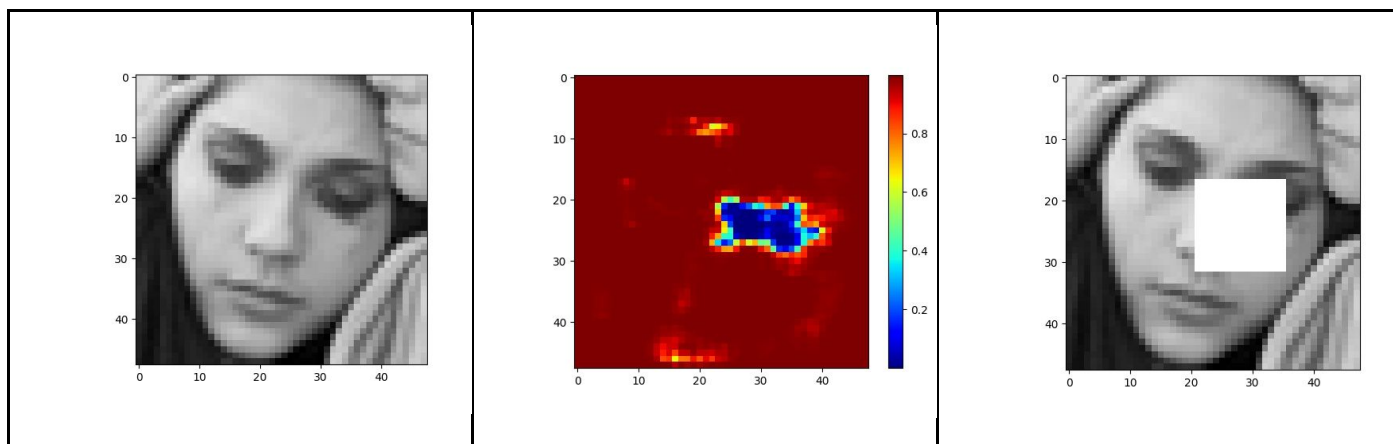
**Class 3: happy**



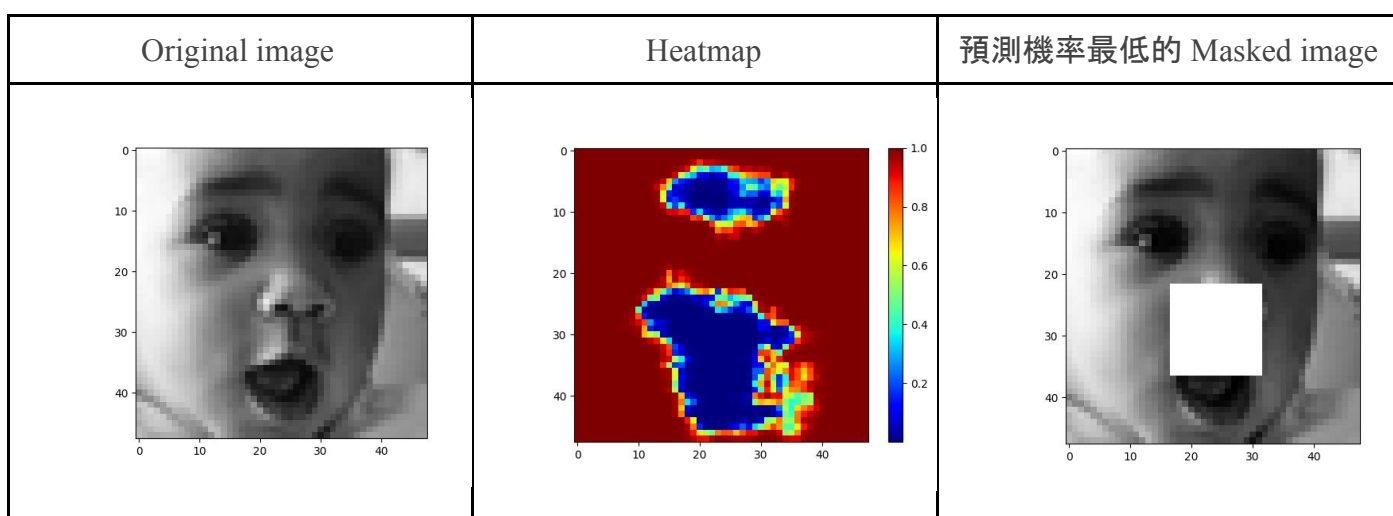
**Class 4: sad**



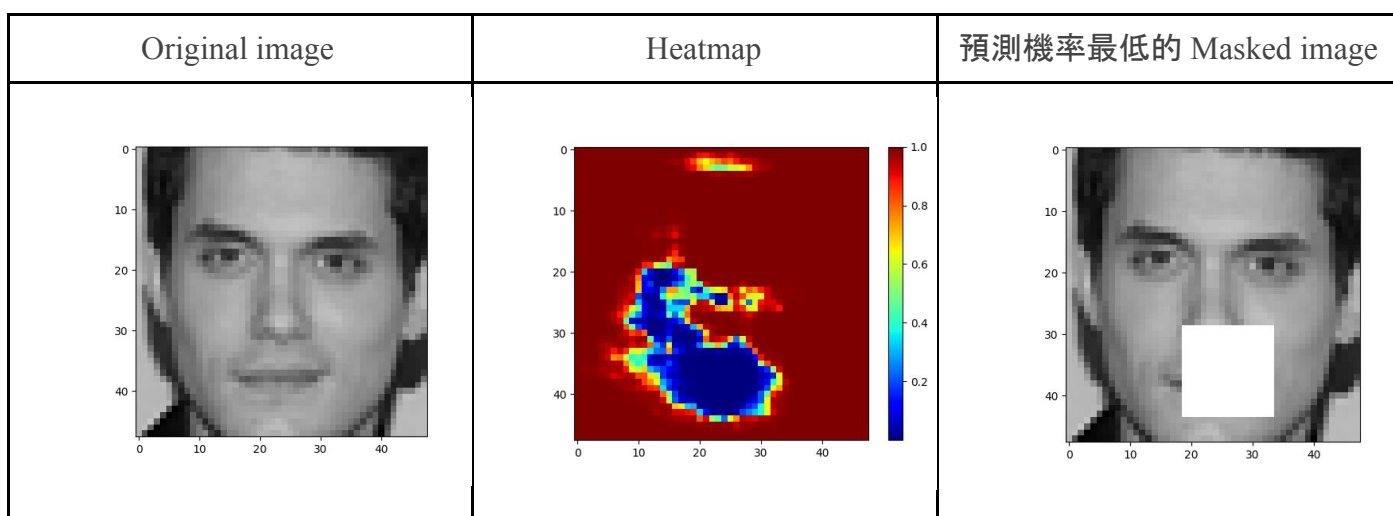




**Class 5: surprise**



**Class 6: neutral**



從 Heatmap 上發現，CNN model 判斷情緒的依據很大是從眼睛、嘴巴這兩個部位，如果這些部位被移除了，則很大的機率會使得 model 判斷錯誤，



而且根據各個 class 的 Heatmap 所畫出的那些會讓 model 輸出的預測機率最低的圖片，也都是罩在嘴巴、眼睛這兩個部位為主。