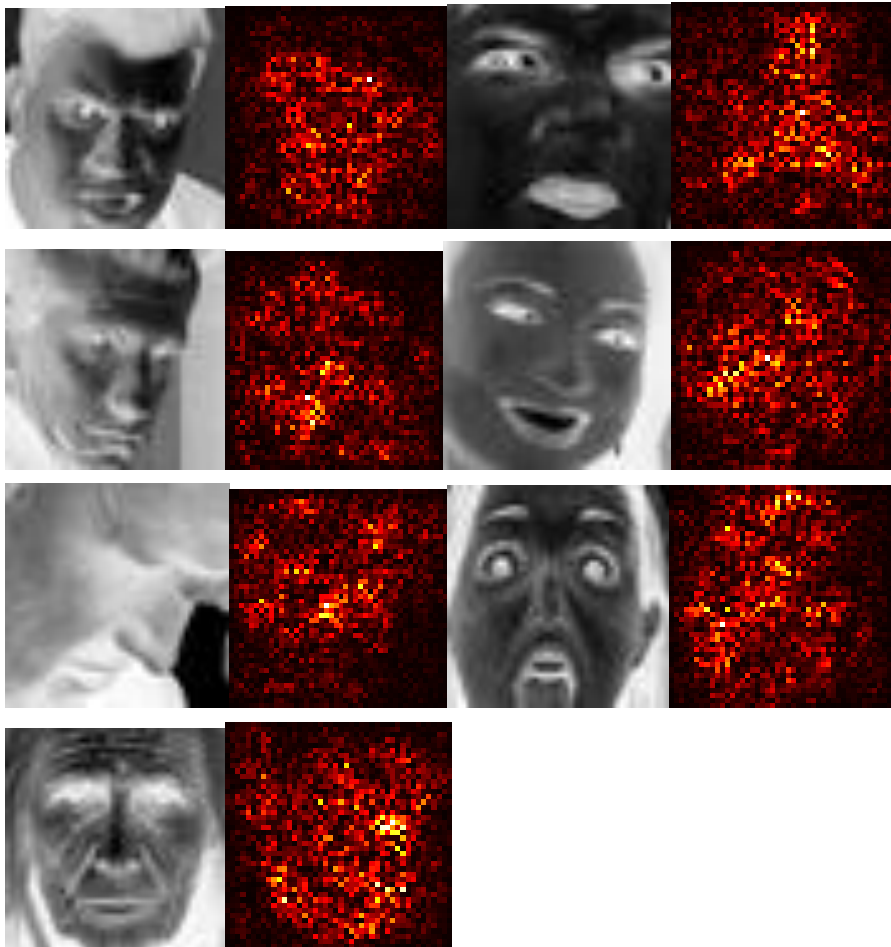


1. (2%) 從作業三可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？
(Collaborators: none)

答：

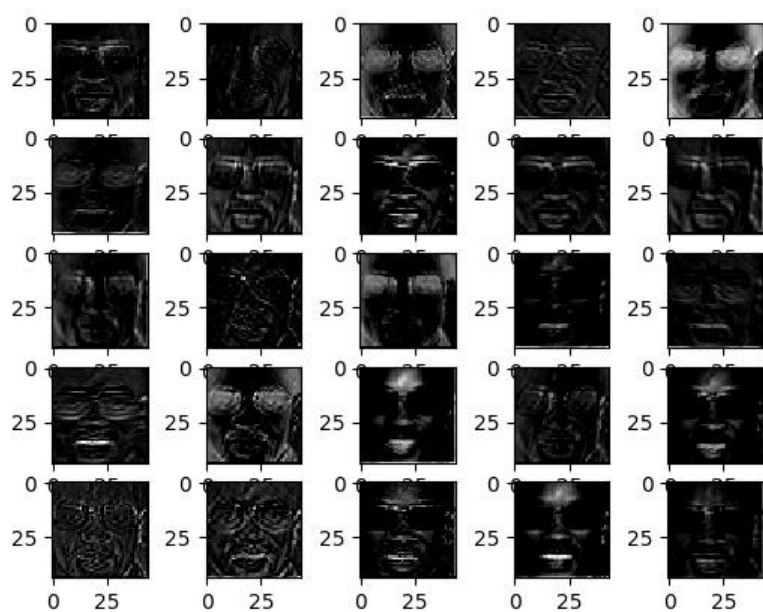
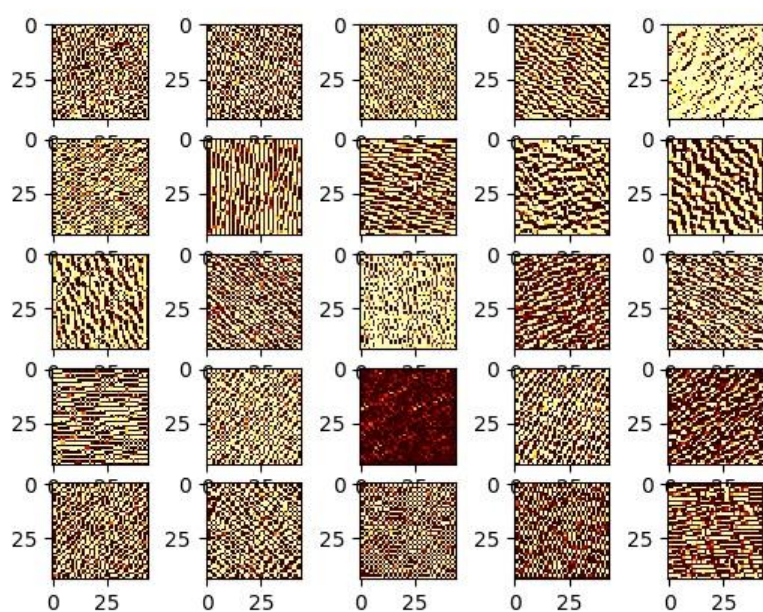


從 saliency map

中，可以發現在臉部的範圍內影響loss比較大的（較亮的區域），表示模型確實是在依據臉部特徵進行分類任務。在圖中的幾個例子又可以發現，模型重視的區域會在鼻子、嘴、眼睛等五官的附近，顯示模型學習的效果不錯。

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: none)

答：



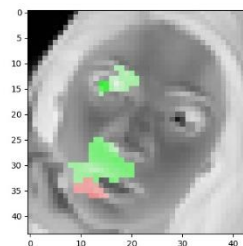
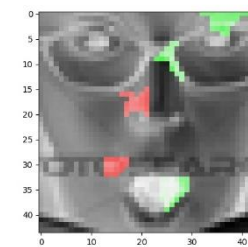
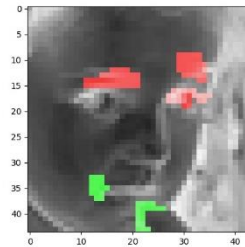
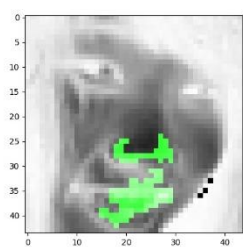
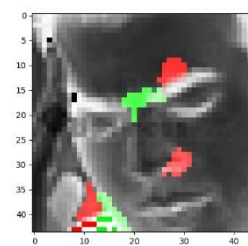
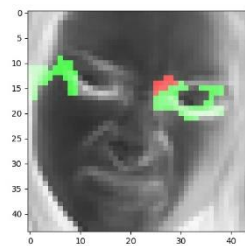
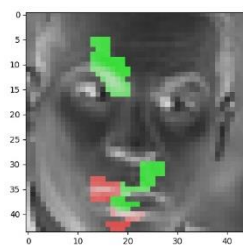
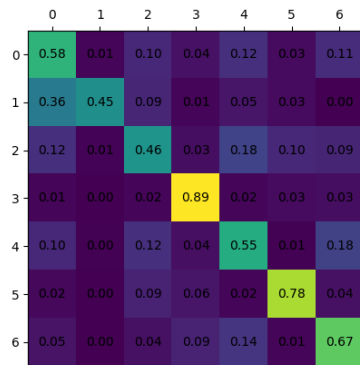
我所觀察的是第一層convolutional layer,

在觀察哪種圖片會使各個filter最有反應的過程中，發現在這層filter對於簡單的幾何線條最有反應，像是橫向、縱向、斜向的線條或是斑點，而機選擇一圖片通過這些filters的結果如圖所示，圖片中不同方向的特徵分別被分離出來。

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：



由上圖中利用 **lime** 套件所得出來的結果，可以看出模型對於圖片中五官的位置（眼睛、鼻子、嘴）最重視，也是影響判斷結果最多的地方。我認為若是圖中人物眼睛閉上，則模型可能傾向於預測為難過或是厭惡，而嘴巴張開則預測為恐懼或是開心。而根據 **confusion matrix** 的結果，模型對於"開心"這個類別表現最好，個人認為除了這個類別的表情較為容易辨別，像是打開嘴巴笑，眼角上揚等，也可能是因為資料本身分配不均所造成（這個類別在訓練資料中數量最多）。

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：

我對於模型所預測的7個類別，分別去找出模型認為最像這個類別的圖案。我都從一張充滿隨機信號的圖片開始，輸入模型，使用**cross entropy loss**計算與該類別的差距，求出在圖片上每個位置的**gradient**，然後加上去，如此週而復始（20個**iteration**），使圖片越來越會被分類到該類別，以下就是我實驗的結果。從圖片中可以看到有許多複雜的**feature**，卻很難人為解釋到底什麼樣的圖片會讓模型認為最像"開心"或是"生氣"，這大概也是深度學習被詬病卻也迷人的地方。



label_0



label_1



label_2



label_3



label_4



label_5



label_6