

Machine Learning - Homework 4 Report

學號：B06902049 系級：資工二 姓名：林首志

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: 無)

答：

表情	原圖（訓練資料中的編號，從1開始）	Saliency Map
Angry	 (2315)	
Disgust	 (1543)	
Fear	 (27936)	
Happy	 (28685)	
Sad	 (18489)	
Surprise	 (8838)	
Neutral	 (22404)	

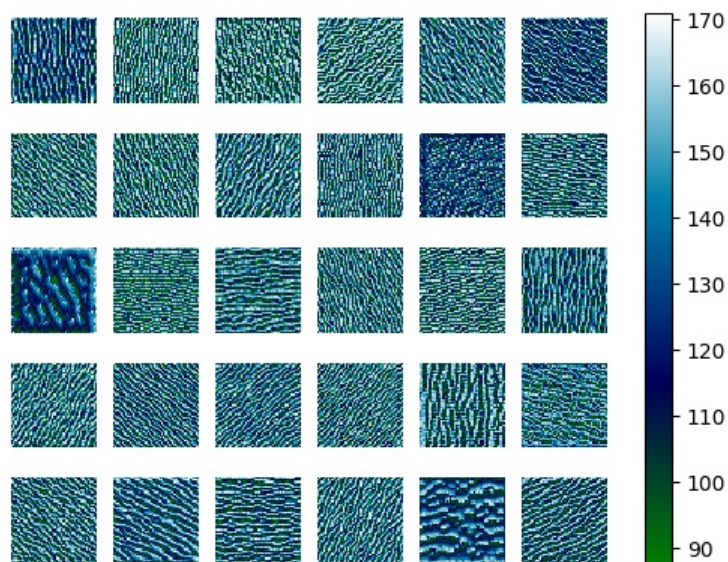
觀察計算出來的 saliency maps，可以猜測我訓練出來的 CNN 判斷表情的其中一個重要依據是臉頰的變化。眼睛、嘴巴的影響反而沒有直覺上認為的那麼重要（或者是微小的改變不會對判定結果造成太大的影響）。

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: B06902029)

答：

下圖是第三個Convolution層（尚未經過任何Pooling層）所有filters的可視化結果。

（右方數值經過縮放、平移處理，並非原始數值）。

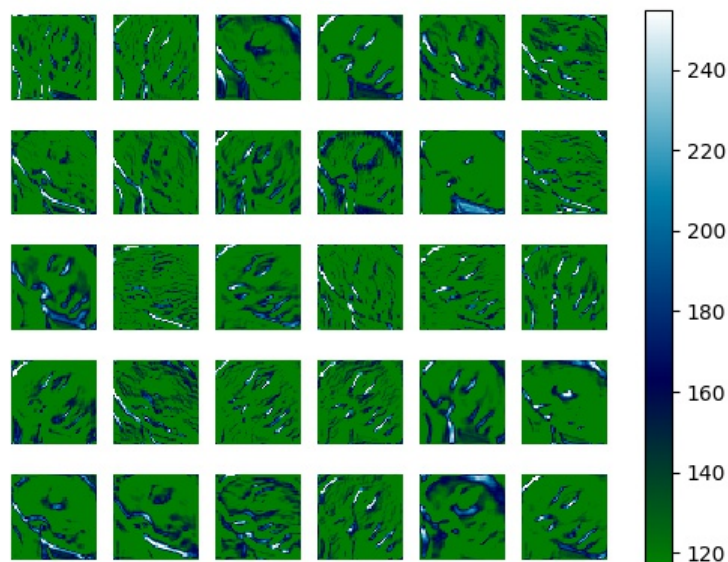


下圖是第18858筆訓練資料。



下圖是第18858筆訓練資料經過該層的輸出。


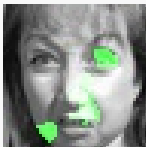
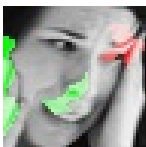

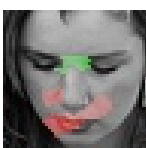
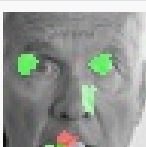
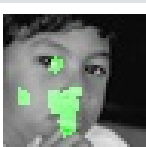
（右方數值經過縮放、平移處理，並非原始數值）。



可以發現這一層filters的主要功能是偵測圖片中不同角度、不同密度的條紋。圖片經過該層的輸出符合filter visualization的結果，例如第四個Row，第一個Column的filter是用來偵測左下一右上的條紋，而圖片的輸出結果確實只有偵測出這個方向的變化。

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)。

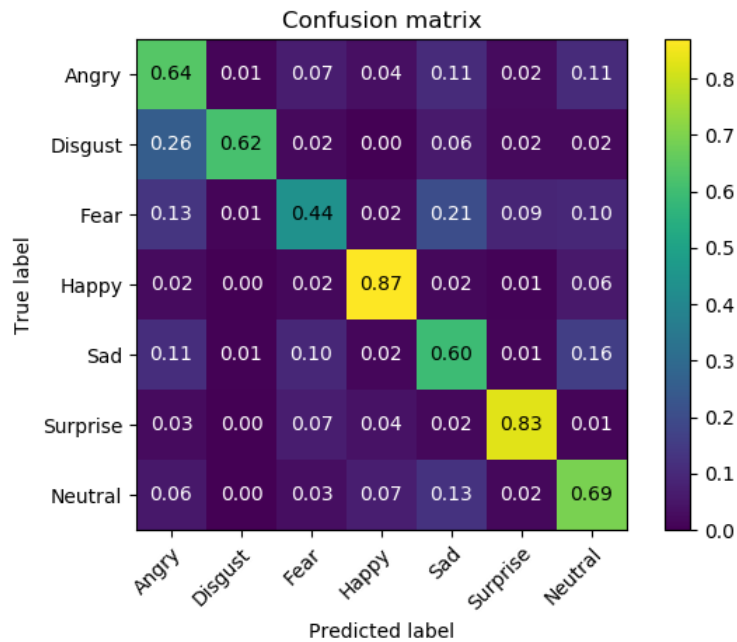
答：

表情	結果 (原圖同第一小題)
Angry	
Disgust	
Fear	
Happy	
Sad	
Surprise	
Neutral	

圖中綠色區塊代表正相關，紅色區塊代表負相關。

根據結果，可以猜測我的模型判斷Angry的依據是牙齒、嘴、眼周，Disgust是眼睛、嘴巴附近，Fear是手、眼睛、口鼻，Happy是嘴附近、臉頰，Sad是鼻子上方、嘴附近，Surprise是眼睛、鼻子附近、嘴和手，Neutral是口鼻、臉頰、和眼睛。

下圖是作業三的Confusion Matrix，

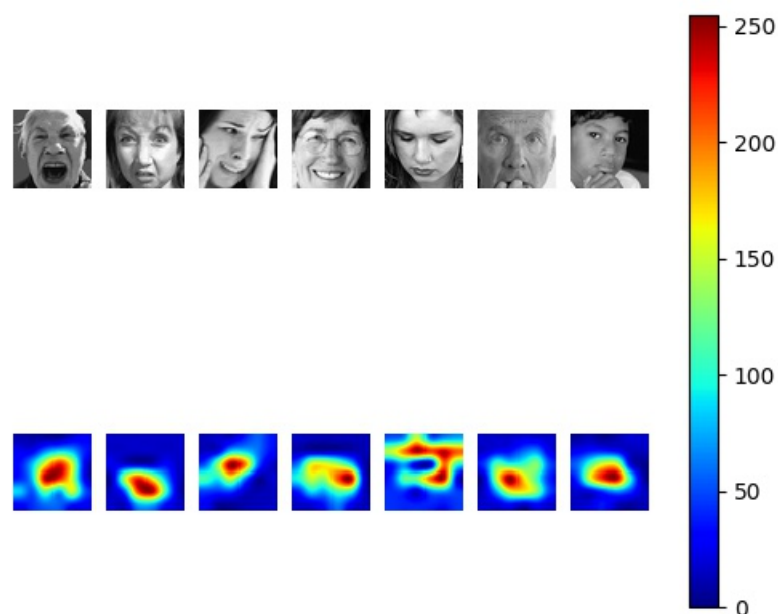


我的模型表現最好的前兩名是Happy和Surprise。根據LIME的結果，我們可以猜測Happy的結果特別好是因為笑的時候，嘴巴附近和臉頰肌肉的改變很劇烈，而且相對於其他表情來說是獨一無二的。而Surprise的結果特別好可能是因為驚訝時眼睛附近的改變也是明顯而獨一無二的。其他表現比較不好的表情可能是因為沒有獨一無二的特徵可以當作判斷依據。

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：

我使用了keras-vis套件裡面的visualize_cam函式（即Grad-CAM演算法）對我的模型做了分析。我先將我的Model最後一層（Activation層）的Softmax改成Linear，並針對最後一層和每個類別的圖片做分析。除了必須指定的參數之外，我沒有改動任何其他visualize_cam的參數。下圖是實驗結果。



觀察結果可以發現，我的Model判斷Angry可能是著重在鼻子附近，Disgust著重在嘴的附近，Fear著重在鼻子附近，Happy著重在鼻子附近，Sad著重在額頭和臉頰，Surprise著重在鼻子附近，Neutral著重在鼻子附近。