

Homework 3 Report - Image Sentiment Classification

學號: R06942018, 姓名: 何適楷, 系級: 電信碩一

1 Problem 1

1. (1%) 請說明你實作的 CNN model，其模型架構、訓練參數和準確率為何？

Layer (type)	Output Shape	Param #
ZeroPadding2D	(50, 50, 1)	0
Conv2D	(48, 48, 32)	320
Conv2D	(48, 48, 32)	9248
MaxPooling2D	(24, 24, 32)	0
Conv2D	(24, 24, 64)	18496
Conv2D	(24, 24, 64)	36928
MaxPooling2	(12, 12, 64)	0
Conv2D	(12, 12, 128)	73856
Conv2D	(12, 12, 128)	147584
Conv2D	(12, 12, 128)	147584
MaxPooling2	(6, 6, 128)	0
Flatten	(4608)	0
Dense	(196)	903364
Dropout(0.5)	(196)	0
Dense	(196)	38612
Dropout(0.5)	(196)	0
Dense	(7)	1379
Total params: 1,377,371		
optimizer: Adam		
Learning Rate: 3×10^{-4}		
epochs: 69(early stopping)		
batch size: 128		
training accuracy: 0.73953		
public accuracy: 0.69155		
private accuracy: 0.68069		

2 Problem 2

(1%) 請嘗試 data normalization, data augmentation, 說明實行方法並且說明對準確率有什麼樣的影響？

	public accuracy	private accuracy
Raw data	0.55363	0.53970
data normalization	0.58818	0.58484
data argmentation	0.66759	0.67205
nor+arg	0.69155	0.68069

在 data normalization 中，我把所有資料減去所有資料平均然後再除以所以資料的標準差，得到的結果馬上就比沒有做 data normalization 的高 3-5%。

在 data augmentation 中，我加入了轉了 10 度、20 度、30 度的 data 然後這三筆 data 與原始資料各 train 20 epochs，得到顯著的成長，可以比原始增加 14% 左右之多，實在非常驚人。

3 Problem 3

(1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析] (Collaborators:)

我們可以發現到 4(難過) 和 2(恐懼) 這兩種表情互相的猜錯率較高於其他情形，這也蠻合理

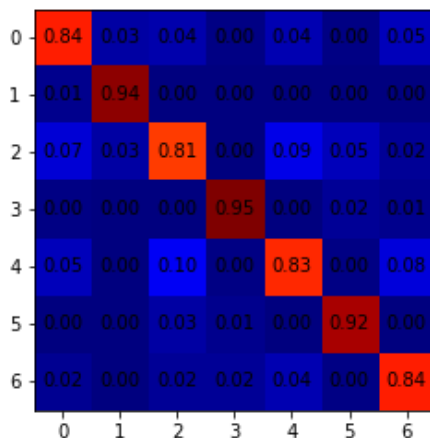


Figure 1: confusion matrix

的，以人類的角度來看，這兩種表情確實還蠻神似的。

4 Problem 4

(1%) 從 (1)(2) 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？

可以看出這個 model 主要都是 focus 在臉上面，似乎嘴巴也有稍微被強化，背景則是被弱化。

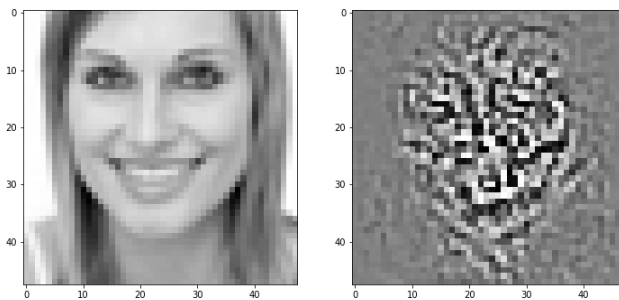


Figure 2: saliency map

5 Problem 5

(1%) 承 (1)(2)，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。

我取的是 conv2D 的第二層 filter，可以觀察到，很多 filter 是在觀察直條紋，斜條紋，有粗有細，方向也不盡相同，有趣的是 filter19 和 filter11 似乎有抓到臉部和眼睛的輪廓，這兩個 filter 是比較接近人類的常識。

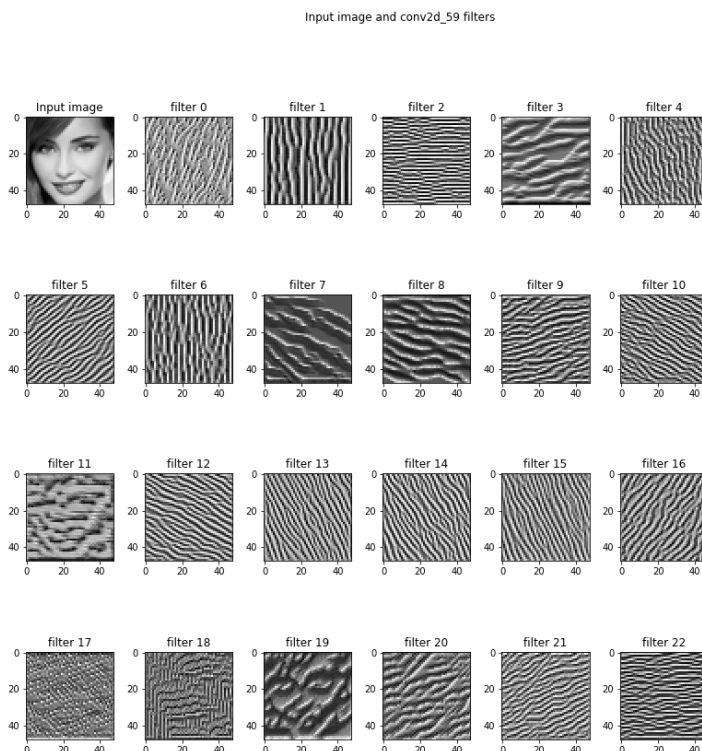


Figure 3: filter activated