

學號：R06521705 系級： 土木系營管組碩一 姓名：陳思愷

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？

(Collaborators:)

模型架構：

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 46, 46, 32)	320
conv2d_2 (Conv2D)	(None, 46, 46, 32)	9248
batch_normalization_1 (Batch Normalization)	(None, 46, 46, 32)	128
max_pooling2d_1 (MaxPooling2D)	(None, 23, 23, 32)	0
dropout_1 (Dropout)	(None, 23, 23, 32)	0
conv2d_3 (Conv2D)	(None, 23, 23, 64)	18496
conv2d_4 (Conv2D)	(None, 23, 23, 64)	36928
batch_normalization_2 (Batch Normalization)	(None, 23, 23, 64)	256
max_pooling2d_2 (MaxPooling2D)	(None, 11, 11, 64)	0
dropout_2 (Dropout)	(None, 11, 11, 64)	0
conv2d_5 (Conv2D)	(None, 11, 11, 128)	73856
conv2d_6 (Conv2D)	(None, 11, 11, 128)	147584
batch_normalization_3 (Batch Normalization)	(None, 11, 11, 128)	512
max_pooling2d_3 (MaxPooling2D)	(None, 5, 5, 128)	0
dropout_3 (Dropout)	(None, 5, 5, 128)	0
flatten_1 (Flatten)	(None, 3200)	0
dense_1 (Dense)	(None, 512)	1638912
batch_normalization_4 (Batch Normalization)	(None, 512)	2048
dropout_4 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 7)	3591
Total params: 1,931,879		
Trainable params: 1,930,407		
Non-trainable params: 1,472		

訓練參數:

Filter: 3*3

Maxpooling: 2*2

Dropout: 0.3

ImageDataGenerator:

rotation_range=10

width_shift_range=0.1

height_shift_range=0.1

horizontal_flip=True

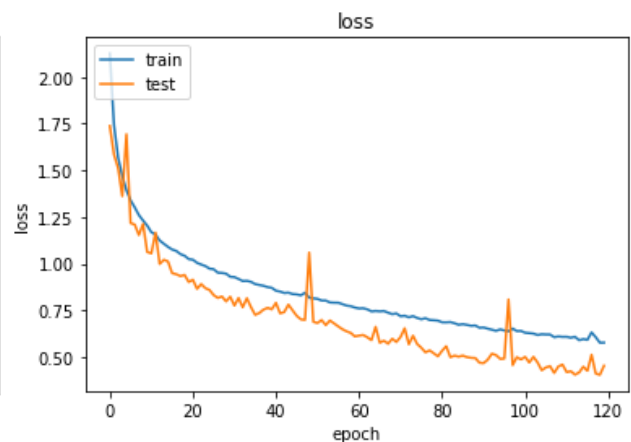
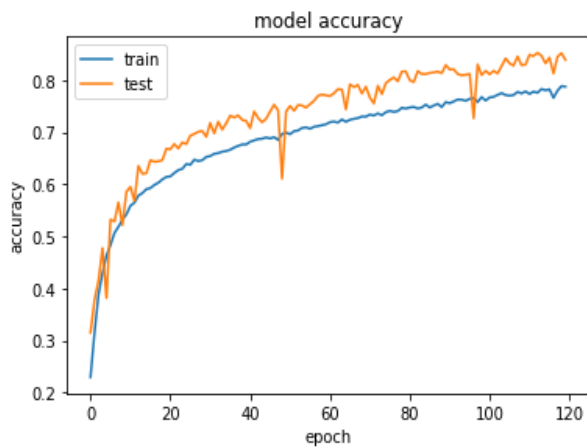
loss: categorical_crossentropy

optimizer: adam

metrics: accuracy

準確率:

Public score : 0.68487



2. 方法並且說明對準確率有什麼樣的影響？

(Collaborators:)

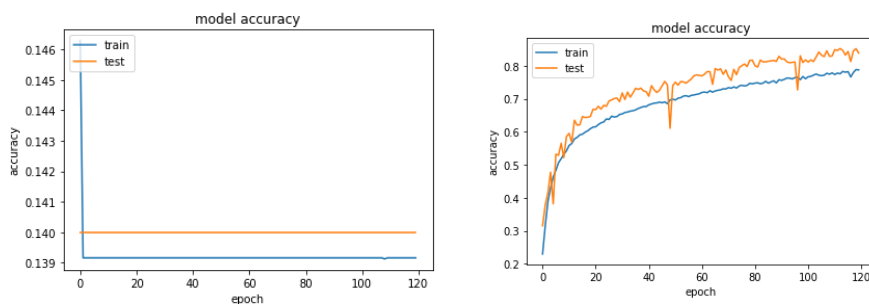
答: (1) data normalization :

對於 data 的 normalization 主要是在最一開始的 input 前，將所有數字都除以 255

讓數字介於 0~1 之間，再來是設置 batch_normalization 在 conv2d 以及在 dense 之前

有助於收斂速度以及避免 overfitting 的發生。

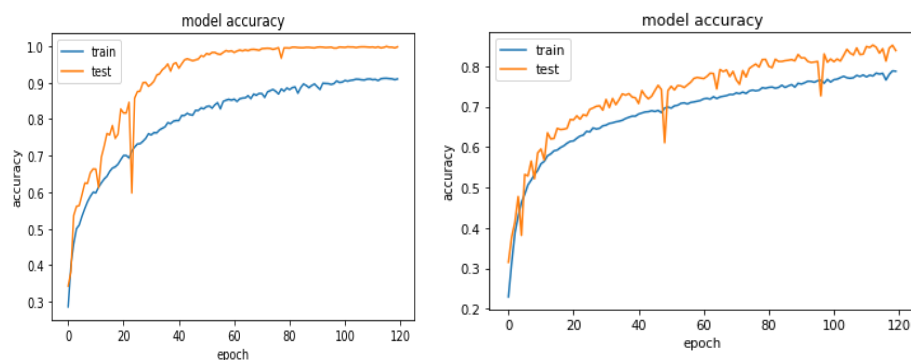
左邊是都沒有做的情況，右邊為有做的：



(2) data augmentation :

利用 keras 內建的 ImageDataGenerator 對圖片旋轉、縮放等等，以此來增加 data 的數量

(左邊是都沒有做的情況，右邊為有做的)



Public: 0.66480

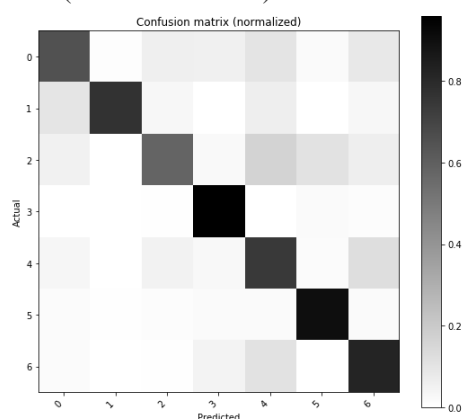
0.68487

Private: 0.65895

0.67957

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]

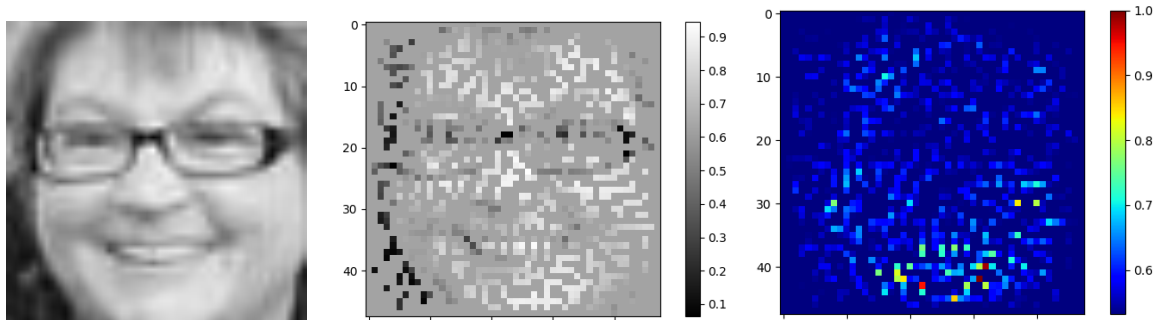
(Collaborators:)



class0(生氣)跟 class2(恐懼)比較容易答錯

推測應該是因為這兩個情緒的嘴巴可能都會打開，而照片裡面的比例本來就不同，機器難以精確辨識大小如何地開口是這兩種情緒因而造成誤差。

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？
(Collaborators:)



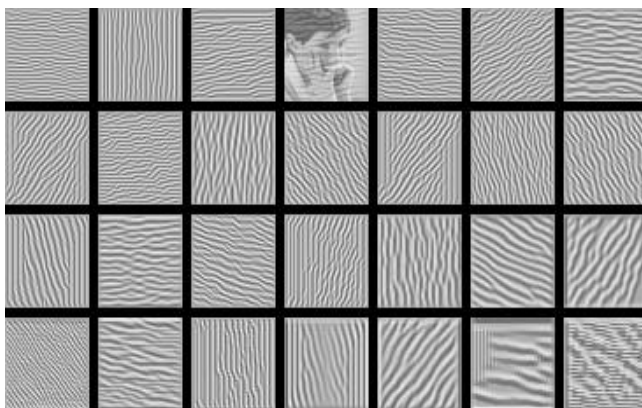
由圖片中可以看到，機器主要 focus 嘴巴的開度或角度，以及些許臉頰的起伏來作為判斷依據。

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。

(Collaborators:)

答：合理說明 test 的層數和觀察到的東西 -> 0.5 分

貼出 filter input and output 的圖片 -> 0.5 分



圖片 input 到各 filter 後 output 出來的圖片主要描繪出了眼睛、鼻子、嘴巴的輪廓，推測應該是這些部份來幫助機器做辨識

Output of layer0 (Given image17)

