機器學習 hw3 report
b04507009 電機三 何吉瑞

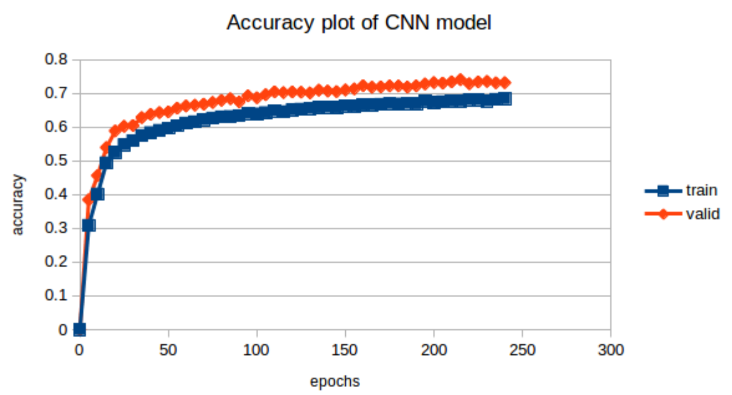
1.Build Convolution Neural Network

best 的結果由多個 model 的預測線性組合而成，此處挑其中一個 model(在 public 分數為 0.65)，以下為其架構
根據實驗結果，疊多層 Convolution 的結果好過於疊多層 Dense，此外，filter 數量逐漸增加的效果較好，和理論上 CNN 觀察的特徵越來越複雜符合

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 44, 44, 64)	1664
zero_padding2d_1 (ZeroPaddin	(None, 48, 48, 64)	0
max_pooling2d_1 (MaxPooling2	(None, 22, 22, 64)	0
zero_padding2d_2 (ZeroPaddin	(None, 24, 24, 64)	0
conv2d_2 (Conv2D)	(None, 22, 22, 64)	36928
zero_padding2d_3 (ZeroPaddin	(None, 24, 24, 64)	0
conv2d_3 (Conv2D)	(None, 22, 22, 64)	36928
average_pooling2d_1 (Average	(None, 10, 10, 64)	0
zero_padding2d_4 (ZeroPaddin	(None, 12, 12, 64)	0
dropout_1 (Dropout)	(None, 12, 12, 64)	0
conv2d_4 (Conv2D)	(None, 10, 10, 96)	55392
zero_padding2d_5 (ZeroPaddin	(None, 12, 12, 96)	0
conv2d_5 (Conv2D)	(None, 10, 10, 128)	110720
zero_padding2d_6 (ZeroPaddin	(None, 12, 12, 128)	0
dropout_2 (Dropout)	(None, 12, 12, 128)	0
conv2d_6 (Conv2D)	(None, 10, 10, 128)	147584
zero_padding2d_7 (ZeroPaddin	(None, 12, 12, 128)	0
average_pooling2d_2 (Average	(None, 5, 5, 128)	0
dropout_3 (Dropout)	(None, 5, 5, 128)	0
flatten_1 (Flatten)	(None, 3200)	0
dense_1 (Dense)	(None, 900)	2880900
activation_1 (Activation)	(None, 900)	0
dropout_4 (Dropout)	(None, 900)	0
dense_2 (Dense)	(None, 900)	810900
activation_2 (Activation)	(None, 900)	0
dropout_5 (Dropout)	(None, 900)	0
dense_3 (Dense)	(None, 7)	6307
activation_3 (Activation)	(None, 7)	0

```
Total params: 4,087,323
Trainable params: 4,087,323
Non-trainable params: 0
```

此 model 的 accuracy 對 epochs 做圖如下：

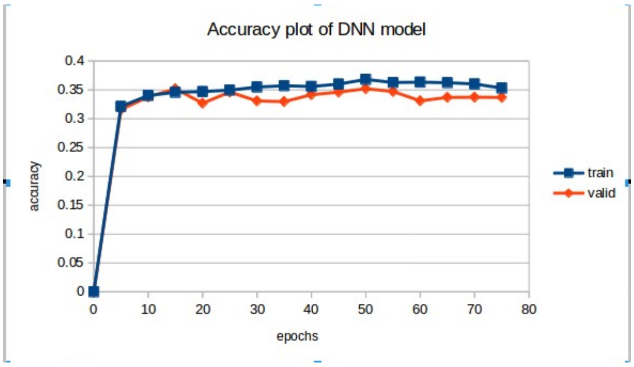


valid 的 accuracy 始終在 train 之上，這可能是因為我在切 valid 之前，就先對所有的 training data 做了 flip，也就是因為 preprocessing 使得 valid 的性質更接近 training，導致這個結果發生，大約差在 5%，但整體而言還是看的出正相關，後來就沒有再去修改，所以 valid 的表現會高過於 test

2. Build DNN
DNN 架構：

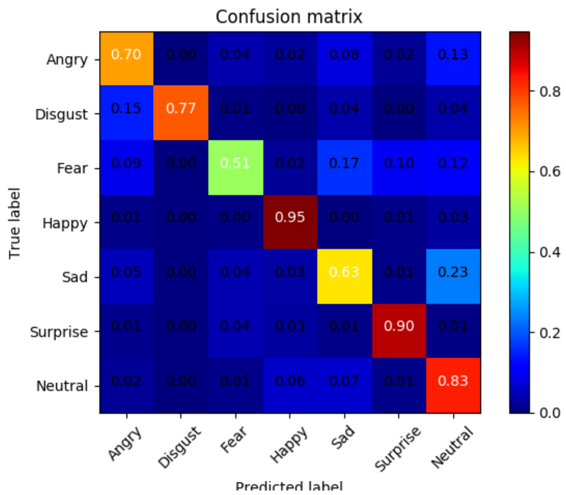
```
Layer (type)                 Output Shape              Param #
=================================================================
dense_1 (Dense)              (None, 1024)              2360320
_________________________________________________________________
activation_1 (Activation)    (None, 1024)              0
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 1024)              1049600
_________________________________________________________________
activation_2 (Activation)    (None, 1024)              0
_________________________________________________________________
dropout_2 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_3 (Dense)              (None, 768)               787200
_________________________________________________________________
activation_3 (Activation)    (None, 768)               0
_________________________________________________________________
dropout_3 (Dropout)          (None, 768)               0
_________________________________________________________________
dense_4 (Dense)              (None, 768)               590592
_________________________________________________________________
activation_4 (Activation)    (None, 768)               0
_________________________________________________________________
dropout_4 (Dropout)          (None, 768)               0
_________________________________________________________________
dense_5 (Dense)              (None, 512)               393728
_________________________________________________________________
activation_5 (Activation)    (None, 512)               0
_________________________________________________________________
dropout_5 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_6 (Dense)              (None, 512)               262656
_________________________________________________________________
activation_6 (Activation)    (None, 512)               0
_________________________________________________________________
dropout_6 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_7 (Dense)              (None, 7)                 3591
_________________________________________________________________
activation_7 (Activation)    (None, 7)                 0
=================================================================
Total params: 5,447,687
Trainable params: 5,447,687
Non-trainable params: 0
_________________________________________________________________
```

參數數量和 CNN 在同一個數量級



因為 earlystop 只記錄到大約 80 個 epoch，可以發現 DNN 的表現遠差於 CNN，且在 epoch 上升後沒有明顯進步，推論因為 feature 不限於某個位置，CNN 在偵測這個現象上勝過 DNN

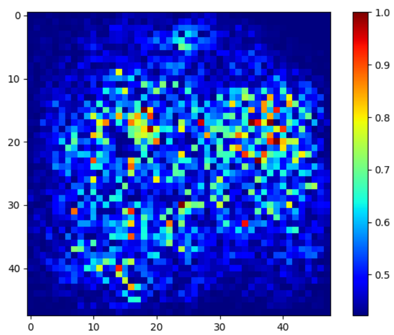
3. Analyze the model with confusion matrix



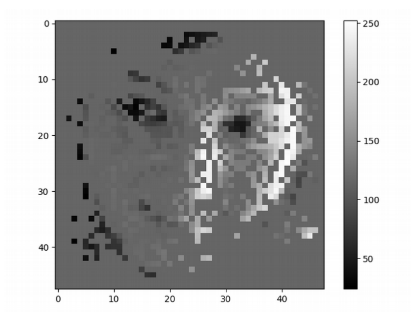
如 P1 所提到，因為做了同樣的 preprocessing，所以 accuracy 相較於 testing set 的表現較高，這裡可以看出 fear 的預測表現最差，但整體來講還是和預測錯誤有差距，因此由 confusion matrix 認為這是可信的預測

4.
input 原圖:

Saliency Map:



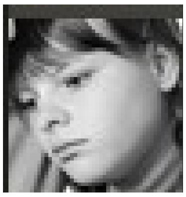
mask 掉 heatmap 小的部份



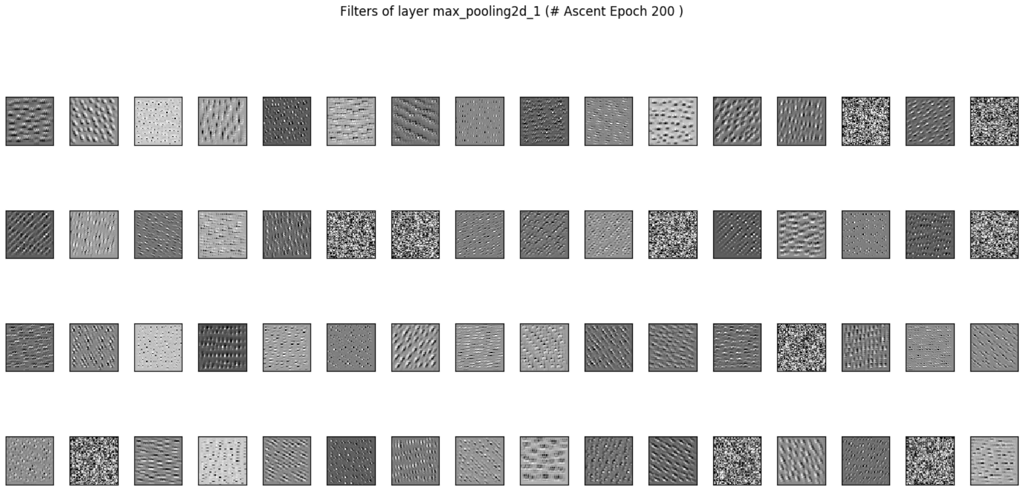
可以看出 silence map 在眼睛和嘴巴附近值較高，雖然還有其他空白，但大致保留這幾個部位去計算，認為 model 的選擇是合理的

5.Analyze the model by visualizing filters

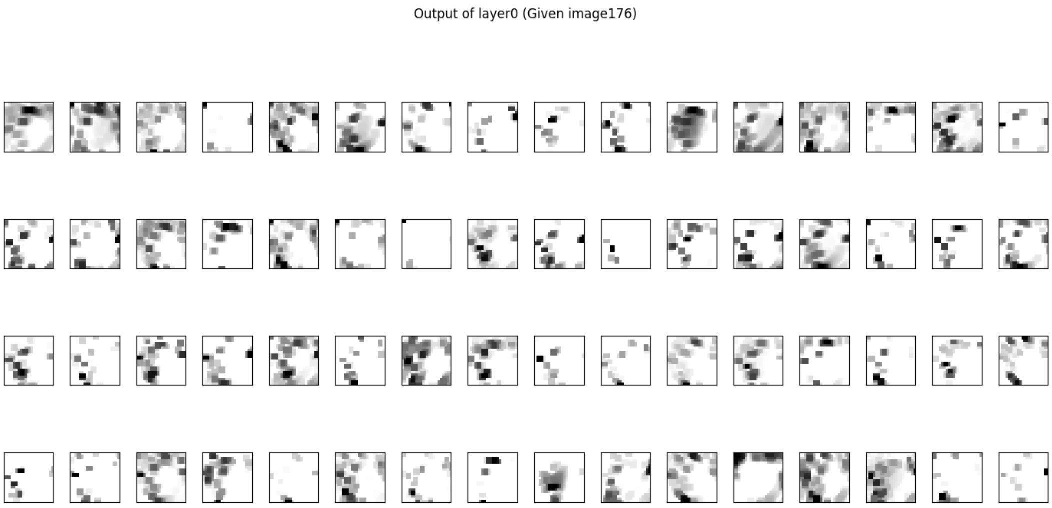
原始圖片：



filters 觀察



圖片 trigger 後的 layer 觀察



(cmap='gray_r')
可以發現在第 1 層 filter 確實是觀察簡單的線條，但還是有些許的雜訊，但大致記錄了圖片的特徵
而在 output layer 方面，大部份確實偵測到臉部的表情特徵如五官，也大概能觀察出臉的輪廓，符合理論上第一層的特性