

Homework 3 Report – Image Sentiment Classification

學號：r06521605

系級：土木所電輔組碩一

姓名：許舜翔

1.

a. 模型架構

根據 Stanford 大學的一篇關於 CNN 影像辨識的論文[1]，及 VGG-Net 的架構，但 padding 設置為'valid'，filter 數量以 2 倍成長，並設置如下的結構：

```
model.summary()
```

Layer (type)	Output Shape	Param #
conv2d_25 (Conv2D)	(None, 46, 46, 64)	640
leaky_re_lu_19 (LeakyReLU)	(None, 46, 46, 64)	0
conv2d_26 (Conv2D)	(None, 44, 44, 64)	36928
leaky_re_lu_20 (LeakyReLU)	(None, 44, 44, 64)	0
max_pooling2d_13 (MaxPooling)	(None, 22, 22, 64)	0
dropout_17 (Dropout)	(None, 22, 22, 64)	0
conv2d_27 (Conv2D)	(None, 20, 20, 128)	73856
leaky_re_lu_21 (LeakyReLU)	(None, 20, 20, 128)	0
conv2d_28 (Conv2D)	(None, 18, 18, 128)	147584
leaky_re_lu_22 (LeakyReLU)	(None, 18, 18, 128)	0
max_pooling2d_14 (MaxPooling)	(None, 9, 9, 128)	0
dropout_18 (Dropout)	(None, 9, 9, 128)	0
conv2d_29 (Conv2D)	(None, 7, 7, 256)	295168
leaky_re_lu_23 (LeakyReLU)	(None, 7, 7, 256)	0
conv2d_30 (Conv2D)	(None, 5, 5, 256)	590080
leaky_re_lu_24 (LeakyReLU)	(None, 5, 5, 256)	0
max_pooling2d_15 (MaxPooling)	(None, 2, 2, 256)	0
dropout_19 (Dropout)	(None, 2, 2, 256)	0
flatten_5 (Flatten)	(None, 1024)	0
dense_9 (Dense)	(None, 2048)	2099200
dropout_20 (Dropout)	(None, 2048)	0
dense_10 (Dense)	(None, 7)	14343
Total params: 3,257,799		
Trainable params: 3,257,799		
Non-trainable params: 0		

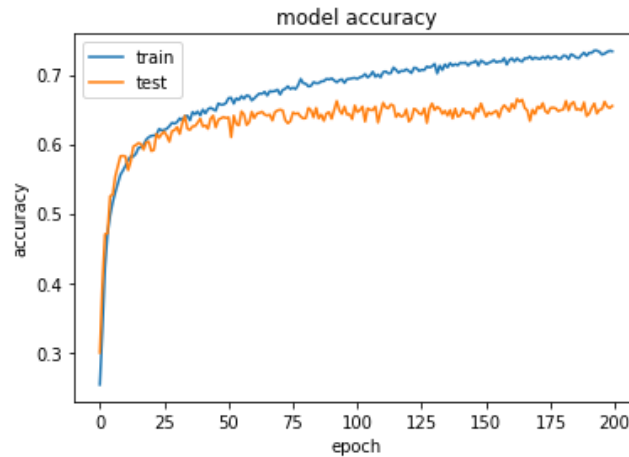
b. 訓練參數

Optimizer = Adam

Learning rate = 0.001

Epochs = 200

c. 準確率



最後 validation accuracy 大概會卡在 0.64-0.65 之間，training accuracy 則是持續上升，有 overfitting 的現象。

2.

a. Data normalization

採用 standard normalization 的方式，採用的是根據全部 image 每個 pixel 取得的平均值及標準差。

b. Data augmentation

利用 keras 套件中的 ImageGenerator，設置 featurewise_center=True, featurewise_std_normalization=True, rotation_range = 30, zoom_range = 0.25, horizontal_flip = true

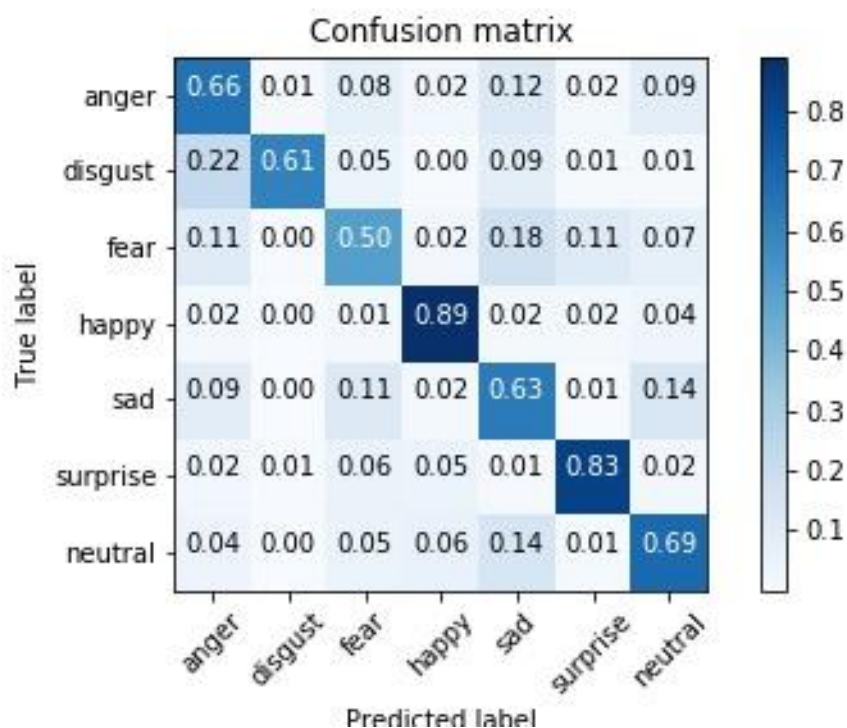
表一、實作前後準確率比較

Data normalization (validation 的結果)	
實作前	實作後
0.60	0.65
Data augmentation (validation 的結果)	
實作前	實作後
0.55	0.65

可以看出這次 augmentation 的影響比較大，應該是影像資料過少，又分為 7 類，導致模型容易產生 overfitting 的現象，故雖然 training accuracy 可以持續上升，但 validation 卻會卡在 0.55-56 左右。

3.

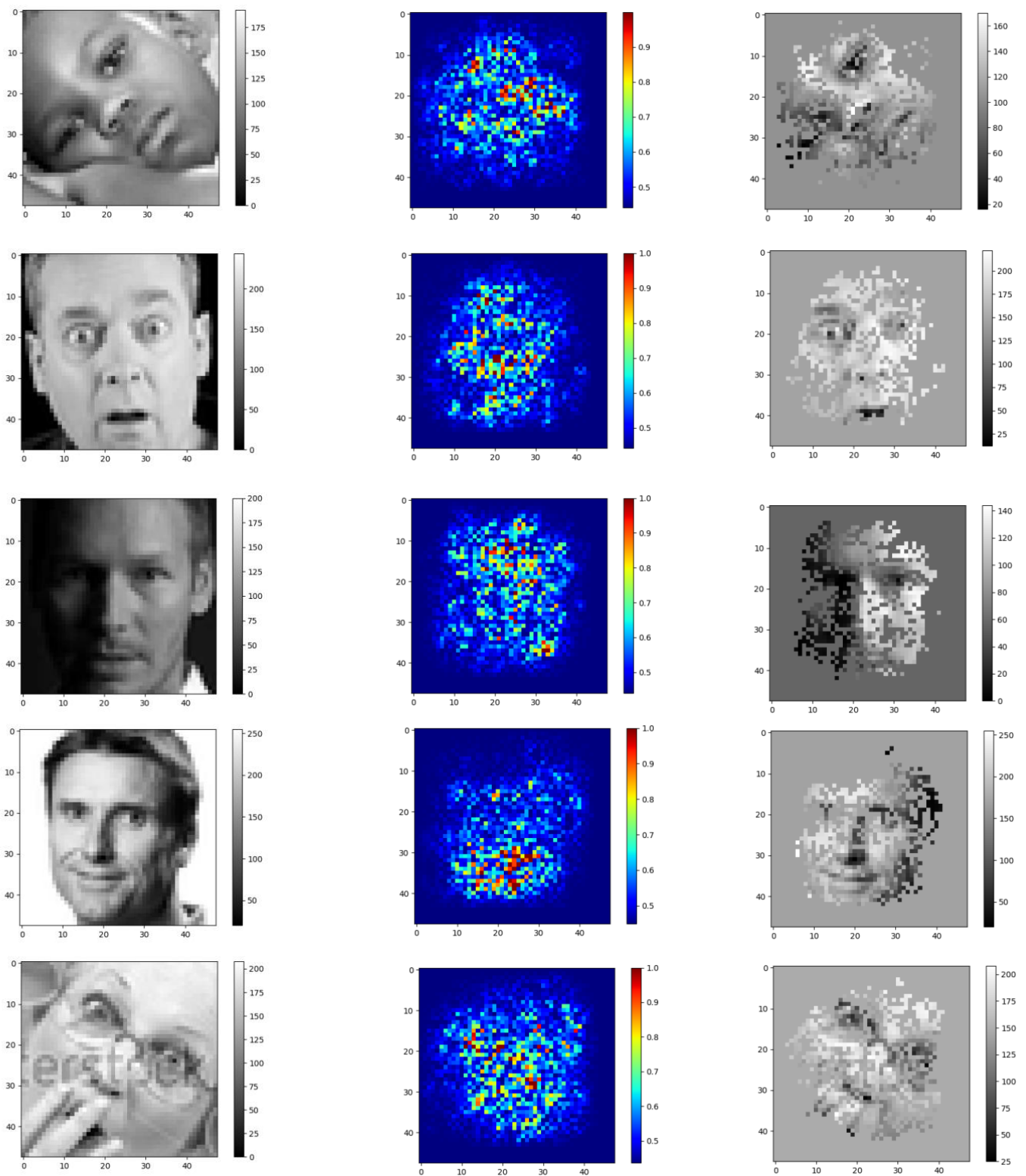
以下為根據 confusion matrix 畫出的比較圖：



可以看出 disgust 容易跟 anger 搞混；fear 容易跟 sad 搞混。

4.

隨機 random 選出 5 張照片製作 saliency maps，結果如下：

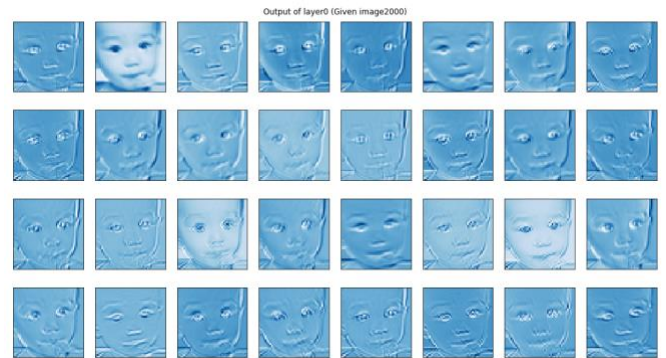
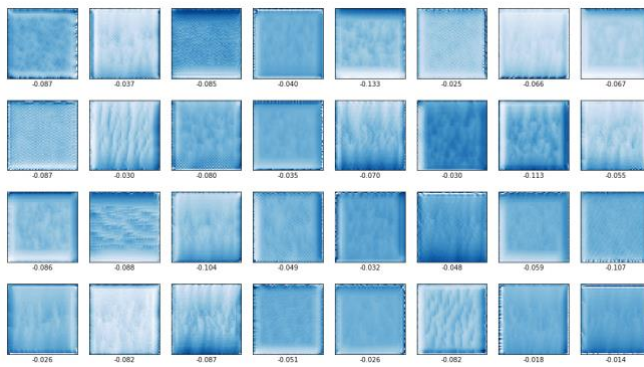


可以看出有抓出重點會在臉部的部分，主要會是鼻子嘴巴眼睛等部位。

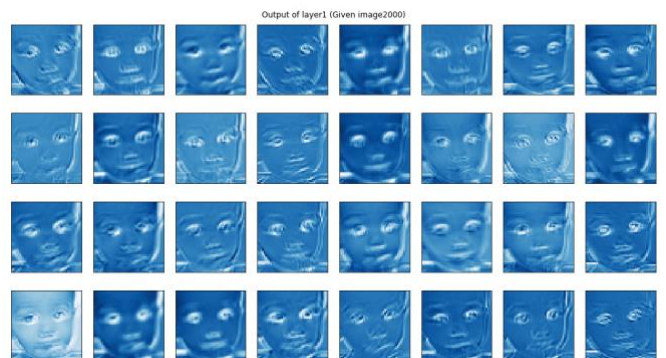
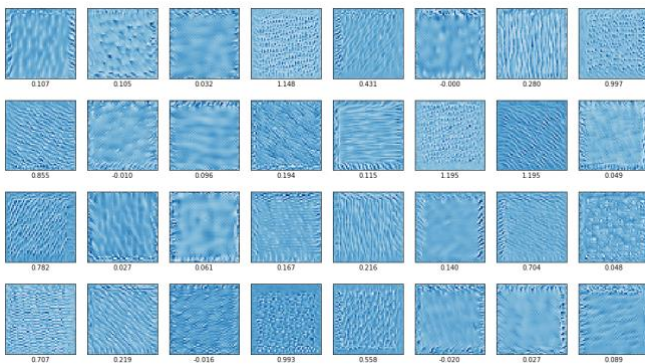
5.

取前兩層的 conv2d layer 的 filter 進行觀察，並套入隨機選擇的圖片，可得到以下結果：

第一層



第二層



可以發現模型能抓到 texture 的特徵，並於圖片中加強這些特徵於圖片中所佔的比例。

Reference

[1] Alizadeh, S., & Fazel, A. (2017). Convolutional Neural Networks for Facial Expression Recognition. *arXiv preprint arXiv:1704.06756*.