Homework3 Report

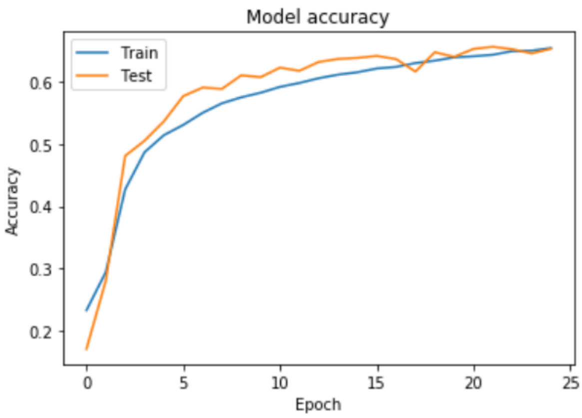
Professor Pei-Yuan Wu
EE5184 - Machine Learning

姓名：盧庭偉
學號：B05602022

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

Conv2D(64)	kernel_size=(3, 3) activation=relu	Dense(1024)	activation=relu
BatchNormalization		BatchNormalization	
MaxPooling2D	pool_size = (2, 2)	Dropout	0.5
Dropout	0.25	Dense(1024)	activation = relu
Conv2D(64)	kernel_size = (7, 7) activation = relu	BatchNormalization	
BatchNormalization		Dropout	0.5
MaxPooling2D	pool_size = (2, 2)	Dense(512)	activation = relu
Dropout	0.3	BatchNormalization	
Conv2D(128)	kernel_size = (5, 5) activation = relu	Dropout	
BatchNormalization		Dense(256)	activation = relu
MaxPooling2D	pool_size = (2, 2)	BatchNormalization	
Dropout	0.35	Dropout	0.5
Conv2D(128)	kernel_size = (3, 3) activation = relu	Dense(128)	activation = relu
BatchNormalization		BatchNormalization	
MaxPooling2D	pool_size=(2, 2)	Dropout	0.5
Dropout	0.4	Dense(7)	activation = softmax
Flatten			

(順序為左邊欄 → 右邊欄)

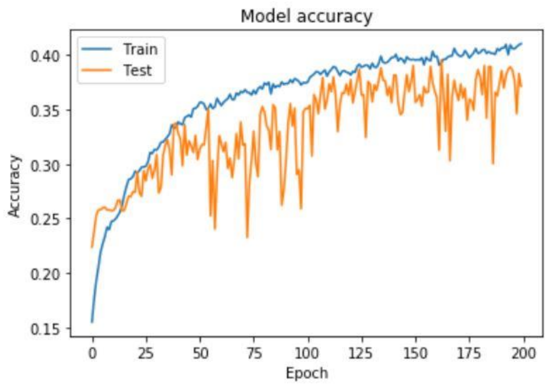


Total params	2,906,055
Epoch	24 (stopped by early stopping)
Early stopping:	Yes (Patience = 3)
Training data	First 23000 data set
Validating data	Remaining 5709 data set
Loss function	Cross Entropy
Optimizer	Adam
public score	0.67679
private score	0.68682

2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model，其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

Dense(512)	activation = relu	Dense(512)	activation = relu
BatchNormalization		BatchNormalization	
Dropout	0.5	Dropout	0.5
Dense(512)	activation = relu	Dense(512)	activation = relu
BatchNormalization		BatchNormalization	
Dropout	0.5	Dropout	0.5
Dense(512)	activation = relu	Dense(256)	activation = relu
BatchNormalization		BatchNormalization	
Dropout		Dropout	
Dense(512)	activation = relu	Dense(128)	activation = relu
BatchNormalization		BatchNormalization	
Dropout	0.5	Dropout	0.5
Dense(512)	activation = relu	Dense(7)	activation = relu
BatchNormalization			
Dropout	0.5		

(順序為左邊欄 → 右邊欄)



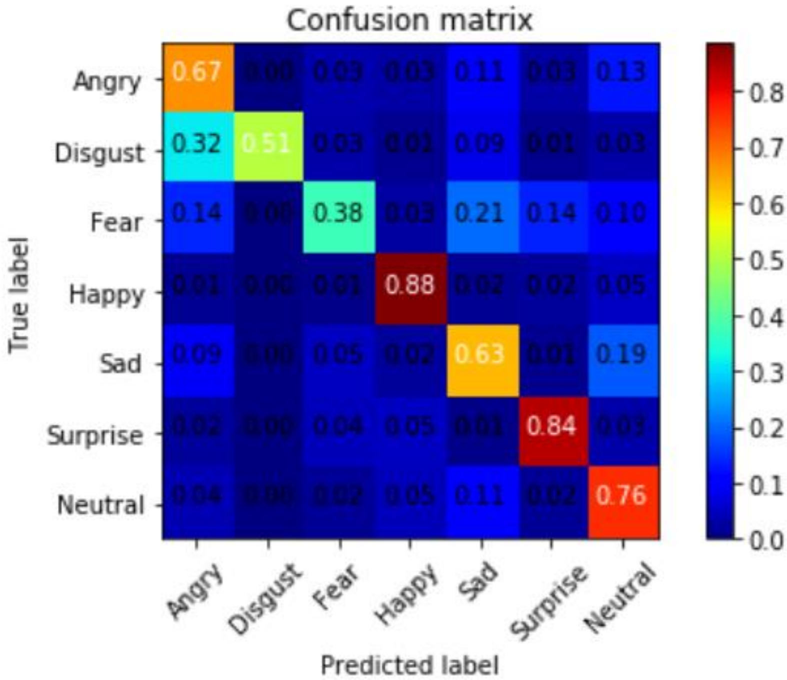
Total params	2,845,895
Epoch	200
Early stopping:	No
Training data	First 23000 data set
Validating data	Remaining 5709 data set
Loss function	Cross Entropy
Optimizer	Adam
public score	0.32432
private score	0.34243

1. DNN 的 performance 明顯比 CNN 差

2. 由於 CNN 較容易收斂，為了避免 overfitting，因此我有設了 early stopping。而 DNN 則是不太會收斂，因此並沒有設 early stopping 而是讓他跑完 200 個 epoch。

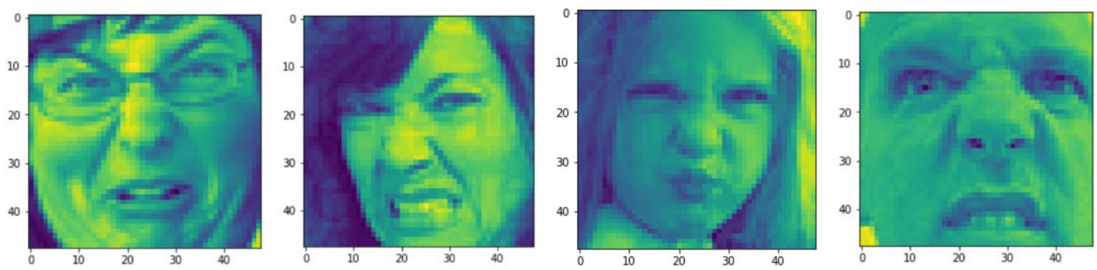
3. 在 training 時可以發現，DNN 了 training 速度比 CNN 快很多，推測是因為 CNN 在做 convolution 的時候， filter 參數是共用的，所以 gradient 是取其對所有 pixel 做 convolution 後的平均。因此在同樣數量級的參數下，CNN 的運算次數是比 DNN 多很多的。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？ 並說明你觀察到了什麼？ [繪出 confusion matrix 分析]



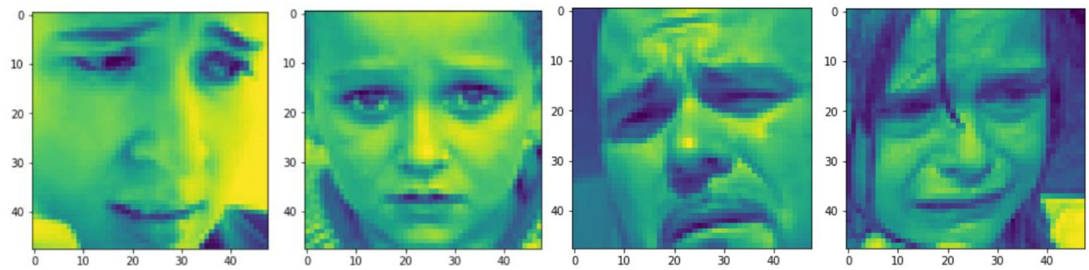
由 confusion matrix 可以發現，Disgust 比要容易搞混成 Angry，而 Fear 比較容易搞混成 Sad。於是我們實際拿一些搞混的例子來看(Label / Predict):

Disgust / Angry:



可以觀察到，導致這兩種狀態判斷困難的相似特稱可能為鼻子皺起來、嘴角向下……

Fear / Sad:



可以觀察到，導致這兩種狀態判斷困難的相似特稱可能為眼角向下、皺眉……

4.

(a) Layer A: $(2\times2\times5+1)\times6=126$ #

Layer B: $(2\times2\times6+1)\times4=100$ #

(b) A: addition $[(2\times2\times5)-1]\times(3\times3)\times6=1026$ #

multiplication: $[(2\times2\times5)]\times(3\times3)\times6=1080$ #

B: addition $[(2\times2\times6)-1]\times(1\times1)\times4=92$ #

multiplication: $[(2\times2\times6)]\times(1\times1)\times4=96$ #

(c) $n_{i+1}=\frac{(n_i-k_i+2P_i)}{S_i}+1$ = output length = # of kernel calculation

length, depth of kernel

for layer i: $\underbrace{[k_i^2\cdot C_{i-1}]}_{\text{one kernel calculation}}\cdot n_{i+1}^2\cdot C_i$

→ # of channel

↳ # of kernel calculation

$\therefore$ total $=\sum_{i=1}^{l}[(k_i^2\cdot C_{i-1})\cdot n_{i+1}^2\cdot C_i]$

where $\begin{cases} n_0 = \text{input length} \\ C_0 = \text{input depth} \end{cases}$

$\Rightarrow$ time complexity: $O\left(\sum_{i=1}^{l}[(k_i\cdot n_{i+1})^2\cdot C_i\cdot C_{i-1}]\right)$ #

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 8 & 5 \\ 3 & 12 & 9 \\ 1 & 8 & 5 \\ 5 & 14 & 2 \\ 7 & 4 & 1 \\ 9 & 8 & 9 \\ 3 & 8 & 1 \\ 11 & 5 & 6 \\ 10 & 11 & 7 \end{bmatrix}$$

$\bar{X}$: 5.4 8 4.8

$\mathrm{cov}(X)$:

$$\begin{bmatrix} 13.38 & 0.56 & 3.64 \\ 0.56 & 13.56 & 3.22 \\ 3.64 & 3.22 & 9.07 \end{bmatrix}$$

V:

$$\begin{bmatrix} -0.62 & -0.68 & -0.4 \\ -0.59 & 0.73 & -0.34 \\ -0.52 & -0.03 & 0.85 \end{bmatrix}$$

D:

$$\begin{bmatrix} 17 & & \\ & 12.92 & \\ & & 6.08 \end{bmatrix} \quad \begin{matrix} 0.47\% \\ 0.36\% \\ 0.17\% \end{matrix}$$

(a) $\begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix} \cdot \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix} \cdot \begin{bmatrix} -0.4 \\ -0.34 \\ 0.85 \end{bmatrix}$ #

(b) $X' = XV$

$$\begin{bmatrix} -3.36 & 0.69 & 1.47 \\ -9.8 & 2.97 & -0.07 \\ -13.62 & 6.45 & 2.37 \\ -7.94 & 5.01 & 1.13 \\ 12.4 & 6.76 & -5.06 \\ -7.22 & -1.87 & -3.31 \\ -14.98 & -0.55 & 1.33 \\ -7.1 & 3.77 & -3.07 \\ -12.89 & -4.01 & -1 \\ -16.33 & 1.02 & -1.79 \end{bmatrix}$$ #

(c) transfer function $P = \begin{bmatrix} -0.62 & -0.68 \\ -0.59 & 0.73 \\ -0.52 & -0.03 \end{bmatrix}$

reconstruction error

$= \| X - XPP^T \|^2$

$= 20.55$ #