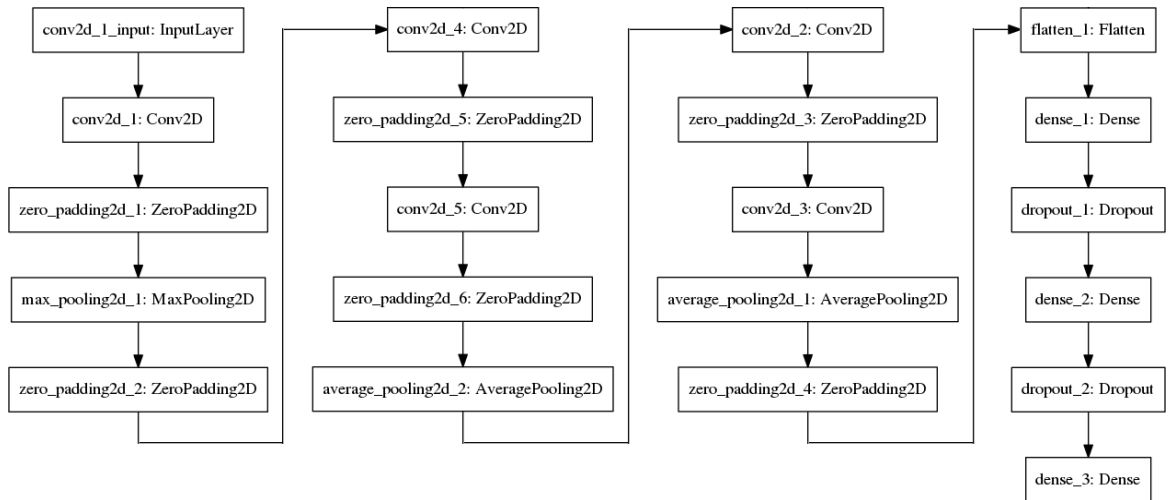
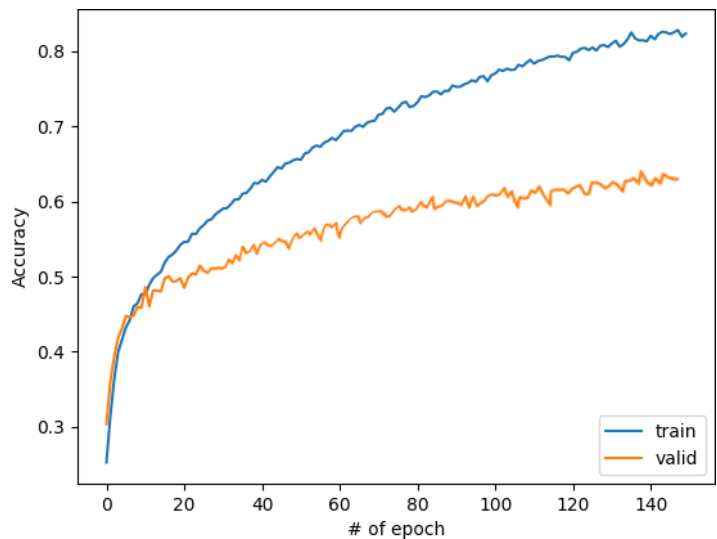


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

答：

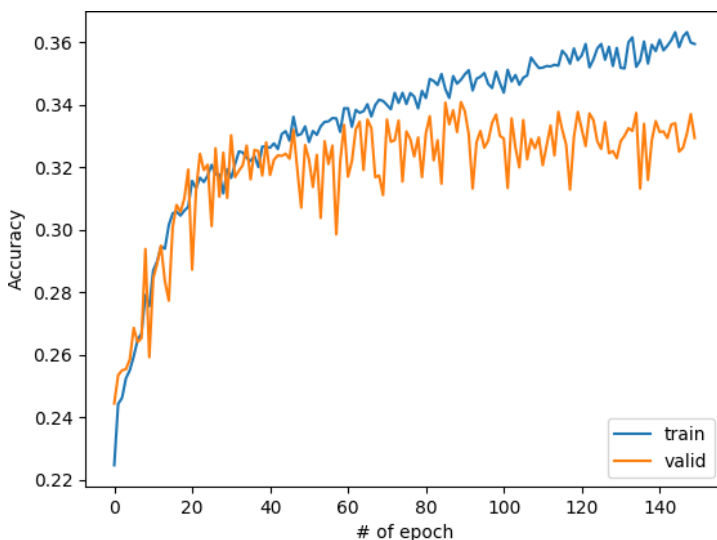


上圖為模型架構圖，是使用助教給的預設模型，而右邊則是在 150 個 epoch 的訓練之下採用 128 batch size，並且使用 0.2 的 validation 和採用 0.2 的 drop rate 來做測試，可以發現到 train 的準確率穩定上升，但是 valid 最後大約都落在 5 到 6 成左右而沒有向上提升的現象。



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？

答：



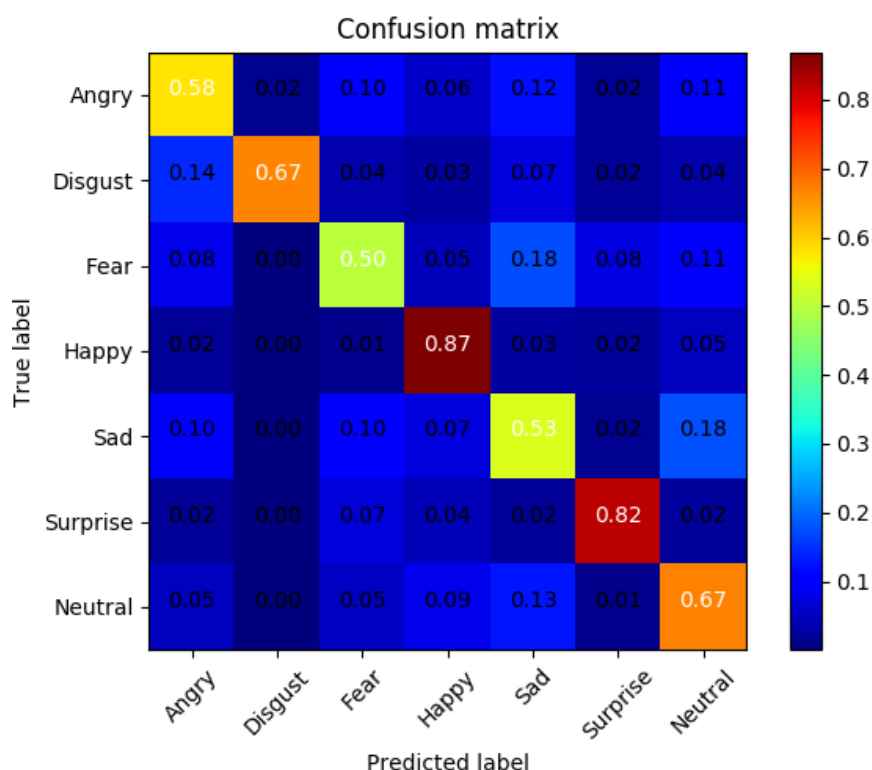
Layer (type)	Output Shape	Param #
max_pooling2d_1 (MaxPooling2D)	(None, 22, 22, 1)	0
flatten_1 (Flatten)	(None, 484)	0
dense_1 (Dense)	(None, 1024)	496640
dropout_1 (Dropout)	(None, 1024)	0
dense_2 (Dense)	(None, 1024)	1049600
dropout_2 (Dropout)	(None, 1024)	0
dense_3 (Dense)	(None, 1024)	1049600
dropout_3 (Dropout)	(None, 1024)	0
dense_4 (Dense)	(None, 1024)	1049600
dropout_4 (Dropout)	(None, 1024)	0
dense_5 (Dense)	(None, 1024)	1049600
dropout_5 (Dropout)	(None, 1024)	0
dense_6 (Dense)	(None, 7)	7175
Total params: 4,702,215		
Trainable params: 4,702,215		
Non-trainable params: 0		

右上為模型架構圖，而左上為則是與上題使用相同的訓練參數。可以發現到，與 CNN 比較之下，準確率不但降低了很多，而且 **train** 的準確率也趨近飽和的狀態。不過相同的是，二個模型大約都在 **20 個 epoch** 就趨近飽和了。因為在沒有透過 **filter** 的情況下，我們所要觀察的特徵就不會很明顯，以至於無法讓機器學的更好。

### 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]

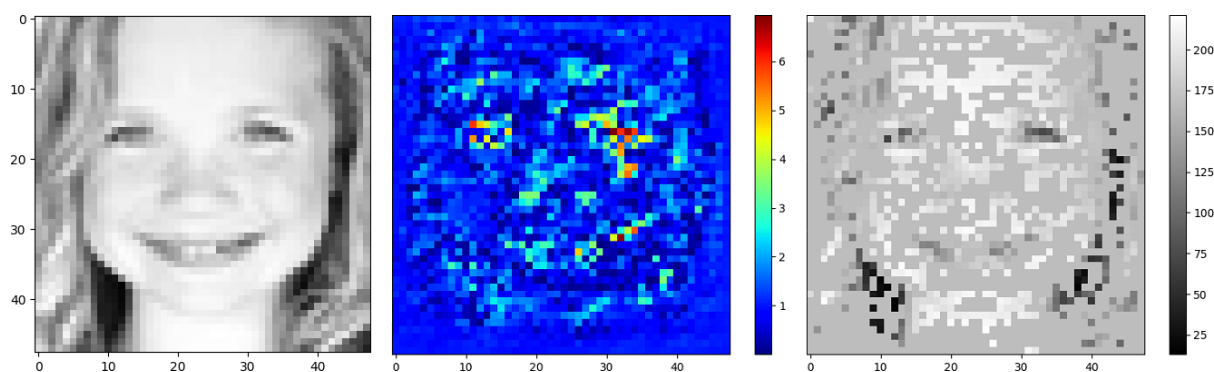
答：

透過右邊的矩陣來看，我們發現到，生氣容易與厭惡容易搞混，而恐懼，難過和中立容易搞混，而快樂跟驚訝則是特徵最明顯的二個分類，準確率都在八成以上。而恐懼與難過則是所有分類中，準確率最低的，我認為恐懼與難過的情緒是相輔相成的。



### 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？

答：



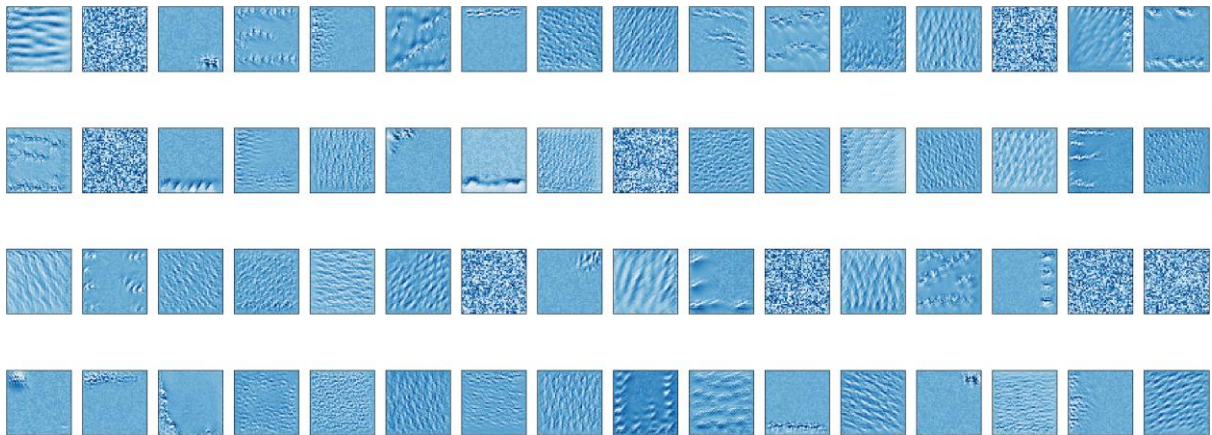
主要著重在五官的部分，用來強化表情的特徵。

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**。

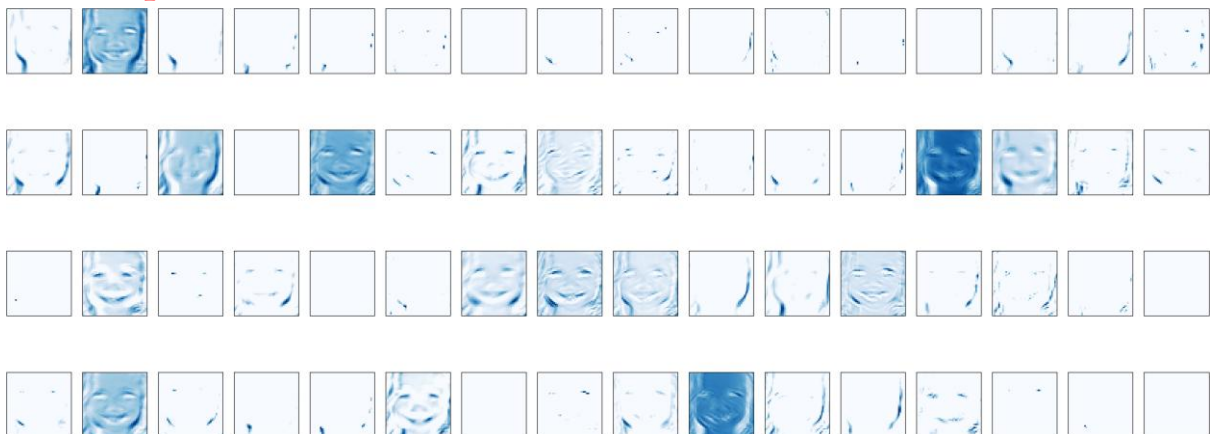
答：

因為 **layer** 之間是互相連接的，因此在低階的 **layer** 中，通常都會過濾出簡單的線條，在高層級的 **layer** 中，則會過濾出更高階的特徵，像是眼睛或嘴巴等特徵，我認為五官明顯的圖片較容易被 **filter** 給過濾出來。

下圖使用 **64** 個 **filter** 來做處理



**Layer1**



**Layer2**

