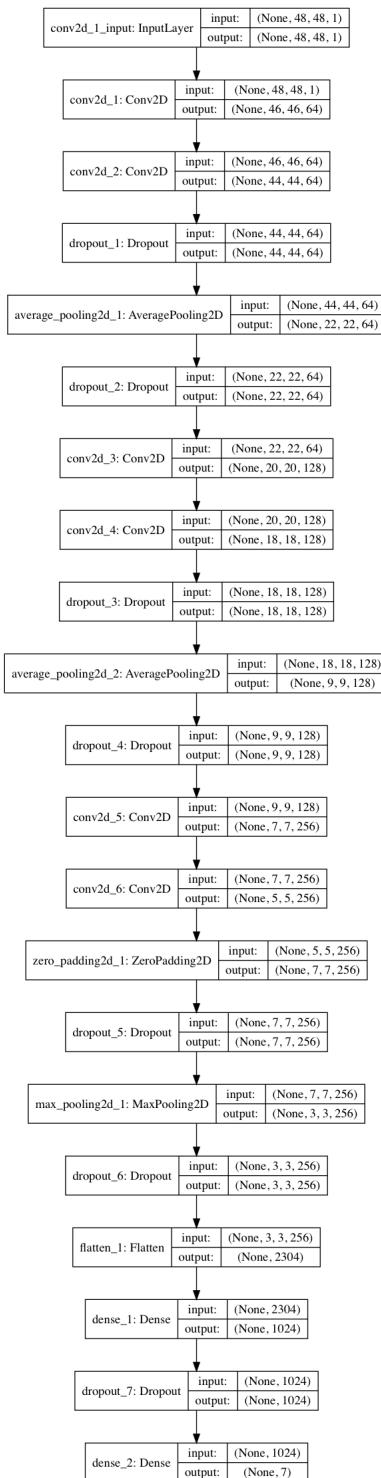


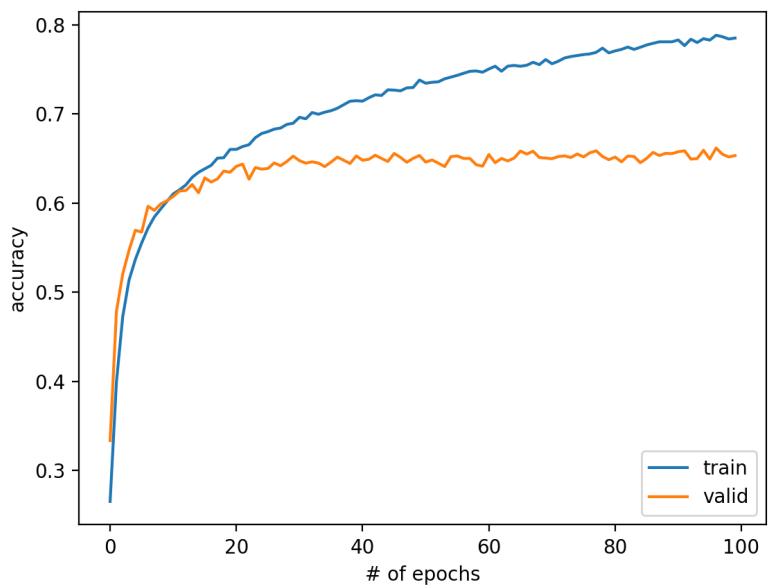
1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

答：

模型架構：



訓練過程：



preprocess: 我將每張圖片左右顛倒，變成兩倍的資料量。

準確率：

validation: 0.6571

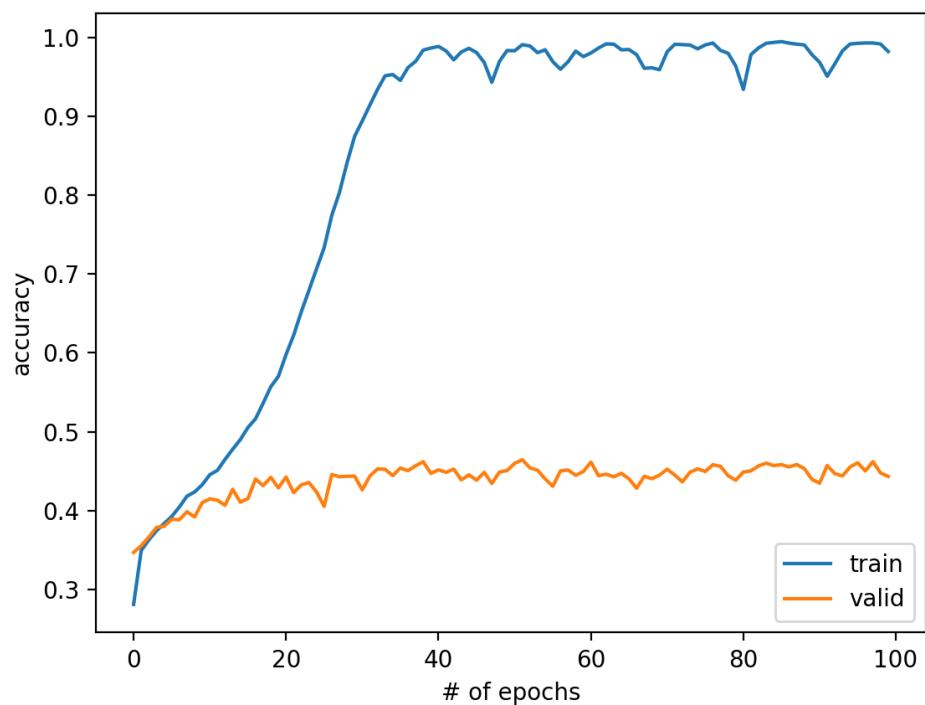
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

答：

模型架構：

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 900)	2074500
dense_2 (Dense)	(None, 900)	810900
dense_3 (Dense)	(None, 900)	810900
dense_4 (Dense)	(None, 7)	6307
<hr/>		
Total params: 3,702,607		
Trainable params: 3,702,607		
Non-trainable params: 0		

訓練過程：



準確率：

validation: 0.443

和 CNN 的比較：（參數：CNN: 約3,500,000；DNN: 約3,700,000）

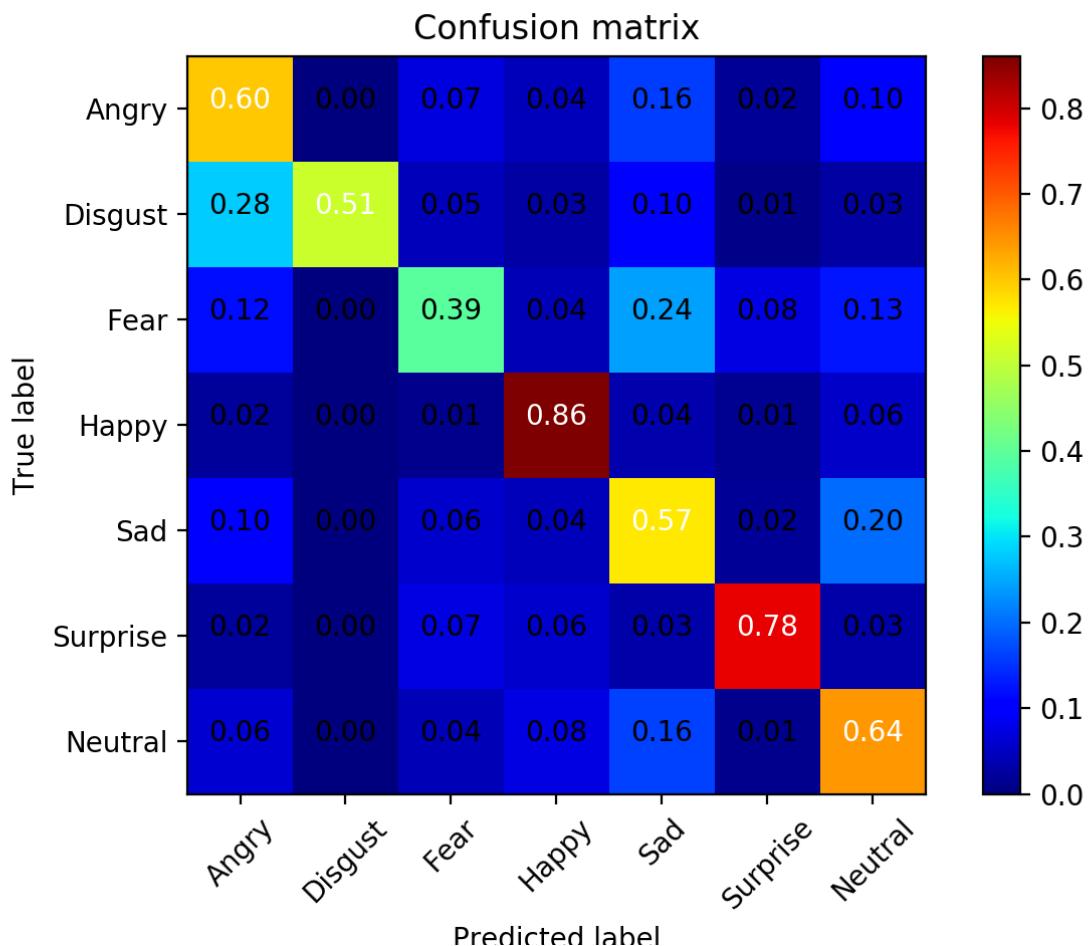
DNN 的訓練時間很快，而 CNN 的訓練時間慢很多

DNN 準確率的上升速度比較慢，而 CNN 準確率上升較快

整體而言，CNN 的準確率較 DNN 為高

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：



Disgust 容易被誤判成 Angry

Fear 容易被誤判成 Sad

Sad, Neutral 兩者容易互相搞混

Sad, Angry 兩者容易互相搞混

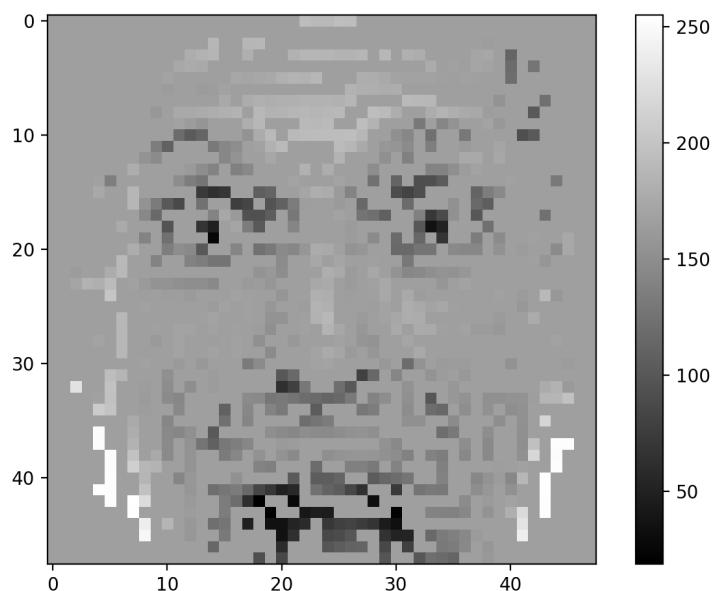
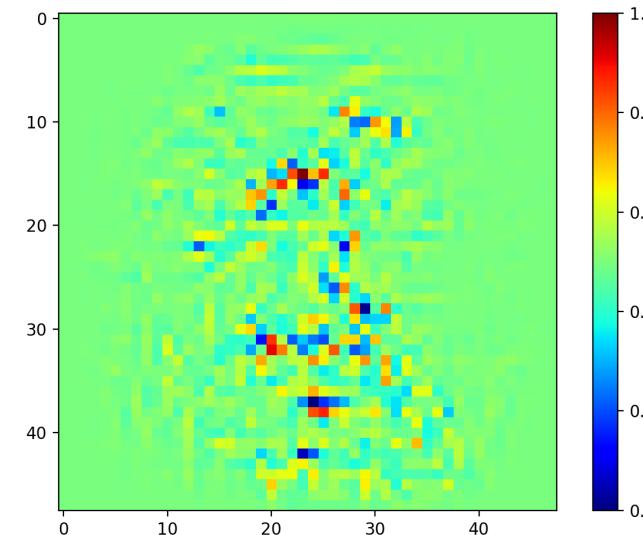
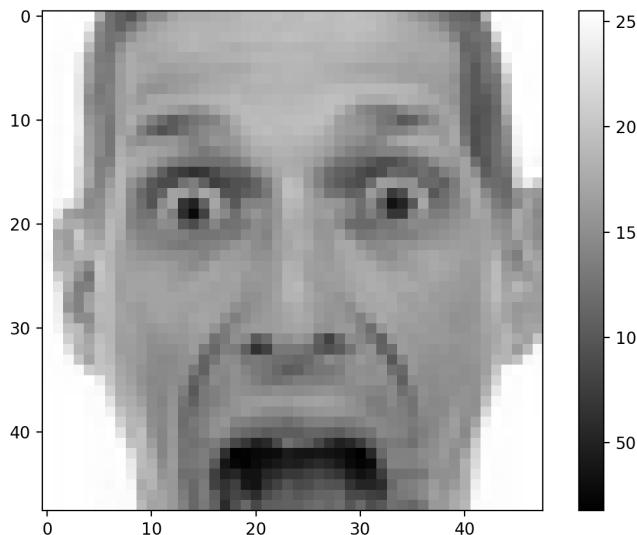
易搞混的表情間有相似的特徵，如 Disgust 和 Angry 會眉毛上揚、嘴巴張開，Fear, Sad 會眼睛較小、眉毛下垂等等。

這個 model 在預測 Fear 時表現最差，只有約 0.39 的正確率

而在預測 Happy 時表現最好，有約 0.86 的正確率

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？

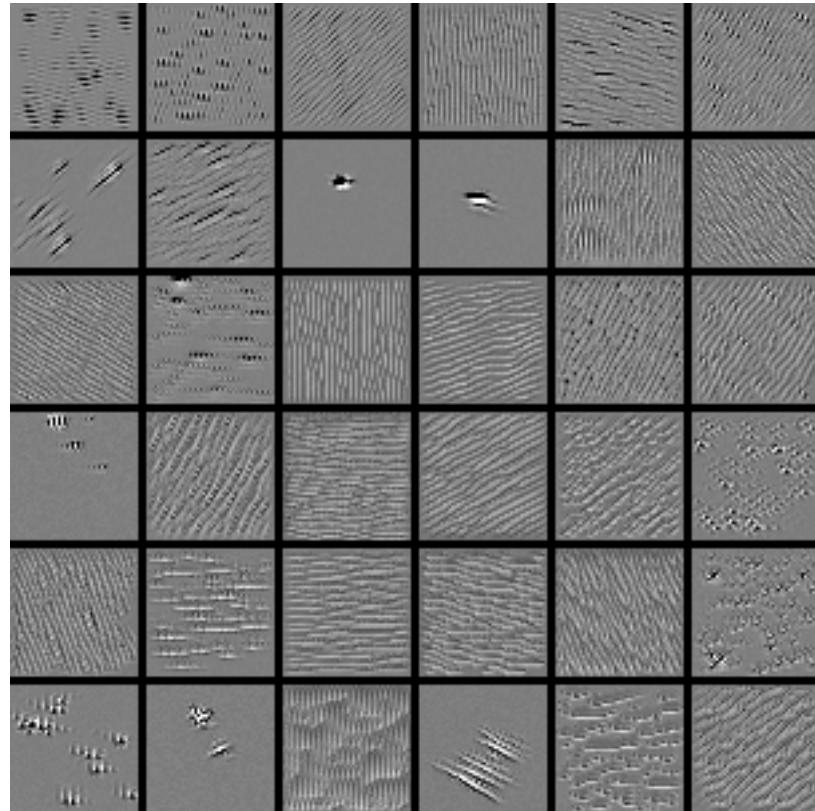
答：



由圖可知，模型主要 focus 在眼睛和嘴巴。

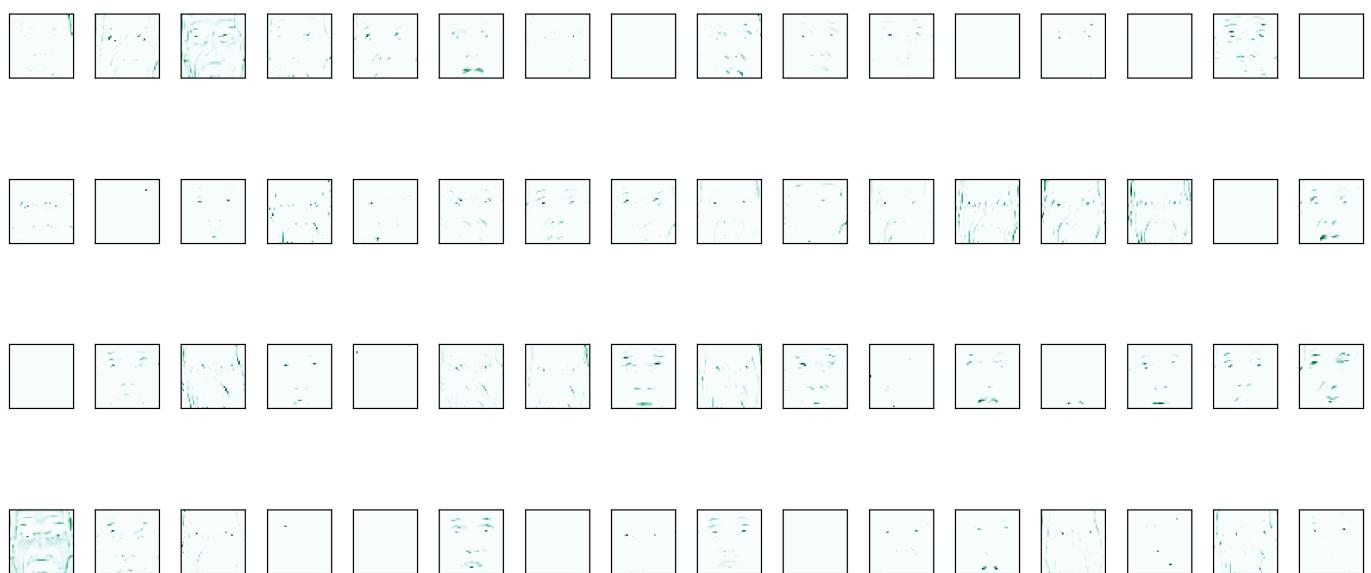
5. (1%) 承(1)(2)，利用上課所提到的 gradient ascent 方法，觀察特定層的filter最容易被哪種圖片 activate。

答：



上圖是 conv2d\_2 中的 filter (擷取其中36個)

可以觀察到，每個 filter 會偏好某種 pattern，例如直的紋路、圓點等等。



上圖是原圖經過 conv2d\_2 的 filter 後的結果（原圖和第四題一樣）

原圖經過 filter 後，每個 filter 會過濾出不同的特徵，例如臉的輪廓、眼睛、嘴巴、全部的五官等等。