

*圖(一)至圖(四)附在最後面

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：圖(一)為我單個模型下能達到最高準確率的 CNN 模型(kaggle 上最佳成績為 ensemble 多個模型)。在 epoch=60, validation_data 4,000 筆的情況下可以達到 valid set: 65.3%, public test:63.9%, private test:64.9%，主要模仿自 K He 的 Residual Network，每兩層 CNN 會有一個來自於前兩層的 shortcut。而此模型總共 Trainable parameter 為 913,095 個。圖(二)為此模型在 training 時候的對於 training data 與 validation data 的 accuracy 變化。

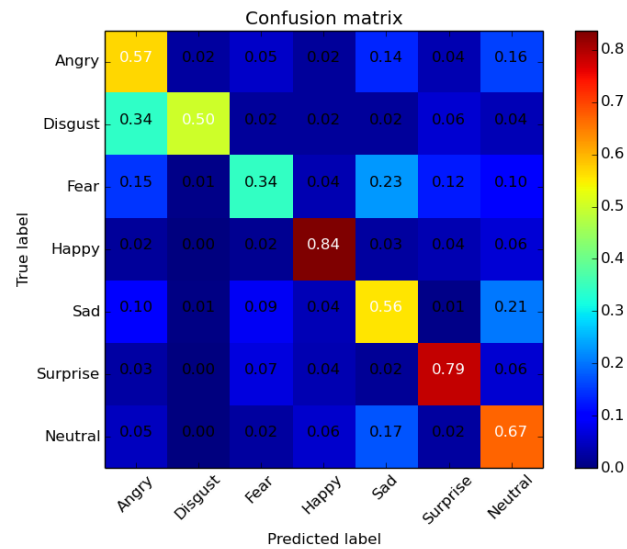
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

答：圖(三)為與上題 CNN 接近的參數量模型(此模型總共的 Trainable parameter 為 908,215 個)。在與上題同樣 epoch=60, validation_data 4,000 筆的情況下可以達到 valid set: 37.9%, public test:36.8%, private test:37.7%。圖(四)為此模型在 training 時候的對於 training data 與 validation data 的 accuracy 變化。我想很明顯地，DNN 的學習效率相較於 CNN 非常低，從圖(二)圖(四)可以看到，在第 10 個 epoch 時，CNN 模型的 train/validation accuracy 已經可以達到 50%以上，反觀 DNN 仍在 30%上下徘徊；且在 60 個迴圈後 CNN 的 accuracy 已經可以達到 65%，而 DNN 仍依舊在 3x%。因此說明參數量並不是絕對與模型準確率成正比，得對於該 task 實施正確的模型建築方式才可以達到事半功倍的效果，否則徒然使用一堆參數下去暴力解也不會有好的成果。

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]

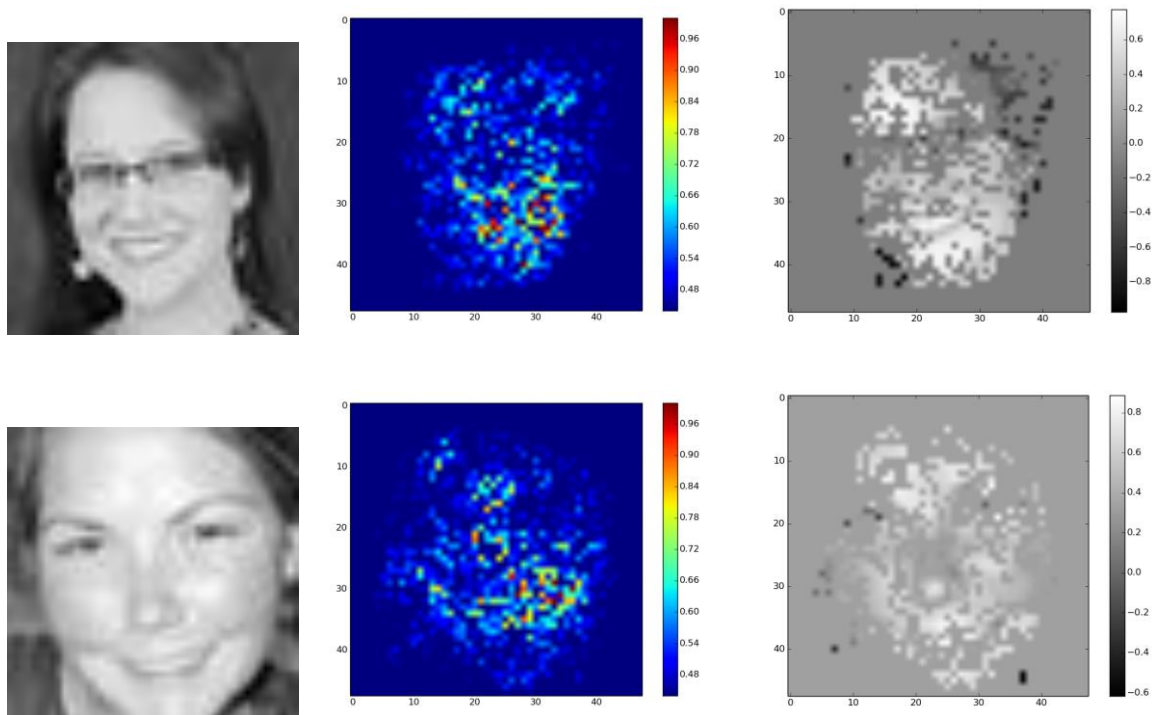
答：從 confusion matrix 中可以發現在我的模型中 Fear 最被難以辨識正確。而 Disgust 最容易被誤認為 Angry，再來則是 Fear 容易被誤認為 Sad，同時 Fear 也較容易被誤認為 Angry。

Confusion Matrix 分析圖



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：



由取樣的兩張 saliency map 可以發現模型在 classification 時主要是 focus 在嘴部、眉心、臉頰等地帶。

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**。

(Collaborators: 周聖荃)

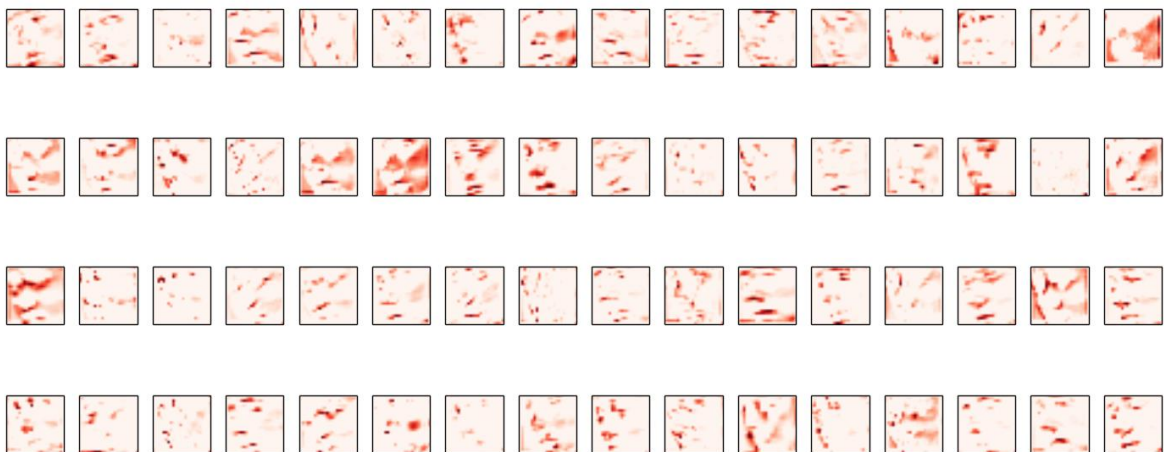
答：

原圖

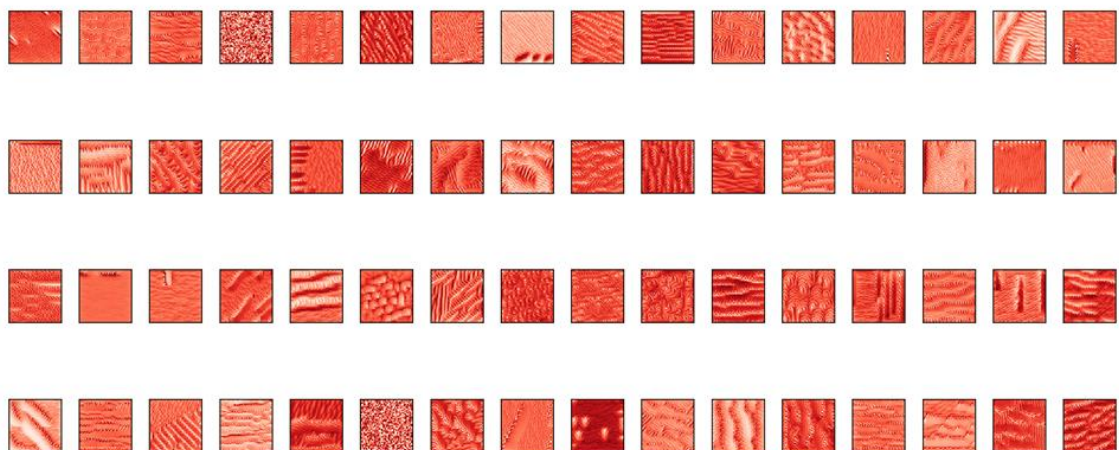


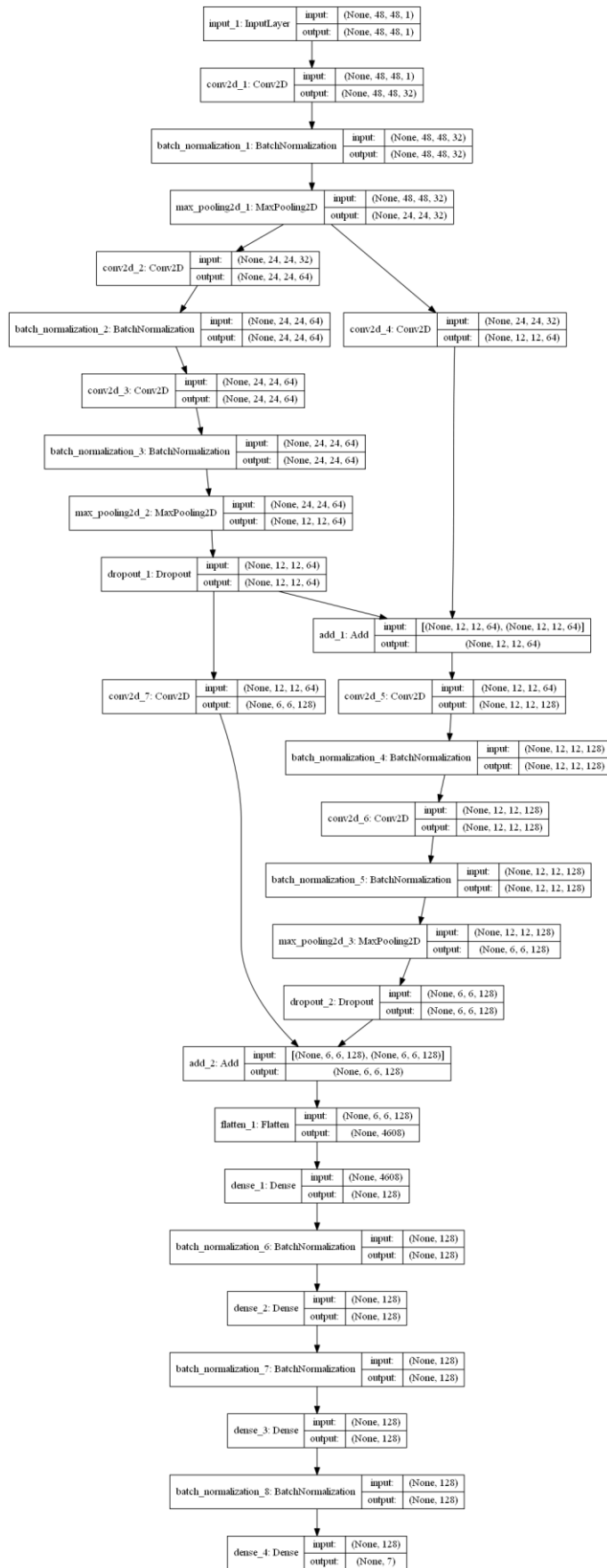
經過 layer: conv2d_2 (64 個 filter)後的圖

Output of layer0 (Given image1)

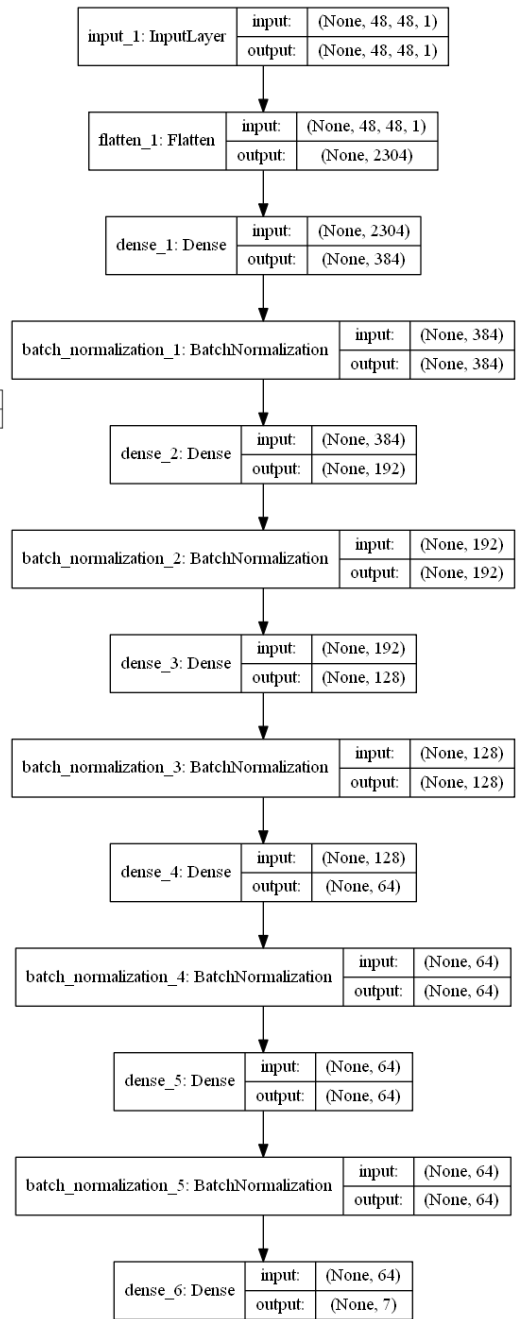


layer: conv2d_2 (64 個 filter)分別最容易被哪種圖片 activate



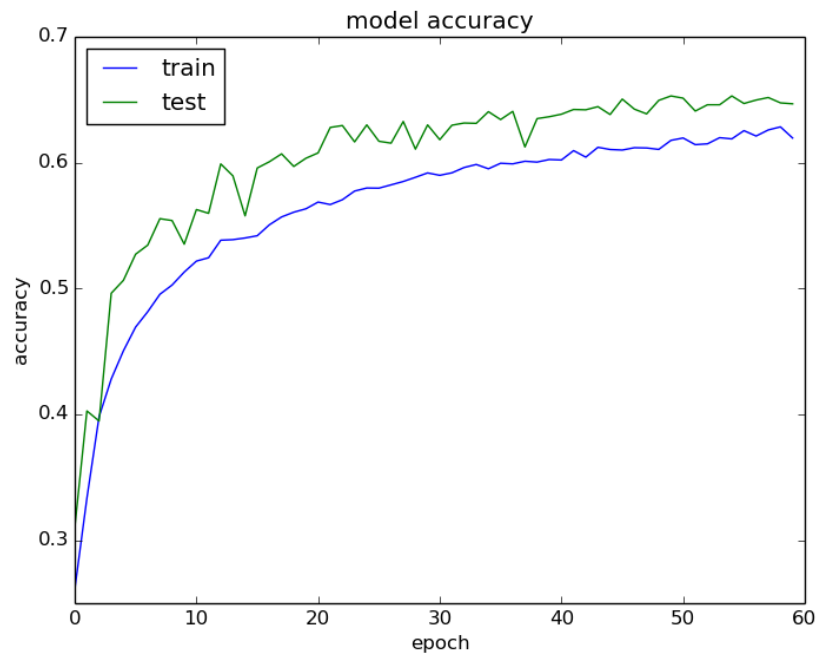


圖(一) CNN 模型架構圖



圖(三) DNN 模型架構圖

圖(二) CNN 模型 training 過程準確率變化



圖(四) DNN 模型 training 過程準確率變化

