

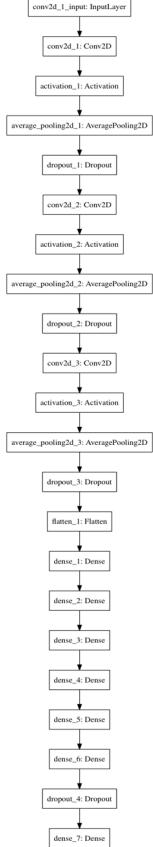
1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

答：

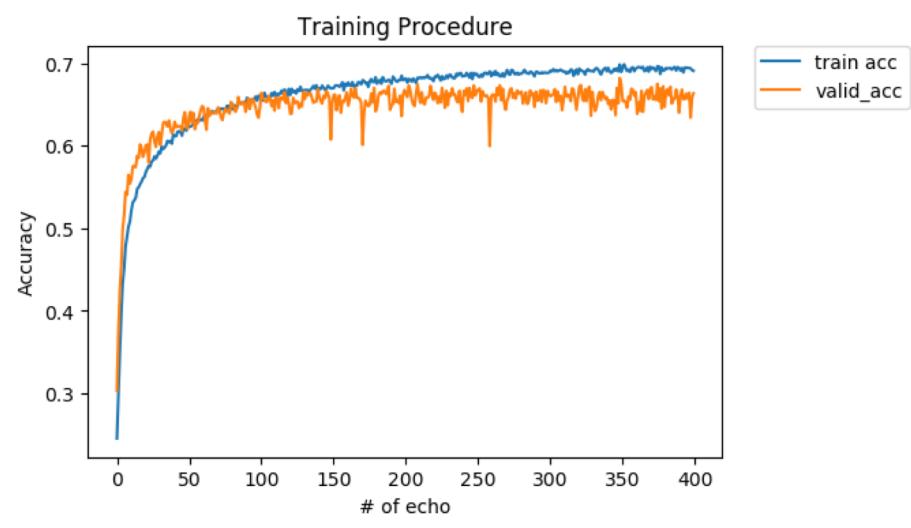
在這次的作業中我嘗試了許多 pre-processing，將 training data 做了 flip, rotation $\pm 1^\circ$, $\pm 2^\circ$, $\pm 15^\circ$ 甚至是 sobel filter 以及參考一篇 paper 採用 LBP[1]，最後的 validation accuracy 以下表整理。上述結果皆無法通過 strong baseline，最後改採用了 keras 的 imageGenerator 去處理 training data augment 後結果有大幅度的進步，如表(1)所示。在處理完資料後，開始調整 model，原本想嘗試 fat v. s. deep，因為既然 Neural Network 可以想像成一個 function，而泰勒展開式只要夠多 neuron 就可以逼近 function，因此我把 model 每層 layer 的 neuron 數設很多，然而效果並不好。最後我將 CNN 以及 DNN 的每層 layer neuron 數目調降並且加深 (3 CNN, 7 DNN) 後 validation accuracy 可以上升至 0.65 左右，最後將 MaxPooling 改為 AveragePooling 以及將 optimizer 從 adam 改為 adadelta 後 validation accuracy 可上升至 0.66。其 model structure 以及 training procedure 如下圖(1, 2)所示。

	flip	rot $\pm 1^\circ$	rot $\pm 2^\circ$	rot $\pm 15^\circ$	sobel	LBP	imageGenerator
val acc	0.58%	0.56%	0.60%	0.53%	0.51%	0.60%	0.64%

表(1). Pre-processing 結果



圖(1). CNN 架構圖

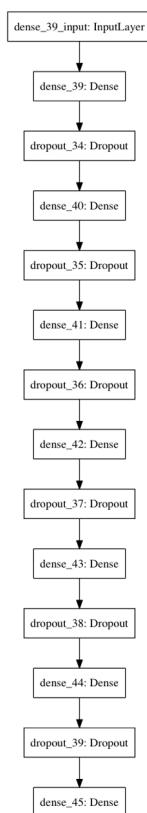


圖(2). CNN History

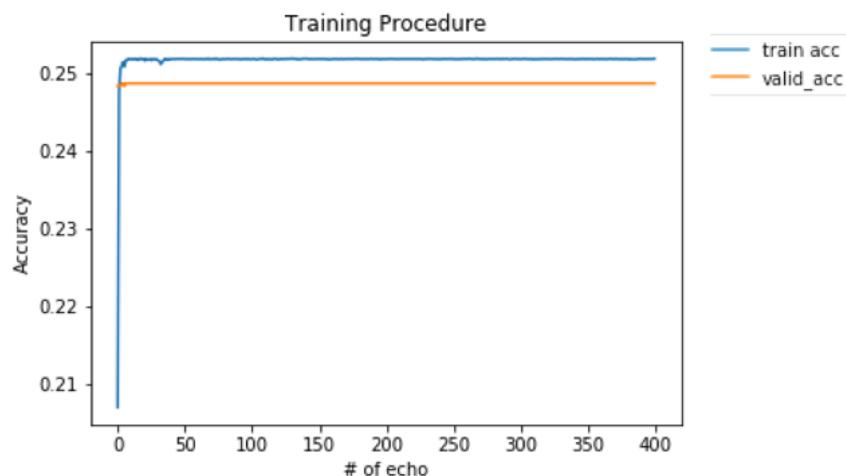
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

答：

題(1)中的model其參數量大約為149萬，本題中實作的DNN參數調整為152萬，一共七層，每層都給dropout(0.25)，在相同的參數量下DNN的training accuracy以及validation accuracy一直都卡在0.24以及0.25上不去，在同樣的epoch情況下顯然CNN的結果比較理想。其model structure以及training procedure如圖(3,4)所示。



圖(3). DNN 架構圖

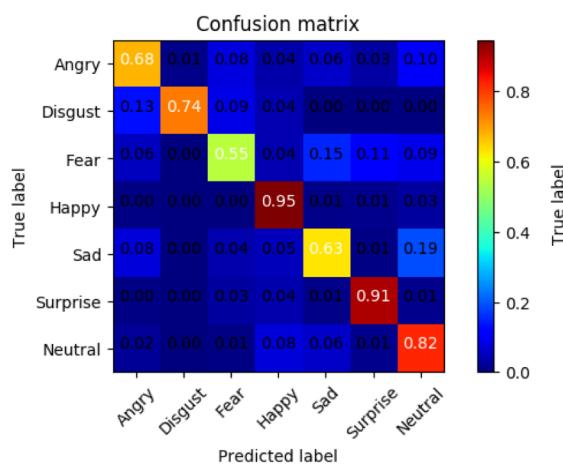


圖(4). DNN History

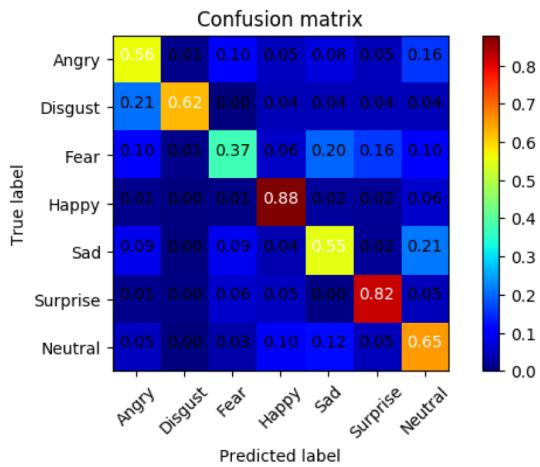
3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]

答：

圖(5)為將 training data 的前 2000 筆資料，圖(6)為 training data 的後面 2000 筆資料（自己切的 validation set），由結果可以看到在 validation data 上的 confusion matrix 結果比較低，不過可以看到一樣的趨勢。我的 model 很容易將 Fear 判斷為 Sad，而 Disgust 也很容易判斷成 Angry，不過這類的情緒是很相近的（都是屬於負面），比較驚訝的是 Happy 以及 Surprise 的嘴巴特徵是相近的（時常張開），然而在 model predict 上的表現還算不錯。



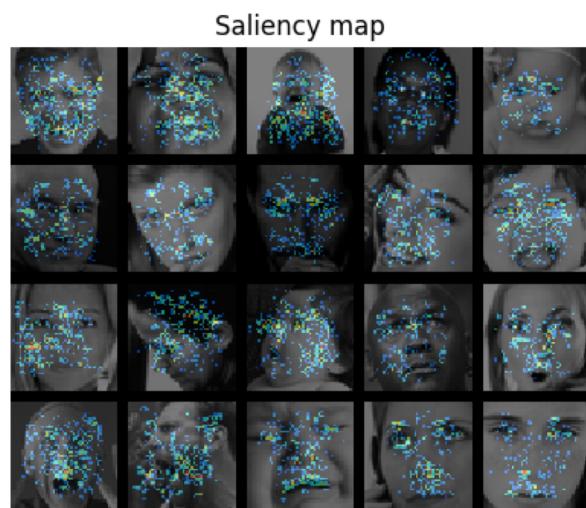
圖(5). Confusion matrix (Train data)



圖(6). Confusion matrix (Validation data)

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：

在一開始寫題目的時候擔心 model 會受到背景影響無法準確抓到人的臉部特徵，還做了許多預處理包括抓取輪廓(sobel filter)，或是採用 LBP 等方法加強 model 對於輪廓的觀察度，但其結果並不理想。透過本題的實作可以發現其實在沒有預處理的情況下仍然能夠抓到臉部的 feature。圖(7)為在第三層 CNN 所輸出的結果，可以明顯的發現 model 能夠針對人的五官去做 classification。



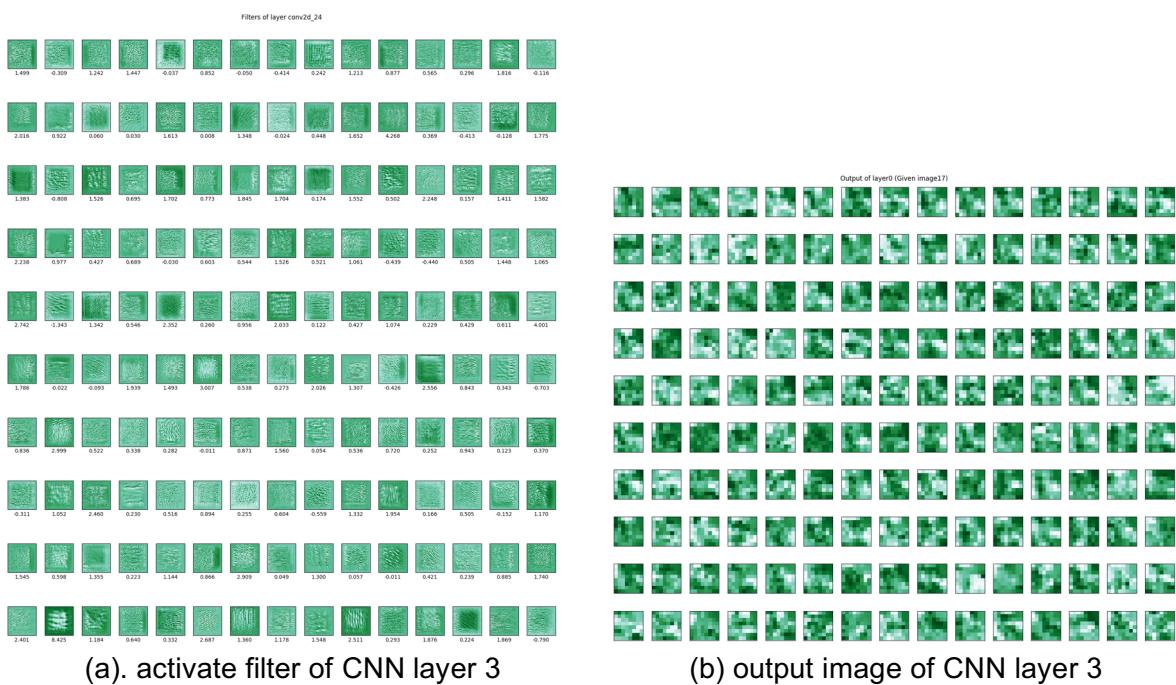
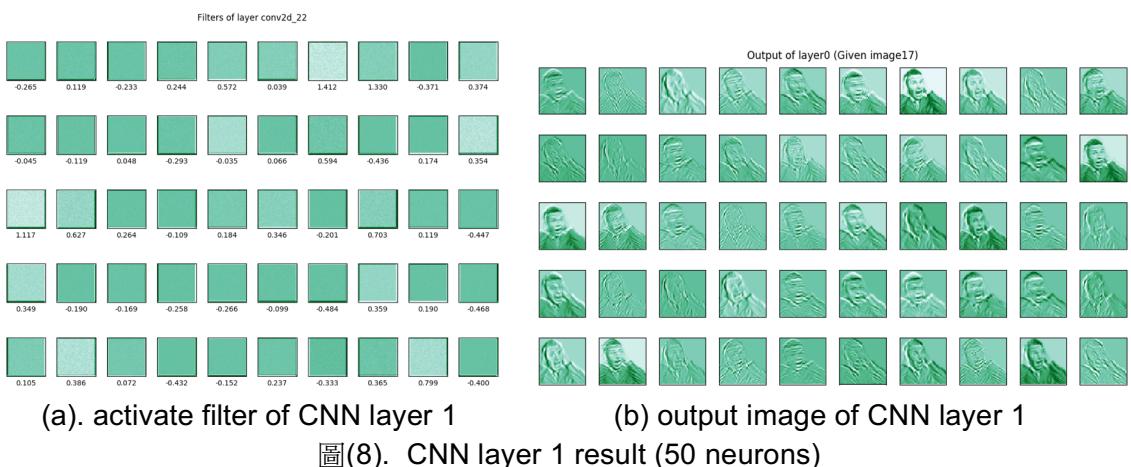
圖(7) 第三層 CNN output 的 Saliency Map

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：

我在本題中觀察了第一層以及第三層的輸出結果，其結果如圖所示。可以發現在第一層的 activate filter 看不出明顯的 feature，且 loss value 也有許多是小於零的，但是在第一層的 image output 可以看到在第一層 layer 後 input image 的背景已經被 filter 濾掉了，如圖(8)所示。

在第三層的 activate filter 可以看出明顯的趨勢，每個 filter 在抓取什麼樣的特徵，例如橫向或是縱向等，然而在通過三層 layer 後其 input image 已經變成 8*8 的 output，因此看起來模糊許多，如圖(9)所示。

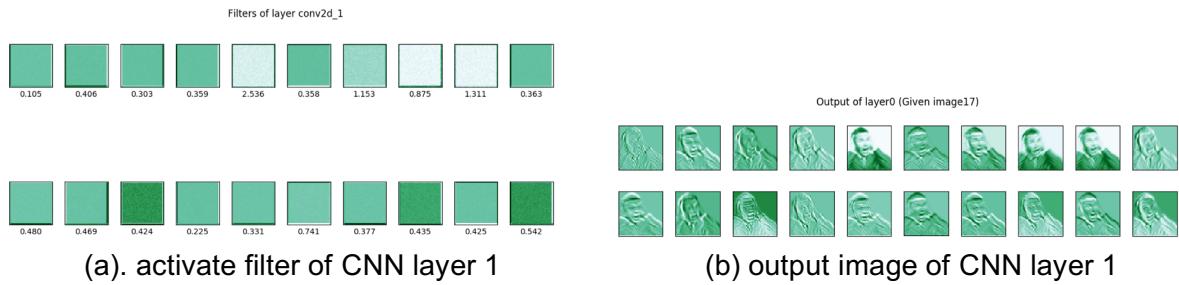


圖(9). CNN layer 3 result (150 neurons)

[Bonus] (1%) 從 training data 中移除部份 label, 實做 semi-supervised learning

[Bonus] (1%) 在 Problem 5 中, 提供了 3 個 hint, 可以嘗試實作及觀察 (但也可以不限於 hint 所提到的方向, 也可以自己去研究更多關於 CNN 細節的資料), 並說明你做了些什麼? [完成 1 個: +0.4%, 完成 2 個: +0.7%, 完成 3 個: +1%]

- 我實作了第 1 個 hint: 利用一個 poor performance model 將他的第一層 layer 的 output image 顯示出來, 而這個 model 的 epoch 只有 10, validation accuracy 只有 0.5 左右, 其結果如圖(10)所示。



圖(10). Poor performance model: CNN layer 1 result (20 neurons)

Reference

- [1] Gil Levi , et al. "Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns" ICMI' 2015. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction Pages 503-510