

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？

(Collaborators: 林哲賢(voting 部分))

CNN 部分：這裡使用 3 層 Convolution：64 個(4,4)，128 個(3,3)，128 個(3,3)。每層皆使用 Valid Padding 和“Relu”函數，Dropout 比例為 0.3。每層 MaxPooling 均為(2,2)。

DNN 部分：輸入層為 CNN 輸出層，1st hidden layer 為 32 個 units，使用函數為 Relu。輸出層為 7 個 units，使用函數為 Softmax。Loss function 使用 Categorical CrossEntropy，優化方法為 adam。

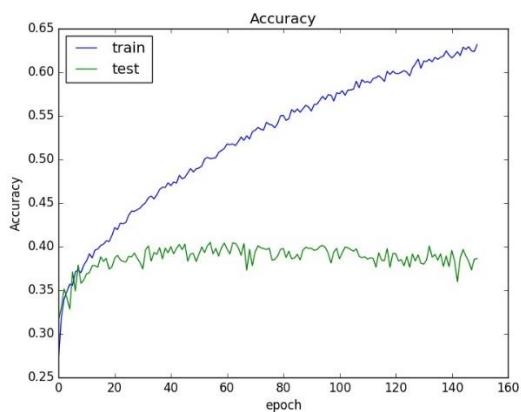
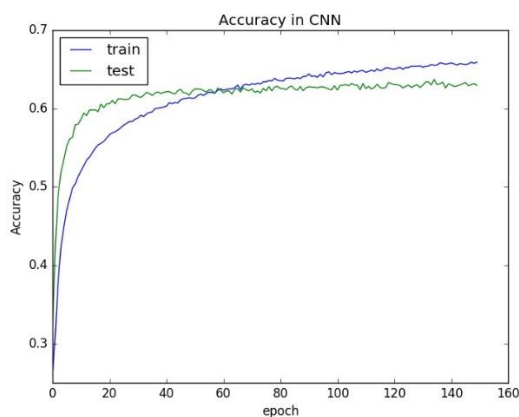
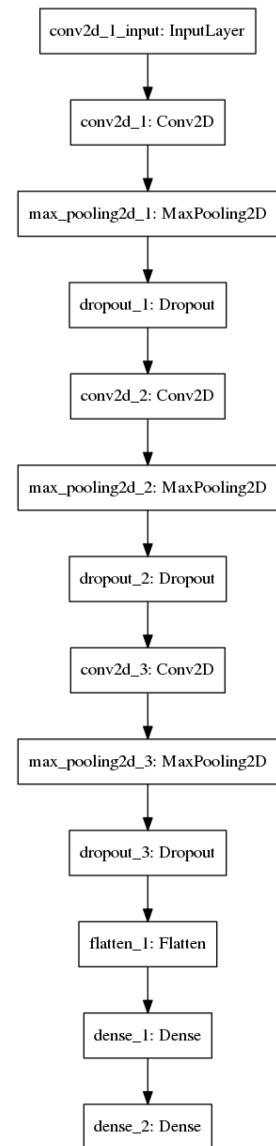
首先將 Training data 的 40%作為 Validation Set，60%用來訓練(約 15,000 張)。再來利用 keras 內建的 ImageDataGenerator 函數，隨機產生 100,000 張經過旋轉、平移、左右翻轉的圖片。再透過 CNN 及 Fully Connect Network 訓練。得到正確率約為 63%

最後，此為所有 model 中正確率最高的單一 model。另外試過，將不同參數的 model 得出來的機率進行相加，取最大值的 label，可以有效將正確率提高為 65.5%。

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？

(Collaborators: 無)

這裡使用的 DNN 架構為：Input layer 為 2304 個單元。Hidden layer 共三層，units 個數分別為 120、48、32，皆使用 relu 函數。Output layer 為 7 個單元，使用 softmax 函數。將參數量相近(約 28~30 萬)的 CNN 和 DNN 做比較，得到 accuracy 和 validation accuracy 如下圖

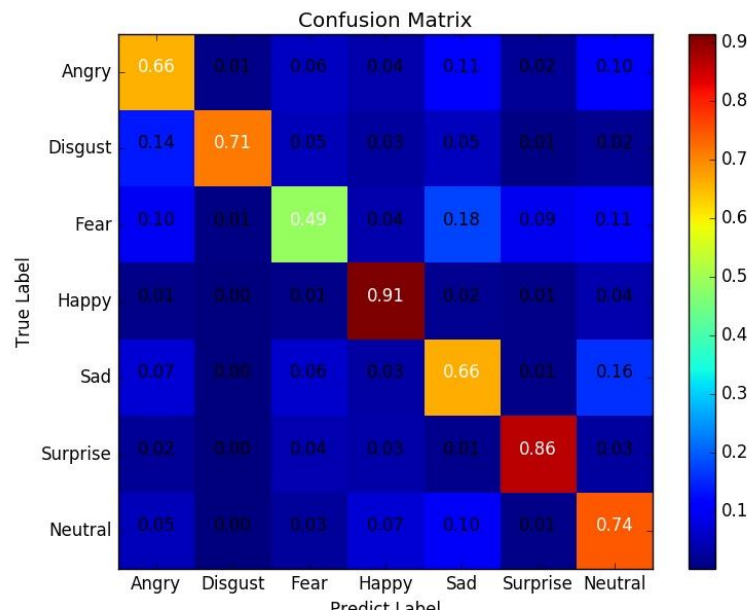


可以明顯得到，當同樣進行 150 個 epoch 時，雖然 training set 的正確率都能達到 65%，但 Validation set 的正確率，CNN 為 60%，DNN 卻只有 35% 左右，明顯差了一截。

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]

(Collaborators: 無)

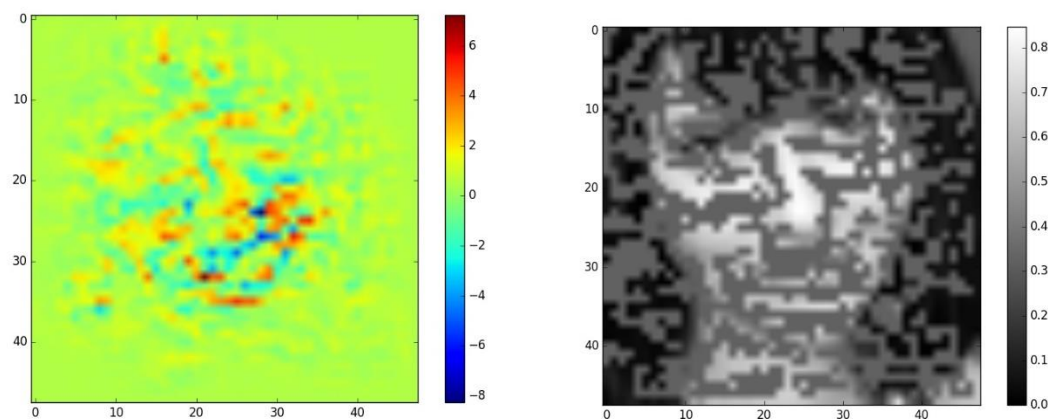
下圖為 Validation Set 的 Confusion Matrix，可以看到 Fear 辨識的正確率最低(只有 49%)。最常被誤認的排名為：Fear 誤認為 Sad、Sad 被認為 Neutral、Disgust 被認為 Angry。



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？

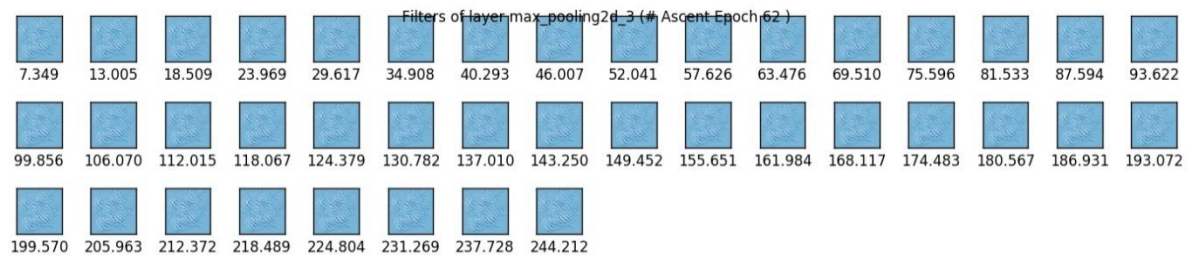
(Collaborators: 無)

由 training set 中取出這張標籤為 happy 的圖片(右圖)，其 saliency maps 如左下，mask 過後的圖片如右下，是將 heatmap 中小於 0.5 的值(有經過 normalize 和 clip)皆設為平均值後，所得的結果。由 mask 過後的圖片，可以得知是由眼睛及鼻子部分判斷其為 happy，但由這個例子，幾乎連全黑的頭髮也考慮進去。



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: 無)

下圖為 **gradient ascent** 方法得到的 **max_pooling2D_3** 層的 **filter**，放大看是一些斜向的紋路。



由上至下分別為第 1 至 3 層 **Convolution** 的 **filter** 做出來的結果。第一層所慮出來的結果多為臉的輪廓；第二層較著重臉部的五官；第三層則較為抽象，比較看不出實體的特徵。

