

HW3 report

學號:B04901009 系級:電機三 姓名:林孟瑾

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

(1) 模型架構: 第一大層為 conv2D+PreLu+BatchNormalization+MaxPooling+Dropout, 接著多加上一層 PreLu, 中間大層為 separable_conv2d+PreLu+BatchNormalization+AveragePooling+Dropout, 再加上 PreLu+MaxPooling, 第三大層為 conv2D+ELu+BatchNormalization+MaxPooling+Dropout。

(2) 訓練參數: 第一層的 conv2D 的 Kernel size 為 (6, 6), 第二和三層的 Kernel size 都是 (5, 5)。Pooling 的部分, pool size 皆為 (2, 2), Dropout 的參數分別是 0.28, 0.35, 0.52。第三大層中的 ELu 層的 alpha 為 0.01。

(3) 準確率: 此 model 在 Kaggle 上的 public score 為 0.70075, private score 為 0.68960

2. (1%) 請嘗試 data normalization, data augmentation, 說明實行方法並且說明對準確率有什麼樣的影響?

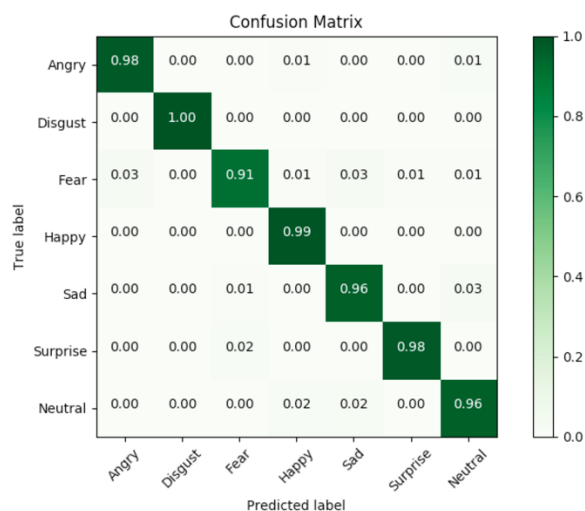
(50 epoch)	Validation accuracy
Normalization 前	0.5180
Normalization 後	0.5481

Normalize (訓練資料減去平均值, 再除以標準差) 後, 準確率明顯提升, 因為如此一來, 訓練的過程中比較不會偏向某一種 data。

(50 epoch)	Validation accuracy
augmentation 前	0.5213
augmentation 後	0.5481

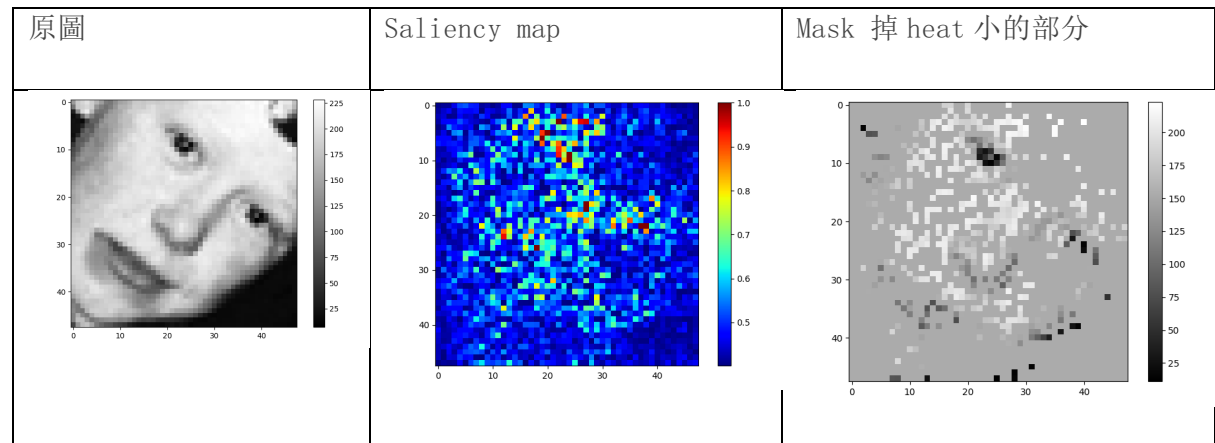
Augmentation 後的準確率也提升了。因為增加了原有圖片的傾斜狀態或位移的變化, 可以增加資料量, 讓電腦更充分的學習其中的特徵。

3. (1%) 觀察答錯的圖片中, 哪些 class 彼此間容易用混?



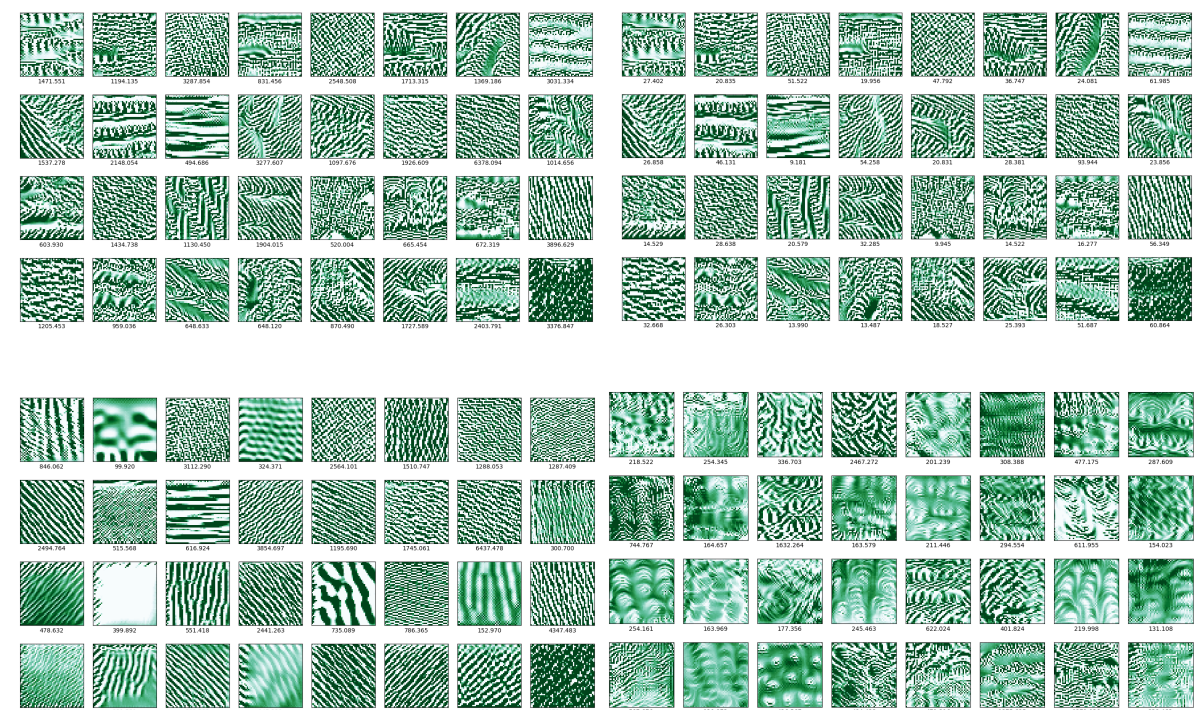
由 confusion matrix 可以看出, Fear 和 “Angry, Sad” 容易搞混, 可能是因為 Fear 的表情其實沒有非常明顯, 不容易讓電腦精確學到 Fear 的特徵。而 Sad 的照片也有可能認成 Fear 或 Neutral, 因為有時候 Sad 的照片其實可能和 Neutral 的表情相似而導致混淆。

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?



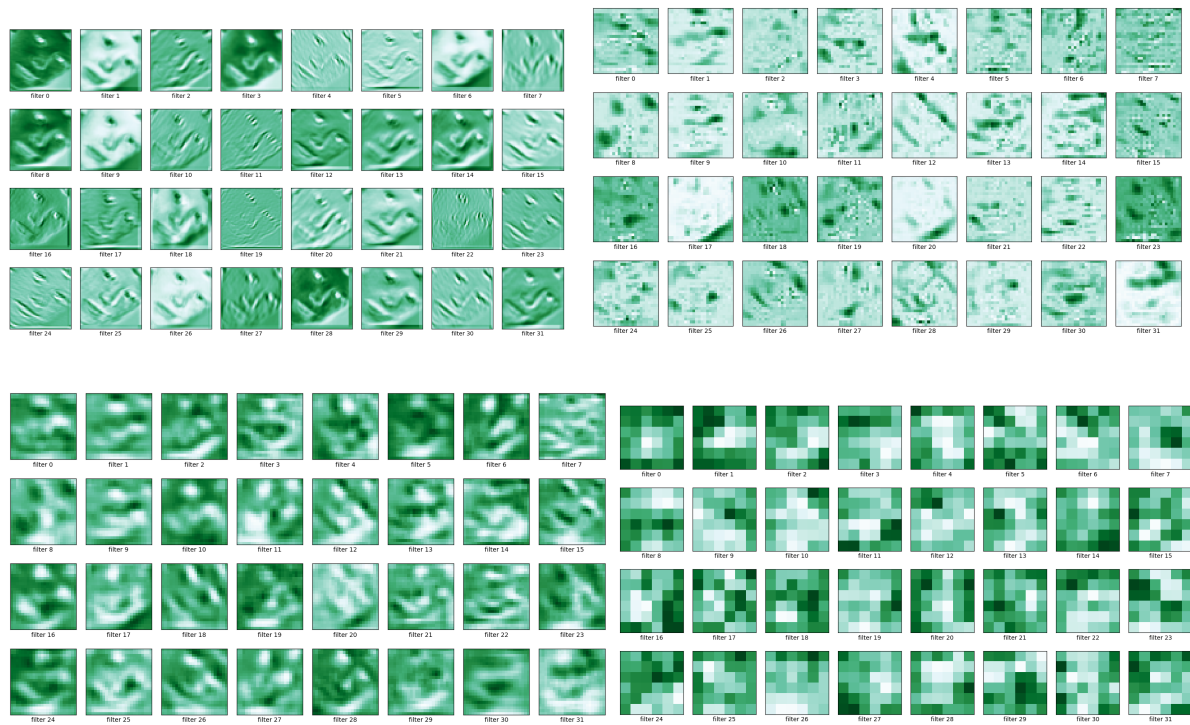
從圖片中可以看出, classify 時是 focus 在五官上 (此圖為眼睛和鼻子), 這張圖的 true label 為 neutral, prediction 也是 neutral。表示 model 在學習時是以五官來判斷分類。

5. (1%) 承(4) 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 與觀察 filter 的 output。



較為淺層的 filter, 容易被較大範圍大區域的輪廓和紋理所 activate, 因為那時候的 filter 所能學習到的資訊大概只有大範圍的。較深入的 filter 則會出現更細節的紋理, 表示這些 filter 可以被圖片中更細微的特徵所 activate, 像是五官之類的部分。

(左上為 PreLu_2, 右上: PreLu_3, 左下: separable_conv2d_1, 右下為 conv2d_2)



淺層的 layer 中，機器所抓到的東西主要是較為表面、不精確的資料，大略的特徵，所以 filter 中的資料都還大概保有原始圖片的樣子，慢慢進入深一點的 layer 時，較為仔細的五官特徵就慢慢的顯現，眼睛鼻子尤其較為明顯。（左上: conv2d_1, 右上: PreLu_2, 左下: separable_conv2d_1, 右下為 conv2d_2）