

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：

A.:CNN 結構如右圖所示，先用 convolution Layer1 將 train data 取出一些較大範圍的部分再經過 maxpooling 降低複雜度之後，之後輸入第二層 3x3 較小的 convolution Layer2 來取出更加細微的部分，而後輸入 120-80 的 DNN 網路 train 出 model，全部的參數約 200,000 左右。

B.:再選用 input 的部分使用前 20000 筆資料，載 train set 的部分得到的準確率為 82%左右，在 test set 的部分得到的正確率為 58%左右，設定的參數為 batch=100,epoch=10。

25xConvolution Layer(4x4)

Maxpooling(2x2)

50xConvolution Layer(3x3)

Maxpooling(3x3)

Flatten Layer

Nueron(units:120)

Activation(relu)

Nueron(units:80)

Activation(relu)

Softmax(units:7)

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：

A. :DNN 結構如右圖所示使用 2x2Maxpooling 後加入 dropout 後輸入 500-250-100 的 neuron 結構中，用 batch=100,epoch=20，因為在使用 DNN 時發現同樣使用 10epoch 整個 model 得到的 train set 正確率只有 23%左右，推測為使用的 iteration 不夠所以後來把 epochs 增加到 20，而後準確率也確實增加到 30%左右，載 test set 也增加到了 20%左右，再選用結構部分原本是用的是只有兩層 neuron 的結構總參數調整為 200,000 左右但是整個準確率非常的低，所以最後改為三層並把總參數調整到 300,000 左右，DNN 因為少了 convolution Layer 整體 train 的

Maxpooling(2x2)

Dropout(rate:0.4)

Flatten Layer

Nueron(units:500)

Activation(relu)

Nueron(units:250)

Activation(relu)

Nueron(units:100)

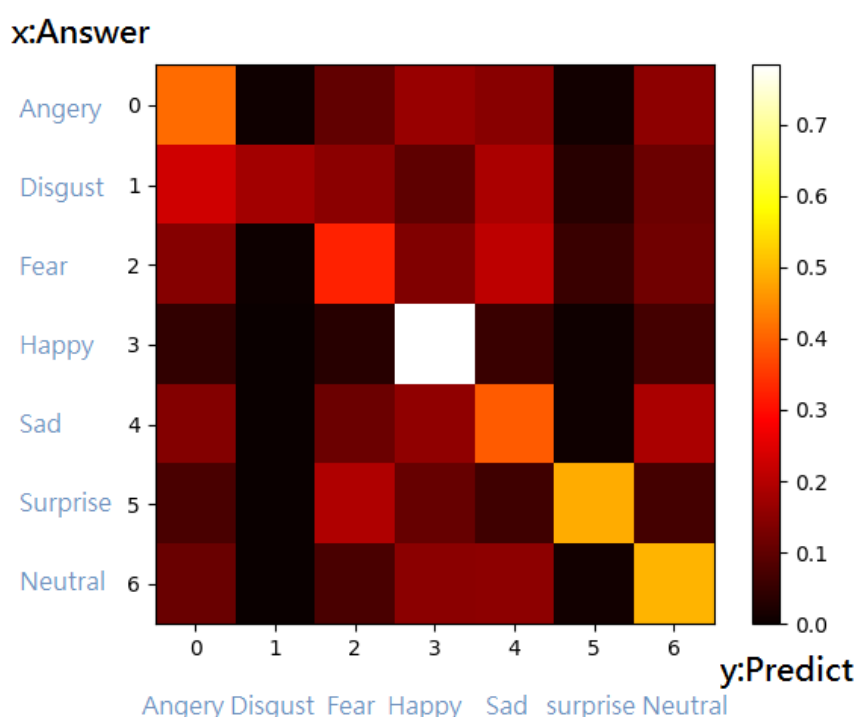
Activation(relu)

Softmax(units:7)

時間比 CNN 快 4-5 倍。

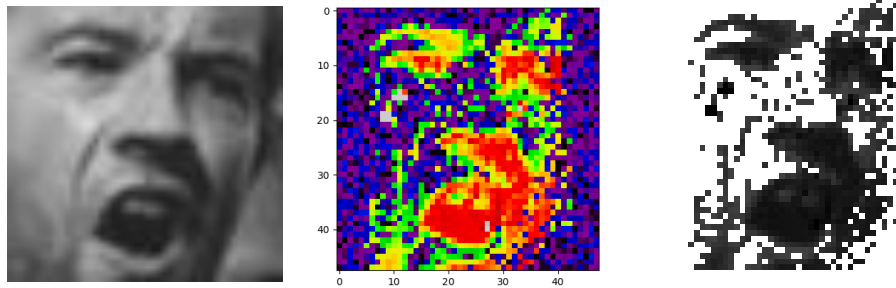
B. :在使用 DNN 時可以觀察到準確率下降的非常多，我認為最主要的原因還是因為少了 **filter** 所以在 **train** 的時候因為是整張圖輸入會導致需要更多的 **epochs** 才能得到比較好的正確率,在判斷方面因為圖片中都有一些小部分可能是判斷的重點,所以在沒了 **filter** 過濾出叫小的部分會導致 **model** 符法精準的找到判斷的重點,造成正確率降低的非常明顯。

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]



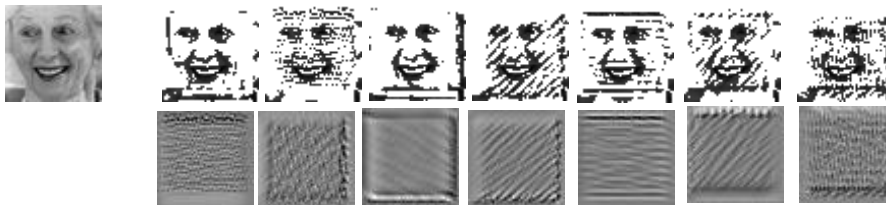
答：這是使用後 20000 筆所得出的 confusion matrix,可以看到在 Happy 的部分正確率非常的高可以達到 80%左右，在 Label3,5,6 的地方可以看到 model 預測的準確率比較高,其中可以看到在 surprise 的部分較容易與 Fear 混淆,其他 1,2,4 的預測準確率都比較差只有在 0.3~0.4 左右而以,其中比較值得注意的地方 Disgust 可以從顏色中發現 Disgust 預測為 Angry 的機率剩至比預測正確的機率還高，在預測錯誤的部分後來去看圖片之後發現有些圖片取的角度,年紀,性別不同可能是因為這些原因導致 model 預測錯誤。

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 saliency maps 可以非常明顯的看到 CNN 可以把焦點聚交在五關處像是眼睛以及嘴巴，比起 DNN 直接是輸入整張沒有經過整理的圖片可以得到更好的結果，這個判斷感覺也比較接近我們平常在判斷一個人的表情時所會注重的表情的，從這個地方來講 CNN 的準確率較高是有道理的。

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：這些是從第一層 50 個 filter 中取出的圖形使用 gradient ascent 所做出來的圖中選出比較容易看得圖，從這些圖中可以很明顯得觀察到不同的 filter 可以取出特定方向或是特徵的圖形，後面的 filter 就可以在由前面 filter 的書出可以進一步的把更加細緻的 feature 取出。

[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%]