

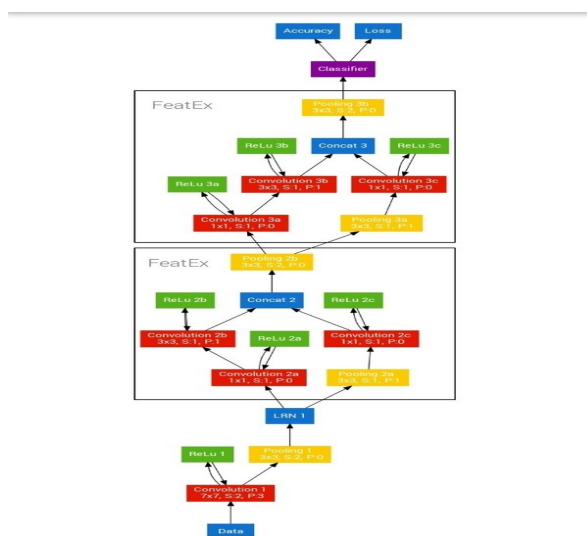
Homework 3 Report - Image Sentiment Classification

學號：b05902008 系級：資工二 姓名：王行健

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？

epoch_num : 10000
learning_rate : 1e-4
batch_size : 100

model : “Dexpression : Deep Convolution Neural Network for Expression Recognition” 的 FeatEx 以下為 Feat net 示意圖(引用自原論文)，最後架構為分別 train 三個 Feat net，將輸出接起來再以三層 dense layer aggregate



accuracy : 對於每個 feat net, validation accuracy 67%~70%
三塊 aggregate, validation accuracy 71%~73%
差別可能來自於不同的初始狀態

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？

data_augmentation : 隨機水頻翻轉，左右旋轉5度以內，上下左右移動0.1，縮放0.1

data_normalization : 將圖片縮到以0為平均，1為標準差

plain [vald_acc 63%~65%]

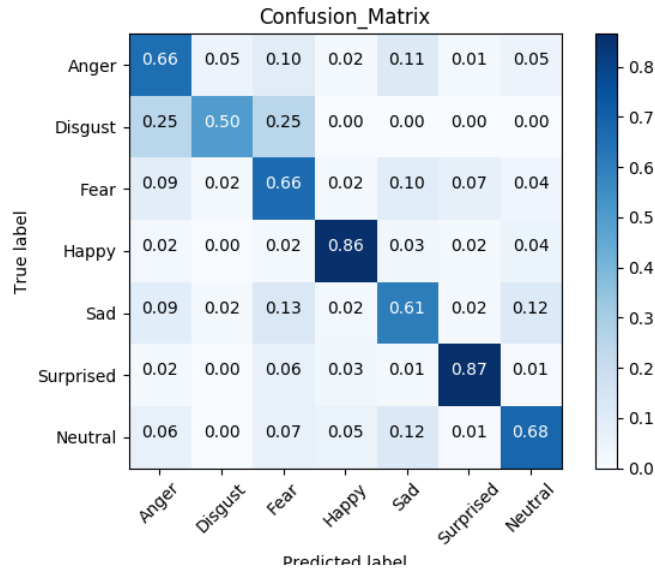
data_augmentation [vald_acc 67~70%]

data_normalization [vald_acc 64~66%]

data_augmentation + data_normalization [vald_acc 67~70%]

在什麼都不做的情況下，overfit非誠嚴重，加入data_augmentation後則使fit變困難，也有效提高validation accuracy。data_normalization則沒有顯著效果

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]

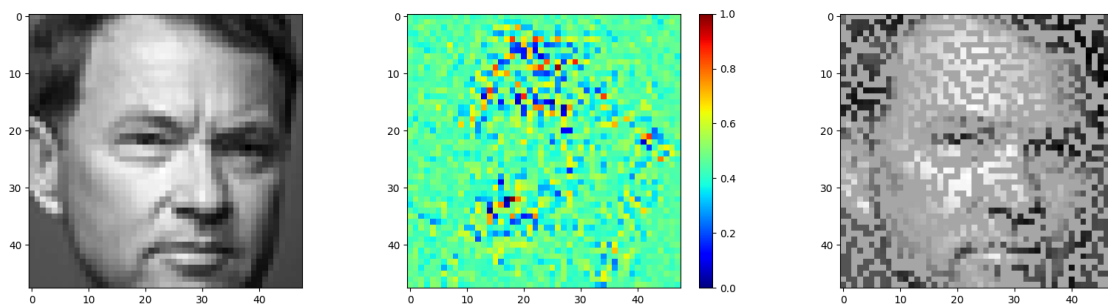


正確率最低的是Disgust，這可能跟其數量有關

Anger, Fear, Sad三者易互相混淆

Neutral會被誤會成Disgust以外的任何東西

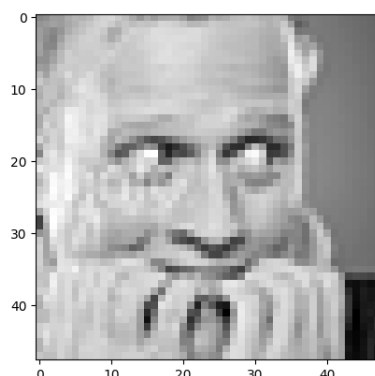
4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？



這是第一個FeatNet的結果，可以看出他姑且能抓到眼睛、嘴巴、鼻子的位置。而這些特徵的確可以幫助判斷一個人的表情。至於其為何不能完全對準這張圖片的關鍵部位，則可能是因為其除了便是這張圖以外，也需要具備辨識其他經過augmentation的圖的能力，因此只能概略性的抓出重點。

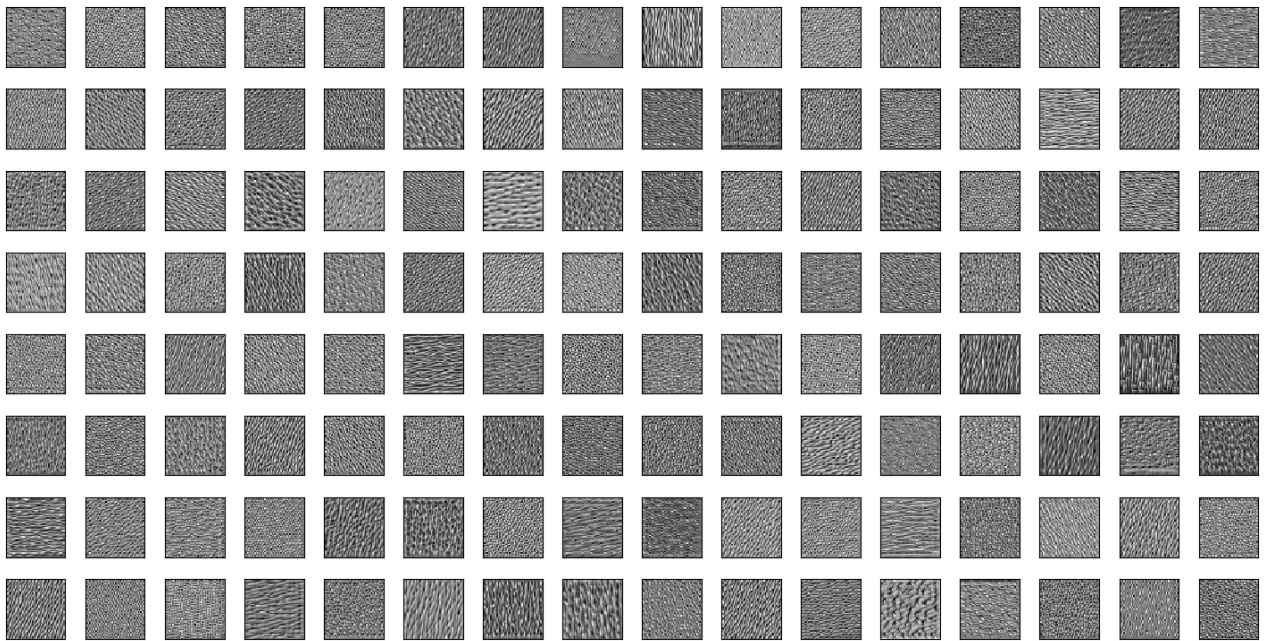
5. (1%) 承(4) 利用上課所提到的 gradient ascent 方法，觀察特定層的filter最容易被哪種圖片 activate與觀察filter的output。

original image





可以看出有部分網路已經抓到眼睛以及嘴巴(position 0,9 1,5 ...), 而另外亦部分則專注於額頭或手部(position 1,10 4,8 ...), 這顯示cnn的確有助於抓出圖片的區域特徵, 而不同filter也可以有效分別處理不同的細節



activation image包含了各種方向的條紋, 可能對應到臉部皺紋, 而各種方向可能來自原本圖片中頭部方向本來就不整齊。另外, 也可以看到一些偏橢圓或長紡錘的結構, 有可能對應到眼睛及嘴吧。和手部有關的形狀, 但應該只是人眼無法辨識而已。