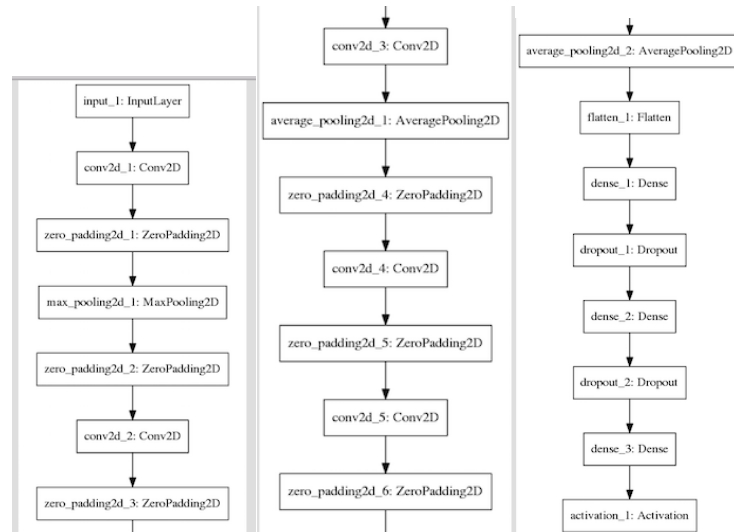


學號：R06922113 系級：資工所碩一 姓名：陳宣伯~

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？

✧ 模型架構：

為求排版方便以下圖順序為由左而右，由上而下。



✧ 訓練過程與辨識率：

■ 資料前處理

◆ 標準化

訓練資料中的圖都經過標準化(normalize)，使每張圖的 pixel 的值都介於-1 到 1 之間

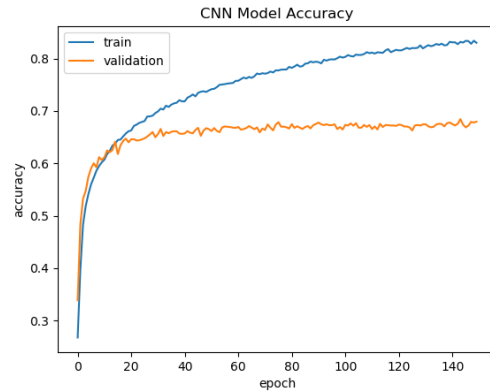
◆ ImageDataGenerator

再將前述經標準化的訓練資料，丟進 keras 提供的 ImageDataGenerator 去產生旋轉(rotation)、翻轉(flip)等等的新的資料，詳細參數如以下：

```
def gen_imgGenerator(X_train, Y_train, batch_size):
    train_datagen = ImageDataGenerator(
        rotation_range=10,
        width_shift_range=0.10,
        height_shift_range=0.10,
        shear_range=0.10,
        zoom_range=0.10,
        horizontal_flip=True,
        fill_mode='nearest')
```

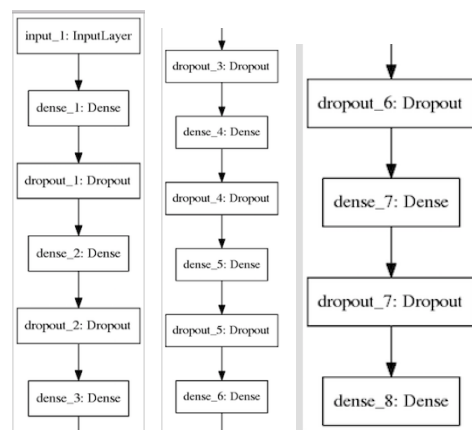
■ 訓練過程

我總共跑 150 次 epoch，每次 batch size 為 256。下圖為 accuracy 對 epoch 做圖的結果，由圖可知大概跑 20 個 epoch 的時候 validation set 的辨識率已經接近飽和(約 0.675)



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？

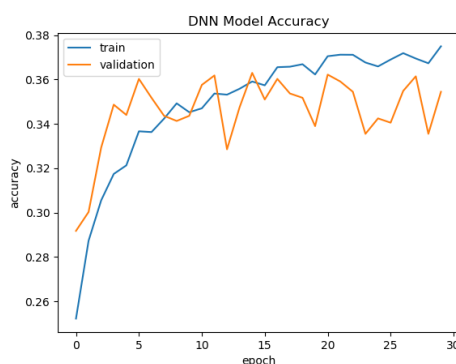
✧ 模型架構：



為求排版方便，上圖順序為由左而右，由上而下。而我在第 1 小題的 CNN 參數個數為 4,631,559 個，本題 DNN 參數個數為 4,467,975

✧ 訓練過程與辨識率：

可以很容易看出大概在跑到 10 次 epoch 時，validation set 的辨識率便已呈現飽和，約 0.35 左右，所以 CNN 的辨識率幾乎是 DNN 的一倍



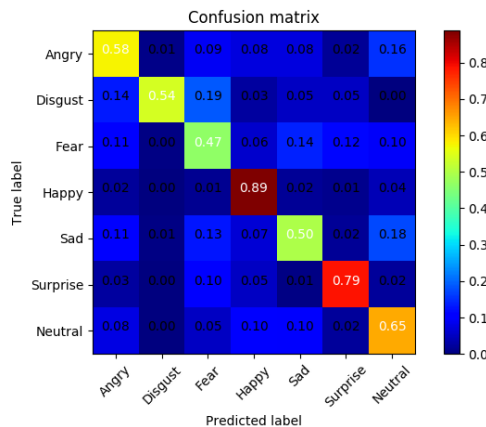
✧ 觀察：

由上圖可見，DNN 訓練到 5 個 epoch 左右就開始收斂，但最後跑完的結果以

辨識率來說還是 CNN 明顯較佳。我認為雖然 DNN 與 CNN 的參數都一樣，但是 CNN 有著老師上課時提到的三種特性，分別為某些 pattern 在一張圖中站的範圍很小、相同的 pattern 會出現在圖中不同區域、subsampling 不會影響整張圖的辨識，光是考慮到圖中各個 pixel 的相對位置這點，就足以說明 CNN 比 DNN 更適合用在圖片分析上

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]

Confusion matrix 如下圖，觀察到難過(Sad)與恐懼(Fear)之間很容易搞混。



☆ 觀察：

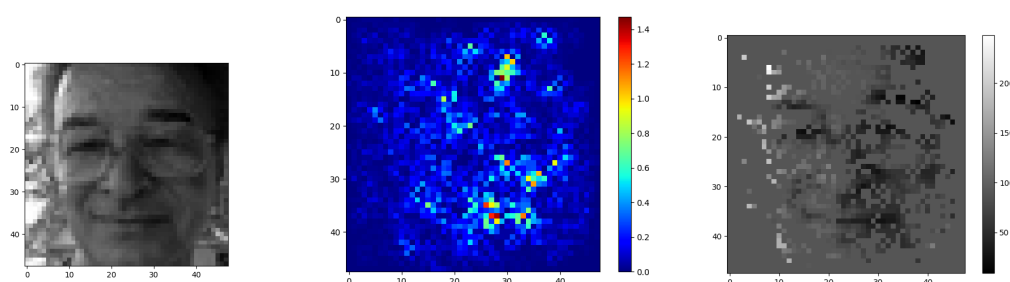
以下列出該分類辨識錯誤比例最高者：

- Angry(生氣)有 16%被錯分到 Neutral(中立)
- Disgust(厭惡)有 19%被錯分到 Fear(恐懼)
- Fear(恐懼)有 14%被錯分到 Sad(難過)
- Happy(高興)有 4%被錯分到 Neutral(中立)
- Sad(難過)有 18%被錯分到 Neutral(中立)
- Surprise(驚訝)有 10%被錯分到 Fear(恐懼)
- Neutral(中立)有 10%被錯分到 Happy(高興)及 Sad(難過)

以下討論分錯比例最高的兩者，Disgust 與 Sad。前者我想當一個人有厭惡感的時候，同時也會有害怕的情緒，例如厭惡吃紅蘿蔔。後者我想一個人難過的時候不一定是表現強烈的，有時候表面上看不出來但其實心裡很難過，這時就會跟面無表情的情緒混淆，以上純粹是我的猜測，與機器學習無關。

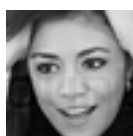
4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？

我選用 id=24 之圖片(GT=3)，如下由左至右為原圖、saliency map 與 mask



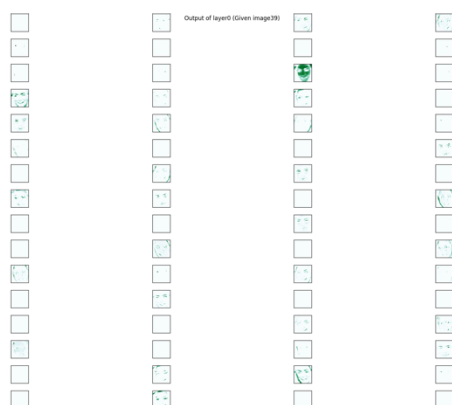
可以看到在做分類時主要是 focus 在人臉的五官上，例如眼睛、嘴巴或眉毛等地方，以本張圖片來說尤其在嘴巴的部分特別明顯。

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。

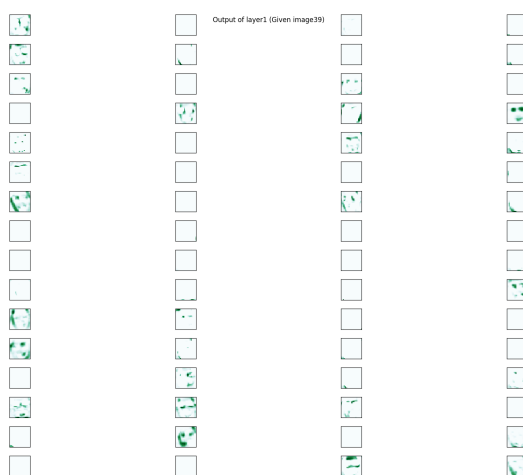


輸入 id=39 的圖片(GT=5)，，可以觀察到 Layer0 的第 3 行、第 3 列的第 35 個 filter 很容易被驚訝的圖片 activate，如下圖

Layer0:



Layer1:



Layer2:



Layer3:

