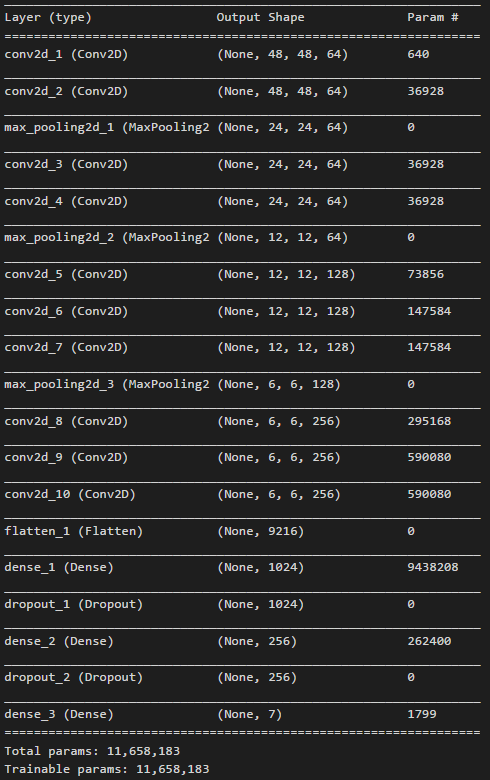
學號：B04901060 系級： 電機三 姓名：黃文璁

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: 無) **答：**

CNN 架構基本為：2至3層conv2d後接一個max\_pooling2d（ conv2d含zero padding）。重複數次直到圖片縮小到一定程度（本模型為6x6），再接上兩層fully-connected layer，dropout皆為0.5，最後是7個類別的softmax。注意到越深層的conv2d有越多filter。

另外由於training data（扣除validation）大約25000筆，相對較少，故在訓練時還使用了Keras內建的ImageDataGenerator來進行data augmentation，使用了旋轉、平移、縮放和水平翻轉。

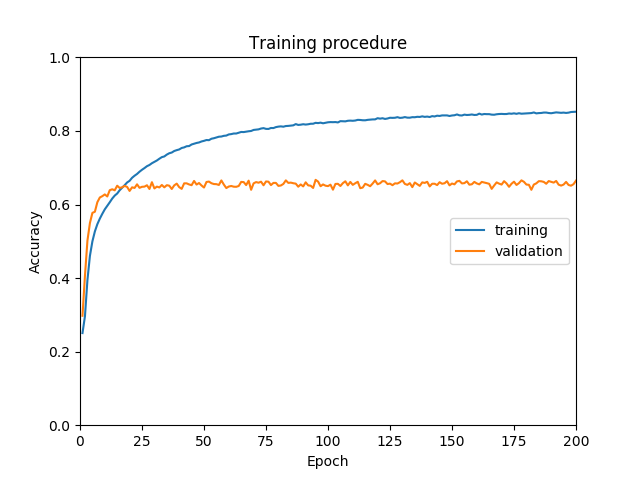
接著可以觀察Training procedure，整體來說，在validation accuracy達到收斂後，training accuracy也逐漸收斂，相較於接下來的DNN來說較不容易overfit。

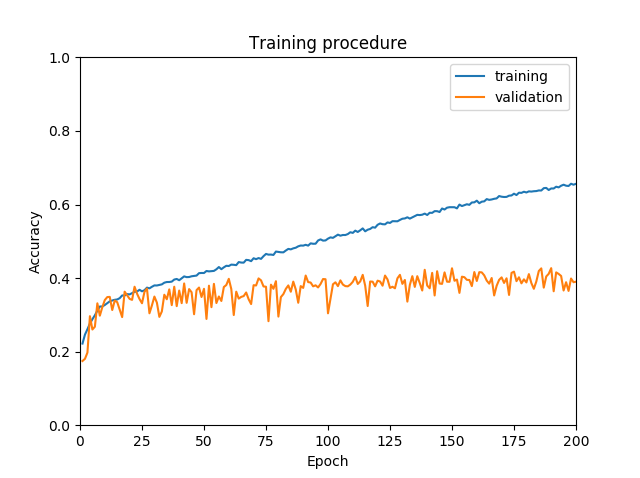
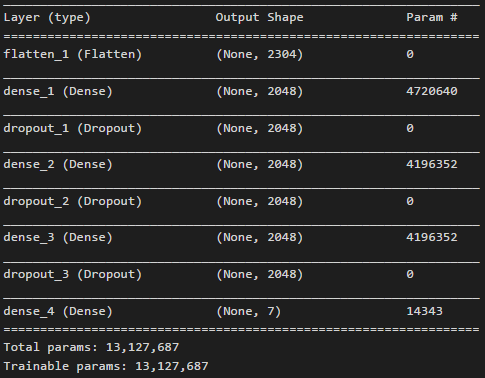
最後是準確率的部分，根據Kaggle：

- Public score: 0.67623

- Private score: 0.68375

皆有通過strong baseline。



1. **(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) **答：**

上題中的CNN約有1200萬個參數，本題使用接近參數的DNN。

DNN架構為三層2048個unit的fully-connected layer，分別都有0.5的dropout。最後同樣是7個類別的softmax。

觀察Training procedure，可以發現training accuracy在200個epoch後仍持續上升，但validation早已收斂，此外也可以觀察出validation accuracy較為浮動，可以猜測DNN相對於CNN更容易有overfitting的情況。另外注意到DNN的validation大約收斂在40%，和CNN的68% 有很大差別。

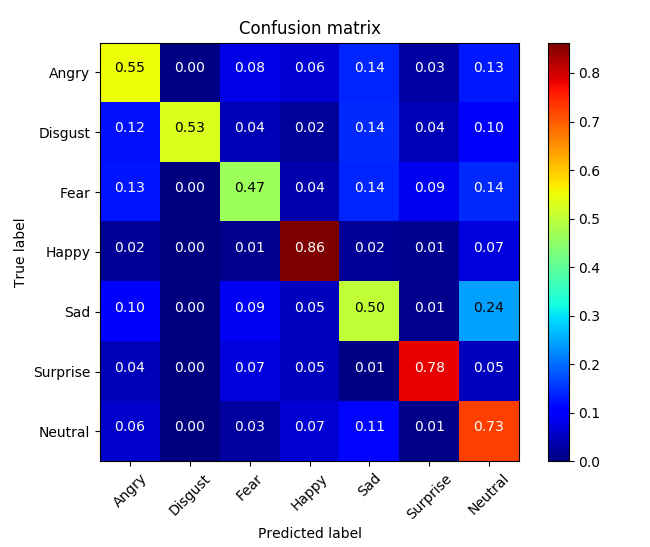
最後是準確率的部分，根據Kaggle比CNN差了約25%：

- Public score: 0.38701

- Private score: 0.40930

1. **(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)**

**答：**

左圖利用10% 資料繪製confusion matrix。

不容易預測錯的表情為Happy和Surprise。

而最容易預測錯的是將Sad預測成Neutral。

此外容易預測錯的還有：

將Angry預測成Sad、Neutral

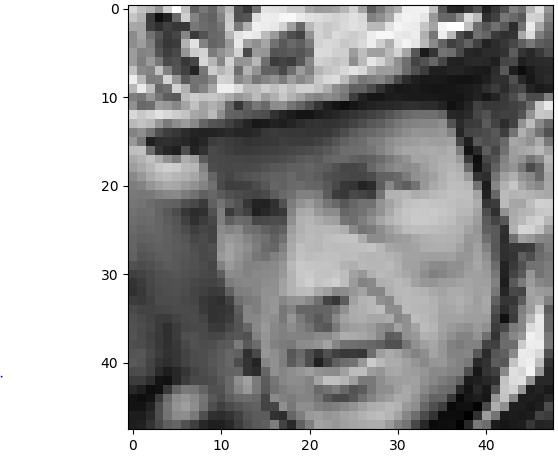
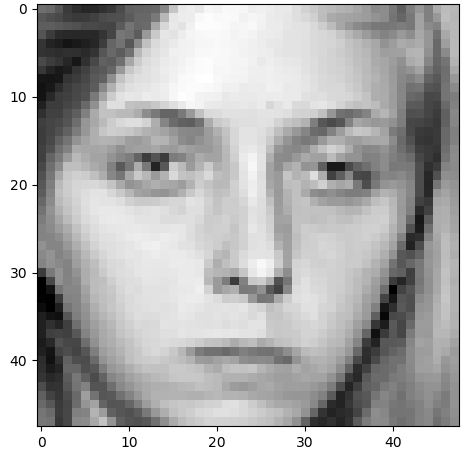
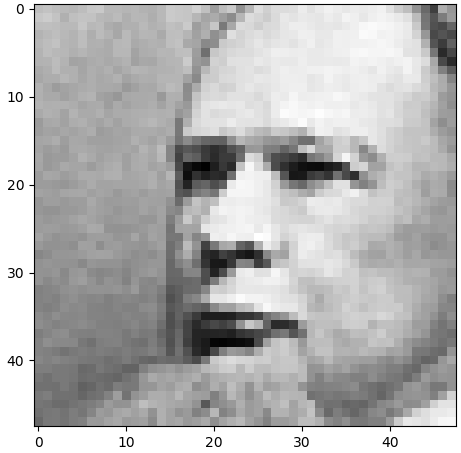
將Disgust預測成Angry、Sad、Neutral

將Fear預測成Angry、Sad、Neutral

將Sad預測成Angry、Neutral

將Neutral預測成Sad

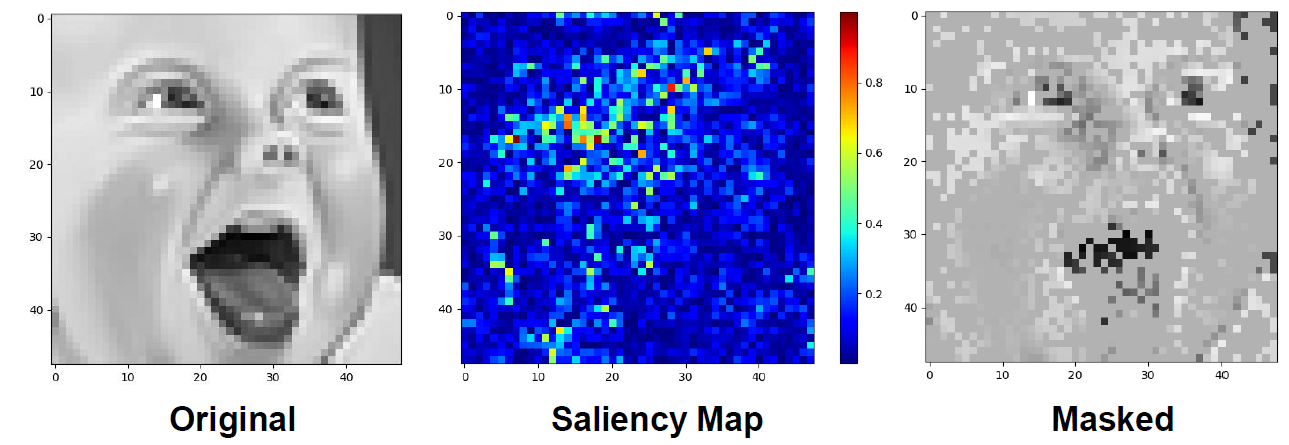
基本上可以理解成情緒分強弱，當初標label時可能會將中性表情理解成「輕微的難過」，舉例來說：

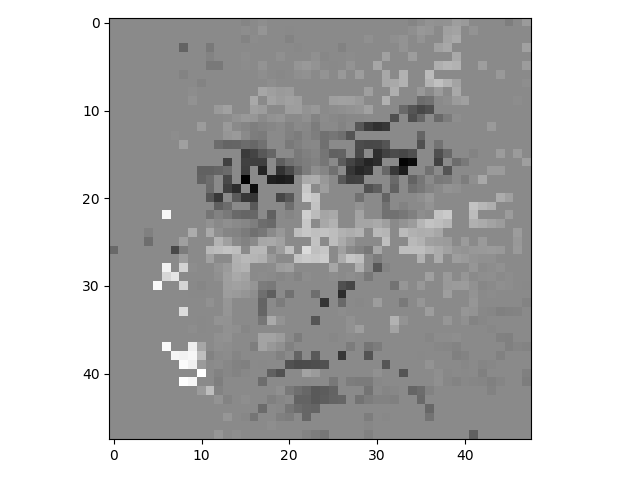
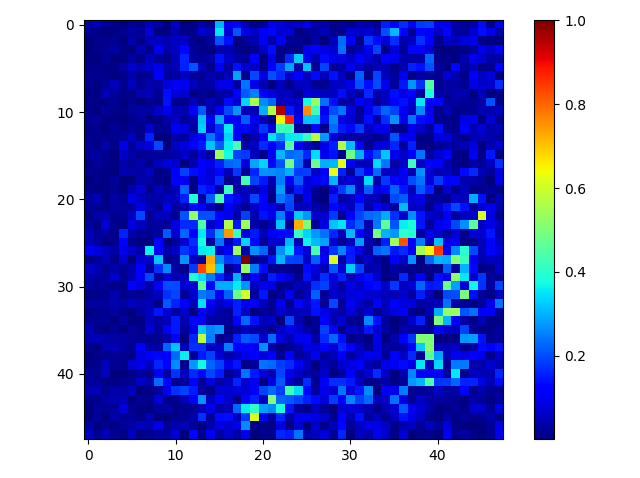
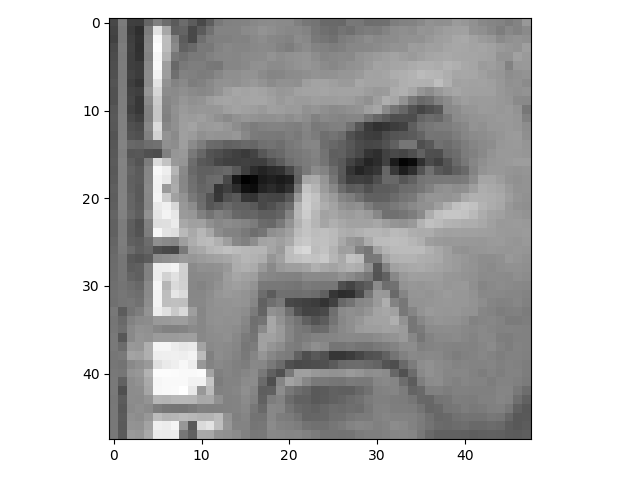
  

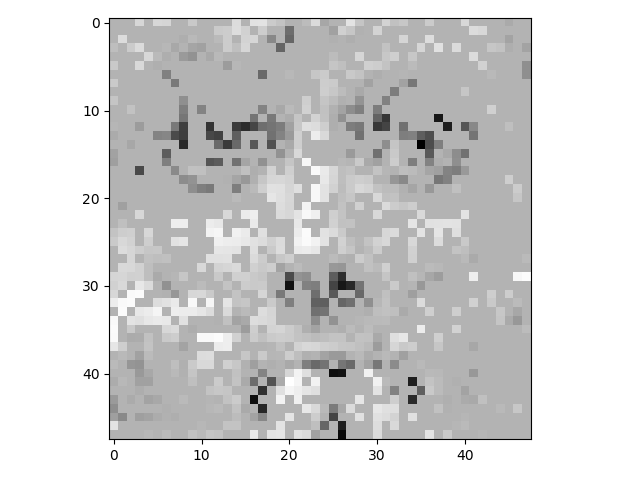
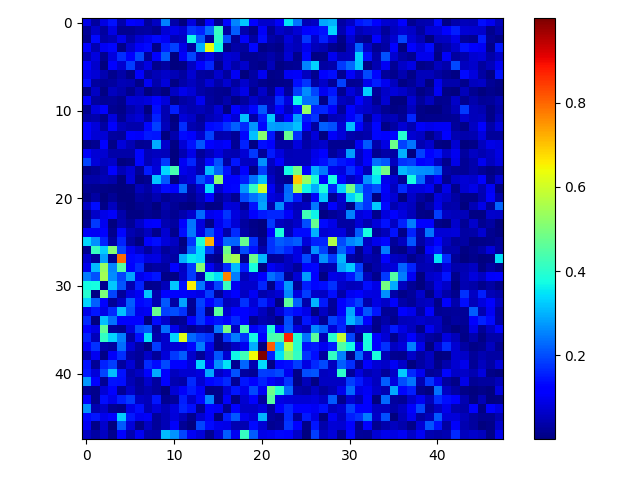
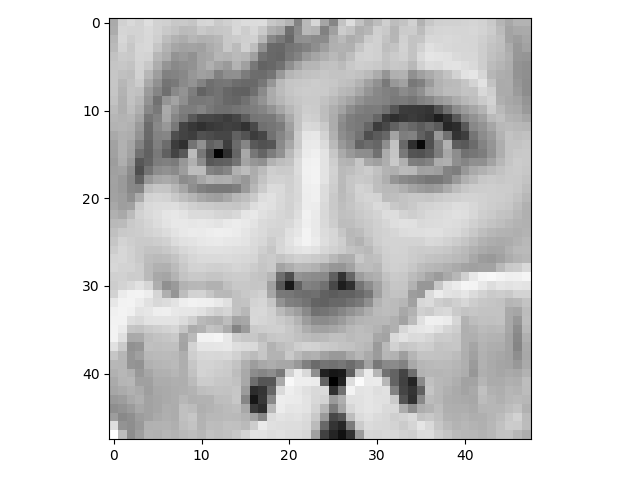
這三張圖在CNN分類後為Neutral，但標記的True label都是Sad。事實上也的確不容易看出三者的sad情緒，故CNN較容易分錯此種情況。

1. **(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) **答：**

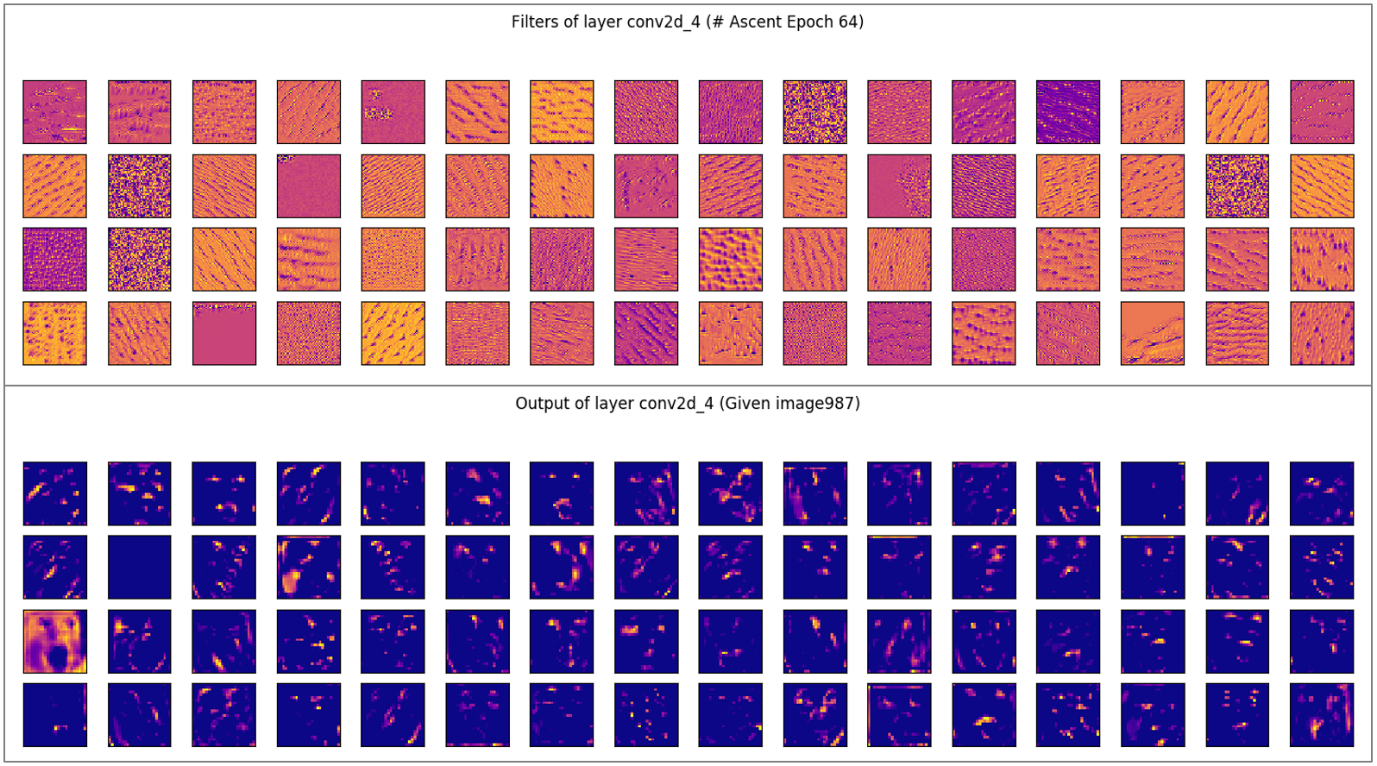








從這幾張圖的Saliency Map中較難看出CNN著重的區域，但若觀察mask後的影像則能比較清楚觀察出：CNN基本上會focus在眼睛、眉毛、鼻子和嘴巴上。

1. **(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) **答：**

取出第四層conv2d的其中64個filter和輸出來觀察，測試圖片同第4題第一張。觀察filter可以發現在這一層中，仍是以較基本的條紋方向為主，以及少數為斑點狀。若比對該filter和對應的輸出，大致可以歸納出：

條紋狀的filter可以用來進行類似edge detection的工作。

斑點狀的filter可以大致找出眼睛、鼻子、嘴巴等臉部特徵的位置。

整體來說各種不同filter都會在圖片梯度變化較大的地方有比較明顯的效果，這點和傳統影像處理中抓feature或corner等作法有點類似。