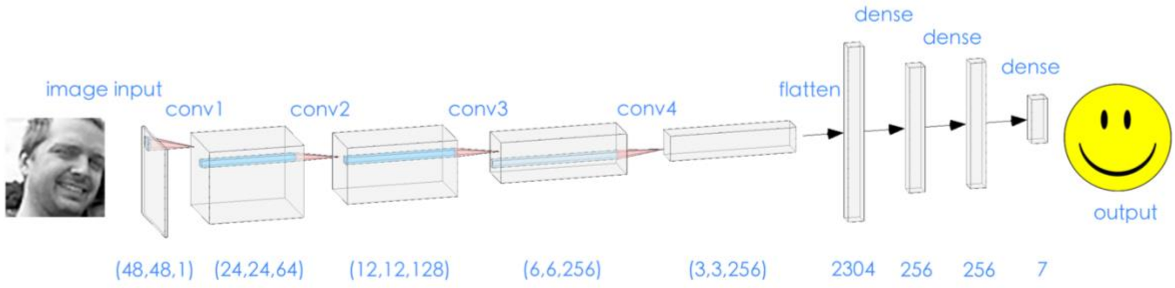
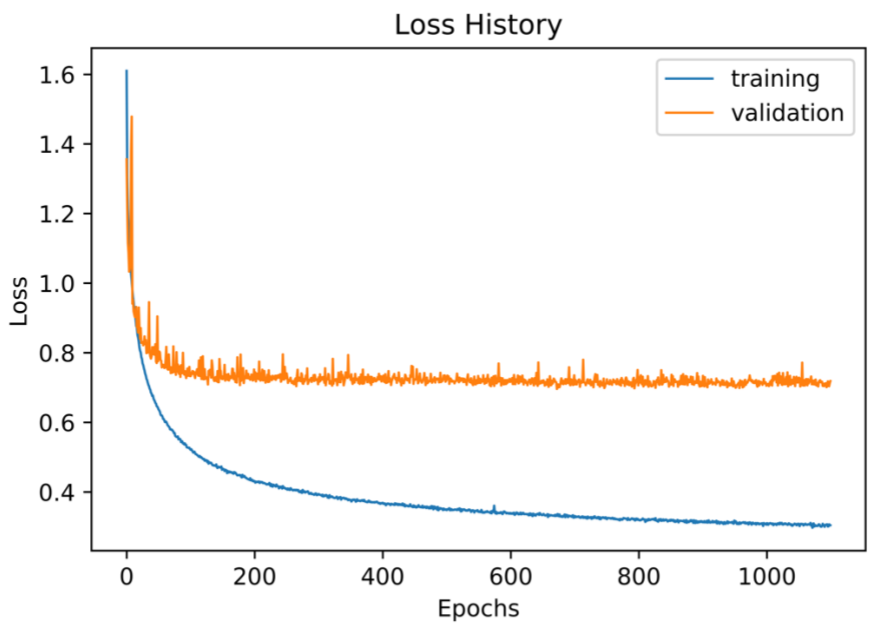
學號：R08725021 系級： 資管四 姓名：王鼎元

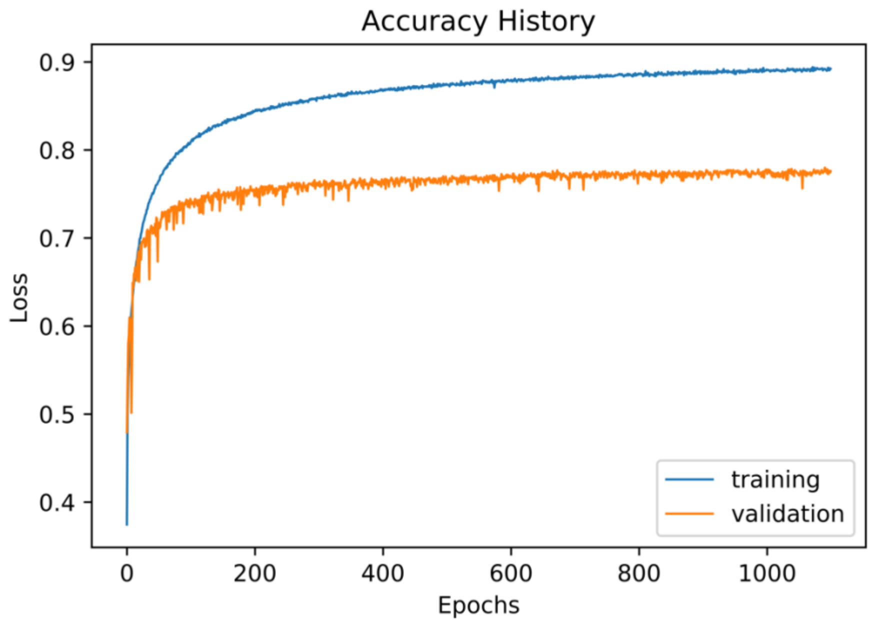
1. (1%) 請說明這次使用的model架構，包含各層維度及連接方式。



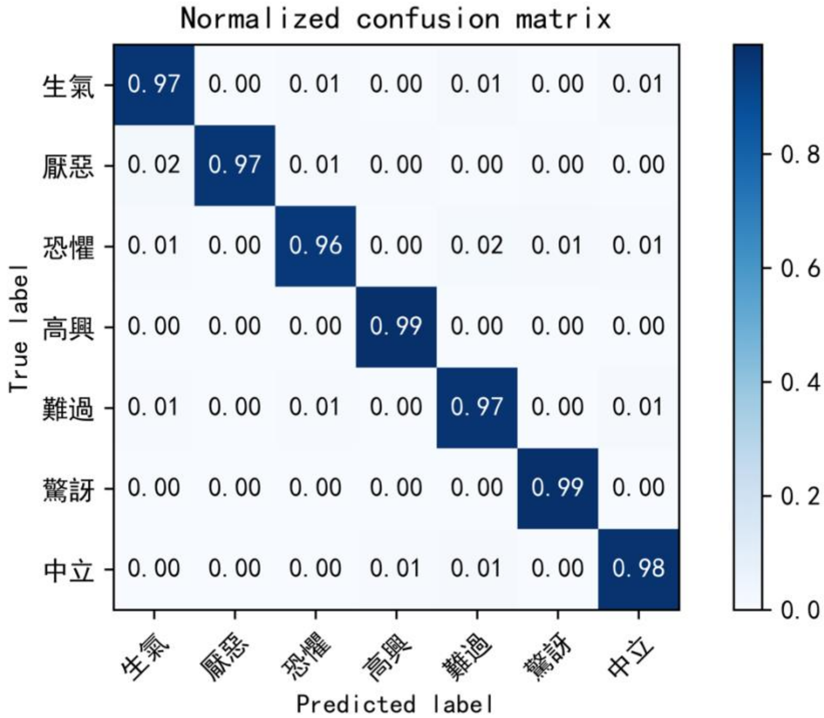
如上圖所示，我的模型架構是用了 4 層的卷積層加上 3 層的全連接層，每層之間都有做 batch normalization 與 dropout, activation function 是 leaky ReLU。

2. (1%) 請附上model的training/validation history (loss and accuracy)。



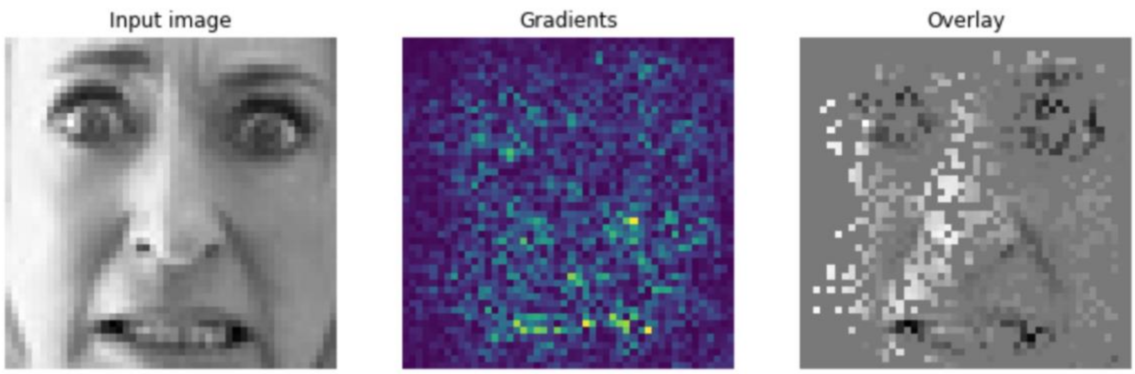


3. (1%) 畫出confusion matrix分析哪些類別的圖片容易使model搞混，並簡單說明。
(ref: https://en.wikipedia.org/wiki/Confusion_matrix)



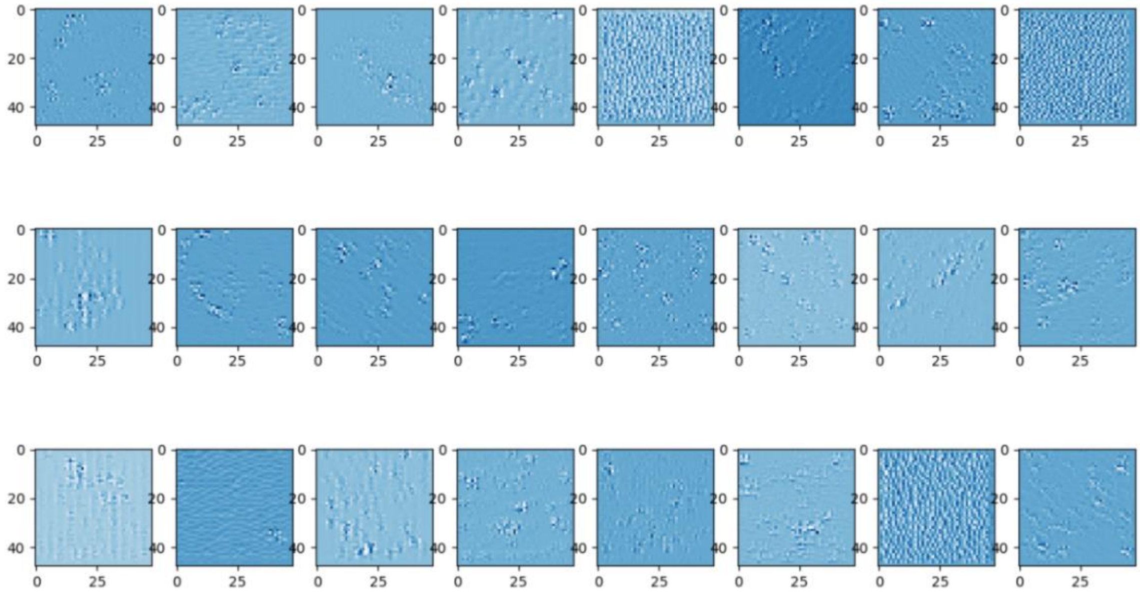
由圖可知，「厭惡」容易被誤判成「生氣」，「恐懼」容易被誤判成「難過」，而「恐懼」是最容易被誤判的類別。

4. (1%) 畫出CNN model的saliency map，並簡單討論其現象。
(ref: https://reurl.cc/Qpjg8b)



由圖可知，在做分類時，嘴巴附近與雙眼是 CNN 模型比較依據的部位。

5. (1%) 畫出最後一層的filters最容易被哪些feature activate。
(ref: https://reurl.cc/ZnrgYg)



由上可知，filter 容易被較簡單的線條與幾何圖形所 activate。

6. (3%)Refer to math problem
   https://hackmd.io/JIZ_0Q3dStSw0t0O0w6Ndw

Q1

- 令 (B', W', H', C') 是經過 convolution layer 後的 output shape
  - 因為 batch size 不會隨著模型架構而改變, 故 $B' = B$
  - $k_1 + (W' - 1) \times s_1 = W + 2p_1$
    $\rightarrow W' = \frac{W+s_1+2p_1-k_1}{s_1}$
  - $k_2 + (H' - 1) \times s_2 = H + 2p_2$
    $\rightarrow H' = \frac{H+s_2+2p_2-k_2}{s_2}$
  - $C' = output\_channels$
- 故 output shape 為 $(B, \frac{W+s_1+2p_1-k_1}{s_1}, \frac{H+s_2+2p_2-k_2}{s_2}, output\_channels)$

Q2

- $\frac{\partial l}{\partial \hat{x}_i}$
  - $\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i}\frac{\partial y_i}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i} \cdot \gamma$
- $\frac{\partial l}{\partial \sigma_B^2}$
  - $\frac{\partial l}{\partial \sigma_B^2} = \Sigma_{i=1}^m \frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = \Sigma_{i=1}^m \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{-1}{2}(x_i - \mu_B)(\sigma_B^2 + \epsilon)$
- $\frac{\partial l}{\partial \mu_B}$
  - $\frac{\partial l}{\partial \mu_B} = \Sigma_{i=1}^m \frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial \mu_B} + \frac{\partial l}{\partial \sigma_B^2}\frac{\partial \sigma_B^2}{\partial \mu_B} = (\Sigma_{i=1}^m \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_B^2+\epsilon}}) + \frac{\partial l}{\partial \sigma_B^2}\frac{\Sigma_{i=1}^m -2(x_i-\mu_B)}{m}$
- $\frac{\partial l}{\partial x_i}$
  - $\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial l}{\partial \sigma_B^2}\frac{\partial \sigma_B^2}{\partial x_i} + \frac{\partial l}{\partial \mu_B}\frac{\partial \mu_B}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_B^2+\epsilon}} + \frac{\partial l}{\partial \sigma_B^2} \cdot \frac{2(x_i-\mu_B)}{m} + \frac{\partial l}{\partial \mu_B} \cdot \frac{1}{m}$
- $\frac{\partial l}{\partial \gamma}$
  - $\frac{\partial l}{\partial \gamma} = \Sigma_{i=1}^m \frac{\partial l}{\partial y_i}\frac{\partial y_i}{\partial \gamma} = \Sigma_{i=1}^m \frac{\partial l}{\partial y_i} \cdot \hat{x}_i$
- $\frac{\partial l}{\partial \beta}$
  - $\frac{\partial l}{\partial \beta} = \Sigma_{i=1}^m \frac{\partial l}{\partial y_i}$
- 由上可得到 $\frac{\partial l}{\partial \gamma}$, $\frac{\partial l}{\partial \beta}$ 的值, 便能用 gradient descent 更新 γ, β

Q3

- $\frac{\partial L}{\partial z_t} = \frac{\partial L}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial z_t}$
  - 因 $L_t(y_t, \hat{y}_t) = -[y_t log\hat{y}_t + (1 - y_t)log(1 - \hat{y}_t)]$
    $\rightarrow \frac{\partial L}{\partial \hat{y}_t} = -\frac{y_t}{\hat{y}_t} + \frac{1-y_t}{1-\hat{y}_t} = \frac{-y_t+\hat{y}_t}{\hat{y}_t(1-\hat{y}_t)}$
  - $\frac{\partial \hat{y}_t}{\partial z_t} = \frac{\partial}{\partial z_t}\frac{e^{z_t}}{\Sigma_i e^{z_i}} = \frac{\Sigma_i e^{z_i} \cdot e^{z_t} - e^{z_t} \cdot e^{z_t}}{\Sigma_i e^{z_i} \cdot \Sigma_i e^{z_i}} = \hat{y}_t - \hat{y}_t^{\,2}$
- 故 $\frac{\partial L}{\partial z_t} = \frac{-y_t+\hat{y}_t}{\hat{y}_t(1-\hat{y}_t)} \cdot (\hat{y}_t - \hat{y}_t^{\,2}) = \hat{y}_t - y_t$