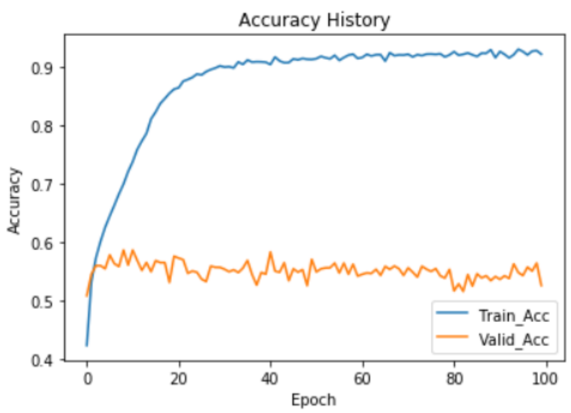
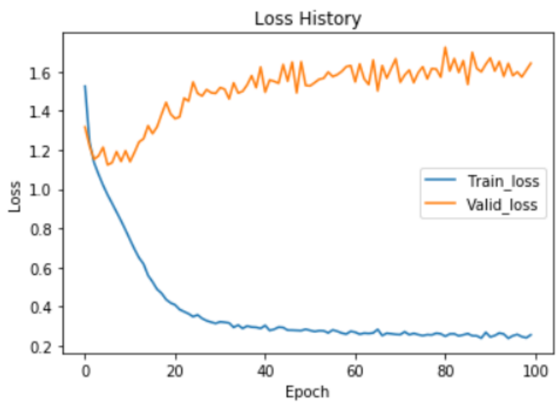
學號：R07725019 系級：資管碩二 姓名：鄒雅雯

1. (1%) 請說明這次使用的 model 架構，包含各層維度及連接方式。

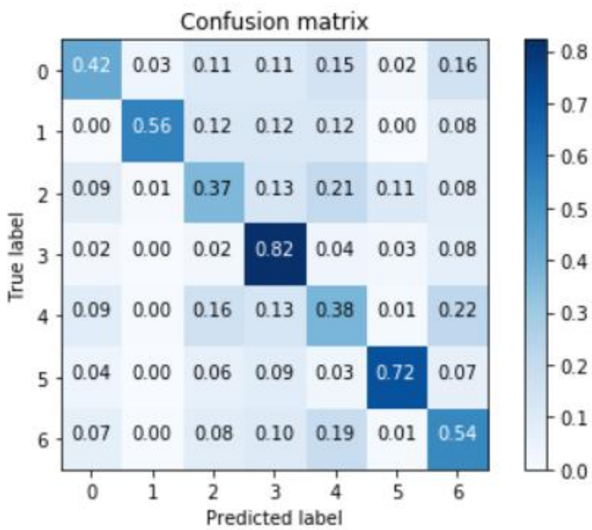
這次 model 總共有 4 層的 Convolution 和 Max pooling，第一層的 filter size 為 5，有 32 個 filter，第二層 filter size 為 3，共 64 個 filter，第三層和第四層的 filter size 皆為 3，且有 128 個 filter，每層都會加一個 LeakyReLU，並且做 Batch normalization，將 Convolution & Max pooling 做完後，將 2d 向量拉成一維(256)，之後加一層 output layer，共 7 個神經元，採 fully-connected，活化函數是 ReLU，得出最後預測結果。

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





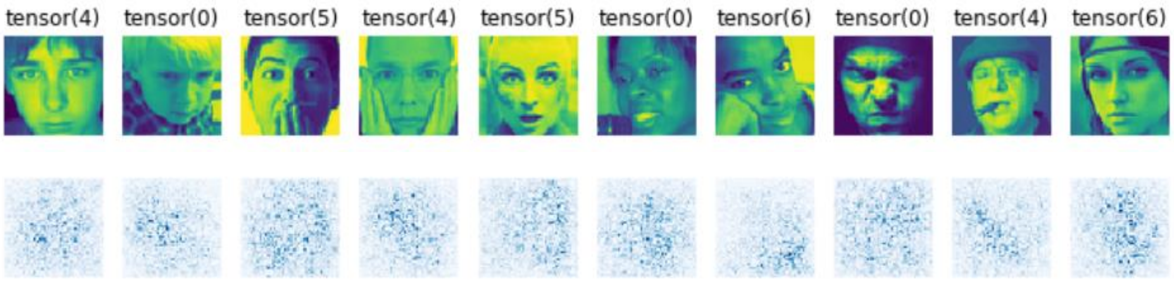
3. (1%) 畫出 confusion matrix 分析哪些類別的圖片容易使 model 搞混，並簡單說明。(ref: https://en.wikipedia.org/wiki/Confusion_matrix)



從 Confusion matrix 結果中，看到類別 2(恐懼)和類別 4(難過)的準確率是較低的，從類別 2 中看出，model 常會把類別 2(恐懼)的表情誤判為類別 4(難過)(0.21%)，而 model 常會把類別 4(難過)誤判為類別 2(恐懼)(0.16%)和類別 6(中立)(0.22%)，因此類別 2、4、6 的圖片較容易使 model 混淆。

[關於第四及第五題]
可以使用簡單的 3-layer CNN model [64, 128, 512] 進行實作。

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



從 saliency map 中看出，算出的 gradient 在圖像的臉頰部位較為顯著，對分類的作用較大

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

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

*1

Input : (B, Win, Hin, input_channels)

Output . (B, Wout, Hout, output_channels)

$$H_{out} = \left\lfloor \frac{H_{in} + 2p_1 - (k_1 - 1) - 1}{s_1} + 1 \right\rfloor = \left\lfloor \frac{H_{in} + 2p_1 - k}{s_1} + 1 \right\rfloor$$

$$W_{out} = \left\lfloor \frac{W_{in} + 2p_2 - (k_2 - 1) - 1}{s_2} + 1 \right\rfloor = \left\lfloor \frac{W_{in} + 2p_2 - k_2}{s_2} + 1 \right\rfloor$$

*2

$$y_i = \gamma \hat{x}_i + \beta = BN_{\gamma,\beta}(x_i)$$

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{B} \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \gamma} = \sum_{i=1}^{B} \frac{\partial l}{\partial y_i} \hat{x}_i$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{B} \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \beta} = \sum_{i=1}^{B} \frac{\partial l}{\partial y_i}$$

$$\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i} \gamma$$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial \mu} + \frac{\partial l}{\partial \sigma^2} \frac{\partial \sigma^2}{\partial \mu} , \quad \hat{x}_i = \frac{(x_i - \mu)}{\sqrt{\sigma^2 + \varepsilon}} , \quad \sigma^2 = \frac{1}{B} \sum_{i=1}^{B} (x_i - \mu)^2$$

$$\frac{\partial \hat{x}_i}{\partial \mu} = \frac{1}{\sqrt{\sigma^2 + \varepsilon}} (-1) , \quad \frac{\partial \sigma^2}{\partial \mu} = \frac{1}{B} \sum_{i=1}^{B} 2(x_i - \mu)(-1)$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{\partial l}{\partial \hat{x}} \frac{\partial \hat{x}}{\partial \sigma^2} \rightarrow \frac{\partial \hat{x}}{\partial \sigma^2} = \sum_{i=1}^{B} (x_i - \mu)\left(\frac{-1}{2}\right)(\sigma^2 + \varepsilon)^{\frac{-3}{2}} = \frac{-1}{2} \sum_{i=1}^{B} (x_i - \mu)(\sigma^2 + \varepsilon)^{\frac{-3}{2}}$$

$$\rightarrow \frac{\partial l}{\partial \sigma^2} = \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{-1}{2} \frac{(x_i - \mu)}{(\sigma^2 + \varepsilon)^{\frac{3}{2}}}$$

$$\frac{\partial l}{\partial \mu} = \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \frac{-1}{\sqrt{\sigma^2 + \varepsilon}} + \frac{\partial l}{\partial \sigma^2} \frac{1}{B} \sum_{i=1}^{B} -2(x_i - \mu) = \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \frac{-1}{\sqrt{\sigma^2 + \varepsilon}} + \frac{\partial l}{\partial \sigma^2} (-2) \left[\frac{1}{B} \sum_{i=1}^{B} x_i - \frac{1}{B} \sum_{i=1}^{B} \mu \right]$$

$$= \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \frac{-1}{\sqrt{\sigma^2 + \varepsilon}} + \frac{\partial l}{\partial \sigma^2} (-2) \left(\mu - \frac{m \cdot \mu}{m} \right) = \sum_{i=1}^{B} \frac{\partial l}{\partial \hat{x}_i} \frac{-1}{\sqrt{\sigma^2 + \varepsilon}}$$

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial l}{\partial \mu}\frac{\partial \mu}{\partial x_i} + \frac{\partial l}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial x_i}$$

$$\frac{\partial \hat{x}_i}{\partial x_i} = \frac{1}{\sqrt{\sigma^2+\varepsilon}}, \quad \frac{\partial \mu}{\partial x_i} = \frac{1}{B}, \quad \frac{\partial \sigma^2}{\partial x_i} = \frac{2(x_i-\mu)}{B}$$

$$\rightarrow \frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i}\frac{1}{\sqrt{\sigma^2+\varepsilon}} + \frac{\partial l}{\partial \mu}\frac{1}{B} + \frac{\partial l}{\partial \sigma^2}\frac{2(x_i-\mu)}{B}$$

$$= \frac{\partial l}{\partial \hat{x}_i}\frac{1}{\sqrt{\sigma^2+\varepsilon}} + \frac{1}{B}\sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j}\frac{-1}{\sqrt{\sigma^2+\varepsilon}} - \frac{1}{2}\sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j}(x_j-\mu)(\sigma^2+\varepsilon)^{-\frac{3}{2}}\frac{2(x_i-\mu)}{m}$$

$$= \frac{\partial l}{\partial \hat{x}_i}(\sigma^2+\varepsilon)^{-\frac{1}{2}} - \frac{(\sigma^2+\varepsilon)^{-\frac{1}{2}}}{B}\sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j} - \frac{(\sigma^2+\varepsilon)^{-\frac{1}{2}}}{B}\frac{(x_i-\mu)}{(\sigma^2+\varepsilon)^{\frac{1}{2}}}\sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j}\frac{(x_j-\mu)}{(\sigma^2+\varepsilon)^{\frac{1}{2}}}$$

$$= \frac{(\sigma^2+\varepsilon)^{-\frac{1}{2}}}{B}\left[B\frac{\partial l}{\partial \hat{x}_i} - \sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j} - \hat{x}_i\sum_{j=1}^{B}\frac{\partial l}{\partial \hat{x}_j}\hat{x}_j\right]$$

*3

$$\frac{\partial L}{\partial z_t} = \frac{\partial L}{\partial y_t}\frac{\partial y_t}{\partial z_t} + \sum_{i\neq t}\frac{\partial L}{\partial y_i}\frac{\partial y_i}{\partial z_t}$$

$$= \frac{-\hat{y}_t}{y_t}\left(\frac{-e^{-z_t}(\sum_j e^{-z_j}) - e^{-z_t}(-e^{-z_t})}{(\sum_j e^{-z_j})^2}\right) + \sum_{i\neq t}\frac{-\hat{y}_i}{y_i}\frac{e^{-z_i}(-e^{-z_t})}{(\sum_j e^{-z_j})^2}$$

$$= \frac{-\hat{y}_t}{y_t}(-y_t + y_t^2) + \sum_{i\neq t}\frac{-\hat{y}_i}{y_i}y_i^2$$

$$= \hat{y}_t - \hat{y}_t y_t + \sum_{i\neq t} -\hat{y}_i y_i$$

$$= \hat{y}_t - \sum_i \hat{y}_i y_i = \hat{y}_t - y_t$$