學號：B04303006 系級：經濟四 姓名：劉傳筠

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

   本次使用 resnet34，以下圖示基本架構：

| 34-layer | |
|---|---|
| conv 7x7x64, stride 2 | |
| max pool 3x3, stride 2 | |
| 3x3, 64<br>3x3, 64 | *3 |
| 3x3, 128<br>3x3, 128 | *4 |
| 3x3, 256<br>3x3, 256 | *6 |
| 3x3, 512<br>3x3, 512 | *3 |
| adaptive avg pool, 7d-fc softmax | |

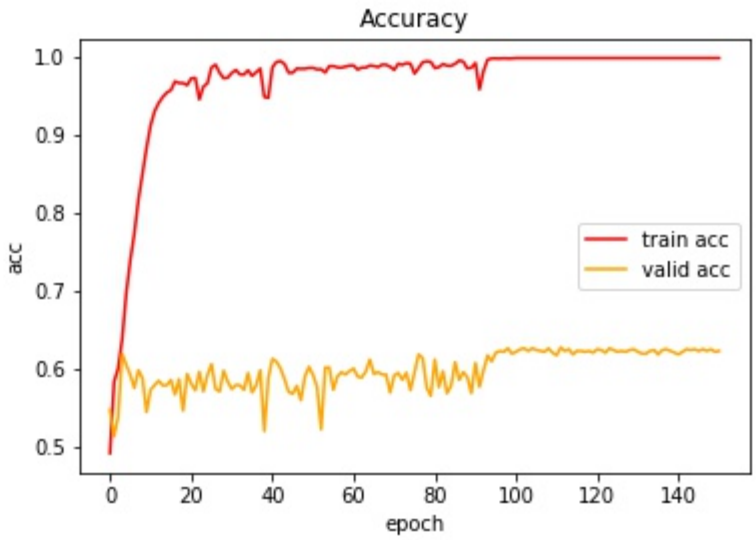
- 最前面先通過一個 convolution layer 和一個 max pooling。
- 中間有 4 個 layer(64,128,256,512)，每個 layer 是由數個相同的 blocks 所組成，每個 block 裡面有兩層 3x3 的 convolution layer, filter 數固定 (64,128,256,512) 。連接方式是 conv1-> batchnorm -> relu -> conv2 -> batchnorm。在同一個 layer 裡面，W 和 H 維持不變。在轉換 layer 時，會在下一個 layer 使用 stride = (2, 2) 以達到 output_channels 變為兩倍的效果。
- 最後，通過一個 adaptive average pool, 然後再通過一個 softmax, 轉換維度為 7。

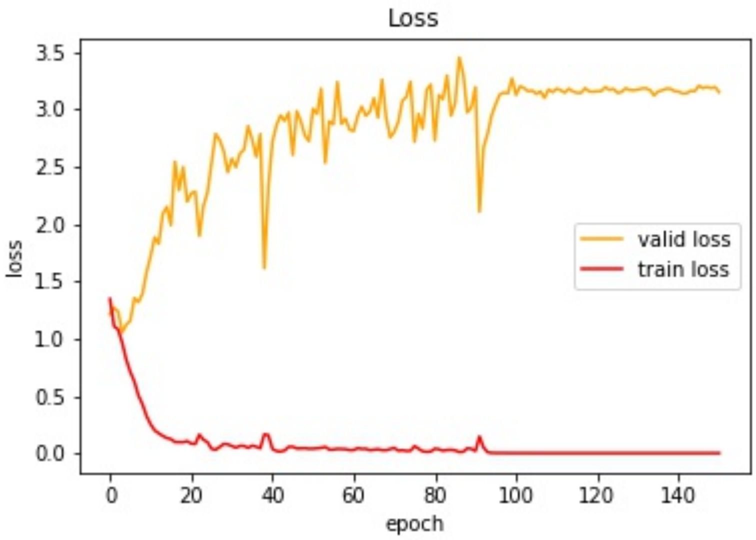
2. (1%) 請附上 model 的 training/validation history (loss and accuracy)。

   # of epochs: 151

   # of training data: 28000

   # of testing data: 888 (the rest)

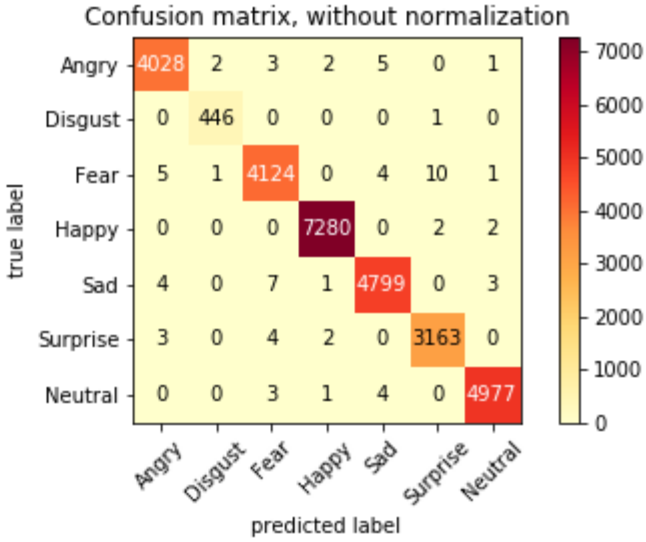


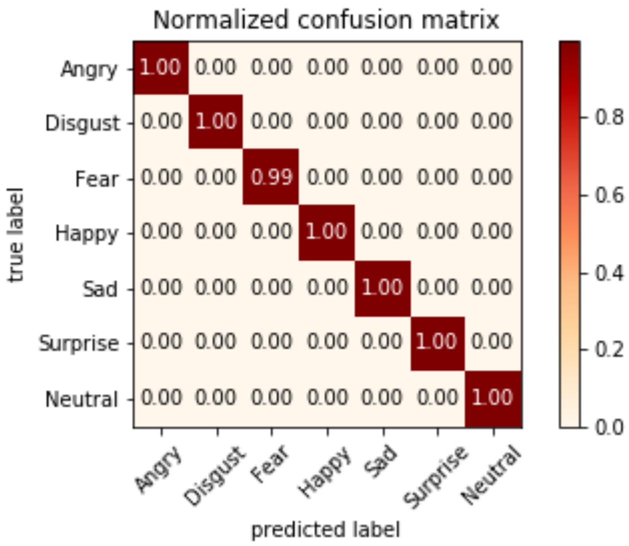




3. (1%) 畫出 confusion matrix 分析哪些類別的圖片容易使 model 搞混，並簡單說明。

   搞混最多的是把恐懼以為是驚訝(10 個)或生氣(5 個), 把悲傷以為是恐懼(7 個), 把生氣以為是傷心(5 個)。這些感覺都是比較負面的情緒(生氣和傷心)，還有瞪大眼睛(生氣和

驚訝)。開心則是最少判錯的表情，只有四個被判錯。




| | Angry | Disgust | Fear | Happy | Sad | Surprised | Neutral |
|---|---|---|---|---|---|---|---|
| Error | 0.322% | 0.223% | 0.507% | 0.055% | 0.312% | 0.284% | 0.160% |

從判錯率來看，錯機機率最小的依序是快樂、中立、厭惡、驚訝、悲傷、生氣、恐懼。感覺仍然是負面情緒比較難判對。

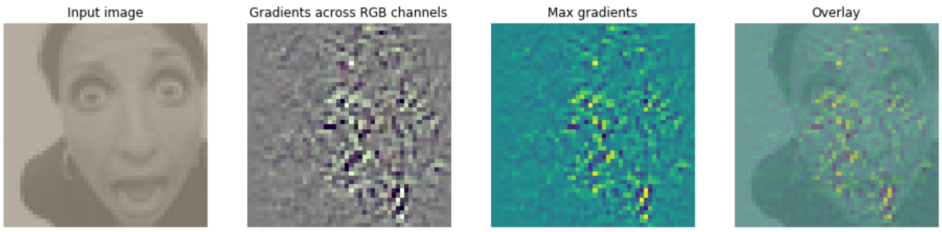
[關於第四及第五題]

可以使用簡單的 3-layer CNN model [64, 128, 512] 進行實作。

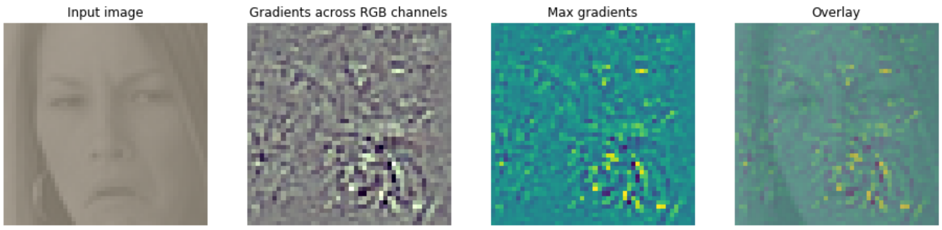
4. (1%) 畫出 CNN model 的 saliency map，並簡單討論其現象。

(ref: https://reurl.cc/Qpjg8b)

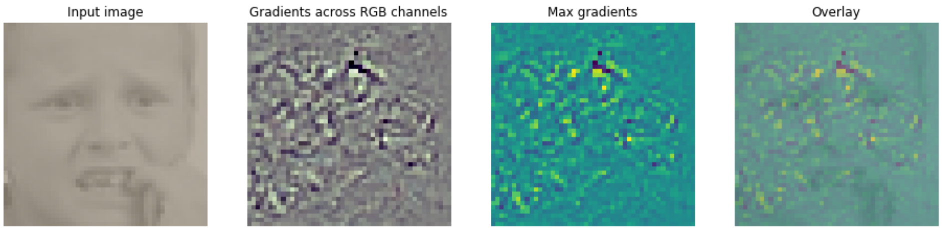
0: angry



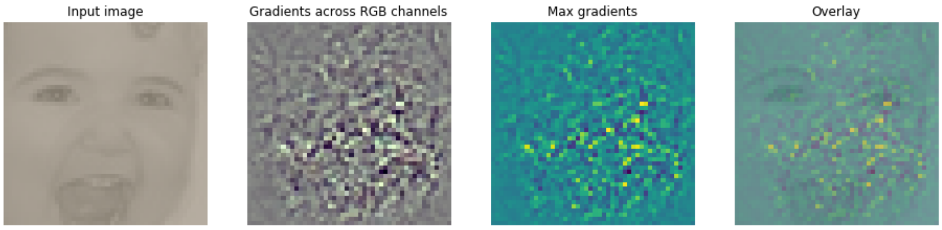

1: disgust



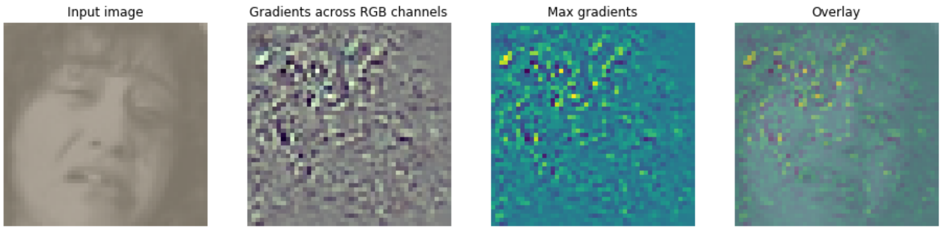

2: fear



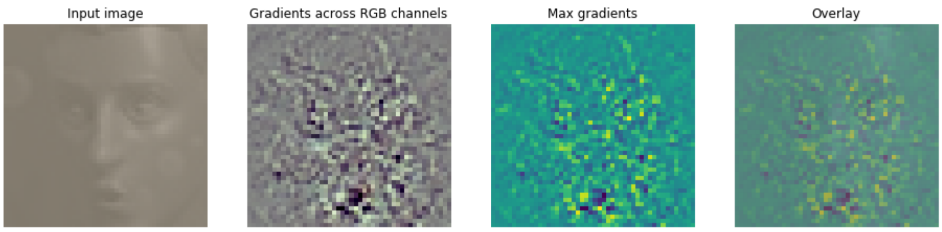

3: happy



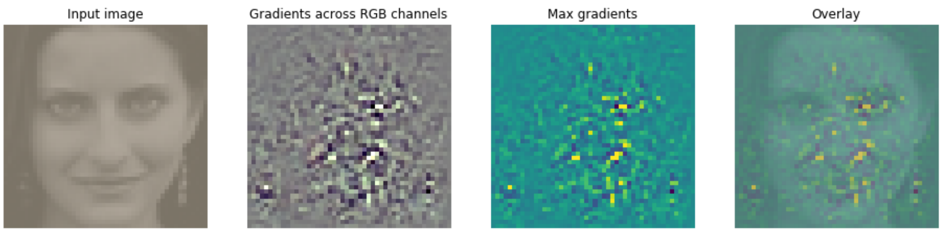
4: sad




5: surprised




6: neutral

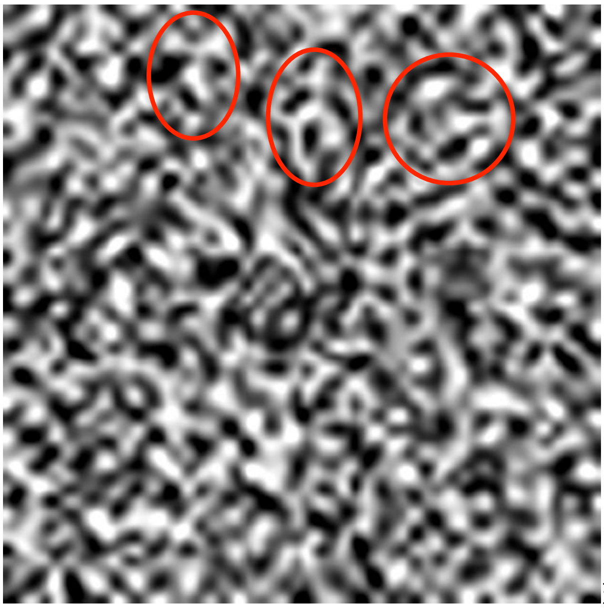



觀察發現，訓練出來的模型好像多為捕捉到五官，尤其是眉毛和嘴巴周遭的表情，來判斷到底是什麼情緒。除了害怕和驚訝有眉毛之外，其他感覺都是看嘴巴。在 2: fear 那張圖中，模型成功忽略照片中小男孩的手。

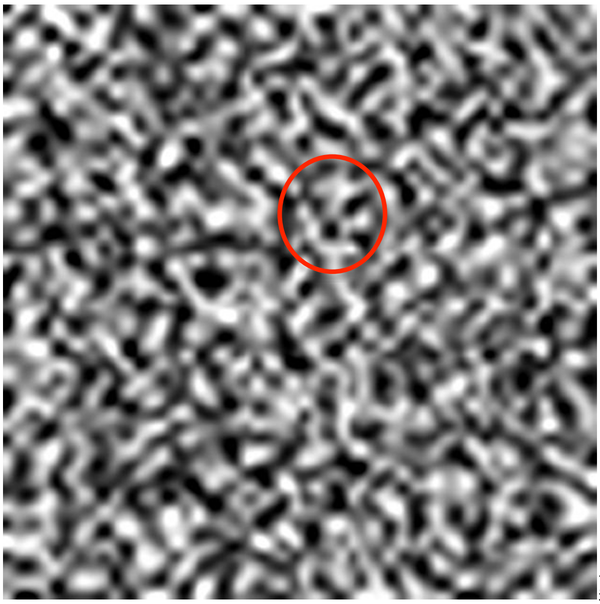
5. (1%) 畫出最後一層的 filters 最容易被哪些 feature activate。

   (ref: https://reurl.cc/ZnrgYg)

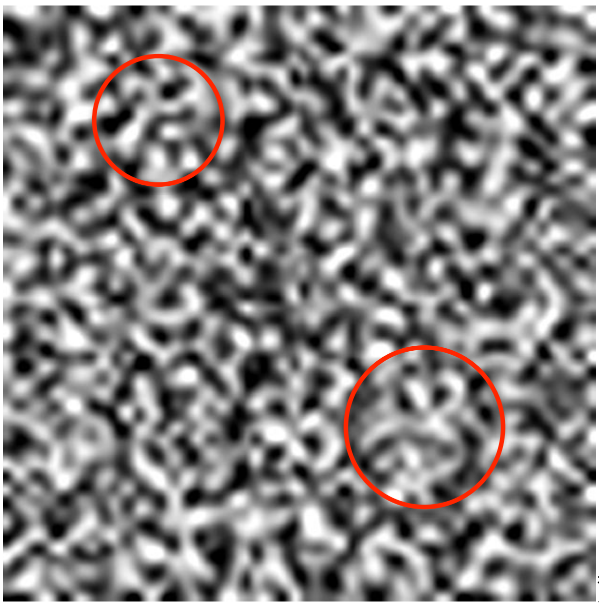
   感覺依稀有捕捉到幾張人臉...但好像不是很清楚。可能是因為最後一層有 512 個 channel 所以最後機器是將很多種不同的 feature 組合起來在判斷表情吧...。



這張裡面感覺有些恐懼的元素



難過



驚訝

Q1: Convolution.

Denote input_channels = C

The shape (B, W, H, C) will change to

$$\begin{cases} H_{out} = \left\lfloor \dfrac{H_{in} + 2 \times P_1 - k_1}{S_1} + 1 \right\rfloor \\ W_{out} = \left\lfloor \dfrac{W_{in} + 2 \times P_2 - k_2}{S_2} + 1 \right\rfloor \\ \text{channel} = \text{output\_channel} \\ B: \text{remains.} \end{cases}$$

Q2: Batch Normalization.

$\because y_i = r\hat{x}_i + \beta$

① $\dfrac{\partial l}{\partial \hat{x}_i} = \dfrac{\partial l}{\partial y_i} \cdot \dfrac{\partial y_i}{\partial \hat{x}_i} = \dfrac{\partial l}{\partial y_i} \cdot r$

$(x_i - \mu_B)(\sigma_B^2 + \varepsilon)^{-\frac{1}{2}}$

② $\dfrac{\partial l}{\partial \sigma_B^2} = \sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{\partial \hat{x}_i}{\partial \sigma_B^2} = \sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} \dfrac{\partial \dfrac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}}{\partial \sigma_B^2}$

$= \sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{-1}{2}(x_i - \mu_B) \cdot (\sigma_B^2 + \varepsilon)^{-\frac{3}{2}}$

③ $\dfrac{\partial l}{\partial \mu_B} = \left[\sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} + \dfrac{\partial \hat{x}_i}{\partial \mu_B}\right] + \dfrac{\partial l}{\partial \sigma_B^2} \cdot \dfrac{\partial \sigma_B^2}{\partial \mu_B}$

$= \left[\sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} (-1) \dfrac{1}{\sqrt{\sigma_B^2 + \varepsilon}}\right] + \dfrac{\partial l}{\partial \sigma_B^2} \cdot \dfrac{1}{m} \cdot 2 \sum_{i=1}^{m} (x_i - \mu_B) \cdot (-1)$

$\frac{1}{m}(x_1 + x_2 + \cdots + x_i + \cdots + x_m)$

④ $\dfrac{\partial l}{\partial x_i} = \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{\partial \hat{x}_i}{\partial x_i} + \dfrac{\partial l}{\partial \sigma_B^2} \cdot \dfrac{\partial \sigma_B^2}{\partial x_i} + \dfrac{\partial l}{\partial \mu_B} \cdot \dfrac{\partial \mu_B}{\partial x_i}$

$= \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{1}{\sqrt{\sigma_B^2 + \varepsilon}} + \dfrac{\partial l}{\partial \sigma_B^2} \cdot \dfrac{2}{m}(x_i - \mu_B) + \dfrac{\partial l}{\partial \mu_B} \cdot \dfrac{1}{m}$

⑤ $\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \cdot \frac{\partial y_i}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \cdot \hat{x}_i$

⑥ $\frac{\partial l}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i}$ ※

Q3: Softmax Entropy

大助教 said:「題目中給的 L 是 $y_t=1$ 的形式」

W.L.O.G, $\frac{\partial L}{\partial z_t} = -\sum_i y_i \frac{\partial \log \hat{y}_i}{\partial z_t}$

$$= \underbrace{-\sum_{i=t} y_i(1-\hat{y}_t)}\ \underline{-\sum_{i\neq t} y_i(\hat{y}_t)}$$ (証明見下方(*))

$$= \underbrace{-y_t(1-\hat{y}_t)}\ \underline{+\sum_{i\neq t} y_i(\hat{y}_t)}$$

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

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

$$= -y_t + \hat{y}_t \quad (\because \sum_i y_i = 1)$$ ※

(*) $\frac{\partial \log \hat{y}_i}{\partial z_t} = \partial \log \frac{e^{z_i}}{\sum_j e^{z_j}} / \partial z_t$

$$= \partial \log e^{z_i}/\partial z_t - \partial \log \sum_j e^{z_j}/\partial z_t$$

$$= \begin{cases} 1 - \frac{e^{z_t}}{\sum_j e^{z_j}} = (1-\hat{y}_t), & \text{if } i=t \\ \frac{e^{z_t}}{\sum_j e^{z_j}} = \hat{y}_t, & \text{if } i\neq t \end{cases}$$