學號：R07946013 系級： 資料科學碩二 姓名：吳泓毅

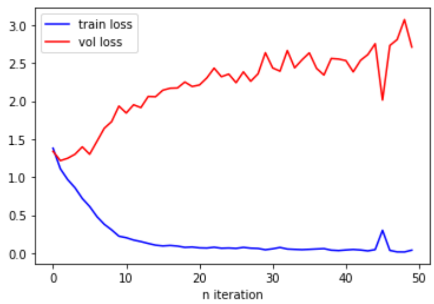
1. (1%) 請說明這次使用的model架構，包含各層維度及連接方式。

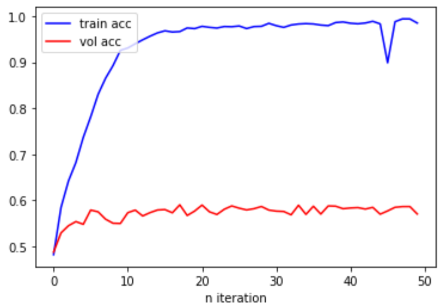
wide_resnet50_2 是用採用 Bottlencek 來當model 基礎的 block，每一層會有富庶個 Bottleneck 而當中第0個會執行 downsample。

```
Basic block:
       Bottleneck:
                 conv1x1 - bn2d - conv3x3 - bn2d - con1x1 - bn2d - relu-(conv2d - bn2d) (if
                                                                        downsampling)

Sequential(
  (0): Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3), bias=False)
  (1): bn2d
  (2): ReLU
  (3): MaxPool2d(kernel_size=3, stride=2, padding=1, dilation=1, ceil_mode=False)
  (4):(layer1): Sequential(3*Bottleneck)
       in_planes = 64, out_planes = 256
  (5):(layer2): Sequential(4*Bottleneck)
       in_planes = 256, out_planes = 512
  (6):(layer3): Sequential(6*Bottleneck)
       in_planes = 512, out_planes = 1024
  (7):(layer4): Sequential(3*Bottleneck)
       in_planes = 1024, out_planes = 2048
  (8): AdaptiveAvgPool2d(output_size=(1, 1))
)
flatten
Linear(in_features=2048, out_features=7, bias=True)
```

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

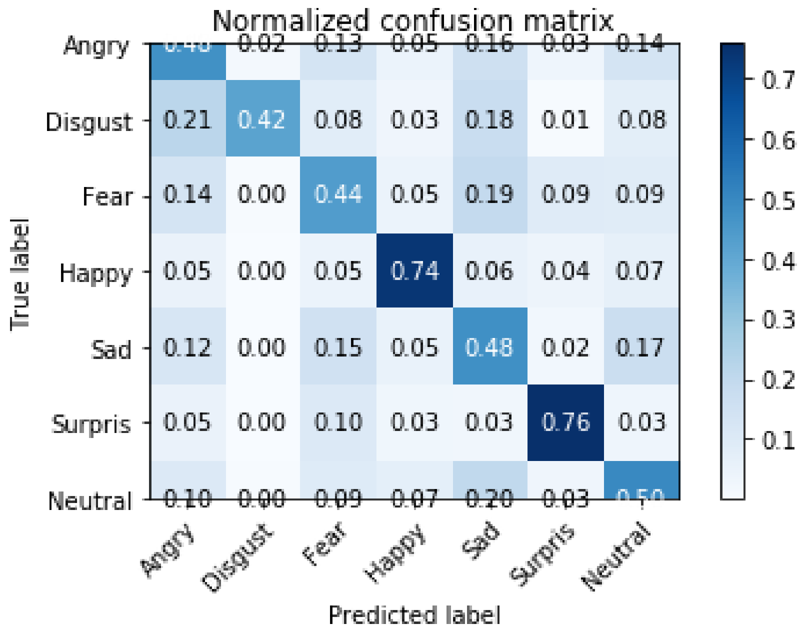




總共28888筆資料，這是把20000比當 training 剩下的8888筆當作 volidation的結果，雖然在 volidation set的準確率只有接近60%，但再把剩下的8888筆都加入 Training 以後，在kaggle 上的準確率卻突然升到 90%左右，猜測是這是suffle 出來的training set 和 volidation set不平均所造成的。

[原來是training set 跟 kaggle test set 一樣的關係]

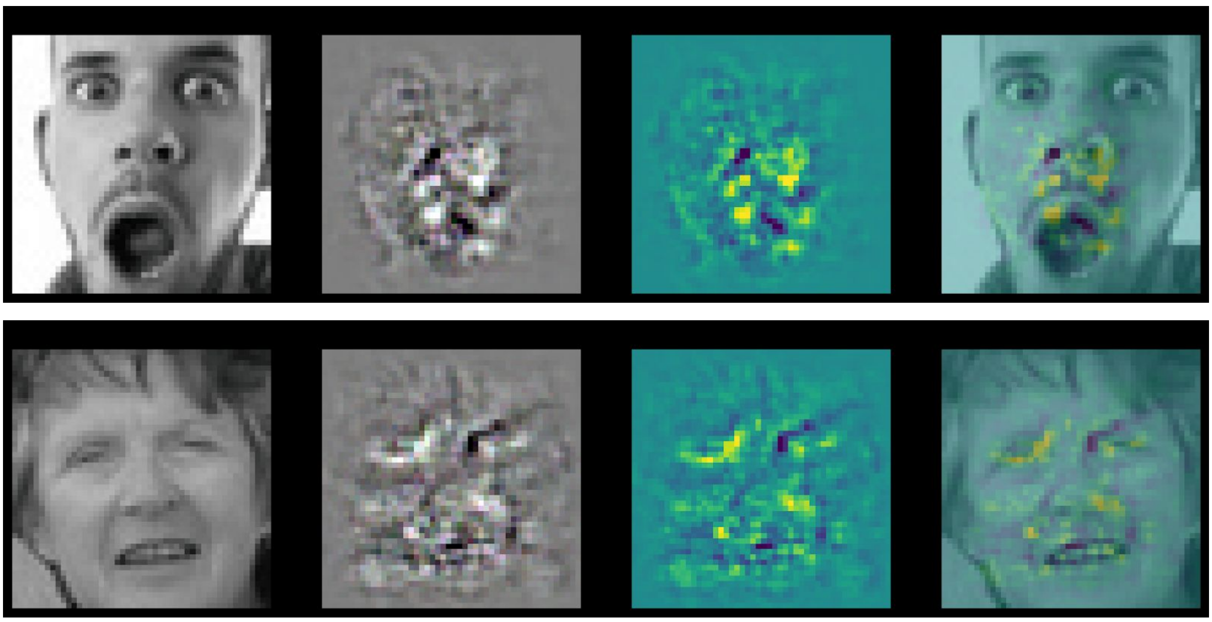
3. (1%) 畫出confusion matrix分析哪些類別的圖片容易使model搞混，並簡單說明。(ref: https://en.wikipedia.org/wiki/Confusion_matrix)

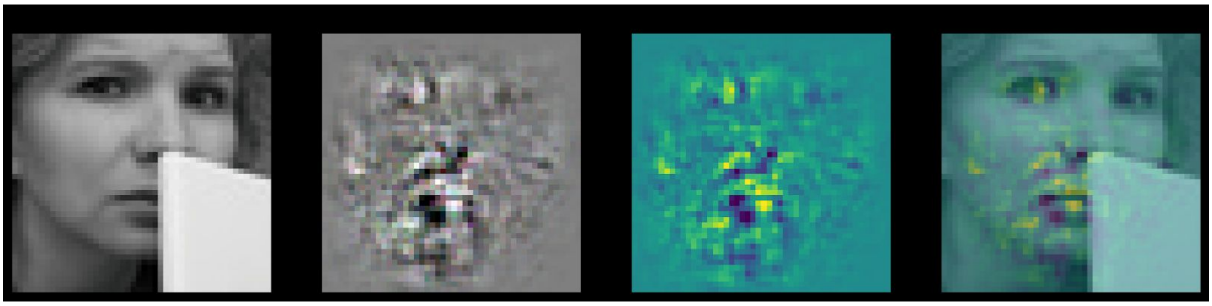


在準確率的部分除了 Happy 、 Surprise 和 Neutral 以外都是低於50%的，除了對角線以外最高的數值是把 Disgust 判斷成 Angry 的 21% 和把 Neutral 判斷成 sad 21%，而資料中判斷最好的是 Happy 及Surprise。

[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。

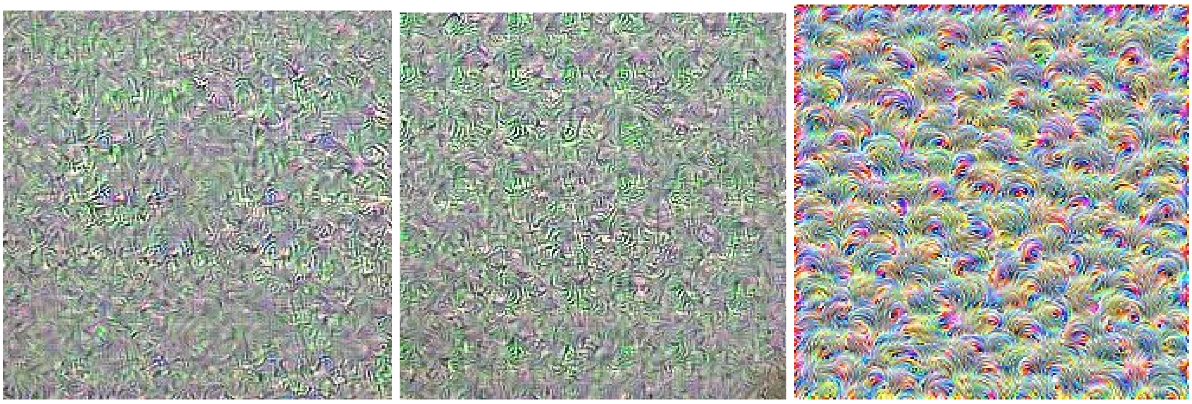
4. (1%) 畫出CNN model的saliency map，並簡單討論其現象。
(ref: https://reurl.cc/Qpjg8b)





這是用3層CNN 跑出來的，可以發現gradient 比較大的地方在於鼻子跟嘴巴，還有這個人臉的輪廓，可以推斷說CNN 嘴巴的形狀對CNN影響最顯著，而這也跟這次Sentiment classification 的目標相符合。

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

1. (B, W, H, input_channels)

$$\downarrow \text{Conv2D}(\text{input\_channels}, \text{output\_channels}, (k_1, k_2), (s_1, s_2), (P_1, P_2))$$

$$\left(B, \left\lfloor \frac{W+2P_1-k_1+1}{s_1} \right\rfloor + 1, \left\lfloor \frac{H+2P_2-k_2+1}{s_2} \right\rfloor + 1, \text{output\_channels}\right).$$

2.

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial y_i} \cdot \gamma$$

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

$$\frac{\partial l}{\partial \mu_B} = \left(\sum_{i=1}^{m} \frac{\partial l}{\partial x_i} \cdot \frac{-1}{\sqrt{\sigma_B^2 + \varepsilon}} \right) + \frac{\partial l}{\partial \sigma_B^2} \cdot \frac{\sum_{i=1}^{m} -2(x_i - \mu_B)}{m}$$

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

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

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

3.

$$\frac{\partial L_t}{\partial z_t} = \frac{\partial}{\partial z_t} - y_t \log \frac{e^{z_t}}{\sum e^{z_i}} = -y_t \left(1 - \frac{\partial}{\partial z_t} \log \sum e^{z_i}\right)$$

$$= -y_t \left(1 - \frac{e^{z_t}}{\sum e^{z_i}}\right) = -y_t (1 - \hat{y}_t)$$

$$= y_t \hat{y}_t - y_t$$

$$= \hat{y}_t - y_t, \quad (y_t = 1).$$