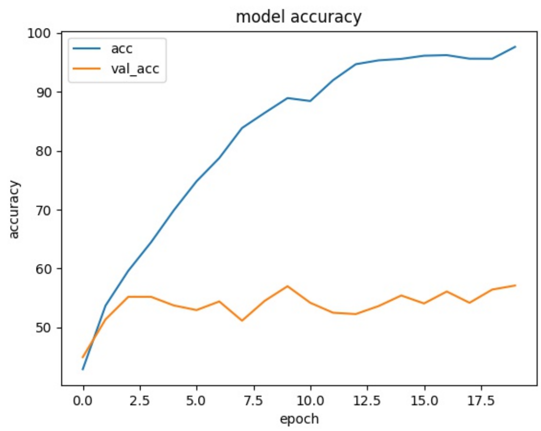
學號：r08942085 系級： 電信所碩一 姓名：陳芃彣

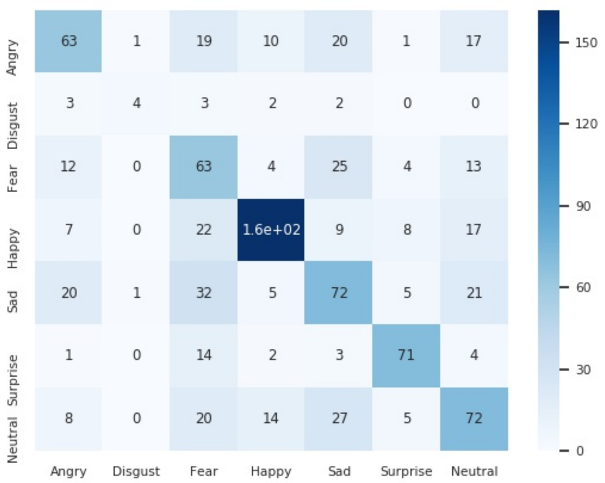
1. (1%) 請說明這次使用的**model**架構，包含各層維度及連接方式。

   我的model架構是四層convolution加一層全連接層，其中四層的convolution filter 數分別是32、64、128、128，convolution的kernel大小只有第一層是5*5，其他皆是3*3，激勵函數使用leakyReLu，此外每層我都有做batch normalize跟maxpooling。

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

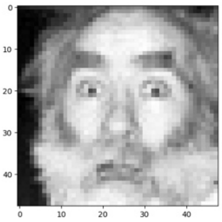
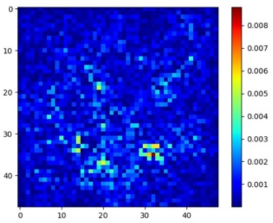


3. (1%) 畫出**confusion matrix**分析哪些類別的圖片容易使**model**搞混， 並簡單說明。



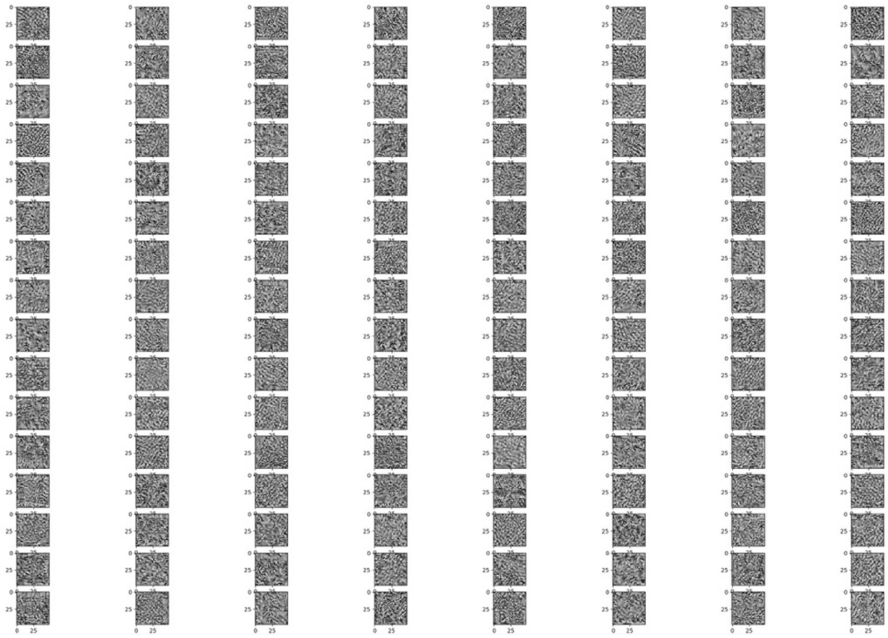
   從confusion matrix 中可以看出”Neutral”比要容易跟其他表情搞混，例如：Neutral被判別成Fear、Happy、Sad的Accuracy都稍微高了點，其中Sad判別成Fear的分數也比較高，但基本上除了Disgust之外，其他心情的判別度大多正確。

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

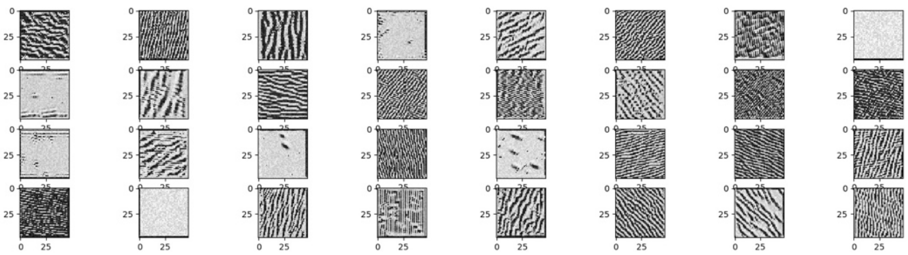
Saliency Map利用圖像分割的方式將相似的pixel分配成同個標籤，所以可以透過Saliency Map看出該圖的特徵，我隨意挑了一張圖畫了Saliency Map，可以看到眼睛的位置有些微的亮點，以及嘴巴鬍子特為明顯。

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



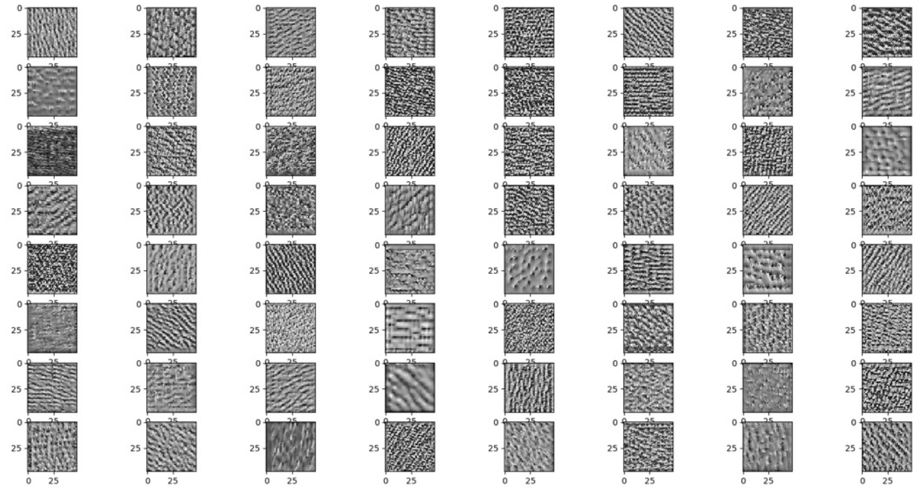
conv4

我架了四層convolution，其中最後一層filter數為128，kernel大小為3*3，filter 的feature activation如上圖所示，可以看出大多filter中央都有些微的扭曲，疑似臉蛋的樣子，但不是很顯著，因此我回頭畫了其他層的filters。

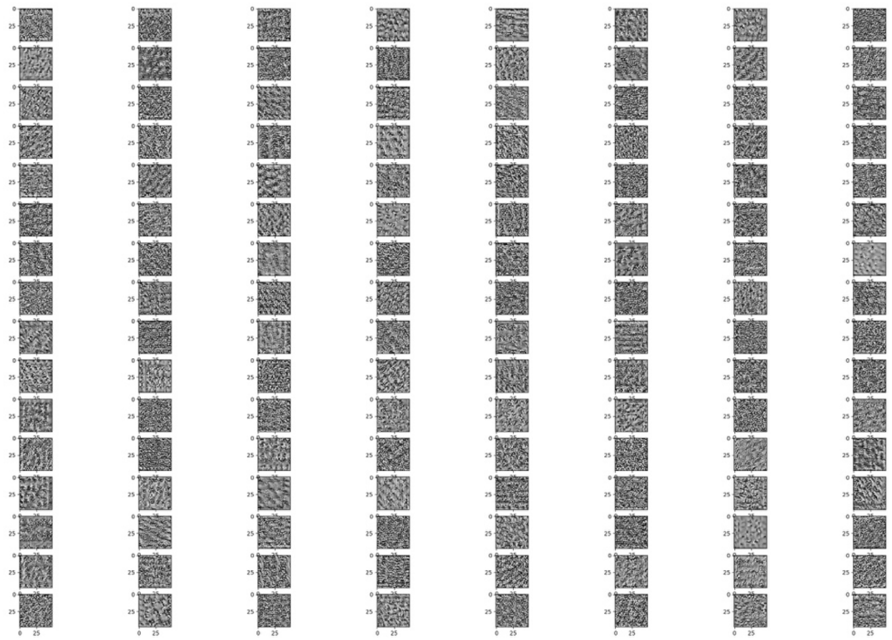


conv1

上圖為conv1的filter，總共有32個，kernelsize為5*5，可以看出conv1主要識別一些紋理圖案，有橫條紋、直條紋、波浪紋等等。



conv2



conv3

6. (3%)Refer to math problem

陳芃彣 R08942085

1. kernel size $= (k_1, k_2)$ original : (B, W, H, input_channels)

stride $= (S_1, S_2)$

padding $= (p_1, p_2)$

$$\begin{cases} \text{width}: & \text{floor}\left(\dfrac{W + 2 \times p_1 - k_1}{S_1}\right) + 1 \quad \# \\[2ex] \text{height}: & \text{floor}\left(\dfrac{H + 2 \times p_2 - k_2}{S_2}\right) + 1 \quad \# \\[2ex] \text{output\_channels}: & \text{the number of filters.} \quad \# \end{cases}$$

2. Input : Values of x over a mini-batch: $\mathcal{B} = \{x_{1 \dots m}\}$;

Parameters to be learned: γ, β

Output : $\{ y_i = BN_{\gamma,\beta}(x_i) \}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_{\mathcal{B}})^2$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \varepsilon}}$$

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

$$\begin{cases} \dfrac{\partial l}{\partial \hat{x}_i} = \dfrac{\partial l}{\partial y_i} \cdot \gamma \\[2ex] \dfrac{\partial l}{\partial \sigma_{\mathcal{B}}^2} = \displaystyle\sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} \cdot (x_i - \mu_{\mathcal{B}}) \\[2ex] \dfrac{\partial l}{\partial \mu_{\mathcal{B}}} = \left(\displaystyle\sum_{i=1}^{m} \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{-1}{\sqrt{\sigma_{\mathcal{B}}^2 + \varepsilon}} \right) + \dfrac{\partial l}{\partial \sigma_{\mathcal{B}}^2} \cdot \dfrac{\sum_{i=1}^{m} -2(x_i - \mu_{\mathcal{B}})}{m} \\[2ex] \dfrac{\partial l}{\partial x_i} = \dfrac{\partial l}{\partial \hat{x}_i} \cdot \dfrac{1}{\sqrt{\sigma_{\mathcal{B}}^2 + \varepsilon}} + \dfrac{\partial l}{\partial \sigma_{\mathcal{B}}^2} \cdot \dfrac{2(x_i - \mu_{\mathcal{B}})}{m} + \dfrac{\partial l}{\partial \mu_{\mathcal{B}}} \cdot \dfrac{1}{m} \\[2ex] \dfrac{\partial l}{\partial \gamma} = \displaystyle\sum_{i=1}^{m} \dfrac{\partial l}{\partial y_i} \cdot \hat{x}_i \\[2ex] \dfrac{\partial l}{\partial \beta} = \displaystyle\sum_{i=1}^{m} \dfrac{\partial l}{\partial y_i} \end{cases}$$

#

3. $\text{softmax}(z_t) = \dfrac{e^{z_t}}{\sum_i e^{z_i}}$

$$\text{cross\_entropy} = L(y, (\hat{y})) = -\sum_i y_i \log(\hat{y}_i)$$

$$\text{cross\_entropy} = L_t(y_t, (\hat{y}_t)) = -y_t \log(\hat{y}_t)$$

$$\hat{y}_t = \text{softmax}(z_t)$$

Derive that $\dfrac{\partial L_t}{\partial z_t} = \hat{y}_t - y_t$

(sol)

$$L_t(y_t, \hat{y}_t) = -y_t \log\left(\frac{e^{z_t}}{\sum_i e^{z_i}}\right) = -y_t\left(\log e^{z_t} - \log \sum_i e^{z_i}\right)$$

$$\frac{\partial L_t}{\partial z_t} = \frac{\partial\left[-y_t \log(\hat{y}_t)\right]}{\partial z_t} = -y_t \cdot \frac{1}{\hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} = -y_t \cdot \frac{1}{\hat{y}_t} \cdot \frac{\partial}{\partial z_t}\left(\frac{e^{z_t}}{\sum_i e^{z_i}}\right)$$

$\because \hat{y}_t = \text{softmax}(z_t)$

$$= -\frac{y_t}{\hat{y}_t}\left[\frac{e^{z_t}\left(\sum_i e^{z_i}\right) - e^{z_t} \cdot e^{z_t}}{\left(\sum_i e^{z_i}\right)^2}\right]$$

$$\hat{y}_t = \frac{e^{z_t}}{\sum_i e^{z_i}}$$

$$= \frac{-y_t}{\hat{y}_t}\left[\hat{y}_t - (\hat{y}_t)^2\right]$$

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

$\because y_t = 1 \quad \therefore -y_t + \hat{y}_t \cdot y_t = \hat{y}_t - y_t$ #

if $y_t = 0 \rightarrow \hat{y}_t - y_t$ #