HW3 Report

學號：B05902017 系級： 資工四 姓名：王盛立

1 請說明這次使用的model架構，包含各層維度及連接方式

這次我總共用了7層的Convolution + 3層的Fully Connected

每一層Convolution都有加BatchNormalizaion，activation function都是LeakyRelu， kernel size都是3x3，stride和padding都是1

- filters = 16
- filters = 32 + maxpooling(2,2) + Dropout(0.1)
- filters = 64 + Dropout(0.2)
- filters = 128 + maxpooling(2,2) + Dropout(0.3)
- filters = 256 + maxpooling(2,2) + Dropout(0.3)
- filters = 512 + maxpooling(2,2) + Dropout(0.5)
- filters = 512 + maxpooling(2,2) + Dropout(0.5)

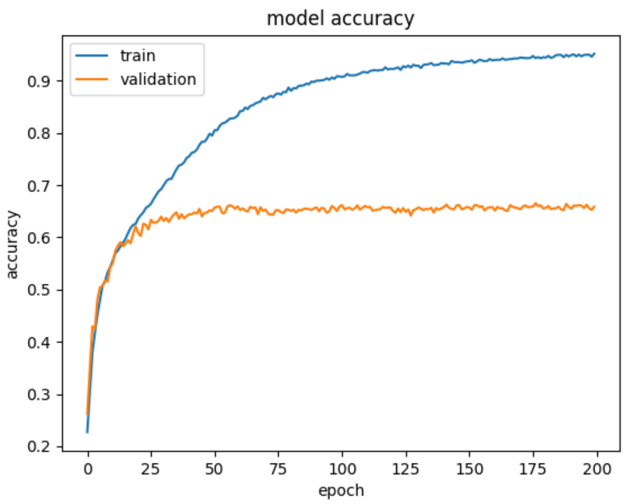
Fully Connected的部分，activation function也是LeakyRelu

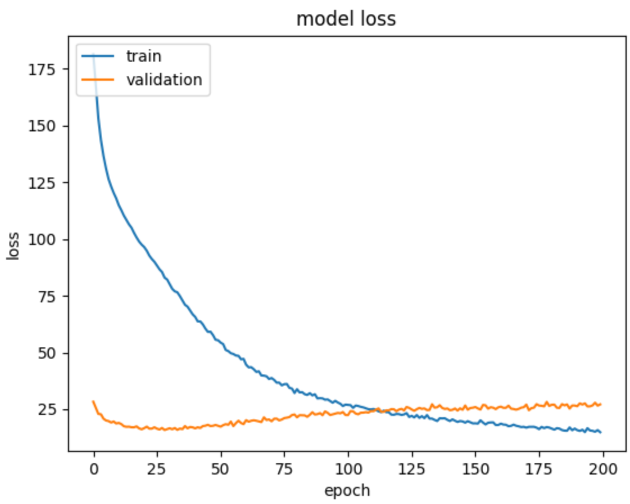
- nn.Linear(512, 256) + Dropout(0.7)
- nn.Linear(256, 128) + Dropout(0.7)
- nn.Linear(128, 7)

另外在我最後作業的testing中，我是用了許多微調一些參數的model來做ensemble來達成比較高的accuracy。

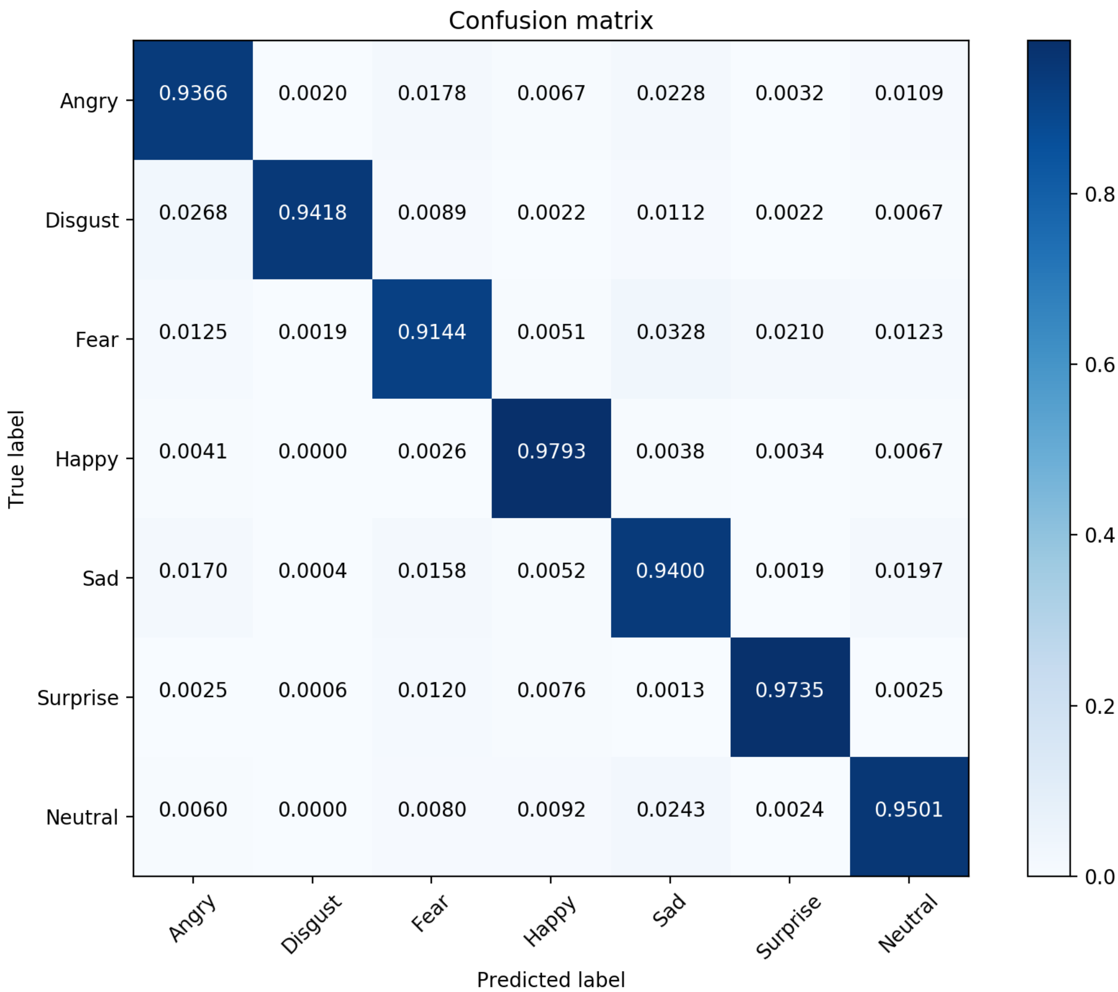
2 請附上model的training/validation history (loss and accuracy)

上圖是model在training和validation上的accuracy，下圖是model在training和validation上的loss，可以看到training的accuracy是不停上升的，但是validation到一定的程度後就停滯了。而training的loss也是不停下降，但是validation到一定程度後反而會慢慢上升。





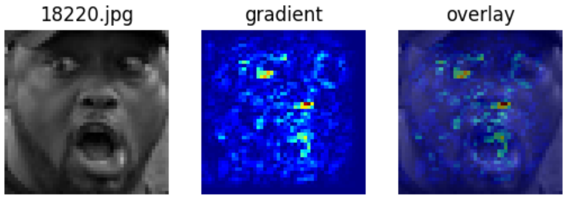
3 畫出confusion matrix分析哪些類別的圖片容易使model搞混，並簡單說明



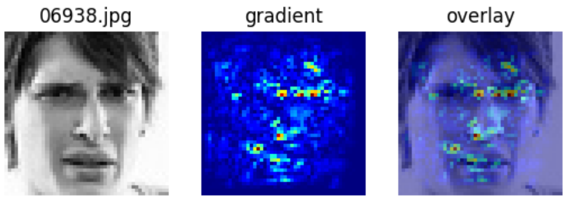
從圖中可以觀察到恐懼很容易被誤認為其他的情緒，像是傷心或是生氣。而驚訝只容易跟恐懼搞混，另外開心和其他表上的情緒都不太一樣，是最不容易被搞混的情緒。

4 畫出CNN model的saliency map，並簡單討論其現象

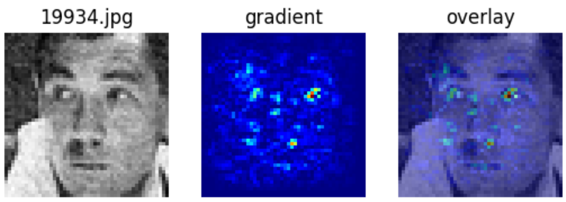
可以看到眼睛和嘴巴的形狀滿容易被認出來的，如果有牙齒的話牙齒的部分也會很明顯。鼻子周圍的皺紋也是容易被觀察到的區域。



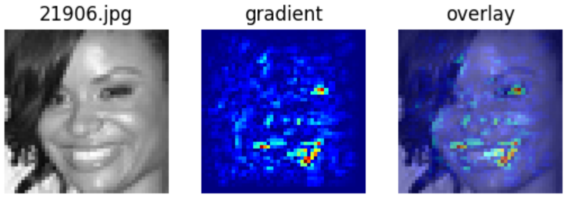
Angry



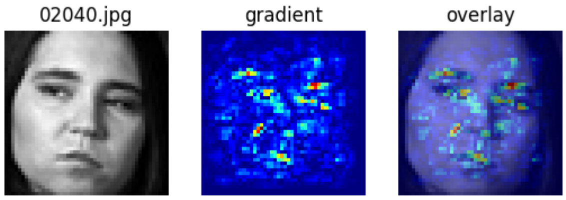
Disgust



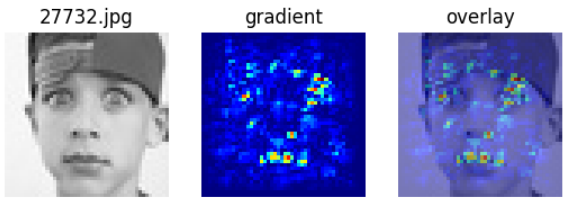
Fear



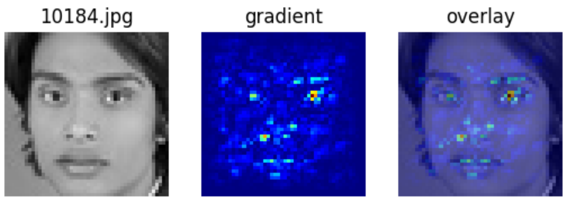
Happy



Sad



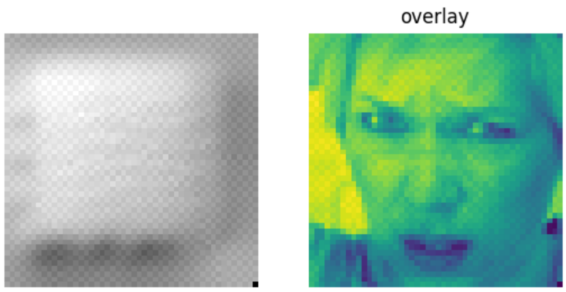
Surprise

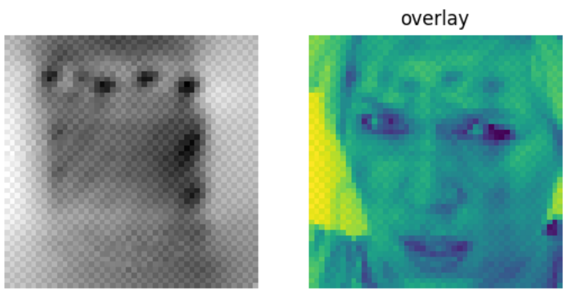


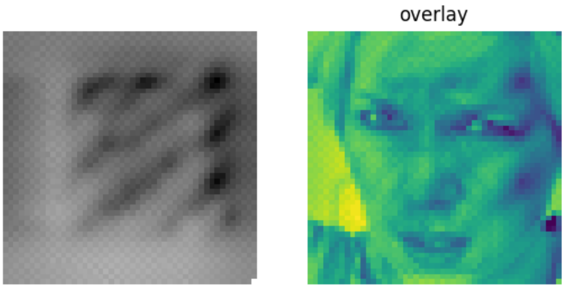
Neutral

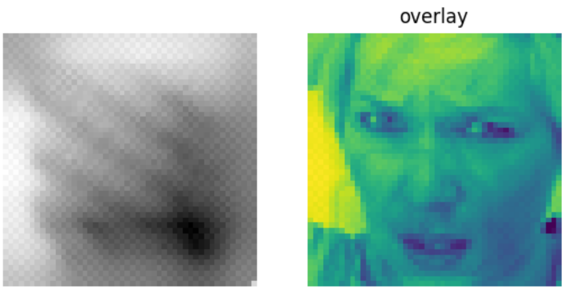
5 畫出最後一層的filters最容易被哪些feature activate

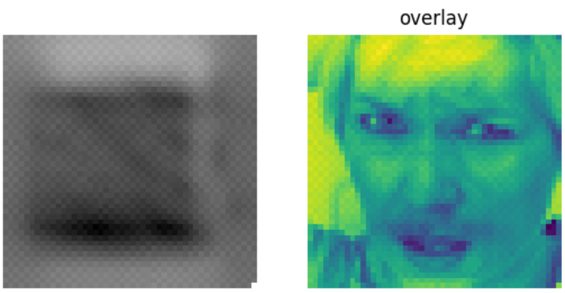
可以看到每一個filter可以學到不同的紋路，有些是斜的線條，有的則比較像棋盤狀的結構。有的則是特定區域的一條線。











6 Refer to math problem

1 Convolution

$$H_{out} = \lfloor \frac{H_{in} + 2*padding[0] - dilation[0]*(kernel\_size[0]-1)-1}{stride[0]} + 1 \rfloor$$

$$W_{out} = \lfloor \frac{W_{in} + 2*padding[1] - dilation[1]*(kernel\_size[1]-1)-1}{stride[1]} + 1 \rfloor$$

$(B, W, H, input\_channels)$ 經過
$Conv2D\ (input\_channels,\ output\_channels,\ kernel\_size = (k_1,\ k_2),$
$\quad stride = (s_1,\ s_2),\ padding = (p_1,\ p_2))$
$\Rightarrow (B, \lfloor \frac{W+2p_2-k_2}{s_2} + 1 \rfloor, \lfloor \frac{W+2p_1-k_1}{s_1} + 1 \rfloor, output\_channels)$

2 Batch Normalization

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

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

$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}}$

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

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

$\frac{\partial l}{\partial \sigma_\beta^2} = \sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i} \cdot (x_i - \mu_\beta) \cdot \frac{-1}{2}(\sigma^2 + \epsilon)^{-3/2}$

$\frac{\partial l}{\partial \mu_\beta} = (\sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma_\beta^2 + \epsilon}}) + \frac{\partial l}{\partial \sigma_\beta^2} \cdot \frac{\sum_{i=1}^{m} -2(x_i - \mu_\beta)}{m}$

$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i} \cdot \frac{1}{\sqrt{\sigma_\beta^2 + \epsilon}} + \frac{\partial l}{\partial \sigma_\beta^2} \cdot \frac{2(x_i - \mu_\beta)}{m} + \frac{\partial l}{\partial \mu_\beta} \cdot \frac{1}{m}$

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

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

3 Softmax and Cross Entropy

$L_t(y_t \hat{y}_t) = -y_t log \hat{y}_t$ is the cross entropy when y_t = 1

$\frac{\partial L_t}{\partial z_t} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial z_t} = \frac{\partial}{\partial \hat{y}_t}(-y_t log \hat{y}_t)\frac{\partial}{\partial z_t}\frac{e^{zt}}{\Sigma_i e^{zi}} = -\frac{y_t}{\hat{y}_t}\frac{e^{zt}\Sigma_i e^{zi} - e^{zt}e^{zt}}{(\Sigma_i e^{zi})^2} = -\frac{y_t}{\hat{y}_t}\frac{e^{zt}}{\Sigma_i e^{zi}}(1 - \frac{e^{zt}}{\Sigma_i e^{zi}}) = -\frac{y_t}{\hat{y}_t}\hat{y}_t(1 - \hat{y}_t)$
$= y_t \hat{y}_t - y_t = \hat{y}_t - y_t$

when y_t = 0, cross entropy is $L_t(y_t \hat{y}_t) = -(1 - y_t)log(1 - \hat{y}_t)$

$\frac{\partial L_t}{\partial z_t} = \frac{\partial L_t}{\partial \hat{y}_t}\frac{\partial \hat{y}_t}{\partial z_t} = \frac{\partial}{\partial \hat{y}_t}(-(1 - y_t)log(1 - \hat{y}_t))\frac{\partial}{\partial z_t}\frac{e^{zt}}{\Sigma_i e^{zi}} = \frac{1-y_t}{1-\hat{y}_t}\frac{e^{zt}\Sigma_i e^{zi} - e^{zt}e^{zt}}{(\Sigma_i e^{zi})^2} = \frac{1-y_t}{1-\hat{y}_t}\frac{e^{zt}}{\Sigma_i e^{zi}}(1 - \frac{e^{zt}}{\Sigma_i e^{zi}}) = \frac{1-y_t}{1-\hat{y}_t}\hat{y}_t(1 - \hat{y}_t)$
$= \hat{y}_t - y_t \hat{y}_t = \hat{y}_t - y_t$