姓名:王嚴 學號:r07323035 系所:經濟所碩二

1.

使用 CNN 總共做了四次 convolution 和 maxpooling

每一層都使用 leakyReLu(0.2)和做標準化

第一層 conv1 使用了 32 個 features, kernel_size=5, padding = 2 maxpooling = 2

48*48 -> ((48 + 4) – 4)*((48 + 4) – 4) -> 32*24*24

第二層 conv2(32, 64, 3, 1), maxpool = 2

32*24*24 -> 64*12*12

第三層 conv3(64, 128, 3, 1), maxpool = 2

64*12*12 -> 128*6*6

第四層 conv4(128, 128, 3, 1), maxpool = 2

128*6*6 -> 128*3*3

在做兩層的 fully connected 並做 leakyReLU(0.2)和 dropout(0.5)

做之前先把 128*3*3 拉長 -> (-1, 128*3*3)

第一層 fc(128*3*3, 256)

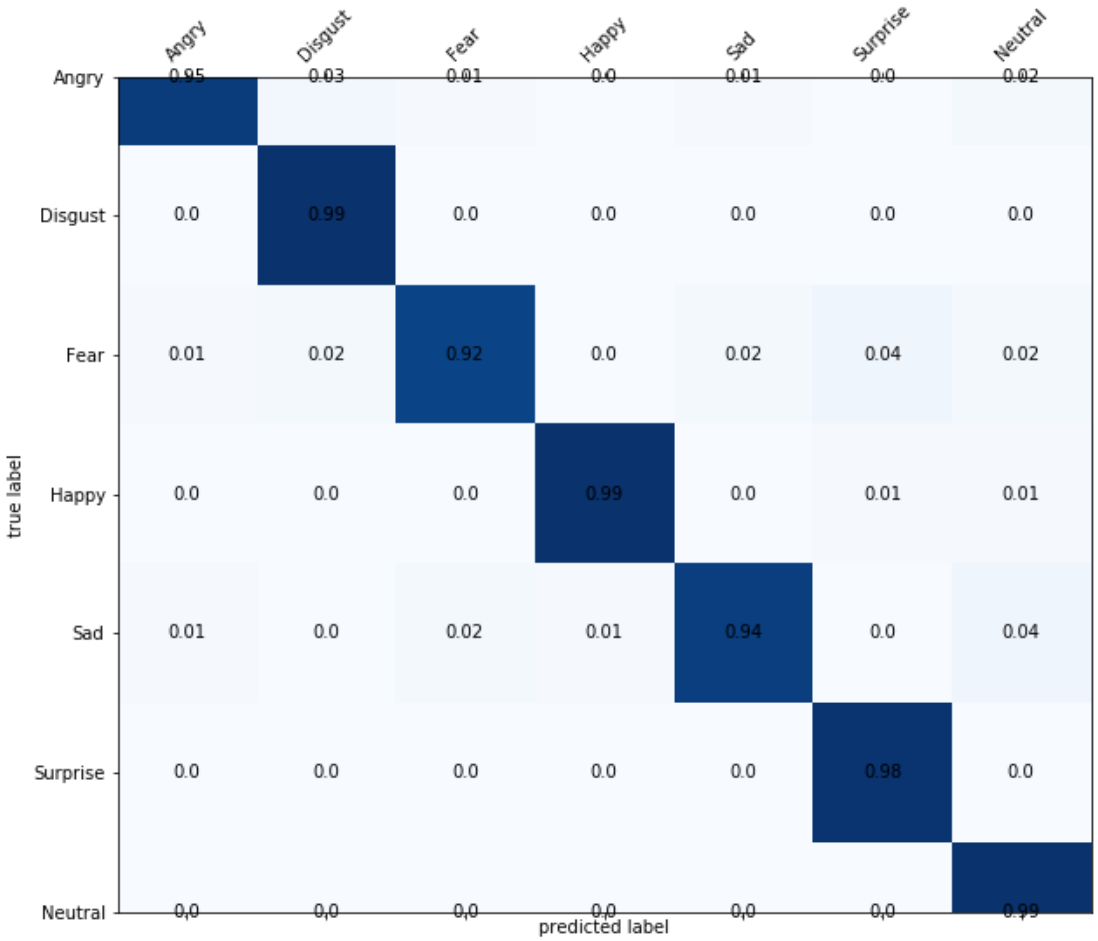
第二層 fc(256, 7)

2.

```
Epoch: 1, train Loss: 1.5787, train Acc: 0.3863
Epoch: 1, valid Loss: 1.2766, valid Acc: 0.5014
Epoch: 2, train Loss: 1.2630, train Acc: 0.5192
Epoch: 2, valid Loss: 1.1329, valid Acc: 0.5588
Epoch: 3, train Loss: 1.1340, train Acc: 0.5701
Epoch: 3, valid Loss: 0.9376, valid Acc: 0.6456
Epoch: 4, train Loss: 1.0171, train Acc: 0.6173
Epoch: 4, valid Loss: 0.9152, valid Acc: 0.6645
Epoch: 5, train Loss: 0.9085, train Acc: 0.6576
Epoch: 5, valid Loss: 0.7597, valid Acc: 0.7205
Epoch: 6, train Loss: 0.7935, train Acc: 0.7043
Epoch: 6, valid Loss: 0.5827, valid Acc: 0.7854
Epoch: 7, train Loss: 0.6594, train Acc: 0.7554
Epoch: 7, valid Loss: 0.4332, valid Acc: 0.8516
Epoch: 8, train Loss: 0.5399, train Acc: 0.7993
```

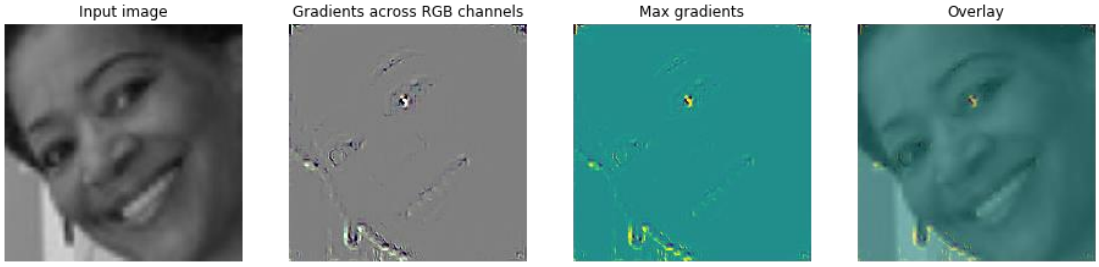
```
Epoch: 8, valid Loss: 0.5642, valid Acc: 0.7983
Epoch: 9, train Loss: 0.4348, train Acc: 0.8425
Epoch: 9, valid Loss: 0.3556, valid Acc: 0.8665
Epoch: 10, train Loss: 0.3380, train Acc: 0.8768
Epoch: 10, valid Loss: 0.2374, valid Acc: 0.9193
Epoch: 11, train Loss: 0.2750, train Acc: 0.9016
Epoch: 11, valid Loss: 0.1907, valid Acc: 0.9290
model saved to model_11.pkl
Epoch: 12, train Loss: 0.2240, train Acc: 0.9218
Epoch: 12, valid Loss: 0.2163, valid Acc: 0.9236
model saved to model_12.pkl
Epoch: 13, train Loss: 0.1875, train Acc: 0.9352
Epoch: 13, valid Loss: 0.1095, valid Acc: 0.9626
model saved to model_13.pkl
Epoch: 14, train Loss: 0.1618, train Acc: 0.9448
Epoch: 14, valid Loss: 0.1052, valid Acc: 0.9671
model saved to model_14.pkl
```

3.



大致上分得不錯 Fear 相對來說最容易混淆

4.



可以看到圖片的主要特徵可能是眼睛和嘴巴

直覺上從這兩者也較容易判斷情緒

6. math problem

Convolution:

$W \to W + 2P_1 \to \frac{W+2P_1-K_1}{S_1} + 1$

$(B, W, H) \to (B, W^{'}, K^{'})$

$w^{'} = \frac{W+2P_1-K_1}{S_1} + 1$

$K^{'} = \frac{K+2P_2-K_2}{S_2} + 1$

$channels = output\_channel$

Batch Normalization:

$$L = \frac{1}{m}\sum_i (y_i - \gamma\widehat{x_i} - \beta)^2$$

$$\frac{\partial L}{\partial \widehat{x_i}} = \frac{-2\gamma}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)$$

$$\frac{\partial L}{\partial \sigma_B^2} = \frac{\partial L}{\partial \widehat{x_i}}\frac{\partial \widehat{x_i}}{\partial \sigma_B^2}$$

$$= \frac{\gamma}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)(\frac{x_i - \mu_B}{(\sigma_B^2 + \epsilon)^{1.5}})$$

$$\frac{\partial L}{\partial \mu_B} = \frac{\partial L}{\partial \widehat{x_i}}\frac{\partial \widehat{x_i}}{\partial \mu_B}$$

$$= \frac{2\gamma}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)(\frac{1}{\sqrt{\sigma_B^2 + \epsilon}})$$

$$\frac{\partial L}{\partial x_i} = \frac{\partial L}{\partial \widehat{x_i}}\frac{\partial \widehat{x_i}}{\partial x_i}$$

$$= \frac{-2\gamma}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)(\frac{1}{\sqrt{\sigma_B^2 + \epsilon}})$$

$$\frac{\partial L}{\partial \gamma} = \frac{-2}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)(\widehat{x_i})$$

$$\frac{\partial L}{\partial \beta} = \frac{-2}{n}\sum_i (y_i - \gamma\widehat{x_i} - \beta)$$

Softmax and Cross Entropy :

$$\frac{\partial L_t}{\partial Z_t} = \frac{\partial(-y_t \ln \widehat{y_t})}{\partial Z_t}$$

$$= \frac{\partial - y_t(Z_t - \ln \sum_i \exp Z_i)}{\partial Z_t}$$

$$= -y_t + \frac{\exp Z_t}{\sum_i \exp Z_i}$$

$$= \widehat{y_t} - y_t$$