ML 2019 Hw3 Report

學號：b05705042 系級：資管四 姓名：皇甫立翔

1. 請說明這次使用的 model 架構，包含各層維度及連接方式。

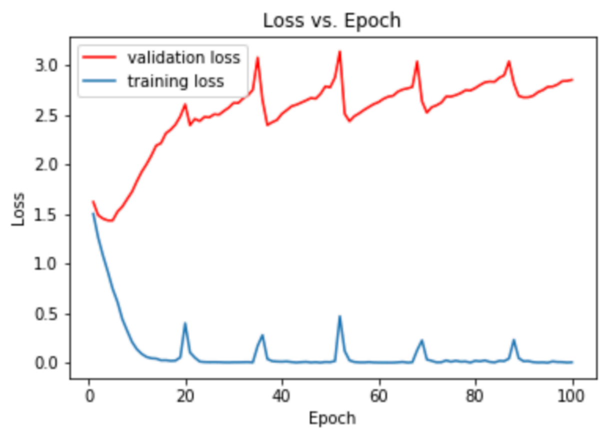
   這次用的架構是 CNN + Linear (dense) 堆疊起來的。

   連續疊了4層 CNN 之後，接著攤平後再疊上三層 Fully connected linear layer，其中最後一層是 output layer。 (每一層 CNN Max Pooling 都是 2)。

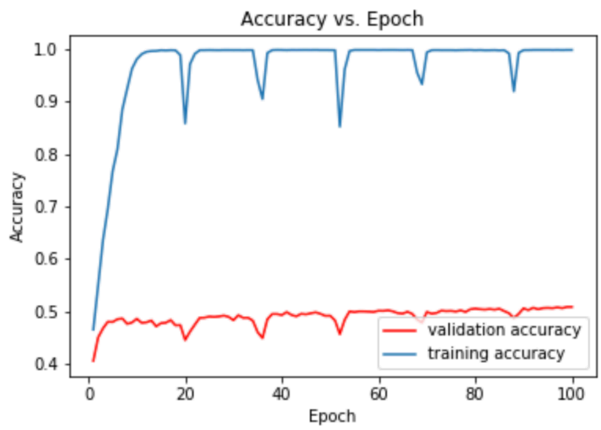
   第一層 CNN：kernel size = 5，output channel = 32，padding = 2
   第二層 CNN：kernel size = 3，output channel = 64，padding = 1
   第三層 CNN：kernel size = 3，output channel = 128，padding = 1
   第四層 CNN：kernel size = 3，output channel = 128，padding = 1
   第一層 Linear：input channel = 3*3*128，output channel = 256
   第二層 Linear：input channel = 256，output channel = 128
   第三層 Linear (Output)：input channel = 128，output channel = 7

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

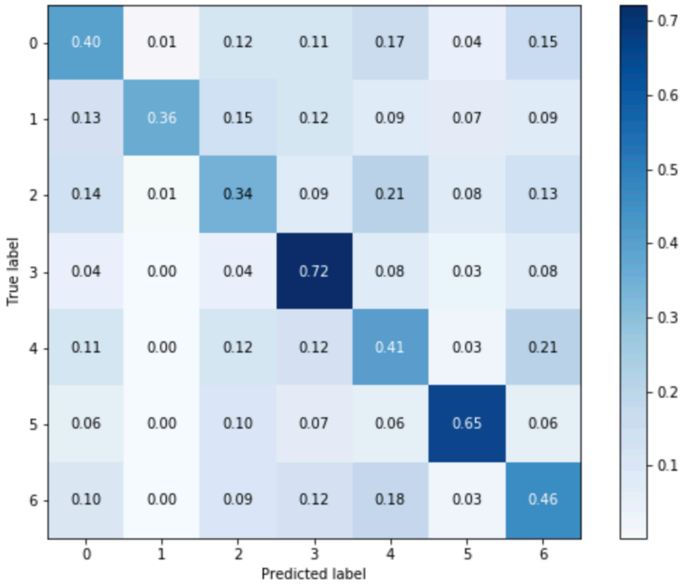
   Loss



   Accuracy



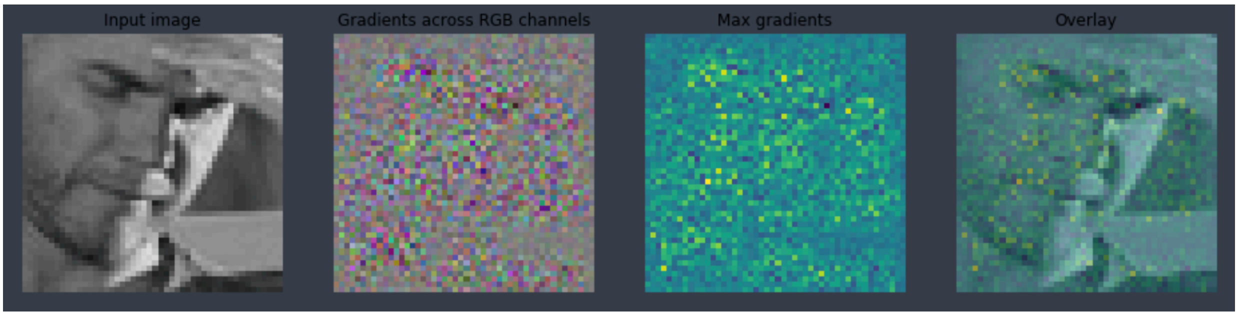
3. 畫出 Confusion Matrix 分析哪些類別的圖片容易使 model 搞混，並簡單說明。



0 = Angry　1 = Disgust　2 = Fear　3 = Happy　4 = Sad　5 = Surprise　6 = Neutral

經由 Confusion Matrix 可以看得出來，(1) Fear vs. Sad (2) Fear vs. Neutral 的圖片，相較其他類別來說，比較容易使 model 搞混，他們在 matrix 中的數字比較大，意即有較高機率使 model 搞混。

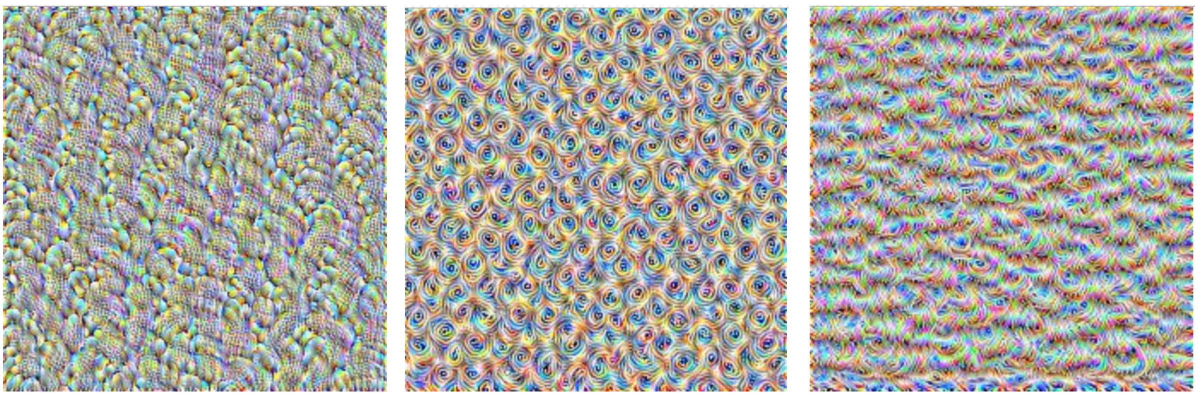
4. 畫出 CNN model 的 saliency map，並簡單討論其現象。



input　gradients　max gradients　overlay

顏色較顯眼的部分即是 feature 影響較大的地方。

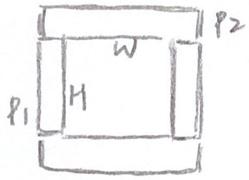
5. 畫出最後一層的 filters 最容易被哪些 feature activate。



6. Refer to math problem。

605705042
皇甫立翔

1、原本的 shape：(B, W, H, input_channels)



寬：$\dfrac{(W+2\cdot P_1)-k_1}{s_1}+1$

長：$\dfrac{(H+2\cdot P_2)-k_2}{s_2}+1$

Ans：經過 convolution layer 後，shape 變成 $\left(B, \dfrac{(W+2\cdot P_1)-k_1}{s_1}+1, \dfrac{(H+2\cdot P_2)-k_2}{s_2}+1, \text{output\_channels}\right)$ *

2、

(1) $\dfrac{\partial l}{\partial \hat{x}_i} = \dfrac{\partial l}{\partial y_i}\gamma$ *

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

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

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

(5) $\dfrac{\partial l}{\partial \gamma} = \sum_{i=1}^{m}\dfrac{\partial l}{\partial y_i}\cdot\hat{x}_i$ *

(6) $\dfrac{\partial l}{\partial \beta} = \sum_{i=1}^{m}\dfrac{\partial l}{\partial y_i}$ *

3、

L_t 為 L 當中 $y_t=1$ 的項，在這邊直接對 L 微分。

$L = -\sum_i y_i \log(\hat{y}_i)$

$\dfrac{\partial L}{\partial z_t} = \dfrac{\partial -\sum_i y_i \log(\hat{y}_i)}{\partial z_t} = -\sum_i y_i \dfrac{1}{\hat{y}_i}\dfrac{\partial \hat{y}_i}{\partial z_t}$

根據 softmax 微分公式可轉化為：$= -y_t(1-\hat{y}_t) - \sum_{i\neq t} y_i \dfrac{1}{\hat{y}_i}(-\hat{y}_i\hat{y}_t)$

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

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

$= \hat{y}_t\cdot\sum_i y_i - y_t$，其中 $\sum_i y_i = 1$

$\Rightarrow \dfrac{\partial L}{\partial z_t} = \hat{y}_t - y_t$ 得證 *