

Homework3 Report – Image Sentiment Classification

B04703117 財金五 謝昊辰

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

(1) Conv2d : input channel: 3, output channel: 64

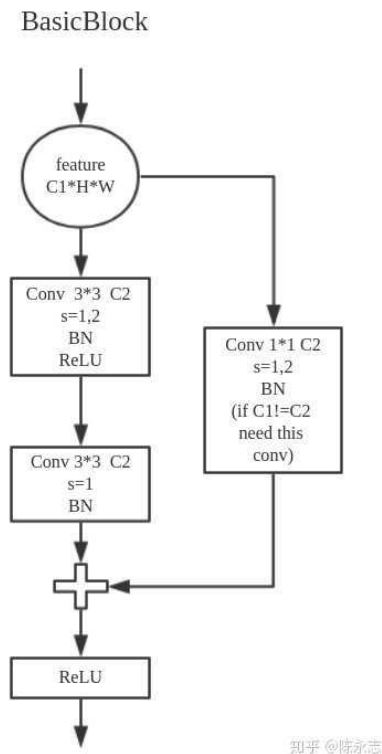
(2) BatchNorm2d : input channel: 64

(3) ReLU

(4) MaxPool2d

接著會經過 4 個由兩個 BasicBlock 組成的 Sequential (需要調用 class ResNet() 中的 function _make_layer())

→ BasicBlock : 一個 BasicBlock 會經過兩個 Conv2d 和 BatchNorm2d 以及一次 ReLU，並將其輸出與原本輸入相加後再經過一次 ReLU function。



→ 第一個 Sequential : input channel: 64, output channel: 64

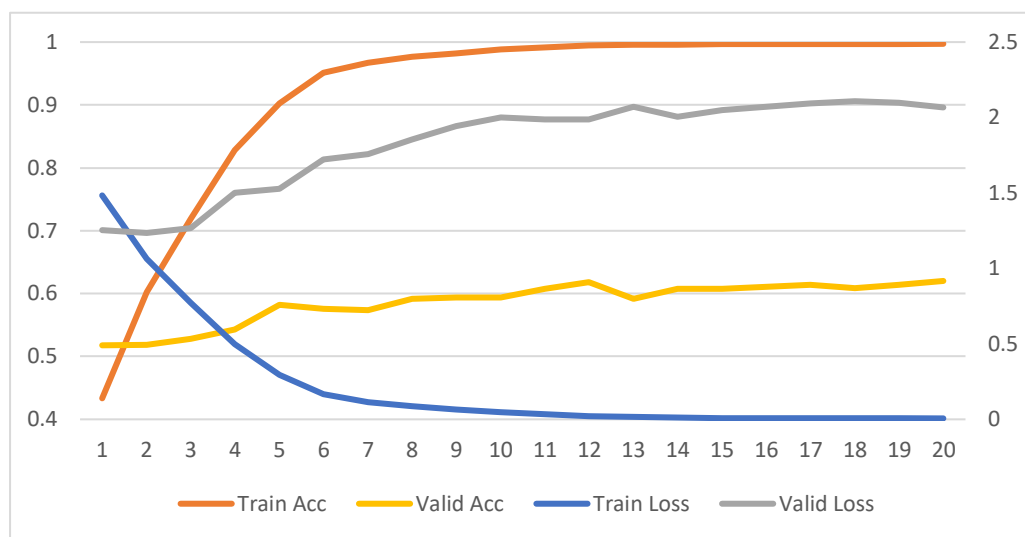
→ 第二個 Sequential : input channel: 64, output channel: 128

→ 第三個 Sequential : input channel: 128, output channel: 256

→ 第四個 Sequential : input channel: 256, output channel: 512

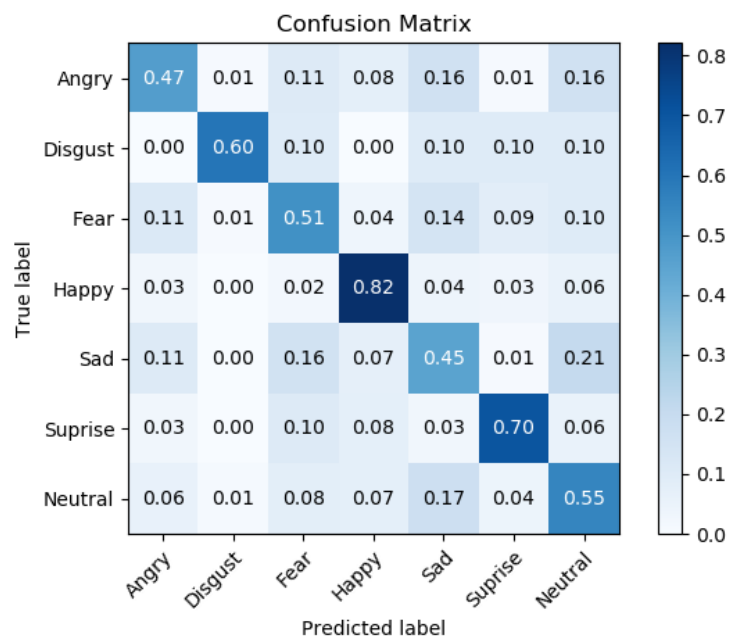
由於第一個 Sequential 中第一個 BasicBlock 的 output channel 和第二個 BasicBlock 的 input channel 皆為 64，不需要調用 downsample 參數，其他的 Sequential 皆需要調用 downsample 參數來使第一個 BasicBlock 的 output channel 和第二個 BasicBlock 的 input channel 一致。

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



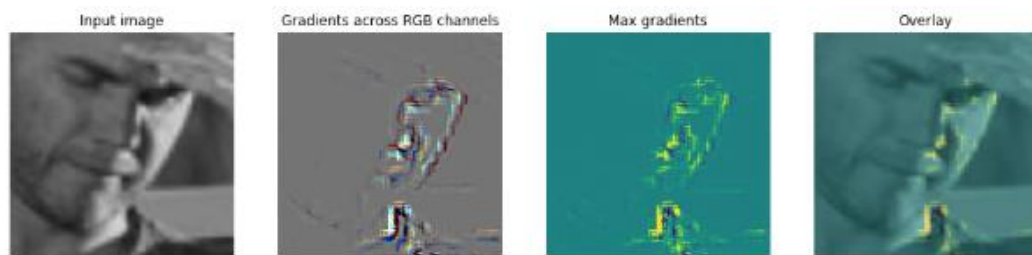
如上圖，左邊為 accuracy 刻度，右邊為 loss 刻度，可以看到在第 10-20 個 epoch 時 training accuracy 就已經接近 100% 了，但 valid accuracy 仍然卡在 60% 左右。training loss 的部分也是在第 10-20 個 epoch 就已經接近 0 了，但 valid loss 卻持續的上升。

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。



此圖為 valid set 所畫出的 confusion matrix，由圖可以看到 Angry、Fear、Sad 和 Neutral 的圖片的準確率較低，可以推斷這四個表情的圖片可能有比較類似的 element 使得機器在學習時搞混，其中又可以看到 Sad 和 Neutral 互相判斷錯誤的比例最高，表示機器在學習時這兩類的圖片是看起來最相近的。

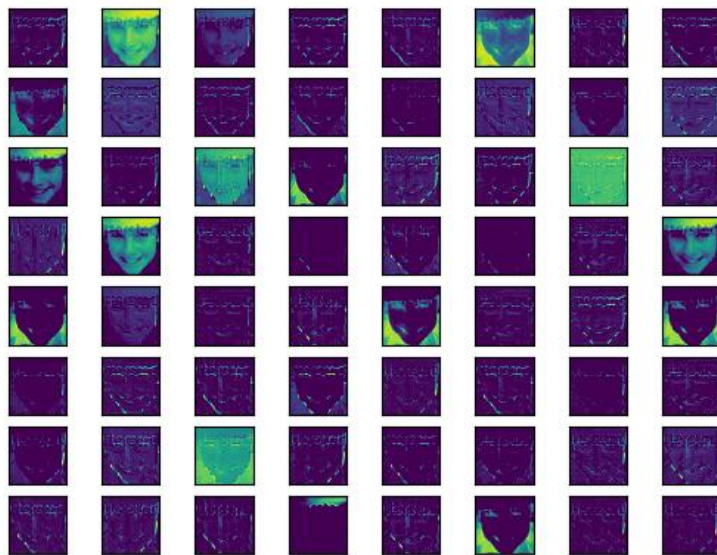
4. (1%) 畫出 CNN model 的 saliency map，並簡單討論其現象。



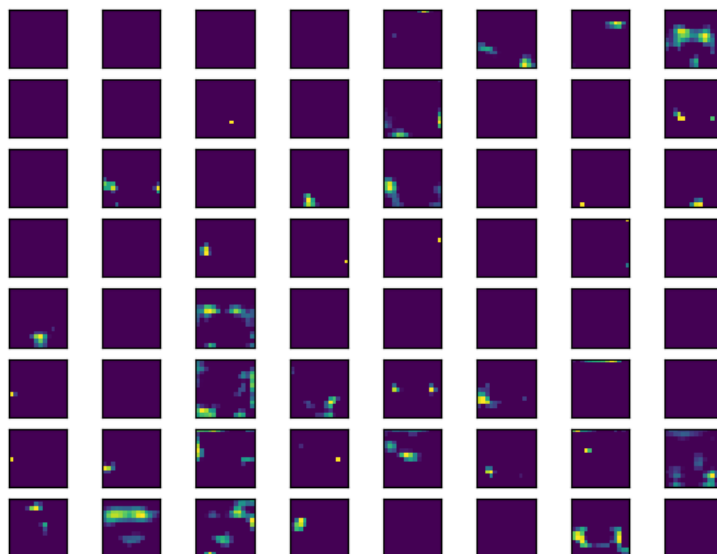
上圖是由"00000.jpg"所畫出的 saliency map，可以看到人臉的輪廓以及一些光影色差較大的地方對於機器是比較顯著的。

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

以下的 model 為 VGG16。第一層 filter：



最後一層 filter：



6. New shape: $(W + p_1 - k_1 + 1)/s_1 * (H + p_2 - k_2 + 1)/s_2$

7.

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

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

$$\frac{\partial \ell}{\partial \mu_\beta} = \left(\sum_{i=1}^m \frac{\partial \ell}{\partial \hat{x}_i} \cdot \frac{-1}{\sqrt{\sigma^2_\beta + \epsilon}} \right) + \frac{\partial \ell}{\partial \sigma^2_\beta} \cdot \frac{\sum_{i=1}^m -2(x_i - \mu_\beta)}{m}$$

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

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

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

8.

$$\frac{\partial L_t}{\partial z_t} = \frac{\partial L_t}{\partial \hat{y}_t} \times \frac{\partial \hat{y}_t}{\partial z_t} = \left(-y_t \times \frac{1}{\hat{y}_t} \right) \times \hat{y}_t (1 - \hat{y}_t) = -y_t \times (1 - \hat{y}_t)$$

$$\because y_t = 1, \therefore \frac{\partial L_t}{\partial z_t} = \hat{y}_t - y_t$$