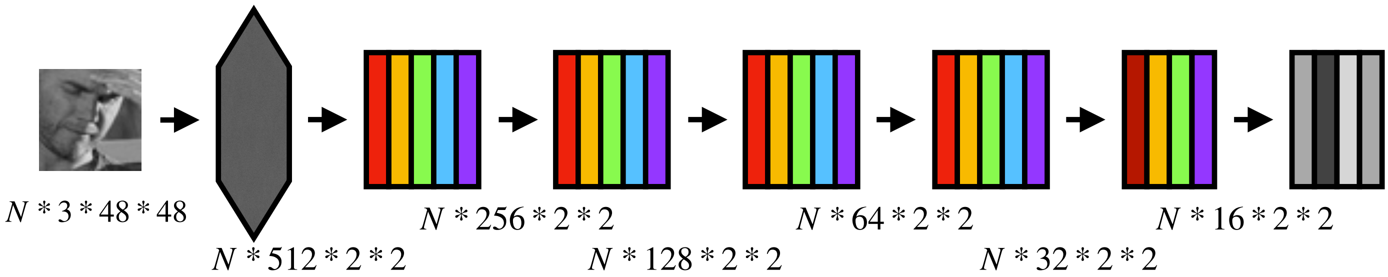
<u>ML Homework #3</u> 學號：B0902120 系級： 資工四 姓名：曾鈺婷

1. 請說明這次使用的 model 架構，包含各層維度的連接方式。

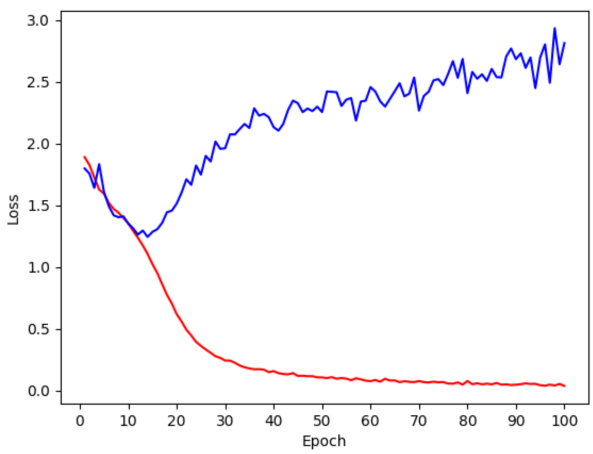
六角深灰色是 Resnet18 (Pretrained = True)

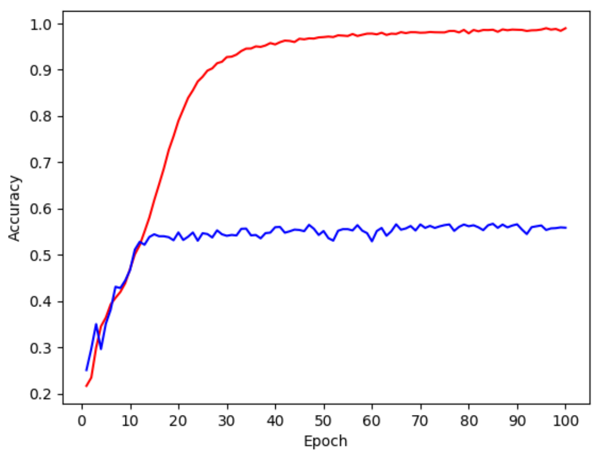
紅色是 ConvTranspose2d	酒紅是 Conv	黃色是 LeakyReLU
綠色是 BatchNorm2d	藍色是 MaxPool2d	紫色是 Dropout
淺灰色是 Linear	深灰色是 ReLU	白色是 BatchNorm1d



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

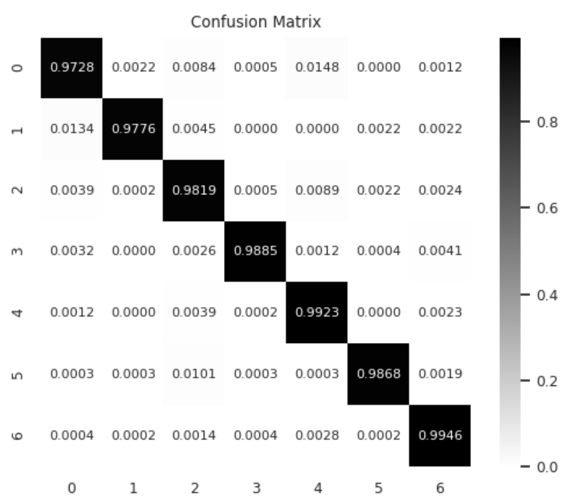
我們可以發現 training and validation data 的 loss 分別收斂在 0.05 與 2.54 左右，這很有可能代表著 overfitting，但 public score 依舊挺高的所以還是決定用 100 個 epoch。另一個指標是 accuracy，他們分別落在 0.9 與 0.5 多，這也印證前述的說法，或許該改用 cross validation 會是比較好的選擇。（紅色是 training、藍色是 validation）



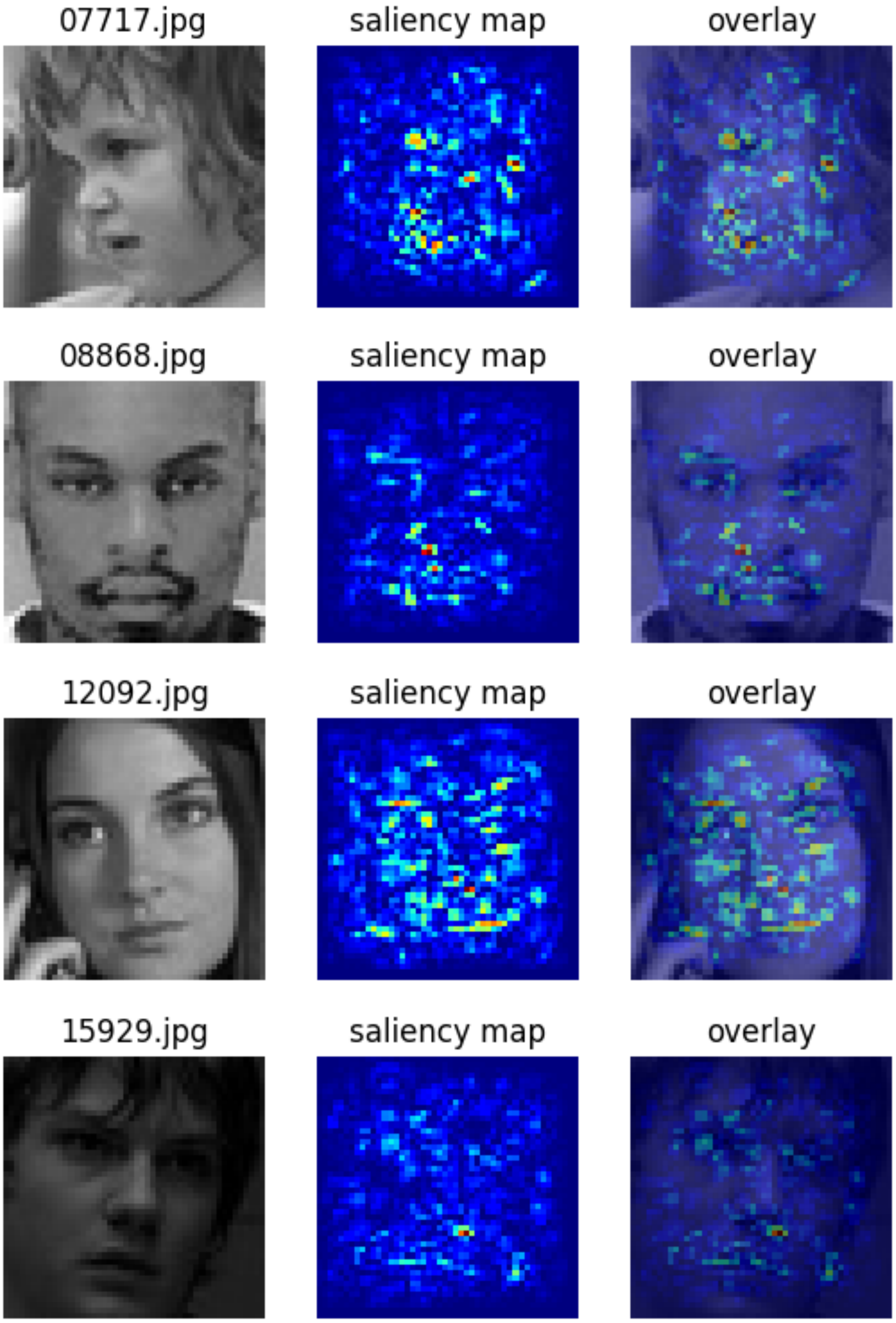


3. 畫出 confusion matrix 分析哪些類別的圖片容易使 model 搞混，並簡單說明。

我們可以發現其實他的混淆程度不算高，只有在 label 0（生氣）及 label 1（厭惡）的時候表現比較不佳。甚至在 label 4（難過）與 label 6（中立）的時候有高達 99% 的準確度！

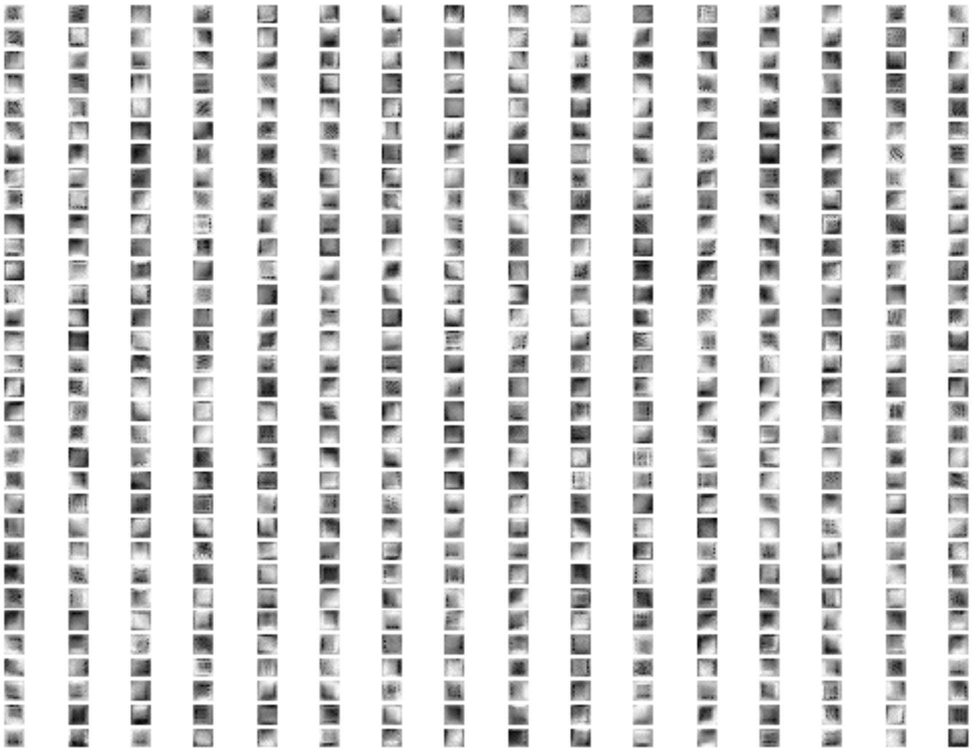


4. 畫出 CNN model 的 saliency map，並簡單討論其現象。



從上面四張圖來看，我們可以發現在眼睛及嘴巴附近呈現比較偏紅黃的顏色，這是因為他在 heatmap 上有較高的值（採用的 colormap 為 jet），這樣的現象也代表著 convolution network 主要學習的重點在這幾個地方。

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



綜觀來看，絕大部分的圖有學習到了三邊深色而中間淺色的地方，這可能代表著嘴角或眼角的部分；有一些有不同方向的直線，這可能代表他學習到線段的特定方向對於判斷是蠻重要的。

6. 手寫題

1. Convolution

以水平方向來看，經過 padding 後的寬度為 $W+p_1$，然後再開始 convolution 計算，得到最終的寬度 $W' = \lfloor \frac{W+2p_1-k_1}{s_1} \rfloor + 1$，同理可得 $H' = \lfloor \frac{W+2p_2-k_2}{s_2} \rfloor + 1$

總結來說，其大小變為 $(B,\ \lfloor \frac{W+2p_1-k_1}{s_1} \rfloor + 1,\ \lfloor \frac{W+2p_2-k_2}{s_2} \rfloor + 1,\ output\ channels)$

2. Batch Normalization

$$\frac{\partial l}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i}\frac{\partial y_i}{\partial \hat{x}_i} = \frac{\partial l}{\partial y_i}\frac{\partial}{\partial \hat{x}_i}[\gamma \hat{x}_i + \beta] = \frac{\partial l}{\partial y_i}\gamma$$

$$\begin{aligned}\frac{\partial l}{\partial \sigma_B^2} &= \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial \sigma_B^2} = \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}\frac{\partial}{\partial \sigma_B^2}[\frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\epsilon}}] \\ &= \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}[-\frac{1}{2}(x_i-\mu_B)(\sigma_B^2+\epsilon)^{-\frac{3}{2}}]\end{aligned}$$

$$\frac{\partial l}{\partial \mu_B} = \sum_{i=1}^{m}\frac{\partial l}{\partial \hat{x}_i}\frac{\partial \hat{x}_i}{\partial \mu_B} + \frac{\partial l}{\partial \sigma_B^2}\frac{\partial \sigma_B^2}{\partial \mu_B}$$

$$= \sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i} \frac{\partial}{\partial \mu_B}[\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}] + \frac{\partial l}{\partial \sigma_B^2} \frac{\partial}{\partial \mu_B}[\frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2]$$

$$= \sum_{i=1}^{m} \frac{\partial l}{\partial \hat{x}_i}[-(\sigma_B^2 + \epsilon)^{-\frac{1}{2}}] + \frac{\partial l}{\partial \sigma_B^2}[\frac{1}{m} \sum_{i=1}^{m} (-2)(x_i - \mu_B)]$$

$$\frac{\partial l}{\partial x_i} = \frac{\partial l}{\partial \hat{x}_i} \frac{\partial \hat{x}_i}{\partial x_i} + \frac{\partial l}{\partial \sigma_B^2} \frac{\partial \sigma_B^2}{\partial x_i} + \frac{\partial l}{\partial \mu_B} \frac{\partial \mu_B}{\partial x_i}$$

$$= \frac{\partial l}{\partial \hat{x}_i} \frac{\partial}{\partial x_i}[\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}] + \frac{\partial l}{\partial \sigma_B^2} \frac{\partial}{\partial x_i}[\frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2] + \frac{\partial l}{\partial \mu_B} \frac{\partial}{\partial x_i}[\frac{1}{m} \sum_{k=1}^{m} x_k]$$

$$= \frac{\partial l}{\partial \hat{x}_i}[(\sigma_B^2 + \epsilon)^{-\frac{1}{2}}] + \frac{\partial l}{\partial \sigma_B^2}[\frac{1}{m}(2)(x_i - \mu_B)] + \frac{\partial l}{\partial \mu_B}[\frac{1}{m}]$$

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \gamma} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \frac{\partial}{\partial \gamma}[\gamma \hat{x}_i + \beta] = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \hat{x}_i$$

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \frac{\partial y_i}{\partial \beta} = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i} \frac{\partial}{\partial \beta}[\gamma \hat{x}_i + \beta] = \sum_{i=1}^{m} \frac{\partial l}{\partial y_i}$$

3. <u>Softmax and Cross Entropy</u>

考量當 $y_t = 1$ 的情況：

$$\frac{\partial L_t}{\partial z_t} = \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} = \frac{\partial}{\partial \hat{y}_t}[-y_t \log \hat{y}_t] * \frac{\partial}{\partial z_t}[\frac{e^{zt}}{\sum_i e^{zi}}]$$

$$= -\frac{y_t}{\hat{y}_t} * \frac{e^{zt} \sum_i e^{zt} - e^{zt} e^{zt}}{(\sum_i e^{zt})^2}$$

$$= -\frac{y_t}{\hat{y}_t} * \frac{e^{zt}}{\sum_i e^{zt}}(1 - \frac{e^{zt}}{\sum_i e^{zt}})$$

$$= -\frac{y_t}{\hat{y}_t} * \hat{y}_t(1 - \hat{y}_t) = y_t \hat{y}_t - y_t = \hat{y}_t - y_t$$

考量當 $y_t = 0$ 的情況：

$$\frac{\partial L_t}{\partial z_t} = \frac{\partial L_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial z_t} = \frac{\partial}{\partial \hat{y}_t}[-(1 - y_t)\log(1 - \hat{y}_t)] * \frac{\partial}{\partial z_t}[\frac{e^{zt}}{\sum_i e^{zi}}]$$

$$= \frac{1 - y_t}{1 - \hat{y}_t} * \frac{e^{zt} \sum_i e^{zt} - e^{zt} e^{zt}}{(\sum_i e^{zt})^2}$$

$$= \frac{1 - y_t}{1 - \hat{y}_t} * \frac{e^{zt}}{\sum_i e^{zt}}(1 - \frac{e^{zt}}{\sum_i e^{zt}})$$

$$= \frac{1 - y_t}{1 - \hat{y}_t} * \hat{y}_t(1 - \hat{y}_t) = \hat{y}_t y_t - \hat{y}_t = \hat{y}_t - y_t$$