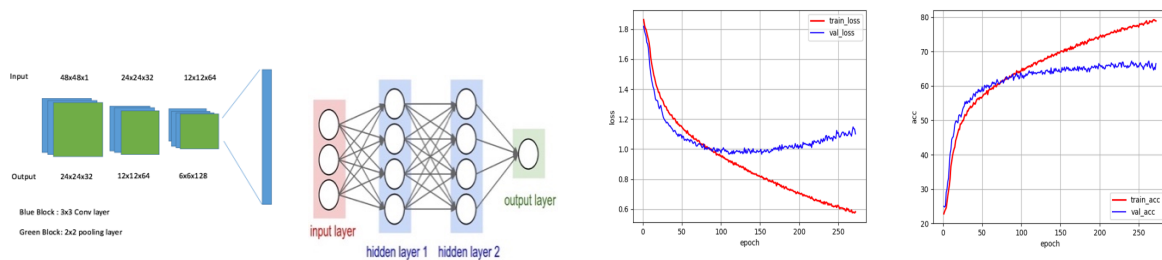


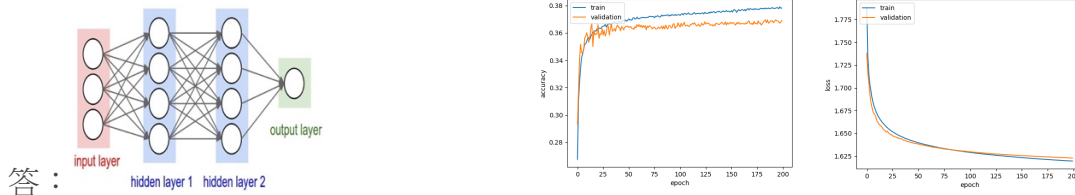
1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

(Collaborators:)

答：使用的架構如下圖所示，使用 conv 後加上 max pooling 的數量分別為 2,2,3，把最終的 feature map 拉成一個 vector 再透過 unit 數分別為 256,256,7 的 fully connected 將特徵一一分開，其中使用的最佳化方法為 Adam，其中參數分別為 learning rate = $3e-4$ dropout = 0.5，對影像另外使用 image augmentation 不僅可以增加 train 難度也可以有效抑制 overfitting，此架構及訓練方式在 kaggle 的 public 上得到 0.668 的準確率。



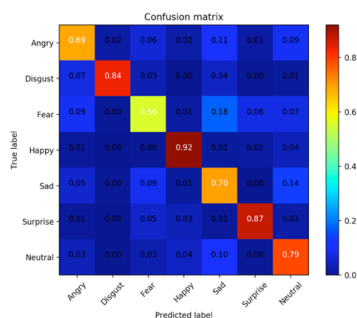
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼



是用共五層的 dense layer 數量分別為 48x48, 1024, 512, 256, 7 為防止 overfitting 所以賦予 dropout 值 0.3 而這樣的結果丟上 kaggle 後只有 0.41 的準確率，可知 CNN 應該是因為具有空間上特徵的關係所以可以得到最佳的準確性。

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？

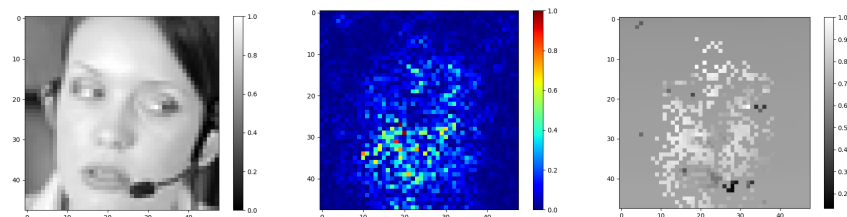
答：從 confusion matrix 可以看出 fear 和 sad 會容易分辨錯誤，此 dataset 目前人的辨識率大約為 67%，這兩個表情人眼也較難分辨，故此種結果是可以接受的



4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？

可以看出使用 CNN 的時候會將重點的部分（五官）提取出來，並將五官與表情做連結

答：



5. (1%) 承(1)(2)，利用上課所提到的 gradient ascent 方法，觀察特定層的 filter 最容易被哪種圖片 activate。

(Collaborators:陳禹齊)

