

學號：R06521601 系級：土木所碩一 姓名：黃伯凱

1. (1%) 請說明你實作的 CNN model，其模型架構、訓練參數和準確率為何？
(1).我使用的模型架構為下圖：共為兩層(CNN+BN+relu+maxpooling)再加一層隱藏層。

Layer (type)	Output Shape
conv2d_1 (Conv2D)	(None, 48, 48, 64)
batch_normalization_1 (Batch Normalization)	(None, 48, 48, 64)
conv2d_2 (Conv2D)	(None, 48, 48, 64)
max_pooling2d_1 (MaxPooling2D)	(None, 24, 24, 64)
dropout_1 (Dropout)	(None, 24, 24, 64)
conv2d_3 (Conv2D)	(None, 24, 24, 128)
batch_normalization_2 (Batch Normalization)	(None, 24, 24, 128)
conv2d_4 (Conv2D)	(None, 24, 24, 128)
max_pooling2d_2 (MaxPooling2D)	(None, 12, 12, 128)
flatten_1 (Flatten)	(None, 18432)
dense_1 (Dense)	(None, 128)
batch_normalization_3 (Batch Normalization)	(None, 128)
dropout_2 (Dropout)	(None, 128)
dense_2 (Dense)	(None, 7)

- (2).訓練參數為使用 adam(learning = 0.001), batch_size=128, epochs=250。
(3).準確率為 66%左右。

2. (1%) 請嘗試 data normalization, data augmentation,說明實行方法並且說明對準確率有什麼樣的影響？
(1).normalization 的做法一開始我最直覺得除以 255，但訓練後的精度大概只有 58%左右，後來我加了 batch_normalization，精度提高到 60%左右。
(2).augmentation 的加入讓我驚天動地了一下.....一開始我覺得兩萬多筆資料應該很充足，但仔細一看，每個表情的訓練資料並不平均，如 label 1 特別少於其他者，因此透過 augmentation 增加訓練資料後，我的精度提高到 66%。

3. (1%) 觀察答錯的圖片中，哪些 **class** 彼此間容易用混？

predict	0	1	2	3	4	5	6
label							
0	2569	7	172	244	194	102	707
1	158	177	19	19	29	10	24
2	594	4	1498	256	468	608	669
3	58	0	26	6686	31	113	301
4	477	2	364	322	2051	65	1549
5	63	1	116	177	17	2685	112
6	156	0	67	381	219	76	4066

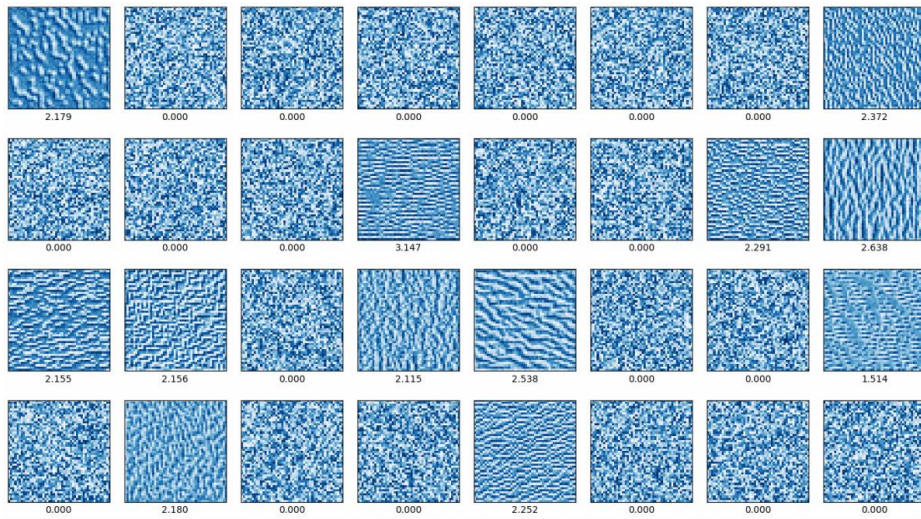
這是我訓練精度為 66% 的預測結果，我發現 label 2 和 label 4 特別容易搞混，如下圖，左邊為恐懼，右邊為難過，連我都快搞不清楚了 XD，所以也是辛苦電腦了。另外 label 1 因為訓練資料的不足，相對其他 label 也錯較多。



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？我選的圖片為 test_data 中的第二張圖，是個面帶笑容的好女孩。我的預測結果也剛好很給力的預測出 label 3 從下面三張圖中可看出，mask 幾乎能把對於表情相當重要的部分記錄下來，如眼睛、笑容以及臉頰上揚；heatmap 的中可以看出機率越高的部分呈現越高的熱度，嘴角附近即成了這張圖片預測的關鍵。



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**。
我選擇的是第一層卷積層 activate 後的 filters(64)，output 的照片則為下圖，似乎我的模型表現不太理想。



Output of layer0 (Given image0)

