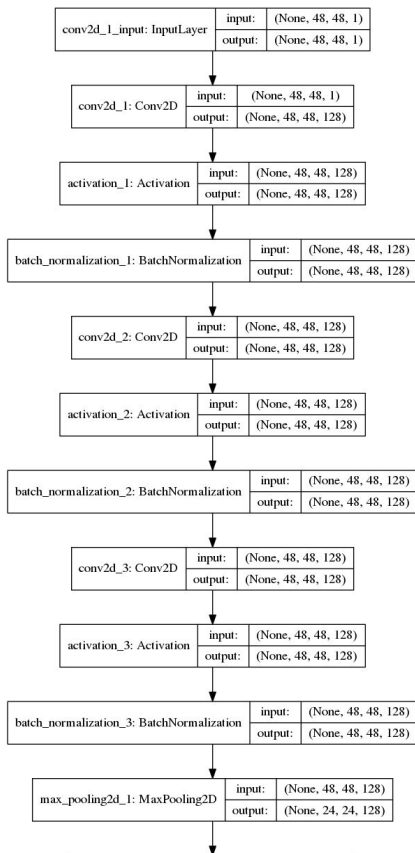
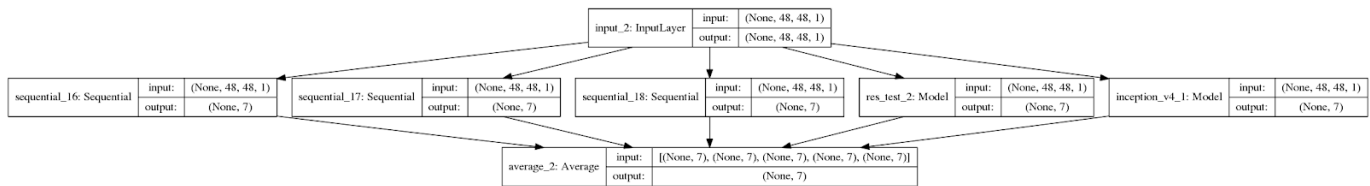


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練參數和準確率為何？

- 模型架構

- Ensemble 5個小model，最終平均每個小模型各class的機率。



- 前3個為相同架構：類似Vgg，但深度不同。
- 第四個，概念來自residual net，差別在於減少模型深度，用了12個residual module。
- 第五個，概念來自inception v4，差別在於在減少模型深度，網路最後加上Dense的分類層。
- 每個模型每層後都加上Batchnormalization，已經在模型尾端的分類層加上Dropout(0.5)。
- 第四，五個模型activation func=relu
- 每個模型接使用Data augmentation。

- 左圖為前參個模型的共同子架構，activation是selu
- 第一個模型有6個子架構，第二個有4個子架構，第3個有2個子架構。
- 共同都有3個Dense layer (unite = 2048, 1024, 512) 在模型尾端。

- 超參數 (每個小model各自訓練，使用一樣的超參數設定)

- Batch Size = 32
- Number of epochs = 100

- 優化器 (每個小model各自訓練，使用一樣的優化器設定)

- Adam
- lr=1e-3, beta\_1=.9, beta\_2=.999
- learning rate 乘以 0.75 如果 val\_loss 4 epochs都沒有沒有下降到新低点

- 準確率

- Public Score = 0.72137
- Private Score = 0.70855

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？

- Data normalization

- 實作方法：

- 針對training set上每個圖片同一個pixel位子，計算其平均與標準差
    - 對training & testing set上每個圖片同一個pixel位子，減去其平均再除以標準差

- 準確率比較：

- 兩個模型都在training set上 準確率大於0.65
    - 此種方式，標準化過後表現非常差，推斷原因是feature wise的標準化，因為對於其他data一起標準化並非針對同一個圖片，沒有考慮到feature”空間上”是有關聯性的。例如，圖片原本比較深的區塊變成有些深有些淺。Normalize造成前面負責抓線條的conv layers抓不到五官輪廓等等重要判斷依據。

	w/o normalization	w/ normalization
public score	.65951	.58679
private score	.65979	.60295

- Data augmentation

- 實作方法：

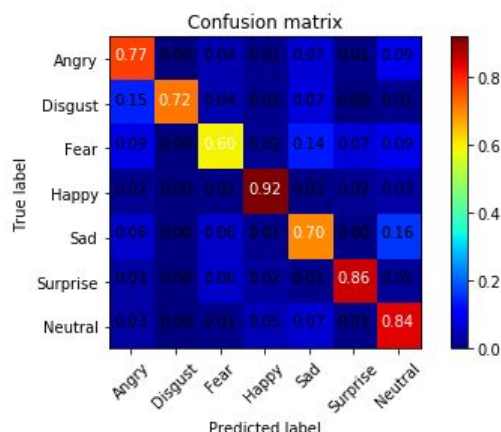
- 0~40度隨機角度旋轉
    - 在X, Y方向隨機平移，最大平移程度為圖片長寬的15%
    - 允許圖片隨機水平翻轉

- 準確率比較：

- 兩個模型都在training set上 準確率大於0.70
    - 但使用data augmentation，增強generalization的能力

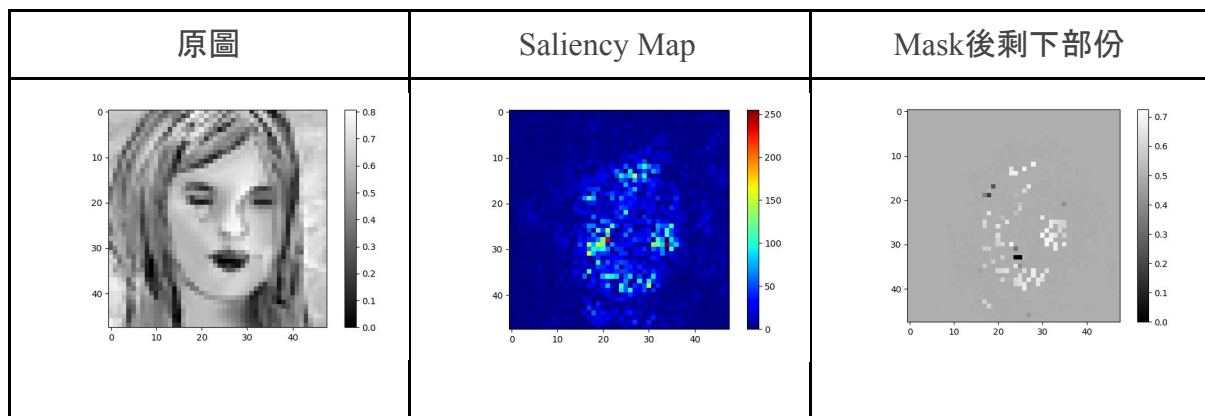
	w/ augmentation	w/o augmentation
public score	.65951	.55001
private score	.65979	.55920

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]



- ‘難過’容易被錯誤分成‘中立’
- ‘厭惡’容易被錯誤分成‘生氣’
- ‘恐懼’容易被錯誤分成‘難過’
- “恐懼”最難被正確分類。

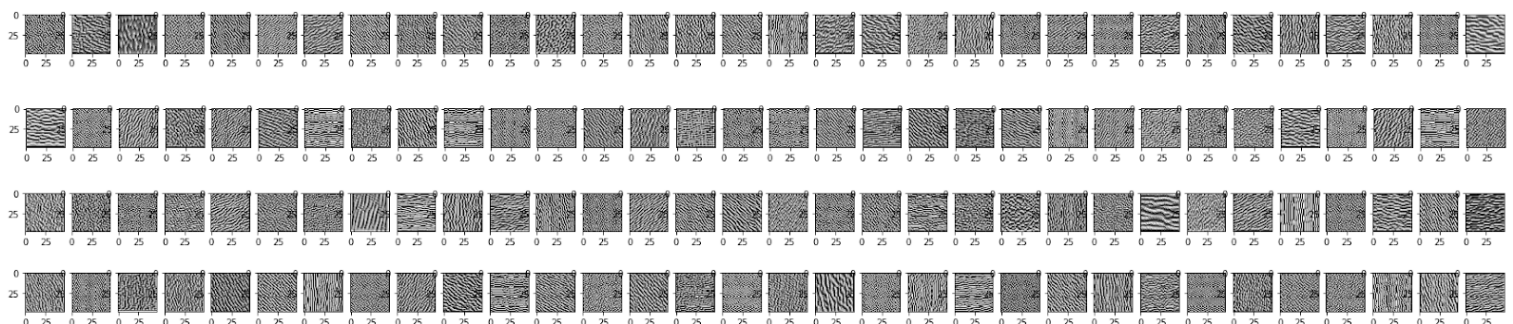
4. (1%) 從(1)(2)可以發現，使用 CNN 的確有些好處，試繪出其 saliency maps，觀察模型在做 classification 時，是 focus 在圖片的哪些部份？



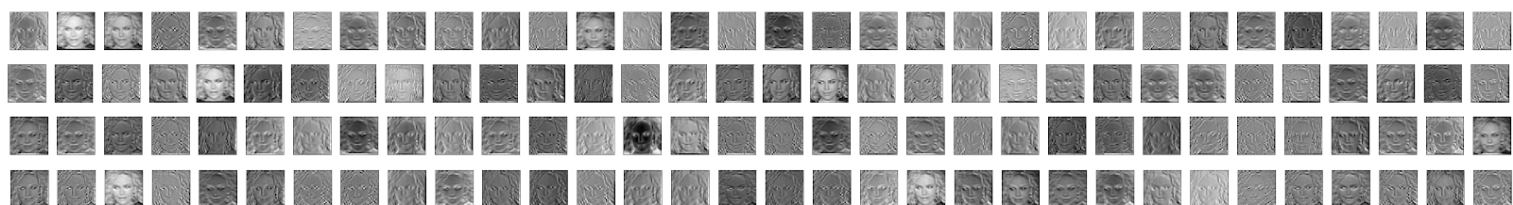
根據Saliency Map與原圖對比，分類器進行分類時，特別著重於眼睛鼻子嘴巴等五官，以及”臉部界線的線條”，分類器會分辨臉的輪廓，並不會注意頭髮或是空白之處。

(本題使用第一個模型的最後一個Dense layer)

5. (1%) 承(4) 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與觀察filter的output。



上圖為最容易activate 第一層conv layer 128個filter的圖片，可由觀察中看到最一開始的filter都是看到一些直線、斜線、波浪線、一些重複出現的特別的幾何形狀。這個觀察，符合老師上課所說，conv layer就像人看東西一樣，會先找一些特定輪廓與圖形。



上圖為第一個conv layer的输出，可以藉由上面最容易activate每一個filter的圖形推斷出每個filter的特徵，相對應的输出也明顯地反映出每個filter著重於不同的幾何圖形。例如，注意水平線條的filter對應的输出，在嘴巴附近就會比較深。