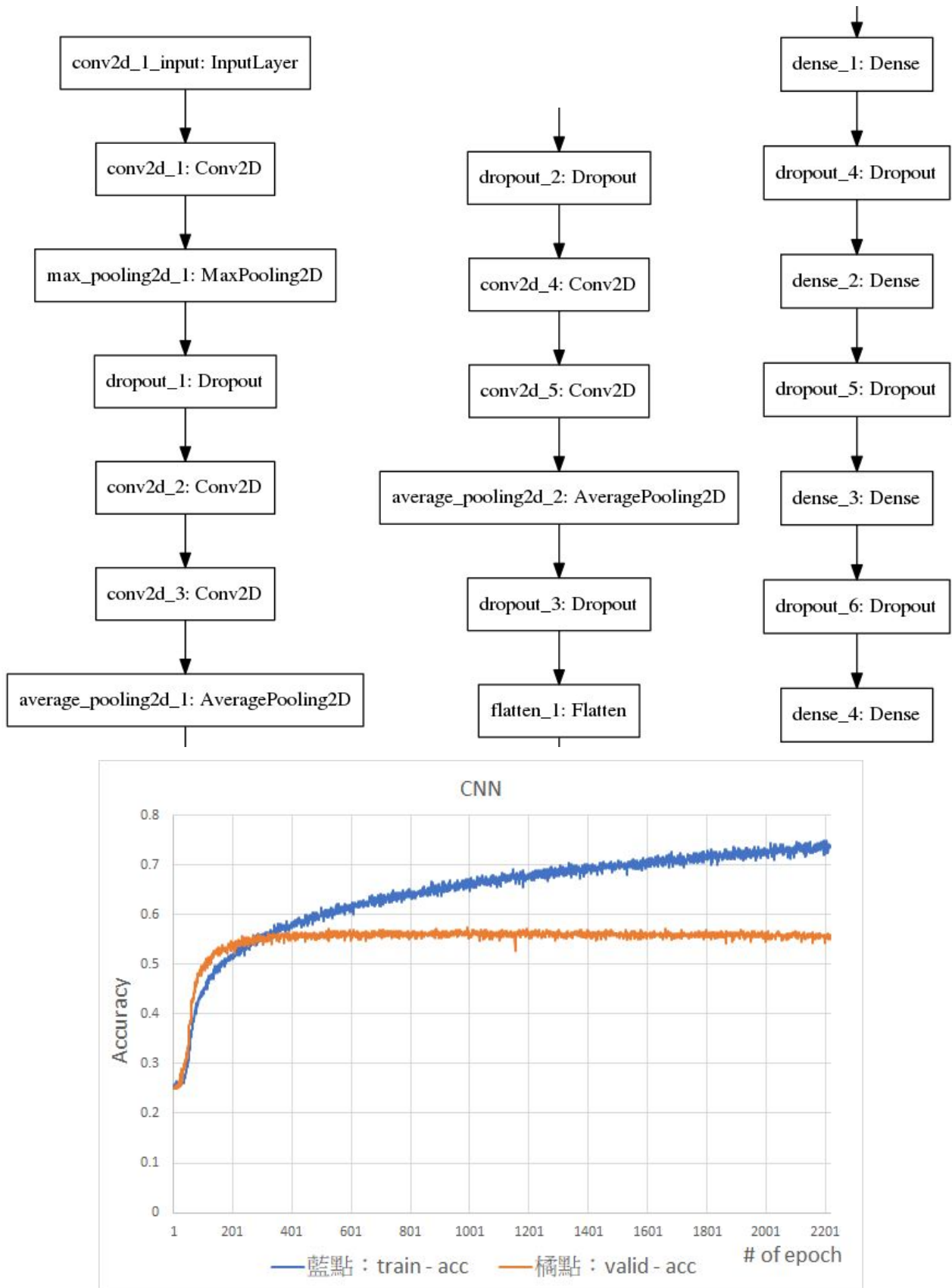


1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？

(Collaborators: b04611038 - 劉記良、b04901146 - 黃禹傑、b04901142 - 陳政曄)

答：準確率：Kaggle Private Score：0.66759；Public Score：0.65979

CNN模型架構和訓練過程如下圖所示。



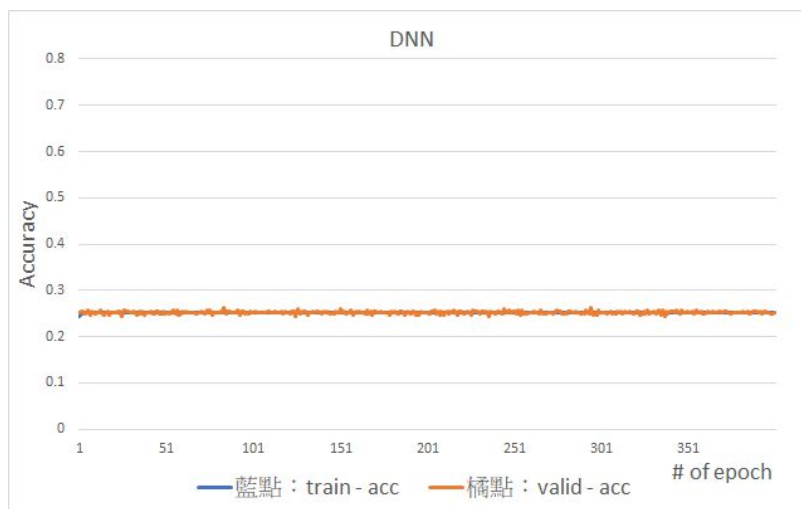
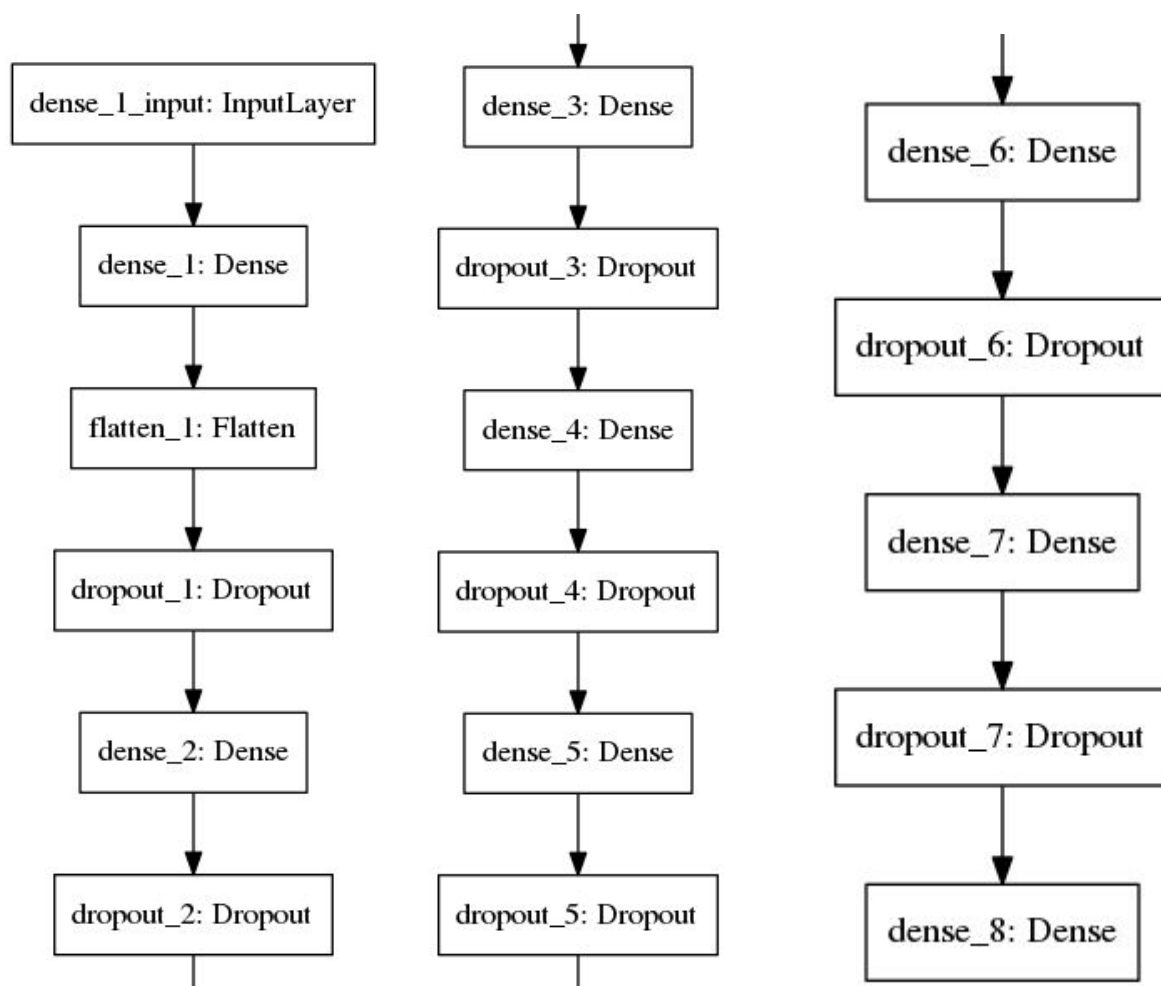
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

(Collaborators: b04611038 - 劉記良、b04901146 - 黃禹傑、b04901142 - 陳政曄)

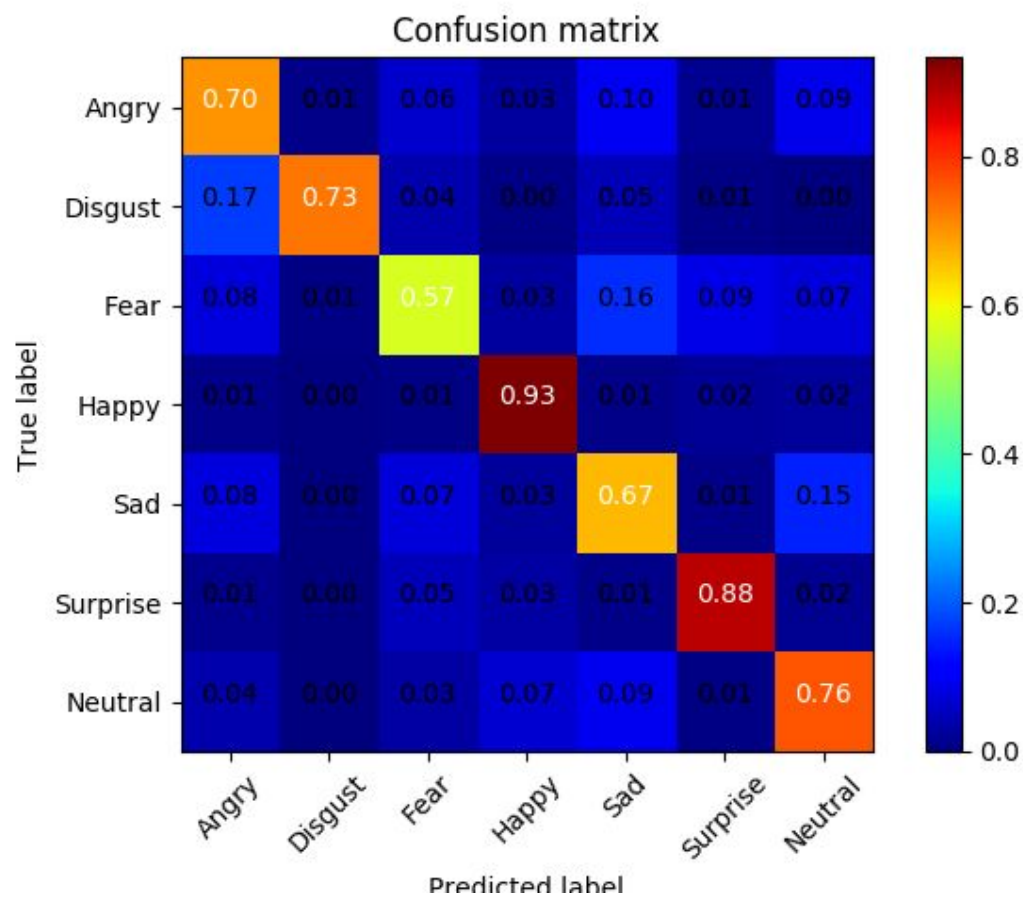
答：準確率0.25，train的時候就直接卡在準確率0.25，嘗試增加或減少層數、參數數量，結果都一樣慘(註)。發現到DNN在做圖形識別時，比較像是用硬記的，而較不會去考慮每個相鄰像素點之關係，而訓練CNN時給予經旋轉映射平移處理的圖片，可以有效提升在valid和Kaggle testing之準確率。因本次data量算少，因此不對圖像做處理的話CNN和DNN都很容易overfitting。

註：因為有對training圖像作旋轉映射平移等處理。

DNN模型架構和訓練過程如下圖所示。



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]  
 (Collaborators: b04611038 - 劉記良、b04901146 - 黃禹傑、b04901142 - 陳政曄)  
 答：跟人類容易誤判之對應形式相近。

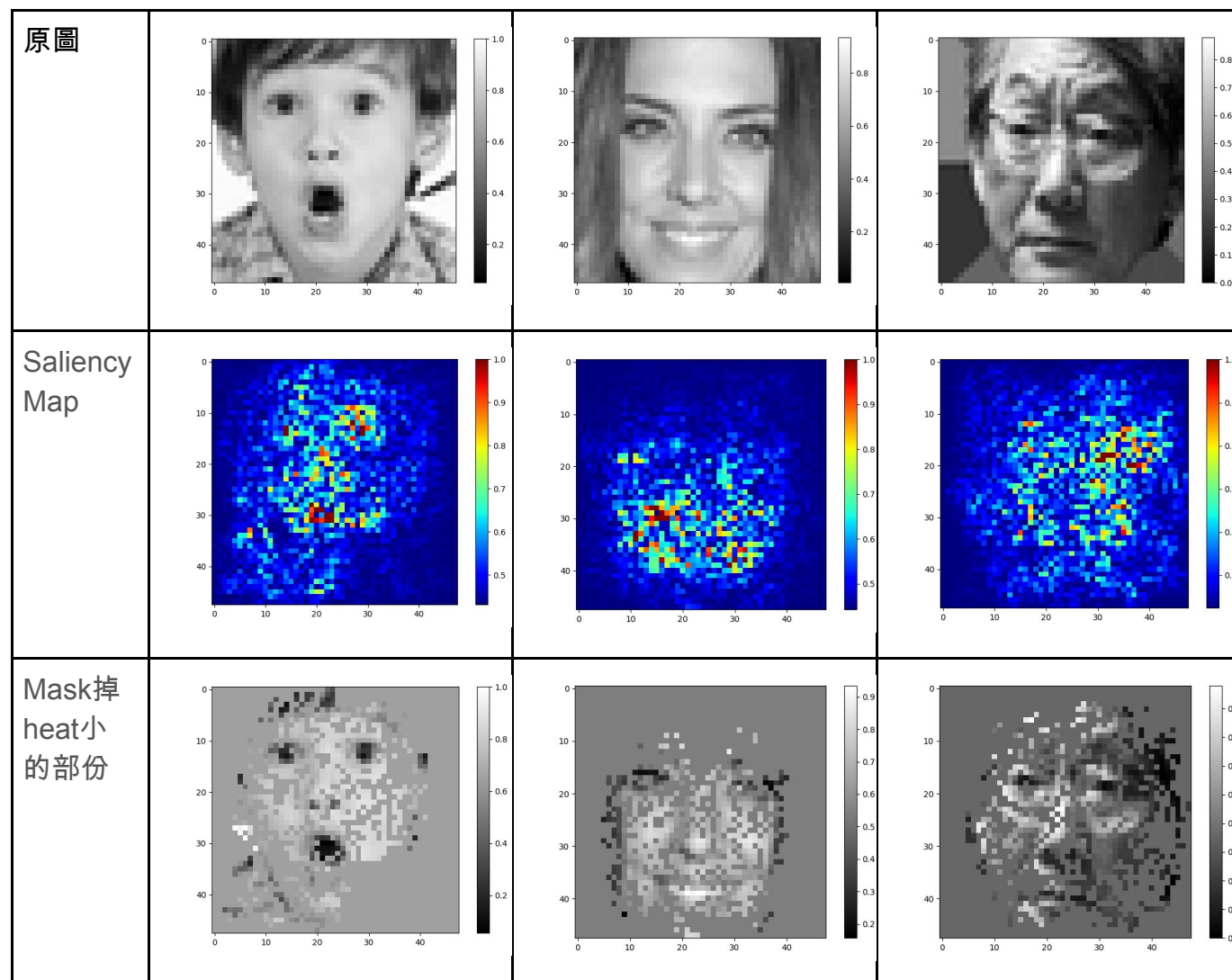


正確：	生氣	厭惡	恐懼	高興	難過	驚訝	中立
誤認為：	難過	生氣	難過	中立	中立	恐懼	難過

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？

(Collaborators: b04611038 - 劉記良、b04901146 - 黃禹傑、b04901142 - 陳政曄)

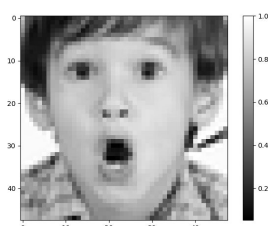
答：第一張測試符合預期，主要是聚焦在眼鼻口；第二張因為臉頰形狀明顯特別，所以受到較多注意；第三張因臉部歪斜且陰影偏多，所以聚焦的地方有點失準。



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: b04611038 - 劉記良、b04901146 - 黃禹傑、b04901142 - 陳政曄)

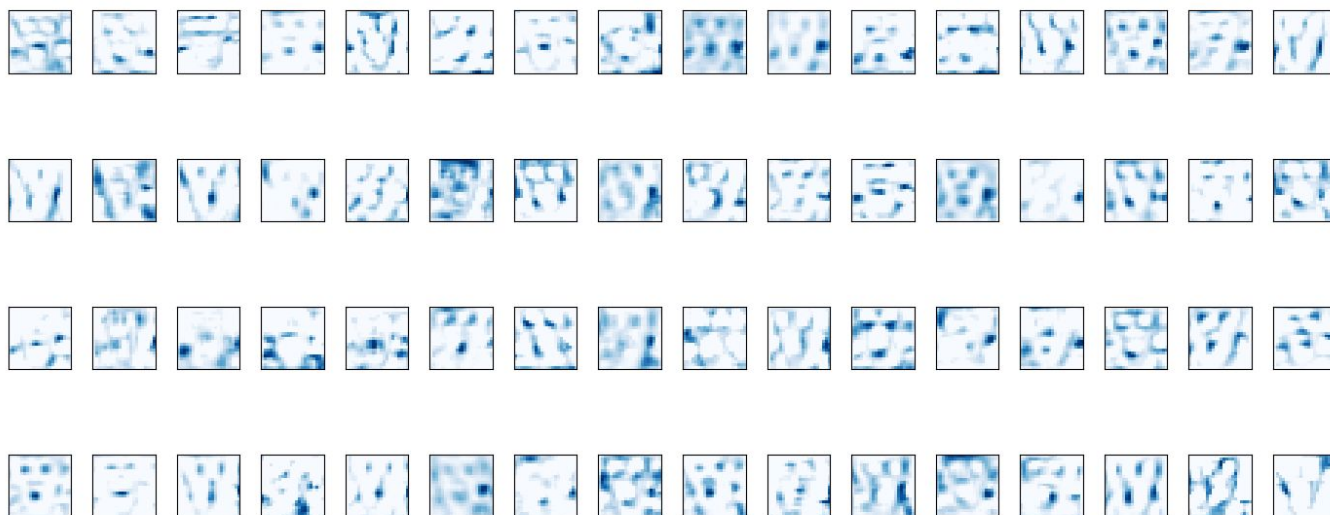
答：



輸入圖像：

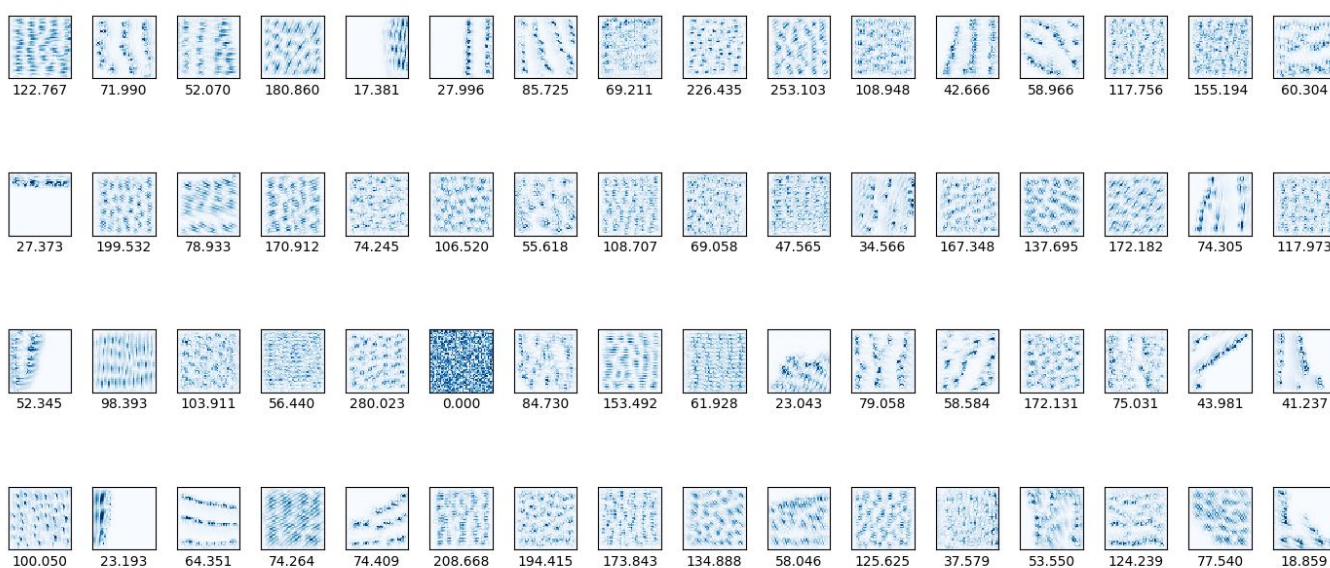
## conv2d\_2之輸出

Output of layer0 (Given image5487)



## conv2d\_2之噪音做gradient ascent之輸出

Filters of layer conv2d\_2 (# Ascent Epoch 500 )



觀察conv2d\_2之輸出，發現filter確實會強化圖片特定特徵之輸出，例如上圖中，嘴巴和眼睛的特徵便被加深很多，可有利於後面的layer學習。

gradient ascent的部分，可看出因為訓練過程有旋轉圖片，所以有些filter的特徵是相當接近(僅僅旋轉)，而因其所蘊含之資訊太high level了，所以人類可能無法看出隱含甚麼辨別技巧。