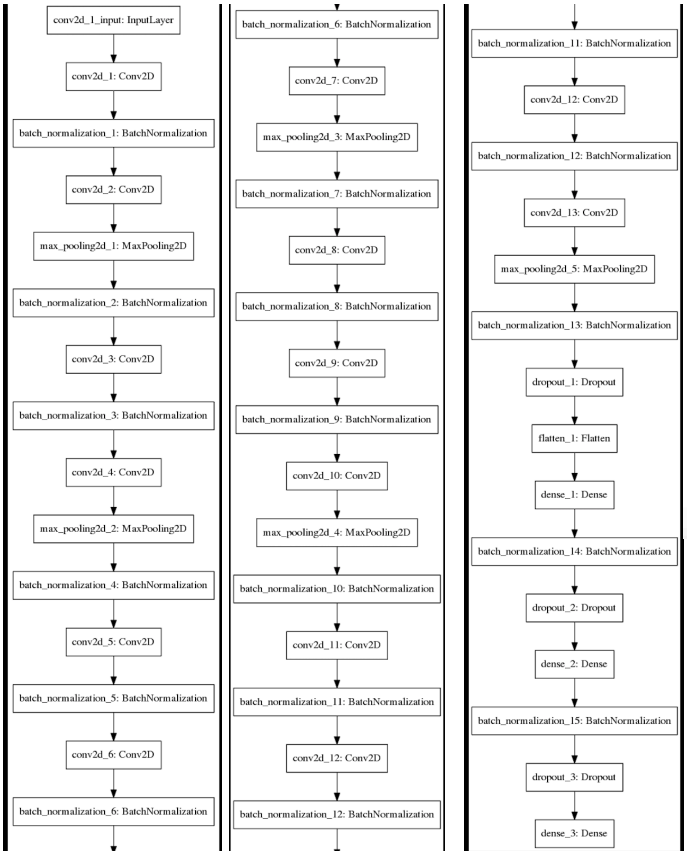
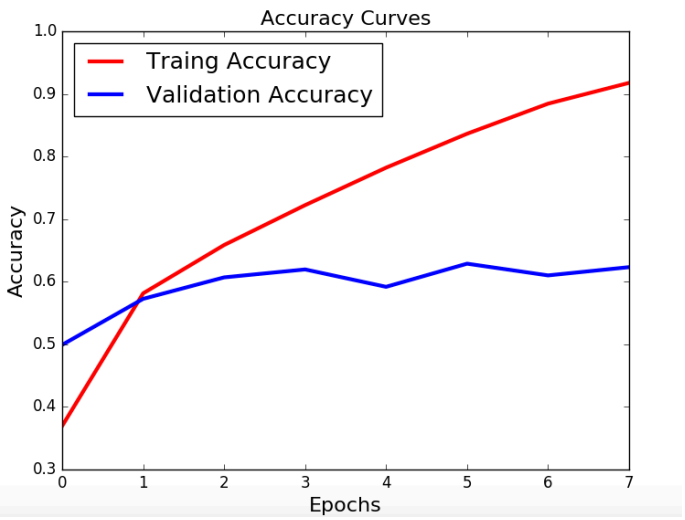
學號： r06946004 系級：資科學程碩一 姓名：蔡尚錡

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？

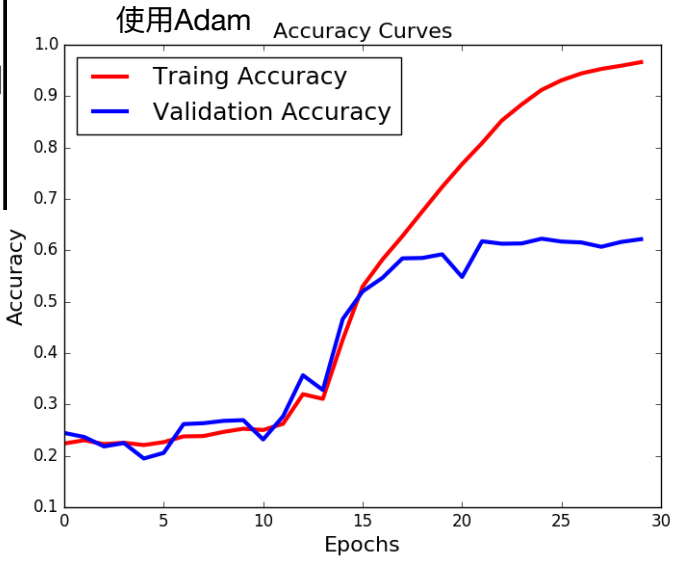
答：我這次所實作的的CNN model 是採用 VGG16的堆疊方式來做架構，一開始有5個blocks來對圖片取出學到的feature，前兩個blocks各別由兩個Conv2D跟一個MaxPooling2D所組成，後三個blocks各別由三個Conv2D跟一個MaxPooling2D所組成，隨著越往上堆疊，每個blocks所使用的channel數會逐漸增加，從第一個block的64到第二個的128、第三個的256、第四第五的512，最後做Flatten以後再傳入兩層各4096的Fully Connected Network最後透過 softmax 得到七種表情的情緒 classification 的結果。這樣的架構我做了另外三件事情來幫助performance，第一個是加入dropout的機制來避免overfitting，dropout rate 選用0.5，第二個是加入batch normalization在relu之後，降低gradient vanishing的可能，提升他的效果，第三個是用data augmentation讓image的data數量透過鏡射、旋轉、移動，數目增加十倍，model也因此可學到多樣的特徵。 這樣的架構參數量相當龐大共有三千三百多萬個參數，訓練過程每個epochs大概需要6分鐘總共長達一至兩小時，我使用了兩個不同的optimizer，Adam() 和 Adagrad() ，Adam 收斂速度比較慢，一開始震盪期間比較久大概十個epochs之後才能開始準確收斂，總共大概需要訓練20到25個epochs，而Adagrad 則相對快速，一兩個epochs之後就能開始收斂總共歷時8到10個epochs完成。兩者都大概收斂到準確率0.65至0.67左右。



使用Adadelta

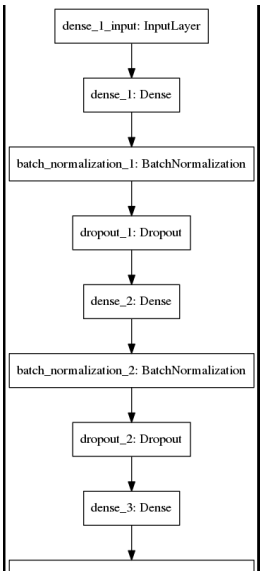


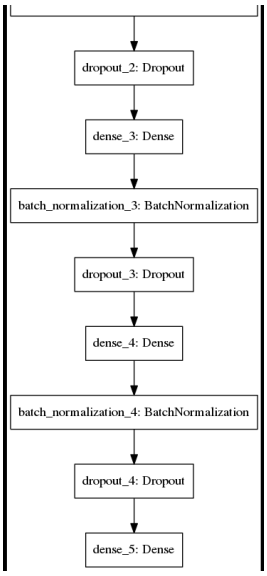
使用Adam

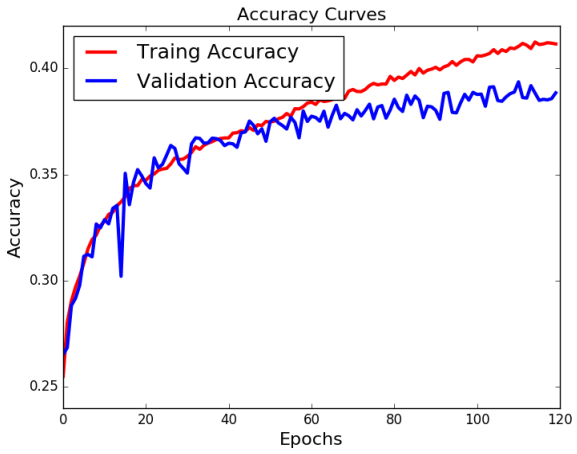


2.(1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model。其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

答：上述的CNN參數量大概是三千三百多萬個，而我實作了大概三千四百多萬參數的DNN model來跟CNN做比較，架構大概是我堆疊了四個hidden layer，size分別是1024、4096、4096、1024，一樣每一層都搭配了batch mormalization 跟 dropout，也透過data augmentation幫助提升performance ，最後再接到一個output layer。我train了大概一百個epochs，雖然每一個epochs的執行速度相當快大概只需一分鐘，但收斂速度相當緩慢，loss下降的幅度也不是很明顯，accuracy不是很高，大概到達0.4左右而已，看趨勢應該是可以繼續train下去但是時間太久。跟CNN比起來效果是差蠻多的，大概都是訓練了一個多小時，CNN使用較少的epochs數就能有明顯的收斂效果以及較高的準確度，accuracy最高到達0.67遠超過DNN的0.42，可見CNN可比DNN能更有效率的學到每個表情的特徵點。

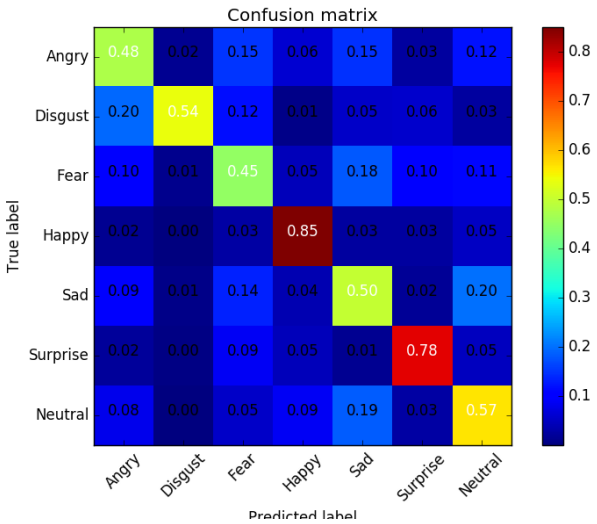






3.(1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？[繪出 confusion matrix 分析]

答：從我得到的confusion matrix觀察中可以得知，七種情緒[Angry、Disgust、Fear、Happy、Surprise、Neutral]裡面，有些種類是可以相當明顯的辨認出來，比如說Happy、Surprise，這幾種準確辨認的機率都可以高達七成以上，其中Happy更可以高達0.85，幾乎沒什麼跟其他情緒搞混的機會，另一種像Disgust、Sad、Neutral的準確辨識機率也可以高達五成以上，但是會有一些可能搞混的機會，例如Disgust有0.20的機會搞混成Angry、0.12的機會搞混成Fear，Sad則是有0.14的機會搞混成Fear和0.20的機會搞混成Neutral，Neutral有0.19的機會搞混成Sad。最後一種是像Angry、Fear這種準確辨識續不到五成，想當然也是很有搞錯機會，例如Angry就有0.15的機會搞混成Fear和Sad，而Fear則有0.18機率搞混成Sad，整體看起來Fear、Sad、Neutral彼此之間是很容易互相弄混的，可能人類在這三種情緒時有可能會出現相類似的表情特徵。



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？

答：由此可見在做classification的時候，CNN會focus在一些容易辨識特徵的部分，例如：眼睛、嘴巴、雙頰、臉的輪廓、鼻子等等，藉由這些部分大小、形狀的變化，來判斷一個人的表情與情緒。

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。

答：