DEEP LEARNING

Homework 2

309511045 黃麟凱

1. MNIST

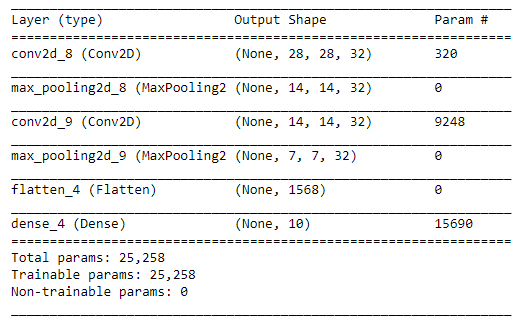
Training data : 50000

Validation data : 5000

Testing data : 10000

Network architecture：

這次使用的架構如下圖所示，首先圖片會經過一層convolution的計算，接著再對其做max\_pooling，讓feature變少，並重複一次。當做完convolution之後將feature攤平，並經過dense層把數量降成10，且activation function選擇softmax，即完成這次的Network architecture

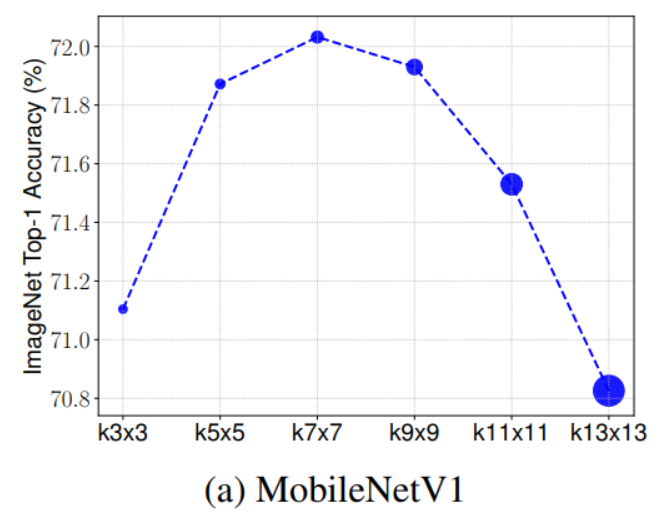


Analysis with different stride size and filter size (all train for 100 epochs)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kernel size | stride | Train acc | Test acc | Val acc | Epochs Train acc > 95% |
| (3,3) | (1,1) | 99.99% | 99.02% | 99.04% | 2nd |
| (5,5) | (1,1) | 99.99% | 99.22% | 99.30% | 2nd |
| (7,7) | (1,1) | 99.99% | 99.28% | 99.18% | 2nd |

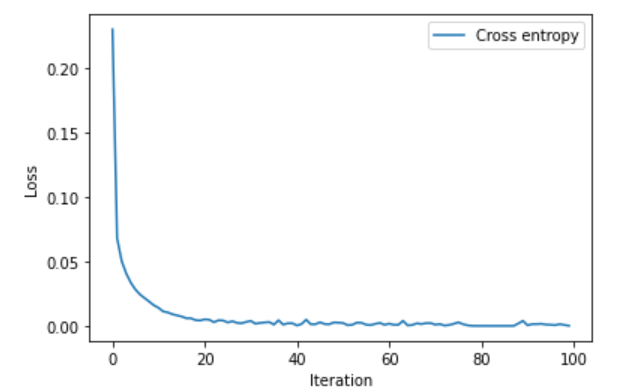
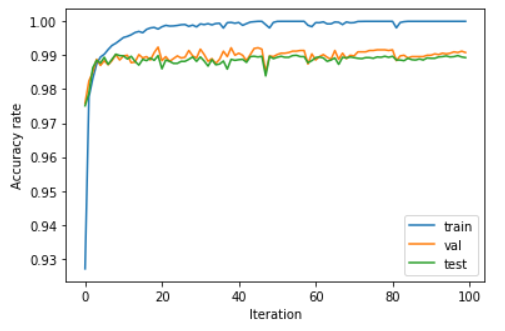
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kernel size | stride | Train acc | Test acc | Val acc | Epochs Train acc > 95% |
| (3,3) | (1,1) | 99.99% | 99.02% | 99.04% | 2nd |
| (3,3) | (2,2) | 99.99% | 98.49% | 98.66% | 2nd |
| (3,3) | (3,3) | 99.63% | 96.65% | 96.38% | 7th |

若調整kernel size 的大小可發現，當size變大整體的performance會變好一點。但若再更大的話，反而會有反效果。上網找了一下data發現kernel size的選擇的確如上述提到那樣，大過一個狀況後，效果就變得不好。從下圖可發現，MobileNet這個網路的kernel size若從3x3 -> 7x7，整體的accuracy有上升，但到9x9，甚至是13x13時，整體的accuracy就一直下降。由此可見kernel size的選擇是很重要的

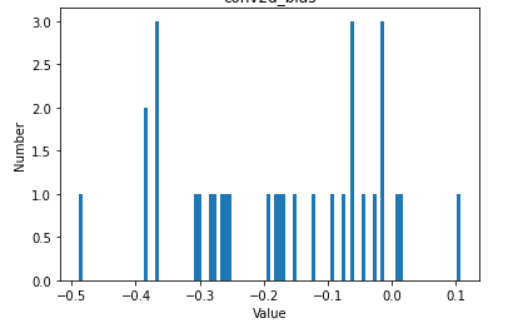
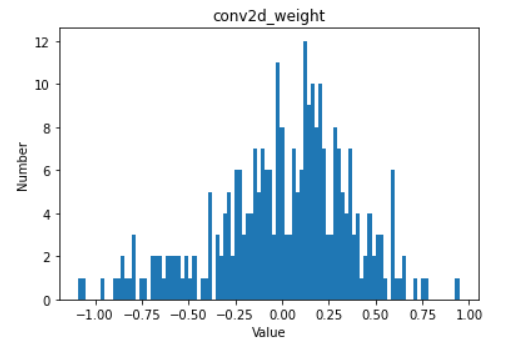


另一方面，若調整strides的大小，也會對整體的accuracy有影響，當stride上升時，accuracy也開始下降。原因可能在於我的kernel size只設定3x3，因此當strides太大時，導致model沒辦法有效的把feature提取出來。

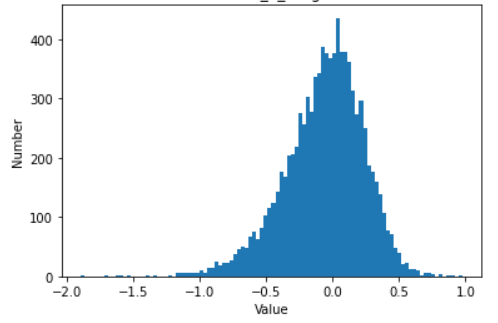
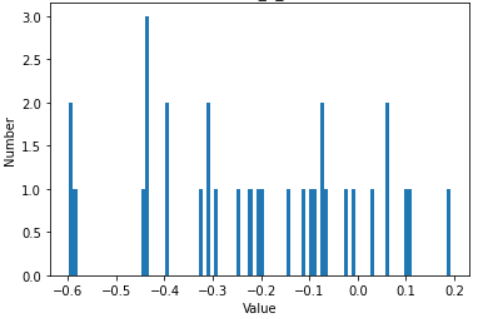
Training Accuracy Learning Curve



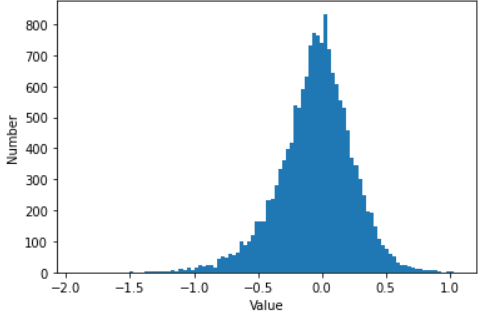
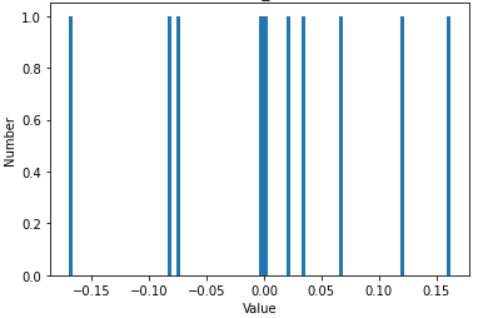
Histogram of Conv1：weights & bias

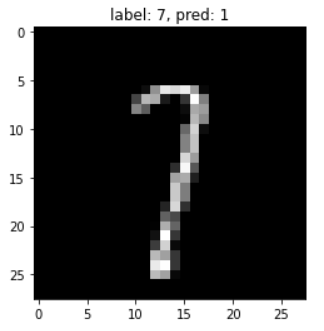
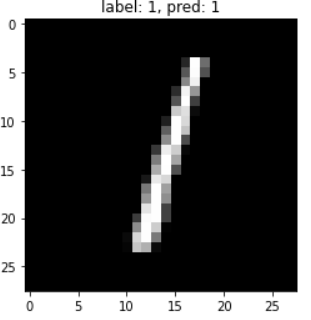


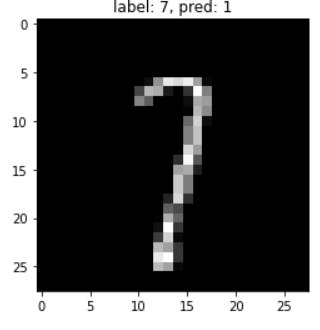
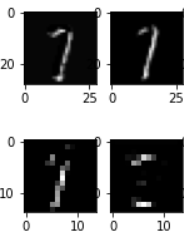
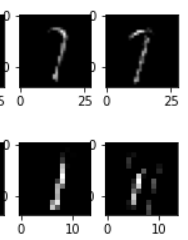
Histogram of Conv2：weights & bias

Histogram of Dense：weights & bias

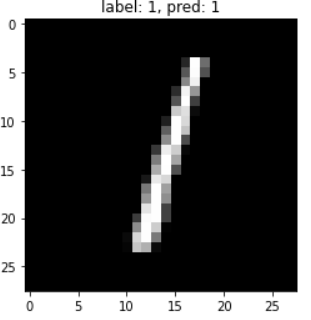
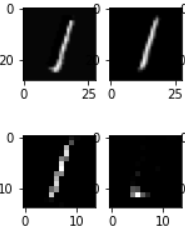
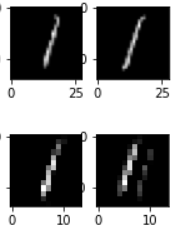
 



Conv2

Conv1

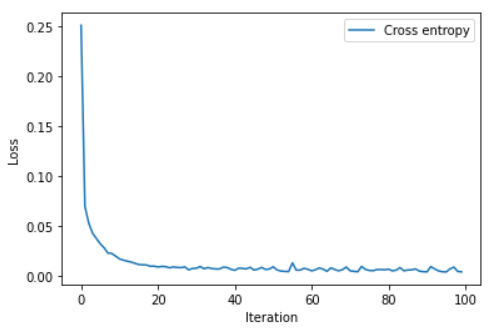
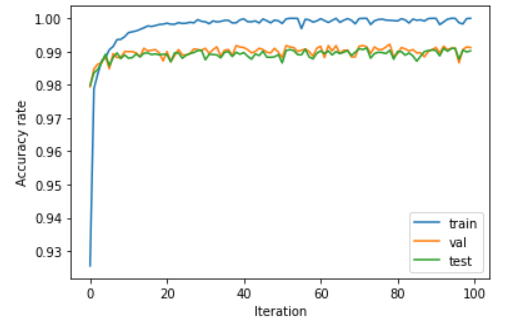


Conv1

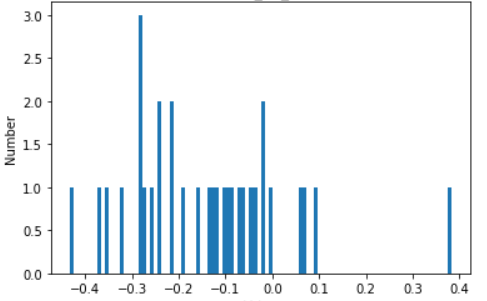
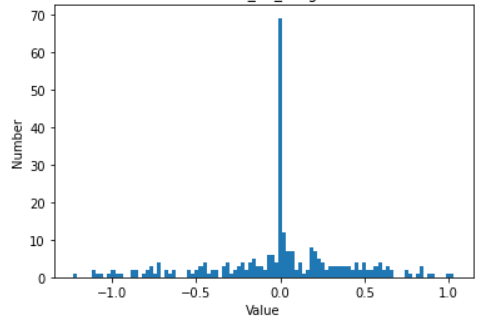
Conv2

在辨識1跟7時，的確有部分是很相似的，從feature map 也可以看出經過convolution之後的結果有部份也很相似(如紅框的地方)，但還是有差異的，所以整體的錯誤辨識結果並沒有很高

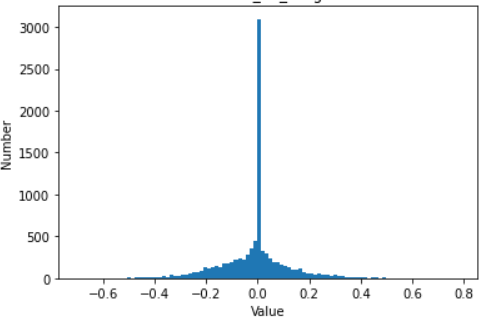
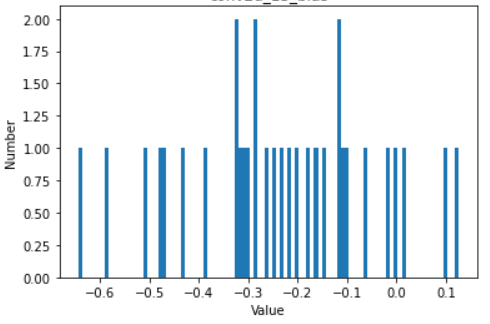
Training Accuracy Learning Curve



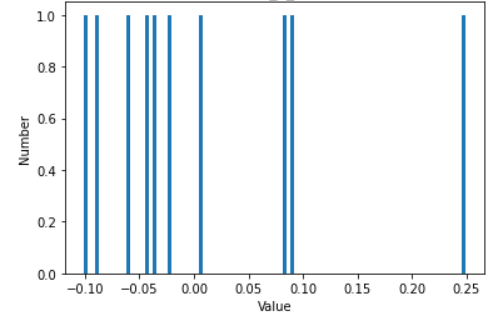
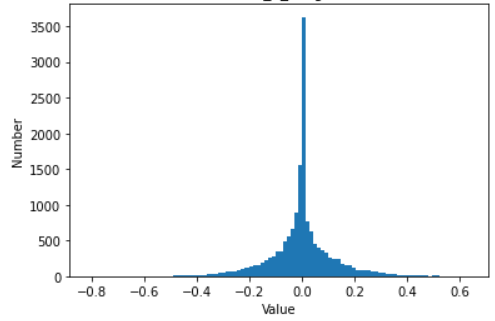
Histogram of Conv1：weights & bias



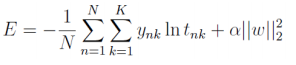
Histogram of Conv2：weights & bias

Histogram of Dense：weights & bias



從結果可以看出，經過L2 regularization後，整體的weight有明顯被限縮的趨勢。原因可以從公式看出，在原本的loss function上，加上了一個權重的參數，如此一來為了讓Error變小，不外乎除了需要讓predict的結果越接近answer之外，還要讓weight的大小變小才行。



1. CIFAR-10

Reference：

<https://medium.com/ai-academy-taiwan/cnn%E8%AB%96%E6%96%87%E5%B0%8E%E8%AE%80-mixconv-mixed-depthwise-convolutional-kernels-4357ccdbfe6c>