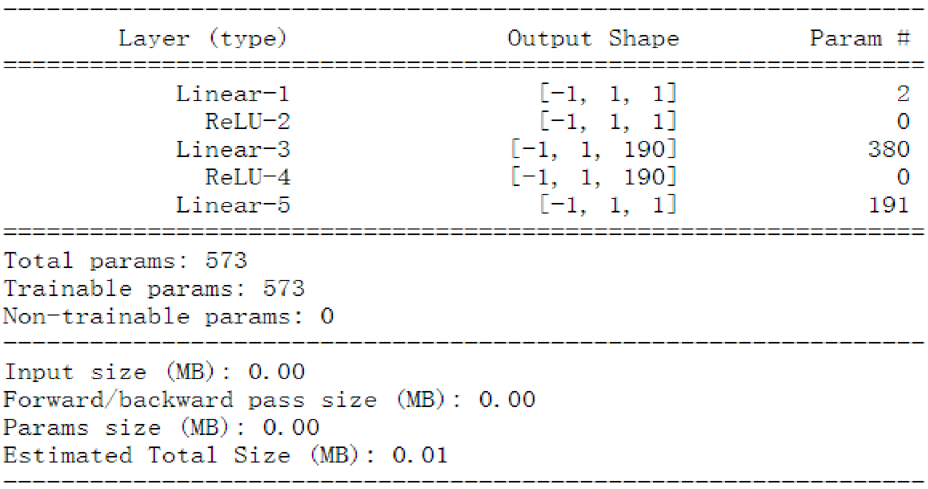
Report

HW1-1

(1) Simulate a Function

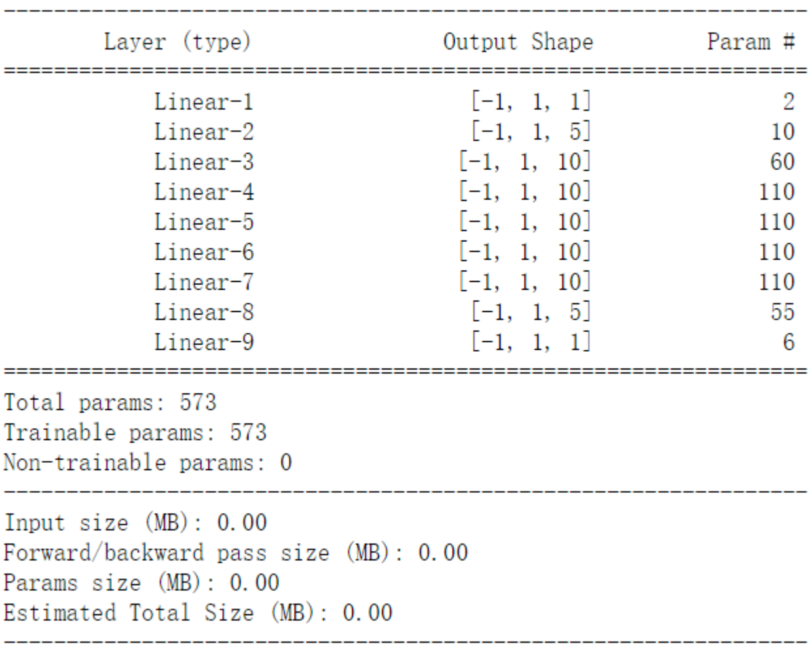
Function sin(5x)/5x

Model 1



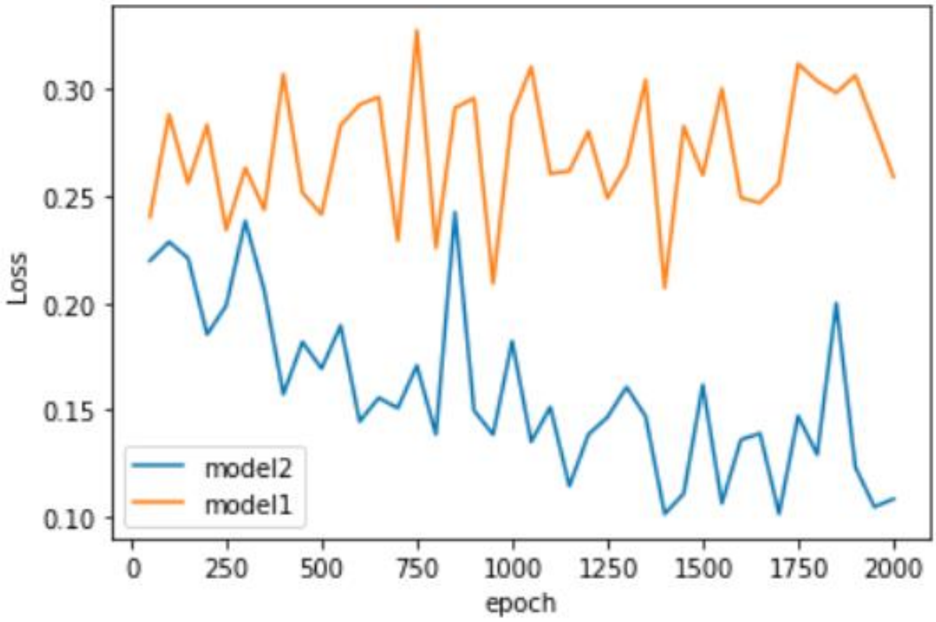
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1                [-1, 1, 1]               2
              ReLU-2                [-1, 1, 1]               0
            Linear-3              [-1, 1, 190]             380
              ReLU-4              [-1, 1, 190]               0
            Linear-5                [-1, 1, 1]             191
================================================================
Total params: 573
Trainable params: 573
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.00
Forward/backward pass size (MB): 0.00
Params size (MB): 0.00
Estimated Total Size (MB): 0.01
----------------------------------------------------------------
```

Model 2

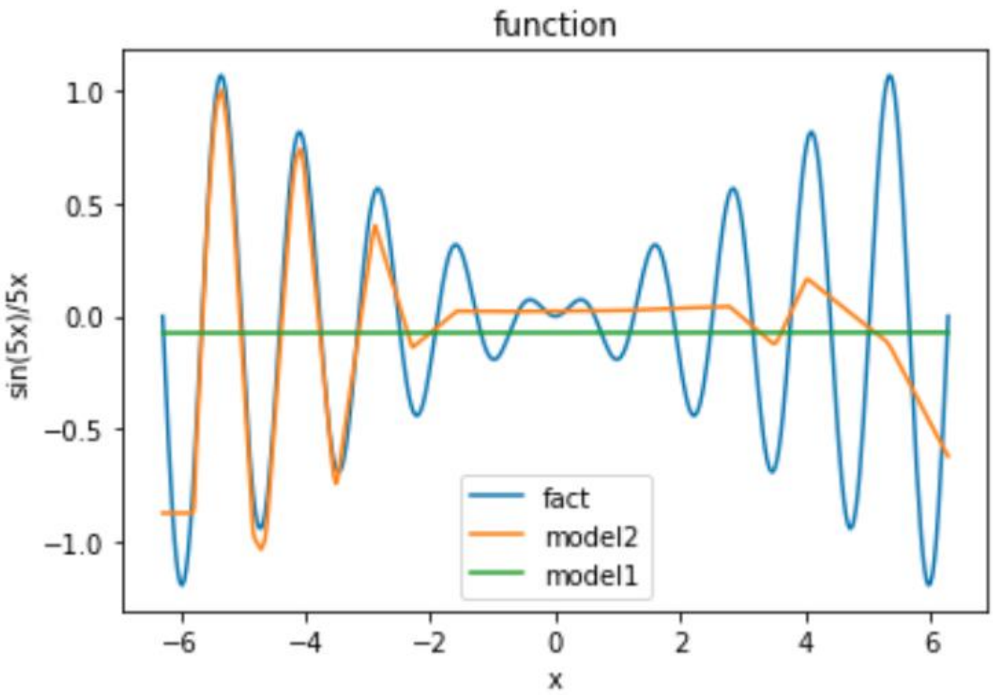


```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Linear-1                [-1, 1, 1]               2
            Linear-2                [-1, 1, 5]              10
            Linear-3               [-1, 1, 10]              60
            Linear-4               [-1, 1, 10]             110
            Linear-5               [-1, 1, 10]             110
            Linear-6               [-1, 1, 10]             110
            Linear-7               [-1, 1, 10]             110
            Linear-8                [-1, 1, 5]              55
            Linear-9                [-1, 1, 1]               6
================================================================
Total params: 573
Trainable params: 573
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.00
Forward/backward pass size (MB): 0.00
Params size (MB): 0.00
Estimated Total Size (MB): 0.00
----------------------------------------------------------------
```

Loss Chart



Function Curve

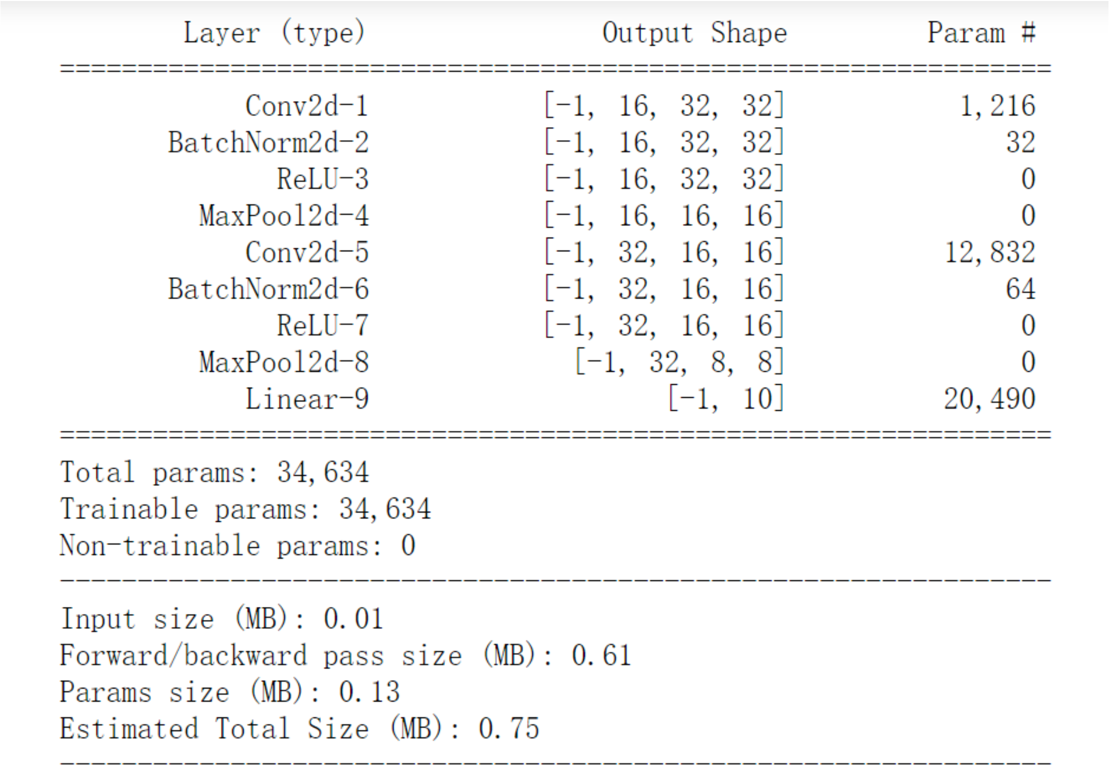


Comment:

According to the chart and the graph, we can see that obviously the deep network is better than the shallow network in simulate function task.

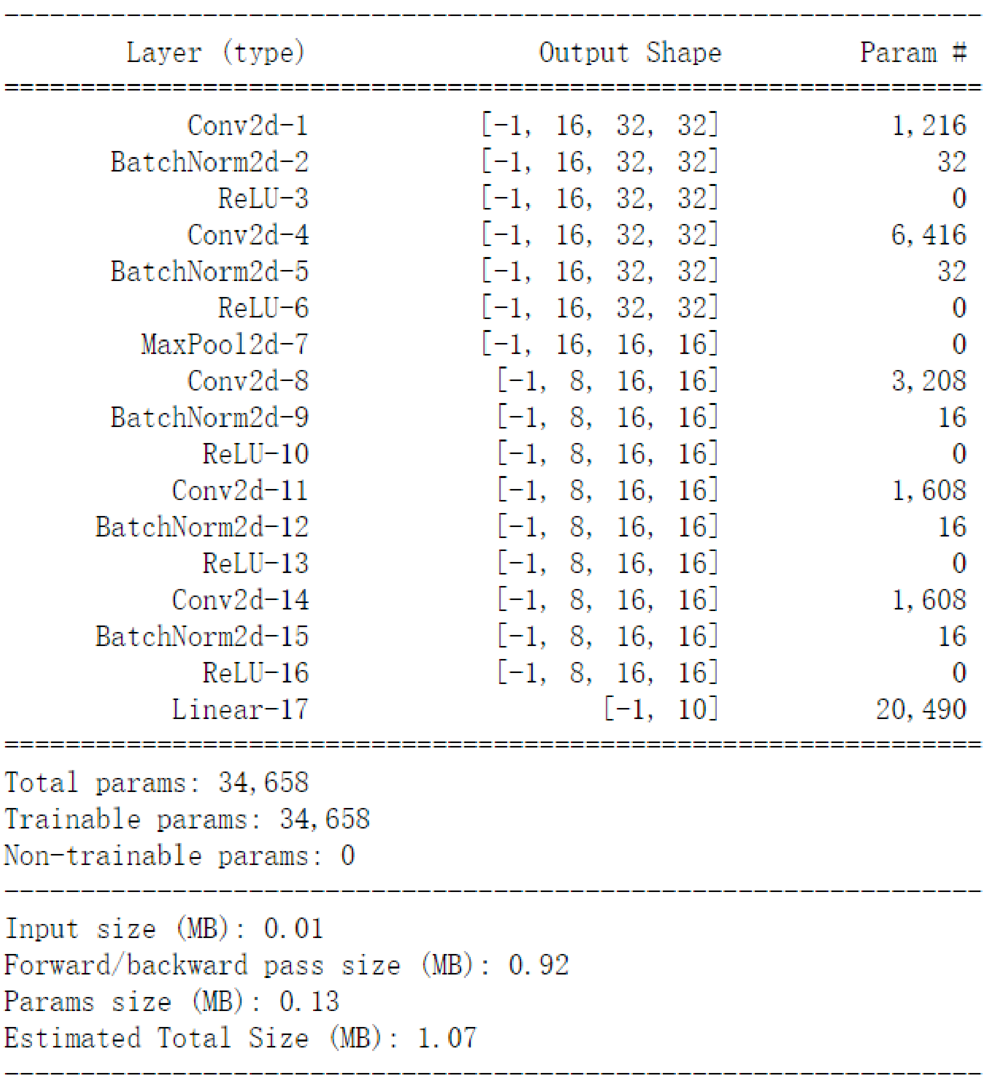
(2) Train on Actual Task

Model1



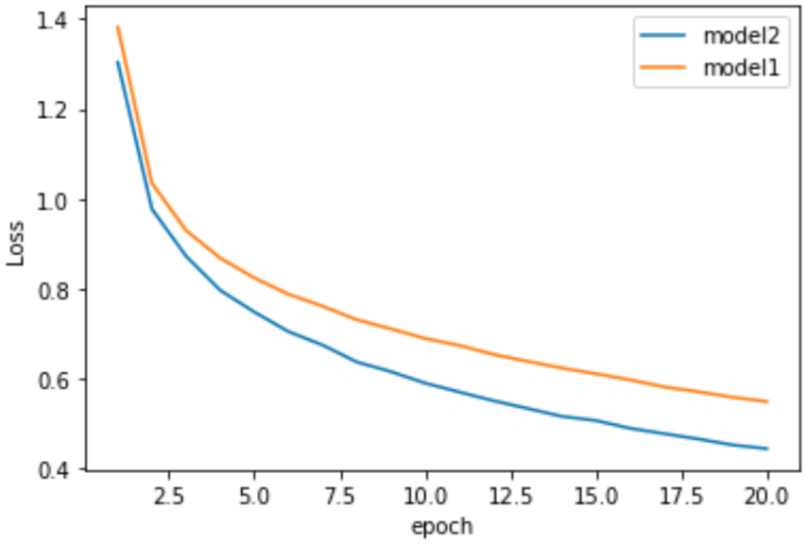
```
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 16, 32, 32]           1,216
       BatchNorm2d-2           [-1, 16, 32, 32]              32
              ReLU-3           [-1, 16, 32, 32]               0
         MaxPool2d-4           [-1, 16, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]          12,832
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 34,634
Trainable params: 34,634
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.61
Params size (MB): 0.13
Estimated Total Size (MB): 0.75
----------------------------------------------------------------
```

Model2

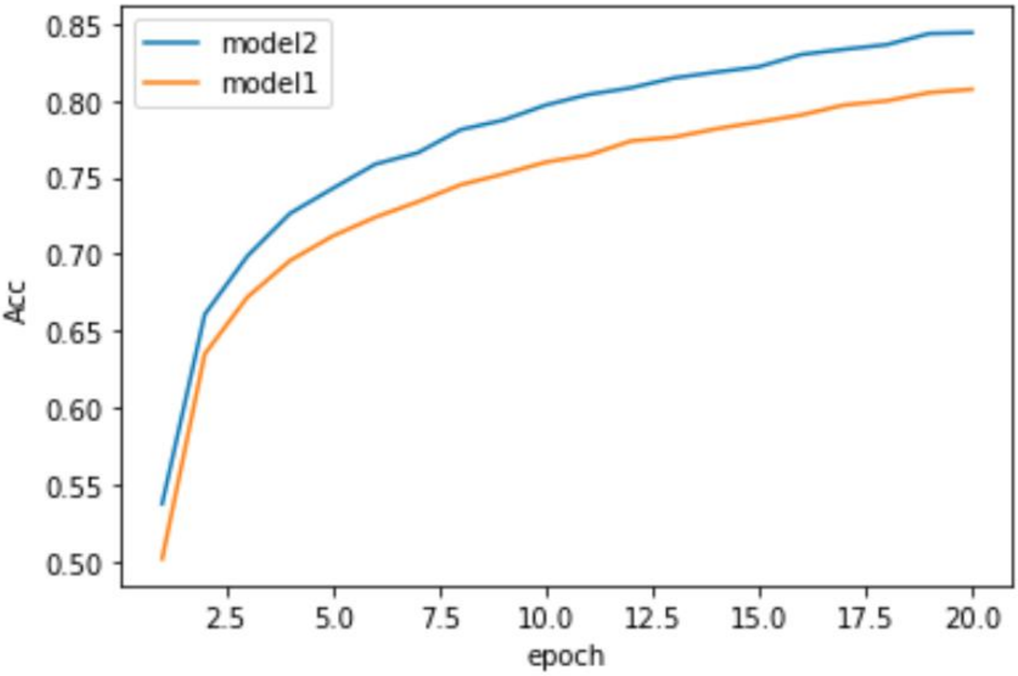


```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 16, 32, 32]           1,216
       BatchNorm2d-2           [-1, 16, 32, 32]              32
              ReLU-3           [-1, 16, 32, 32]               0
            Conv2d-4           [-1, 16, 32, 32]           6,416
       BatchNorm2d-5           [-1, 16, 32, 32]              32
              ReLU-6           [-1, 16, 32, 32]               0
         MaxPool2d-7           [-1, 16, 16, 16]               0
            Conv2d-8            [-1, 8, 16, 16]           3,208
       BatchNorm2d-9            [-1, 8, 16, 16]              16
             ReLU-10            [-1, 8, 16, 16]               0
           Conv2d-11            [-1, 8, 16, 16]           1,608
      BatchNorm2d-12            [-1, 8, 16, 16]              16
             ReLU-13            [-1, 8, 16, 16]               0
           Conv2d-14            [-1, 8, 16, 16]           1,608
      BatchNorm2d-15            [-1, 8, 16, 16]              16
             ReLU-16            [-1, 8, 16, 16]               0
           Linear-17                   [-1, 10]          20,490
================================================================
Total params: 34,658
Trainable params: 34,658
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.92
Params size (MB): 0.13
Estimated Total Size (MB): 1.07
----------------------------------------------------------------
```

Loss Chart


1.4
1.2
1.0
0.8
0.6
0.4
Loss
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
epoch
model2
model1

ACC Graph


0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Acc
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
epoch
model2
model1

Comment:

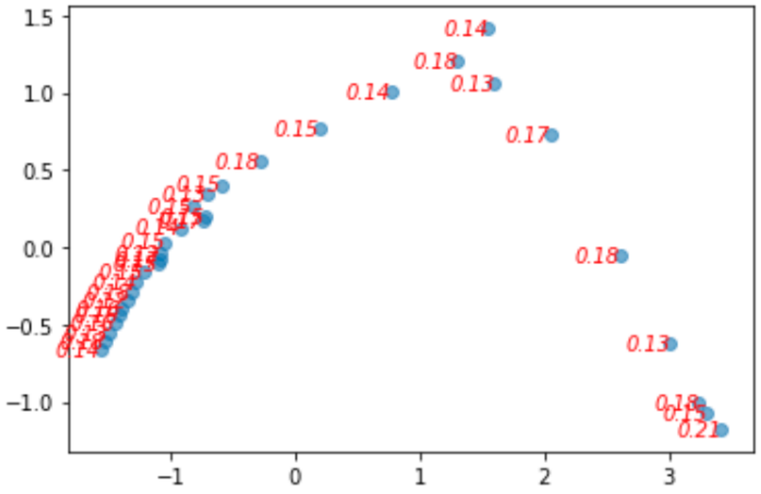
According to the result , we can see that the deep cnn performs better than the shallow cnn. From this we can see that building a deeper network can improve the fitting ability of the model.

HW1-2 Optimizer

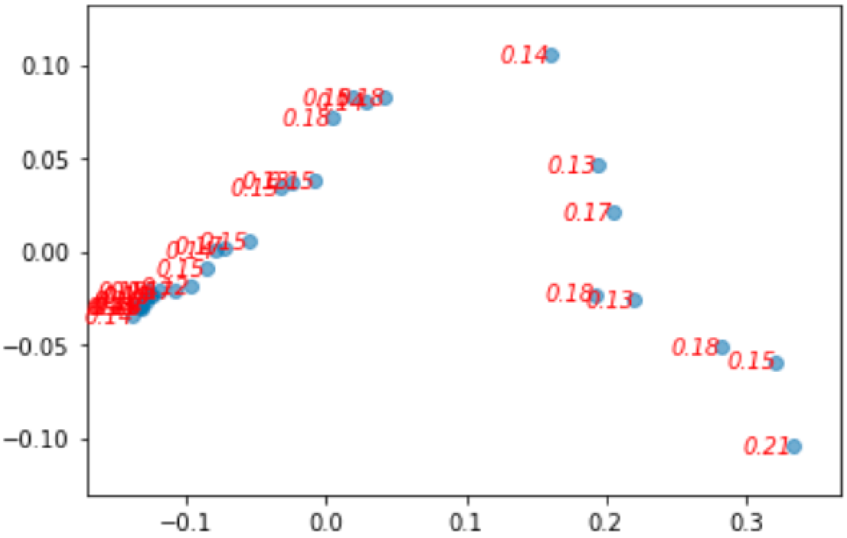
（1）Visualize the Optimization Process

Collect the weights every 5 epochs, and train 8 times. Reduce the dimension of weights to 2 by PCA

Whole Model

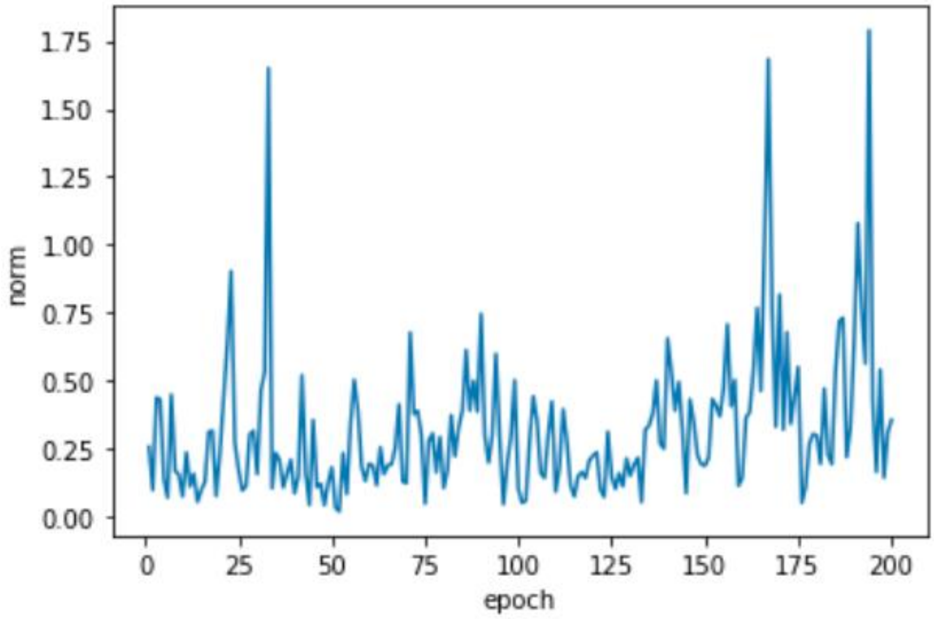


Layer 1



(2) Observe Gradient Norm During Training

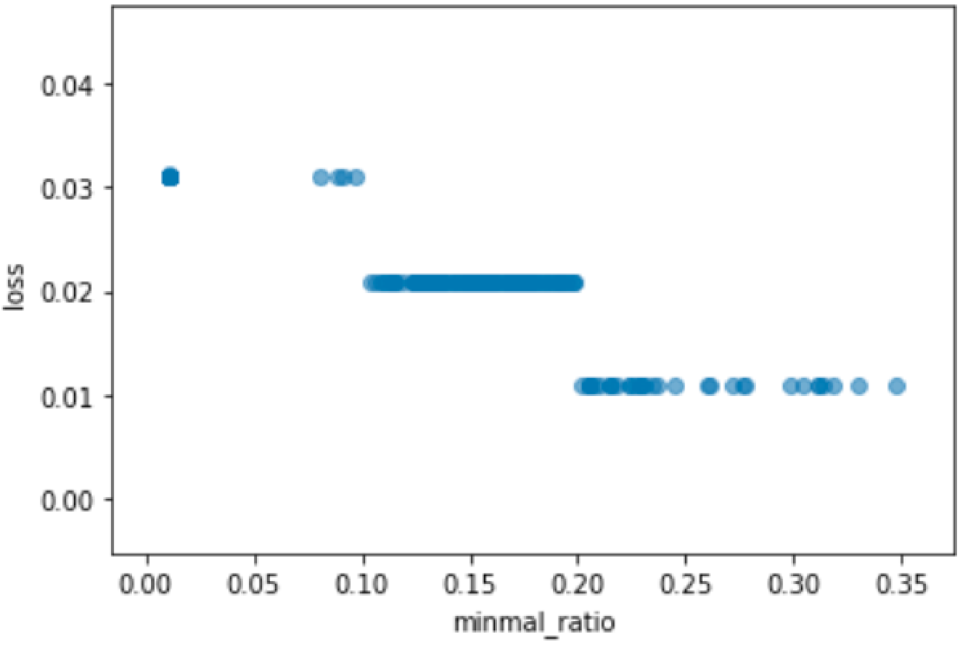
Function sin5x/5x



(3) What Happened When Gradient is Almost Zero

1. To find the find the weights of the model when the gradient norm is zero, I changed the objectibe function. The original function is mse loss and I add L1 norm to this loss function. L1 norm can make the weight of network to zero

2. Figure



HW1-3

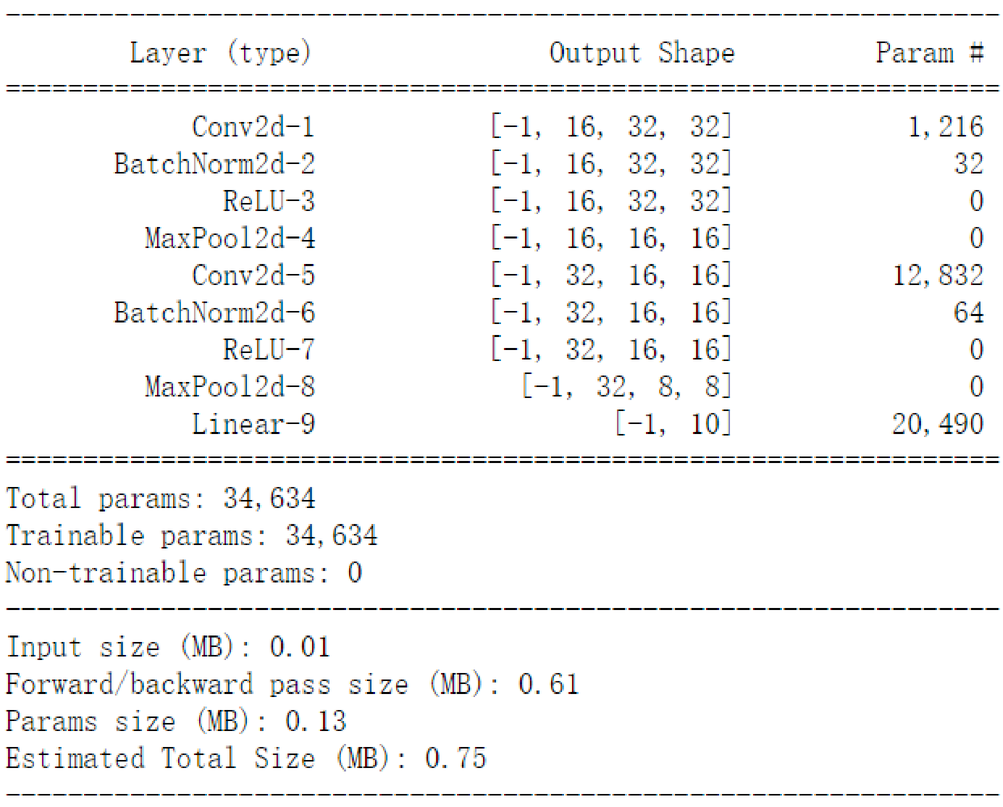
1. Fit on random labels

Task : Cifar 10

Learning Rate: 0.001

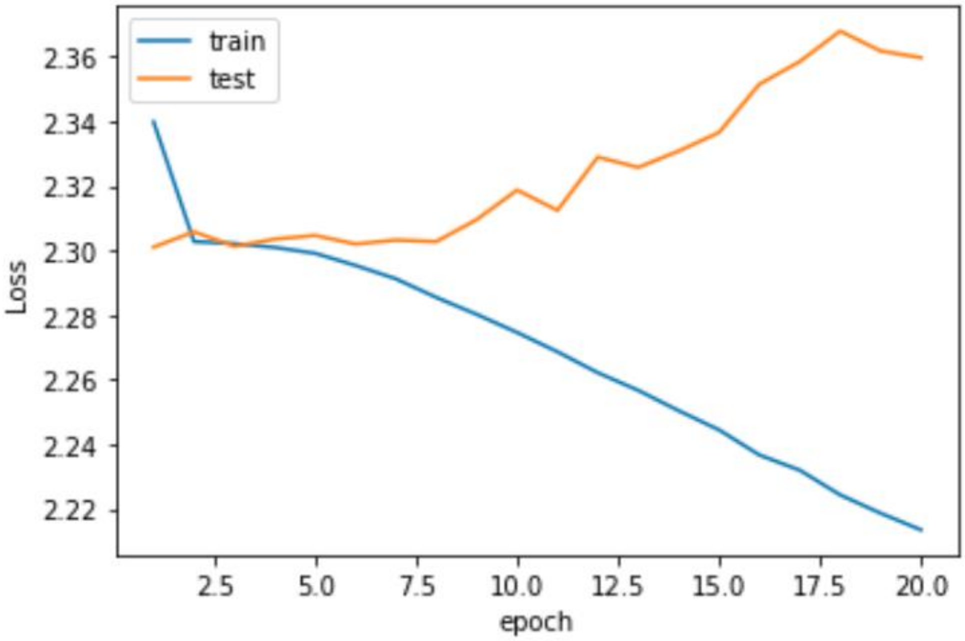
Optimizer ; Adam

Model Summary:

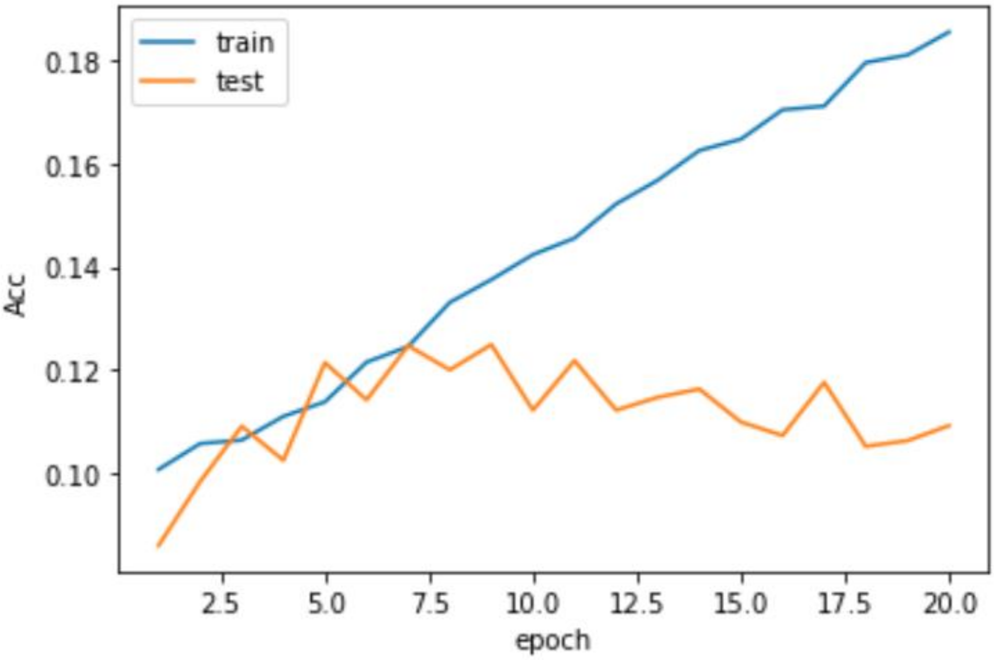


```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 16, 32, 32]           1,216
       BatchNorm2d-2           [-1, 16, 32, 32]              32
              ReLU-3           [-1, 16, 32, 32]               0
         MaxPool2d-4           [-1, 16, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]          12,832
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 34,634
Trainable params: 34,634
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.61
Params size (MB): 0.13
Estimated Total Size (MB): 0.75
----------------------------------------------------------------
```

Loss chart


train
test
Loss
2.36
2.34
2.32
2.30
2.28
2.26
2.24
2.22
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
epoch

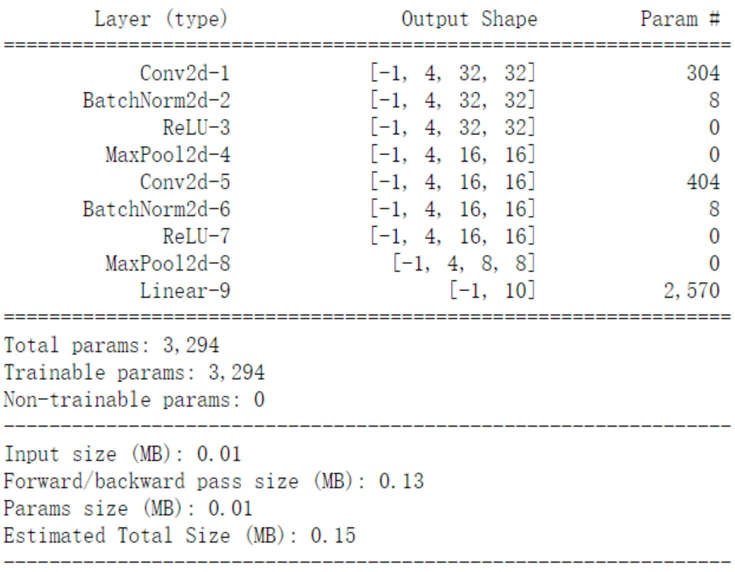
ACC chart


train
test
Acc
0.18
0.16
0.14
0.12
0.10
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
epoch

2. Number of parameters v.s. Generalization

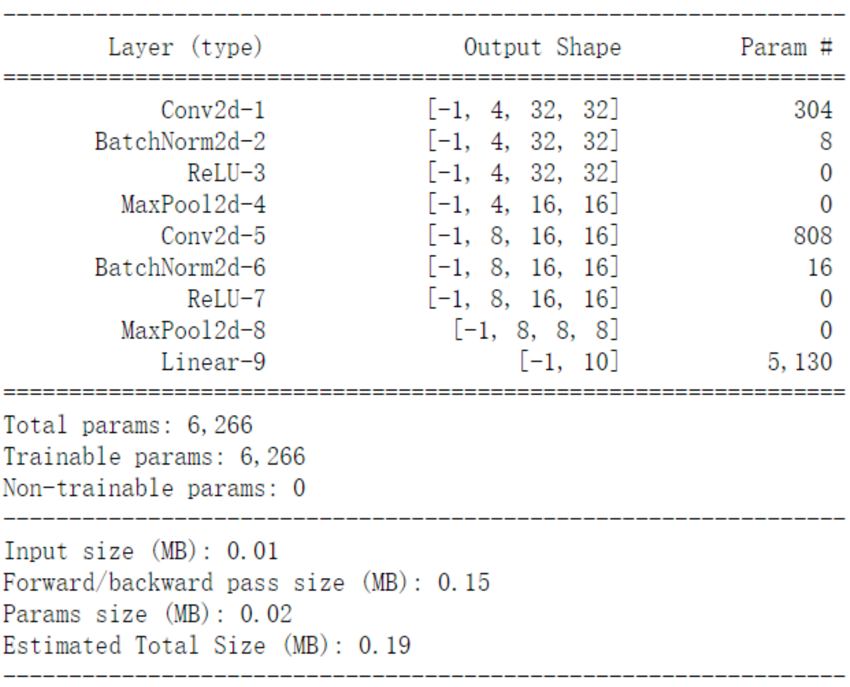
10 models

Model1



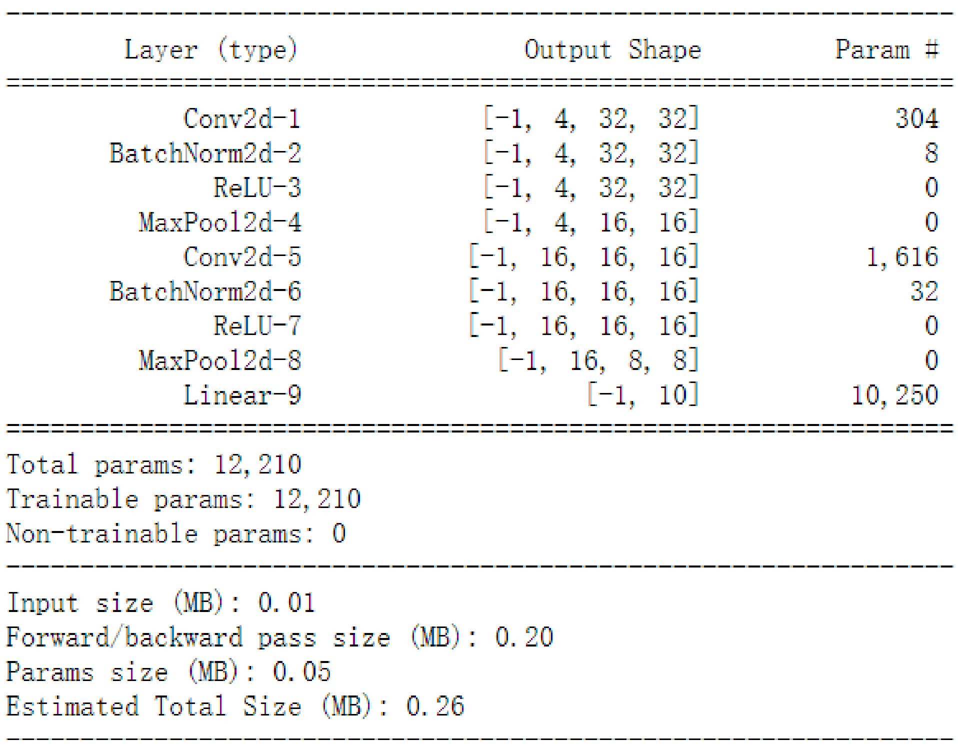
```
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1            [-1, 4, 32, 32]             304
       BatchNorm2d-2            [-1, 4, 32, 32]               8
              ReLU-3            [-1, 4, 32, 32]               0
         MaxPool2d-4            [-1, 4, 16, 16]               0
            Conv2d-5            [-1, 4, 16, 16]             404
       BatchNorm2d-6            [-1, 4, 16, 16]               8
              ReLU-7            [-1, 4, 16, 16]               0
         MaxPool2d-8              [-1, 4, 8, 8]               0
            Linear-9                   [-1, 10]           2,570
================================================================
Total params: 3,294
Trainable params: 3,294
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.13
Params size (MB): 0.01
Estimated Total Size (MB): 0.15
----------------------------------------------------------------
```

Model 2



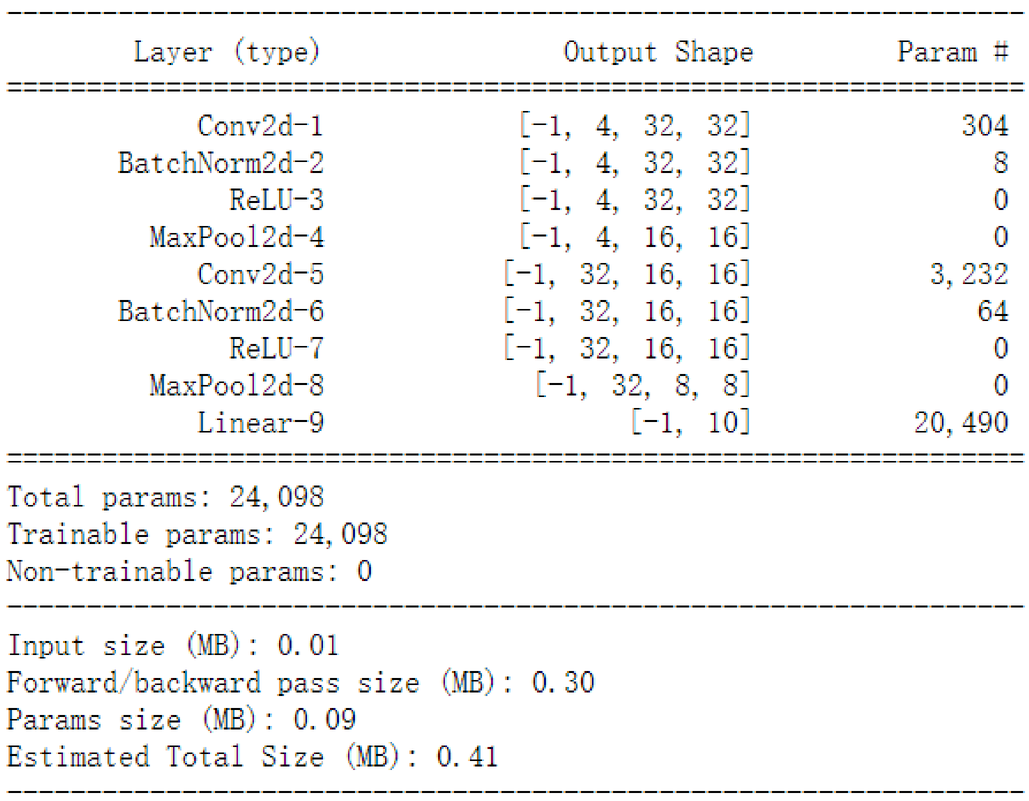
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1            [-1, 4, 32, 32]             304
       BatchNorm2d-2            [-1, 4, 32, 32]               8
              ReLU-3            [-1, 4, 32, 32]               0
         MaxPool2d-4            [-1, 4, 16, 16]               0
            Conv2d-5            [-1, 8, 16, 16]             808
       BatchNorm2d-6            [-1, 8, 16, 16]              16
              ReLU-7            [-1, 8, 16, 16]               0
         MaxPool2d-8              [-1, 8, 8, 8]               0
            Linear-9                   [-1, 10]           5,130
================================================================
Total params: 6,266
Trainable params: 6,266
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.15
Params size (MB): 0.02
Estimated Total Size (MB): 0.19
----------------------------------------------------------------
```

Model 3



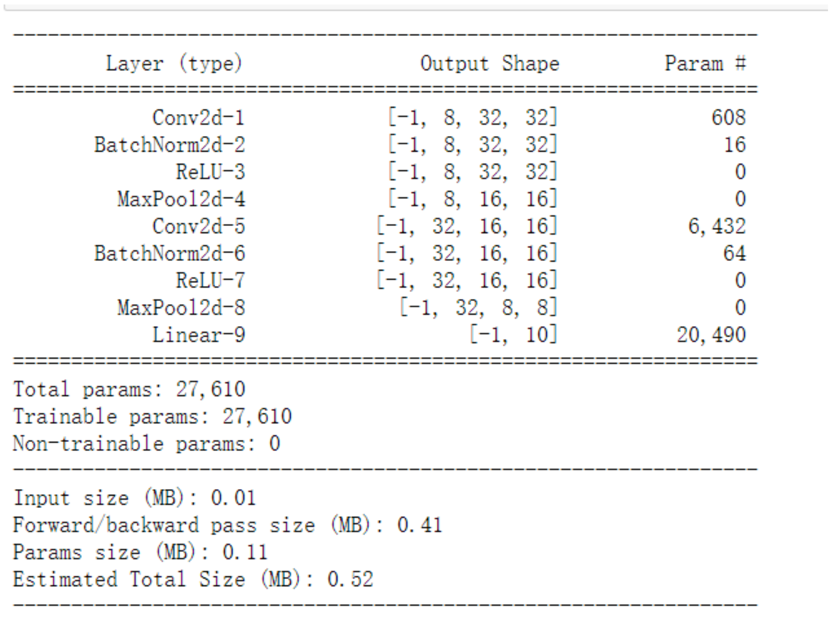
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1            [-1, 4, 32, 32]             304
       BatchNorm2d-2            [-1, 4, 32, 32]               8
              ReLU-3            [-1, 4, 32, 32]               0
         MaxPool2d-4            [-1, 4, 16, 16]               0
            Conv2d-5           [-1, 16, 16, 16]           1,616
       BatchNorm2d-6           [-1, 16, 16, 16]              32
              ReLU-7           [-1, 16, 16, 16]               0
         MaxPool2d-8             [-1, 16, 8, 8]               0
            Linear-9                   [-1, 10]          10,250
================================================================
Total params: 12,210
Trainable params: 12,210
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.20
Params size (MB): 0.05
Estimated Total Size (MB): 0.26
----------------------------------------------------------------
```

Model 4



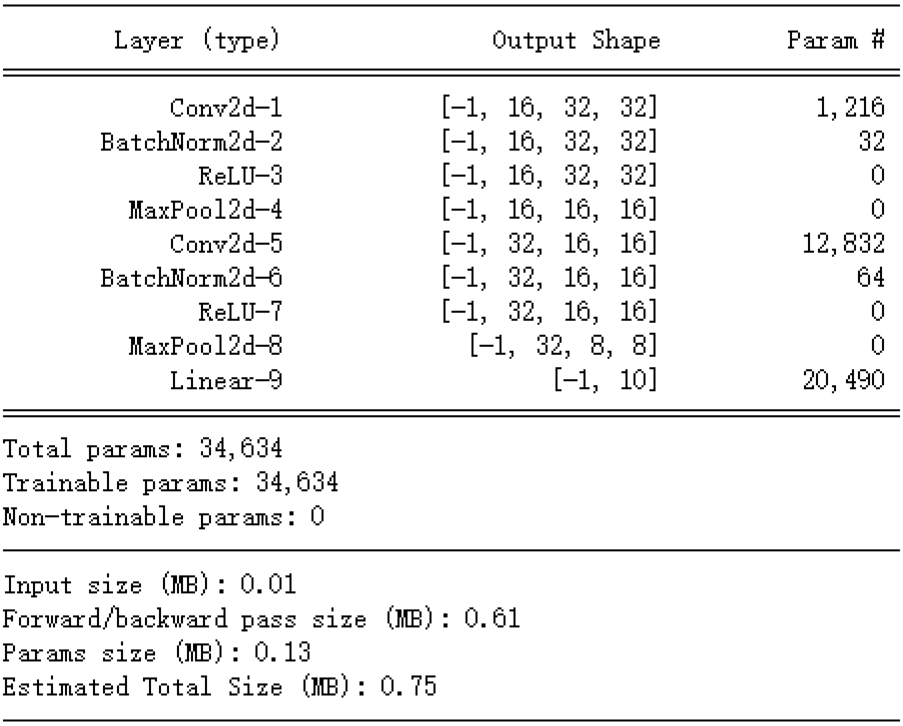
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1            [-1, 4, 32, 32]             304
       BatchNorm2d-2            [-1, 4, 32, 32]               8
              ReLU-3            [-1, 4, 32, 32]               0
         MaxPool2d-4            [-1, 4, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]           3,232
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 24,098
Trainable params: 24,098
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.30
Params size (MB): 0.09
Estimated Total Size (MB): 0.41
----------------------------------------------------------------
```

Model 5



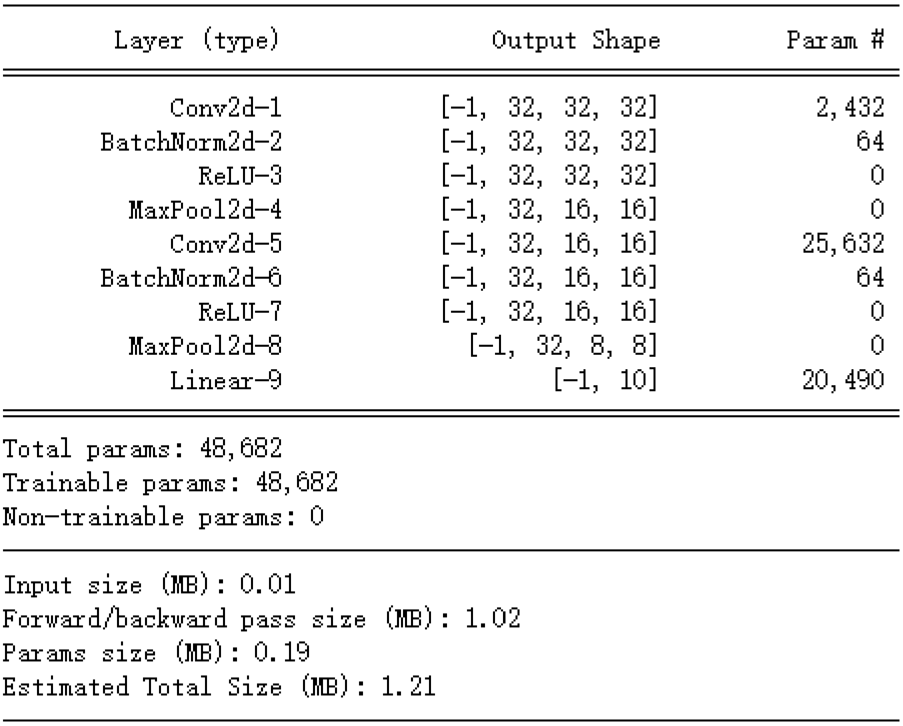
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1            [-1, 8, 32, 32]             608
       BatchNorm2d-2            [-1, 8, 32, 32]              16
              ReLU-3            [-1, 8, 32, 32]               0
         MaxPool2d-4            [-1, 8, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]           6,432
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 27,610
Trainable params: 27,610
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.41
Params size (MB): 0.11
Estimated Total Size (MB): 0.52
----------------------------------------------------------------
```

Model 6



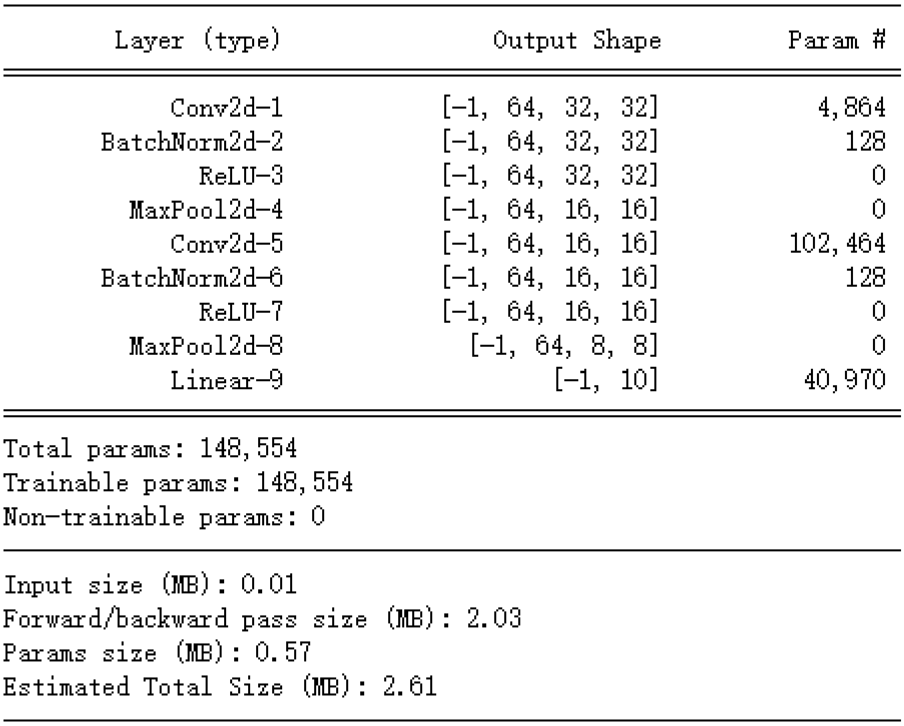
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 16, 32, 32]           1,216
       BatchNorm2d-2           [-1, 16, 32, 32]              32
              ReLU-3           [-1, 16, 32, 32]               0
         MaxPool2d-4           [-1, 16, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]          12,832
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 34,634
Trainable params: 34,634
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 0.61
Params size (MB): 0.13
Estimated Total Size (MB): 0.75
----------------------------------------------------------------
```

Model 7



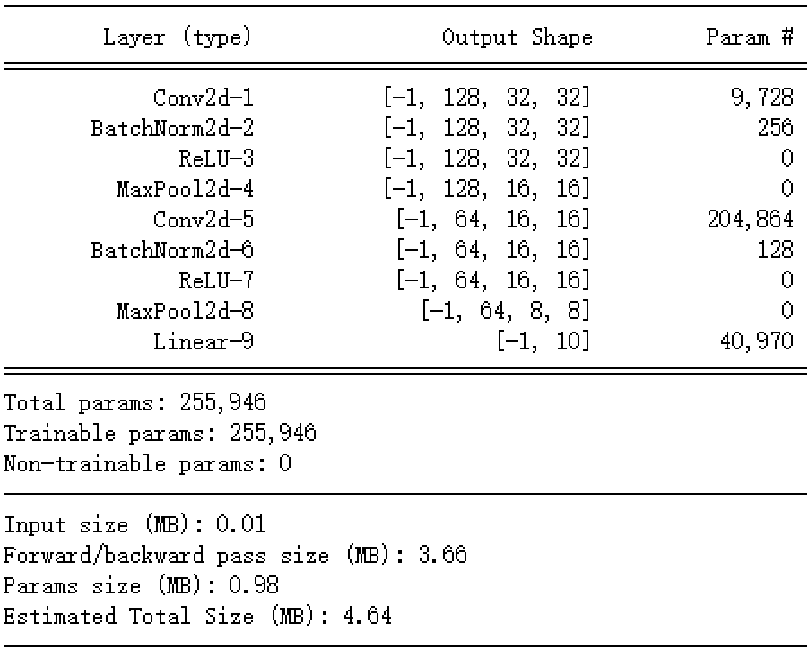
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 32, 32, 32]           2,432
       BatchNorm2d-2           [-1, 32, 32, 32]              64
              ReLU-3           [-1, 32, 32, 32]               0
         MaxPool2d-4           [-1, 32, 16, 16]               0
            Conv2d-5           [-1, 32, 16, 16]          25,632
       BatchNorm2d-6           [-1, 32, 16, 16]              64
              ReLU-7           [-1, 32, 16, 16]               0
         MaxPool2d-8             [-1, 32, 8, 8]               0
            Linear-9                   [-1, 10]          20,490
================================================================
Total params: 48,682
Trainable params: 48,682
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 1.02
Params size (MB): 0.19
Estimated Total Size (MB): 1.21
----------------------------------------------------------------
```

Model 8



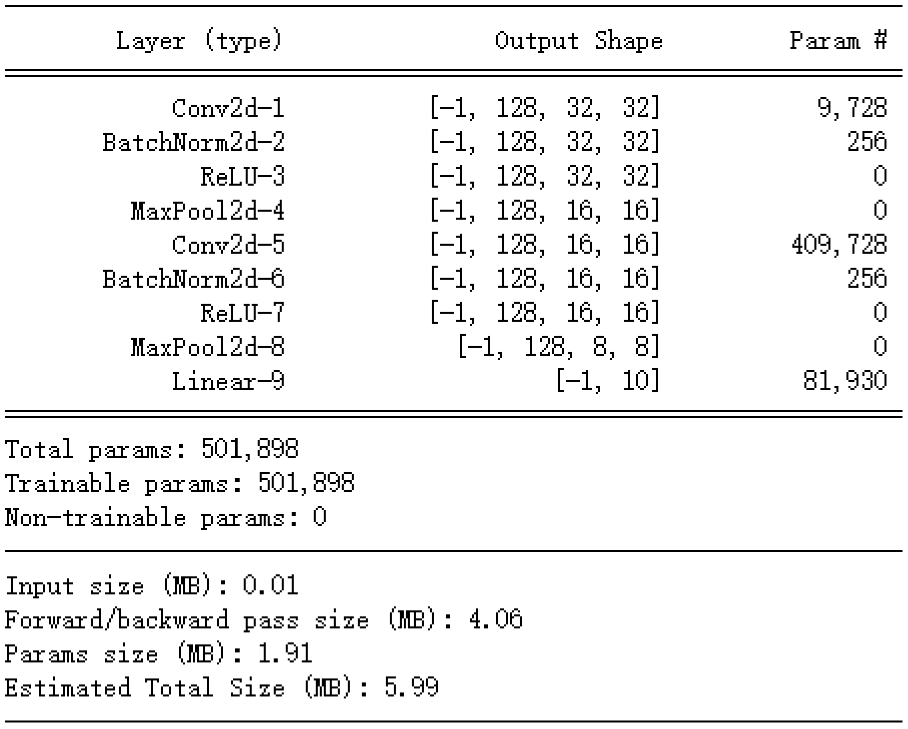
```
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1           [-1, 64, 32, 32]           4,864
       BatchNorm2d-2           [-1, 64, 32, 32]             128
              ReLU-3           [-1, 64, 32, 32]               0
         MaxPool2d-4           [-1, 64, 16, 16]               0
            Conv2d-5           [-1, 64, 16, 16]         102,464
       BatchNorm2d-6           [-1, 64, 16, 16]             128
              ReLU-7           [-1, 64, 16, 16]               0
         MaxPool2d-8             [-1, 64, 8, 8]               0
            Linear-9                   [-1, 10]          40,970
================================================================
Total params: 148,554
Trainable params: 148,554
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 2.03
Params size (MB): 0.57
Estimated Total Size (MB): 2.61
----------------------------------------------------------------
```

Model 9



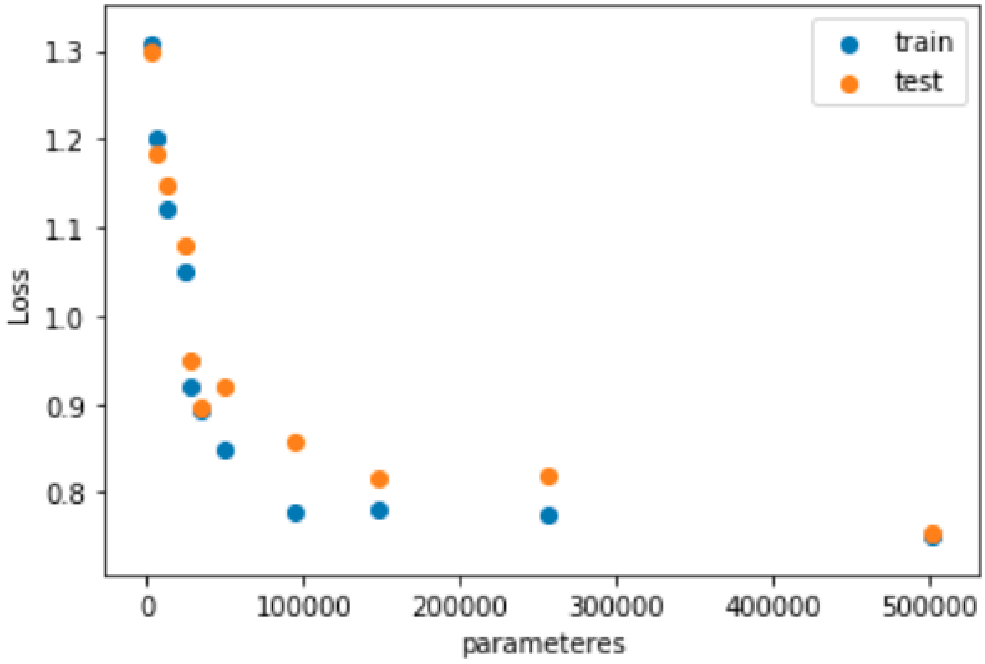
```
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1          [-1, 128, 32, 32]           9,728
       BatchNorm2d-2          [-1, 128, 32, 32]             256
              ReLU-3          [-1, 128, 32, 32]               0
         MaxPool2d-4          [-1, 128, 16, 16]               0
            Conv2d-5           [-1, 64, 16, 16]         204,864
       BatchNorm2d-6           [-1, 64, 16, 16]             128
              ReLU-7           [-1, 64, 16, 16]               0
         MaxPool2d-8             [-1, 64, 8, 8]               0
            Linear-9                   [-1, 10]          40,970
================================================================
Total params: 255,946
Trainable params: 255,946
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 3.66
Params size (MB): 0.98
Estimated Total Size (MB): 4.64
----------------------------------------------------------------
```

Model 10

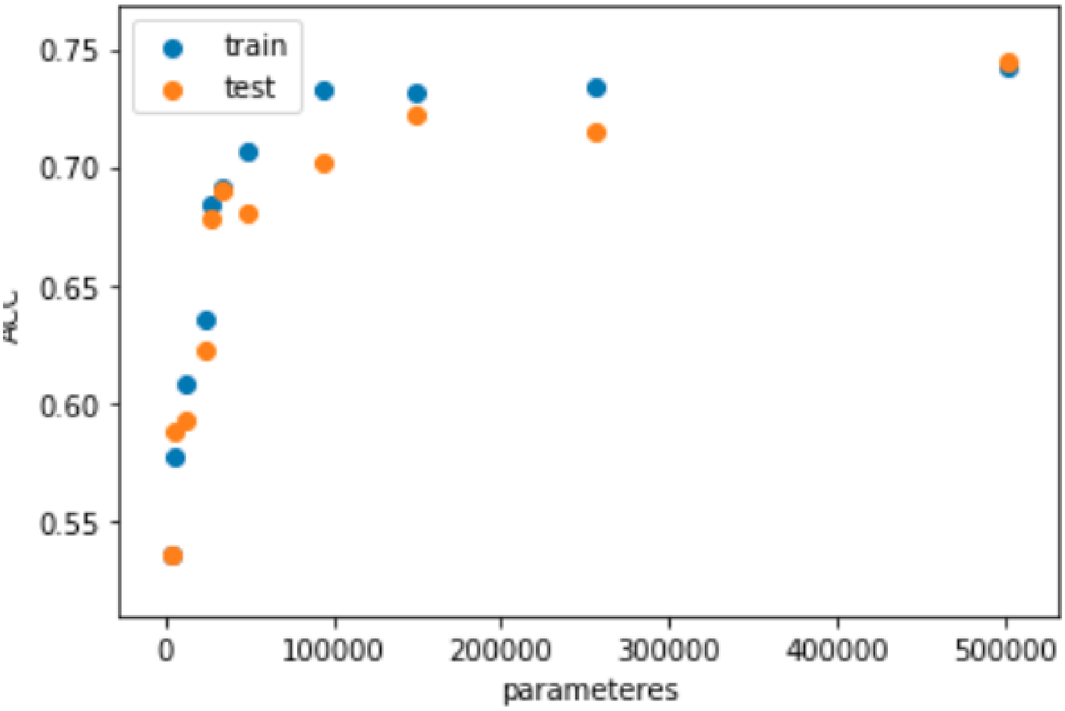


```
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1         [-1, 128, 32, 32]           9,728
       BatchNorm2d-2         [-1, 128, 32, 32]             256
              ReLU-3         [-1, 128, 32, 32]               0
         MaxPool2d-4         [-1, 128, 16, 16]               0
            Conv2d-5         [-1, 128, 16, 16]         409,728
       BatchNorm2d-6         [-1, 128, 16, 16]             256
              ReLU-7         [-1, 128, 16, 16]               0
         MaxPool2d-8           [-1, 128, 8, 8]               0
            Linear-9                  [-1, 10]          81,930
================================================================
Total params: 501,898
Trainable params: 501,898
Non-trainable params: 0
----------------------------------------------------------------
Input size (MB): 0.01
Forward/backward pass size (MB): 4.06
Params size (MB): 1.91
Estimated Total Size (MB): 5.99
----------------------------------------------------------------
```

Train loss Test Loss Chart



Train ACC test ACC chart

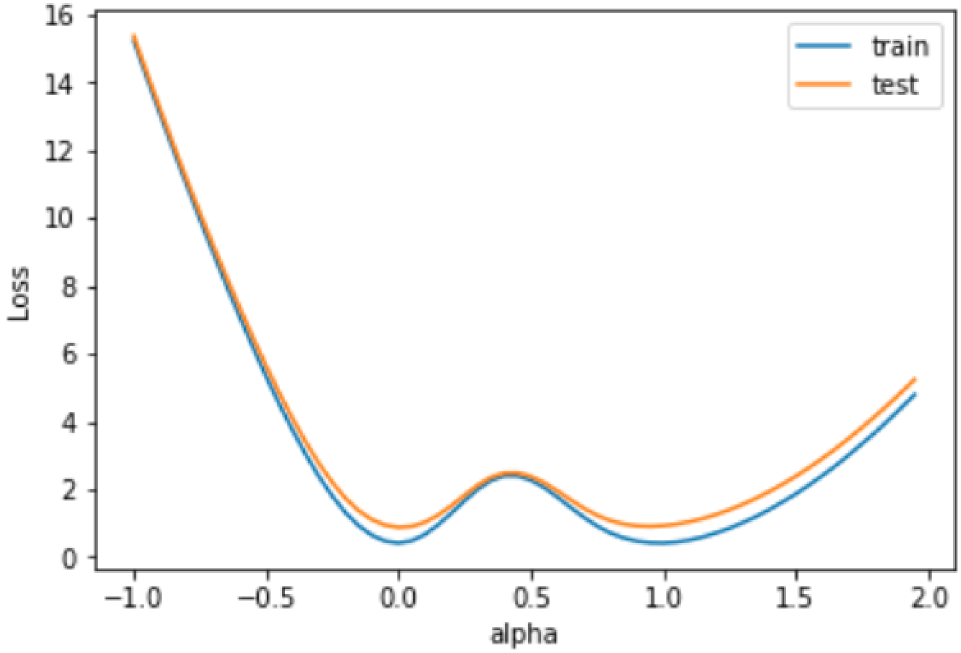


HW1-3

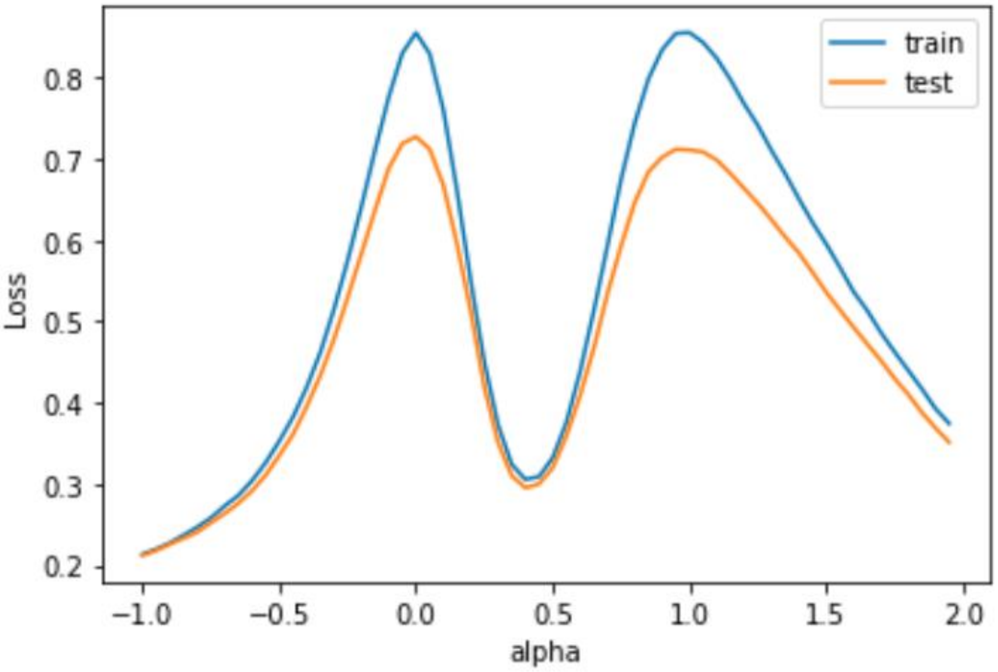
Subtask 1

Batchsize 32 vs 64

Train and Loss Chart



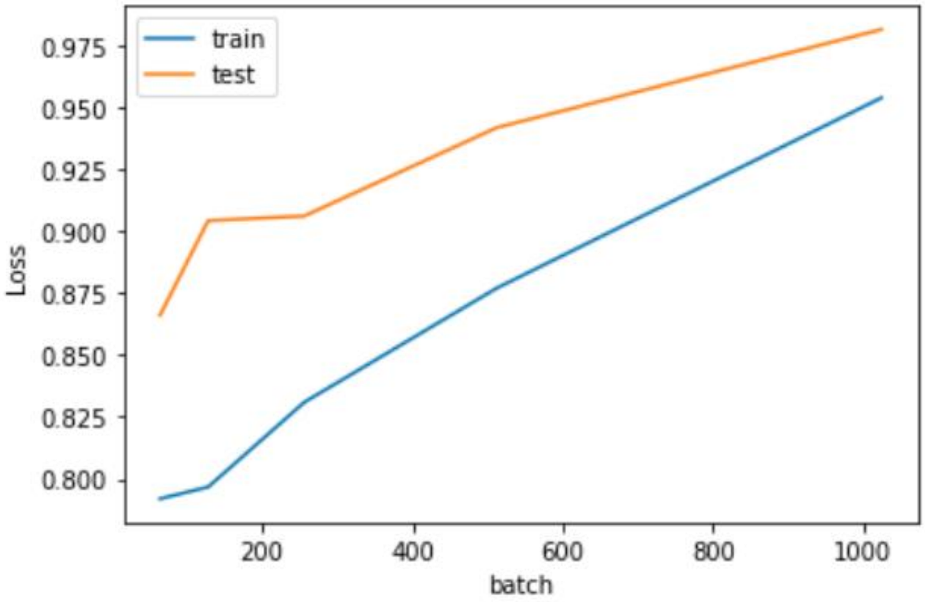
Train and Test Acc Chart



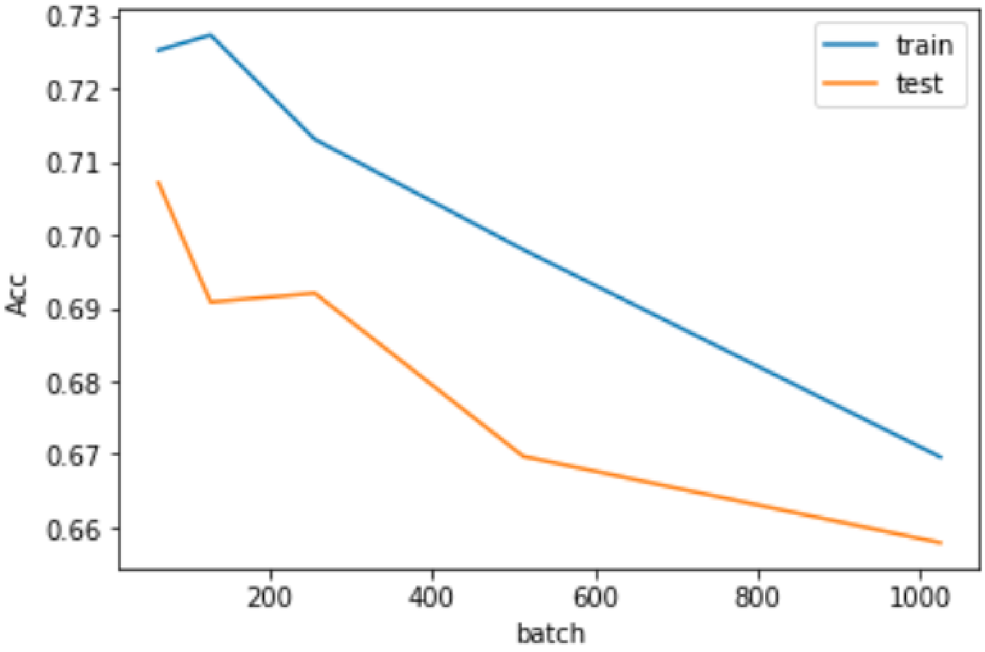
SubTask 2

5 differnet models with batchseize(64 ,128,256,512,1024)

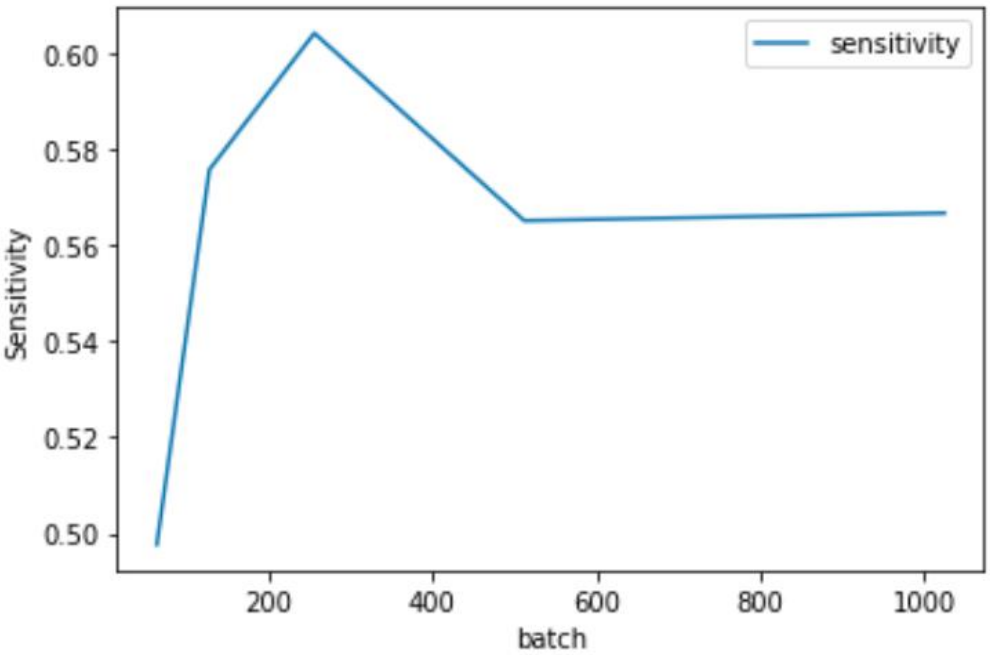
Train Test Loss Chart



Train Test ACC Chart



Sensitivity Chart


sensitivity
Sensitivity
0.50
0.52
0.54
0.56
0.58
0.60
200
400
600
800
1000
batch