



Figure 1: Mnist Visualization

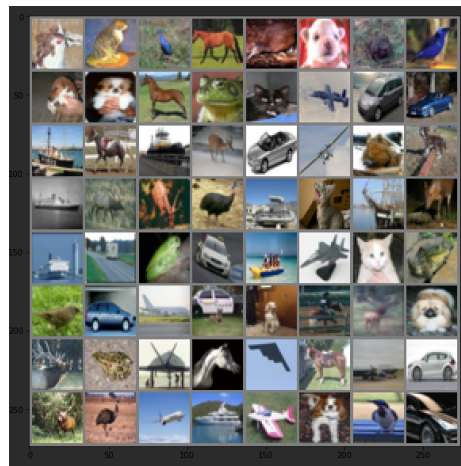


Figure 2: Cifar Visualization



Figure 3: Linear Network Weights: 10 by 874

```

Train Epoch: 1 (1000000/0%) Loss: 3.768914
Test set: Average loss: 3.6394, Accuracy: 1666/10000 (16%)
Train Epoch: 2 (1000000/0%) Loss: 4.122225
Test set: Average loss: 3.9362, Accuracy: 866/10000 (8%)
Train Epoch: 3 (1000000/0%) Loss: 5.032263
Test set: Average loss: 3.8315, Accuracy: 1142/10000 (11%)
Train Epoch: 4 (1000000/0%) Loss: 4.602869
Test set: Average loss: 3.1895, Accuracy: 1606/10000 (16%)
Train Epoch: 5 (1000000/0%) Loss: 3.838709
Test set: Average loss: 3.8312, Accuracy: 1477/10000 (14%)
Train Epoch: 6 (1000000/0%) Loss: 4.077387
Test set: Average loss: 2.8818, Accuracy: 2281/10000 (22%)
Train Epoch: 7 (1000000/0%) Loss: 3.097869
Test set: Average loss: 2.7273, Accuracy: 2637/10000 (26%)
Train Epoch: 8 (1000000/0%) Loss: 0.823004
Test set: Average loss: 2.0885, Accuracy: 3755/10000 (37%)
Train Epoch: 9 (1000000/0%) Loss: 0.062023
Test set: Average loss: 2.0895, Accuracy: 3748/10000 (37%)
Train Epoch: 10 (1000000/0%) Loss: 0.096867
Test set: Average loss: 2.7006, Accuracy: 3743/10000 (37%)

```

Figure 4: Small Dataset(50): We see a huge difference between the training accuracy and test accuracy as a result of overfitting on a small dataset. The test accuracy doesn't decrease because learning via backpropagation requires lots of data.

```

Train Epoch: 7 [0/60000 (0%)] Loss: 0.148668
Train Epoch: 7 [6400/60000 (11%)] Loss: 0.173039
Train Epoch: 7 [12800/60000 (21%)] Loss: 0.237750
Train Epoch: 7 [19200/60000 (32%)] Loss: 0.199186
Train Epoch: 7 [25600/60000 (43%)] Loss: 0.224152
Train Epoch: 7 [32000/60000 (53%)] Loss: 0.176284
Train Epoch: 7 [38400/60000 (64%)] Loss: 0.238162
Train Epoch: 7 [44800/60000 (75%)] Loss: 0.180425
Train Epoch: 7 [51200/60000 (85%)] Loss: 0.185717
Train Epoch: 7 [57600/60000 (96%)] Loss: 0.107280

Test set: Average loss: 0.1946, Accuracy: 9442/10000 (94%)

Train Epoch: 8 [0/60000 (0%)] Loss: 0.126643
Train Epoch: 8 [6400/60000 (11%)] Loss: 0.152318
Train Epoch: 8 [12800/60000 (21%)] Loss: 0.086708
Train Epoch: 8 [19200/60000 (32%)] Loss: 0.109821
Train Epoch: 8 [25600/60000 (43%)] Loss: 0.169002
Train Epoch: 8 [32000/60000 (53%)] Loss: 0.209483
Train Epoch: 8 [38400/60000 (64%)] Loss: 0.227308
Train Epoch: 8 [44800/60000 (75%)] Loss: 0.136739
Train Epoch: 8 [51200/60000 (85%)] Loss: 0.190427
Train Epoch: 8 [57600/60000 (96%)] Loss: 0.141736

Test set: Average loss: 0.1840, Accuracy: 9483/10000 (95%)

Train Epoch: 9 [0/60000 (0%)] Loss: 0.233190
Train Epoch: 9 [6400/60000 (11%)] Loss: 0.137709
Train Epoch: 9 [12800/60000 (21%)] Loss: 0.186599
Train Epoch: 9 [19200/60000 (32%)] Loss: 0.113850
Train Epoch: 9 [25600/60000 (43%)] Loss: 0.156787
Train Epoch: 9 [32000/60000 (53%)] Loss: 0.234822
Train Epoch: 9 [38400/60000 (64%)] Loss: 0.183888
Train Epoch: 9 [44800/60000 (75%)] Loss: 0.308970
Train Epoch: 9 [51200/60000 (85%)] Loss: 0.100252
Train Epoch: 9 [57600/60000 (96%)] Loss: 0.240479

Test set: Average loss: 0.1715, Accuracy: 9513/10000 (95%)

Train Epoch: 10 [0/60000 (0%)] Loss: 0.133540
Train Epoch: 10 [6400/60000 (11%)] Loss: 0.393999
Train Epoch: 10 [12800/60000 (21%)] Loss: 0.153392
Train Epoch: 10 [19200/60000 (32%)] Loss: 0.302003
Train Epoch: 10 [25600/60000 (43%)] Loss: 0.165758
Train Epoch: 10 [32000/60000 (53%)] Loss: 0.170846
Train Epoch: 10 [38400/60000 (64%)] Loss: 0.050763
Train Epoch: 10 [44800/60000 (75%)] Loss: 0.115670
Train Epoch: 10 [51200/60000 (85%)] Loss: 0.079023
Train Epoch: 10 [57600/60000 (96%)] Loss: 0.092462

Test set: Average loss: 0.1631, Accuracy: 9538/10000 (95%)

```

Figure 5: Multi-layer on Mnist

```

Train Epoch: 7 [25600/60000 (43%)] Loss: 69.565025
Train Epoch: 7 [32000/60000 (53%)] Loss: 371.484039
Train Epoch: 7 [38400/60000 (64%)] Loss: 273.383514
Train Epoch: 7 [44800/60000 (75%)] Loss: 124.724365
Train Epoch: 7 [51200/60000 (85%)] Loss: 244.720657
Train Epoch: 7 [57600/60000 (96%)] Loss: 91.018661

Test set: Average loss: 121.0272, Accuracy: 7597/10000 (76%)

Train Epoch: 8 [0/60000 (0%)] Loss: 116.724335
Train Epoch: 8 [6400/60000 (11%)] Loss: 282.993805
Train Epoch: 8 [12800/60000 (21%)] Loss: 33.215286
Train Epoch: 8 [19200/60000 (32%)] Loss: 69.531570
Train Epoch: 8 [25600/60000 (43%)] Loss: 114.619385
Train Epoch: 8 [32000/60000 (53%)] Loss: 536.734375
Train Epoch: 8 [38400/60000 (64%)] Loss: 132.940460
Train Epoch: 8 [44800/60000 (75%)] Loss: 202.454193
Train Epoch: 8 [51200/60000 (85%)] Loss: 253.879196
Train Epoch: 8 [57600/60000 (96%)] Loss: 576.171814

Test set: Average loss: 171.3837, Accuracy: 6779/10000 (68%)

Train Epoch: 9 [0/60000 (0%)] Loss: 143.128876
Train Epoch: 9 [6400/60000 (11%)] Loss: 161.464569
Train Epoch: 9 [12800/60000 (21%)] Loss: 202.136993
Train Epoch: 9 [19200/60000 (32%)] Loss: 267.131104
Train Epoch: 9 [25600/60000 (43%)] Loss: 167.889282
Train Epoch: 9 [32000/60000 (53%)] Loss: 73.554016
Train Epoch: 9 [38400/60000 (64%)] Loss: 197.981659
Train Epoch: 9 [44800/60000 (75%)] Loss: 437.450439
Train Epoch: 9 [51200/60000 (85%)] Loss: 91.220932
Train Epoch: 9 [57600/60000 (96%)] Loss: 112.899933

Test set: Average loss: 364.0168, Accuracy: 6346/10000 (63%)

Train Epoch: 10 [0/60000 (0%)] Loss: 320.816193
Train Epoch: 10 [6400/60000 (11%)] Loss: 228.525543
Train Epoch: 10 [12800/60000 (21%)] Loss: 250.340927
Train Epoch: 10 [19200/60000 (32%)] Loss: 128.350937
Train Epoch: 10 [25600/60000 (43%)] Loss: 107.277092
Train Epoch: 10 [32000/60000 (53%)] Loss: 107.814934
Train Epoch: 10 [38400/60000 (64%)] Loss: 138.165482
Train Epoch: 10 [44800/60000 (75%)] Loss: 175.019958
Train Epoch: 10 [51200/60000 (85%)] Loss: 84.999344
Train Epoch: 10 [57600/60000 (96%)] Loss: 271.819702

```

Figure 6: As a result of changing the learning rate, When setting learning rate too high we get a classic divergence scenario whereby the loss function grows very high and we dont see any improvement in the test set

```

Train Epoch: 12 [0/60000 (0%)] Loss: 6.896702
Train Epoch: 12 [6400/60000 (11%)] Loss: 6.896702
Train Epoch: 12 [12800/60000 (21%)] Loss: 6.896702
Train Epoch: 12 [19200/60000 (32%)] Loss: 6.896702
Train Epoch: 12 [25600/60000 (43%)] Loss: 6.896702
Train Epoch: 12 [32000/60000 (53%)] Loss: 6.896702
Train Epoch: 12 [38400/60000 (64%)] Loss: 6.896702
Train Epoch: 12 [44800/60000 (75%)] Loss: 6.896702
Train Epoch: 12 [51200/60000 (85%)] Loss: 6.896702
Train Epoch: 12 [57600/60000 (96%)] Loss: 6.896702

Test set: Average loss: 0.8453, Accuracy: 8315/10000 (83%)

Train Epoch: 13 [0/60000 (0%)] Loss: 6.896702
Train Epoch: 13 [6400/60000 (11%)] Loss: 6.896702
Train Epoch: 13 [12800/60000 (21%)] Loss: 6.896702
Train Epoch: 13 [19200/60000 (32%)] Loss: 6.896702
Train Epoch: 13 [25600/60000 (43%)] Loss: 6.896702
Train Epoch: 13 [32000/60000 (53%)] Loss: 6.896702
Train Epoch: 13 [38400/60000 (64%)] Loss: 6.896702
Train Epoch: 13 [44800/60000 (75%)] Loss: 6.896702
Train Epoch: 13 [51200/60000 (85%)] Loss: 6.896702
Train Epoch: 13 [57600/60000 (96%)] Loss: 6.896702

Test set: Average loss: 0.8453, Accuracy: 8315/10000 (83%)

Train Epoch: 14 [0/60000 (0%)] Loss: 6.896702
Train Epoch: 14 [6400/60000 (11%)] Loss: 6.896702
Train Epoch: 14 [12800/60000 (21%)] Loss: 6.896702
Train Epoch: 14 [19200/60000 (32%)] Loss: 6.896702
Train Epoch: 14 [25600/60000 (43%)] Loss: 6.896702
Train Epoch: 14 [32000/60000 (53%)] Loss: 6.896702
Train Epoch: 14 [38400/60000 (64%)] Loss: 6.896702
Train Epoch: 14 [44800/60000 (75%)] Loss: 6.896702
Train Epoch: 14 [51200/60000 (85%)] Loss: 6.896702
Train Epoch: 14 [57600/60000 (96%)] Loss: 6.896702

Test set: Average loss: 0.8453, Accuracy: 8315/10000 (83%)

Train Epoch: 15 [0/60000 (0%)] Loss: 6.896702
Train Epoch: 15 [6400/60000 (11%)] Loss: 6.896702
Train Epoch: 15 [12800/60000 (21%)] Loss: 6.896702
Train Epoch: 15 [19200/60000 (32%)] Loss: 6.896702
Train Epoch: 15 [25600/60000 (43%)] Loss: 6.896702
Train Epoch: 15 [32000/60000 (53%)] Loss: 6.896702
Train Epoch: 15 [38400/60000 (64%)] Loss: 6.896702
Train Epoch: 15 [44800/60000 (75%)] Loss: 6.896702
Train Epoch: 15 [51200/60000 (85%)] Loss: 6.896702
Train Epoch: 15 [57600/60000 (96%)] Loss: 6.896702

Test set: Average loss: 0.7790, Accuracy: 8571/10000 (86%)

Results filename: cifar100_10000_0

```

Figure 7: Cifar convolution Network



Figure 8: First Layer filter of the convolution Network

```

over all number of parameters are 8
network structure is Net (
  (conv1): Conv2d(3, 16, kernel_size=(5, 5), stride=(1, 1))
  (pool1): MaxPool2d (size=(2, 2), stride=(2, 2), dilation=(1, 1))
  (conv2): Conv2d(16, 128, kernel_size=(5, 5), stride=(1, 1))
  (pool2): MaxPool2d (size=(2, 2), stride=(2, 2), dilation=(1, 1))
  (linear1): Linear (3200 -> 64)
  (linear2): Linear (64 -> 10)
)
0 th parameter has a length of torch.Size([16, 3, 5, 5])
1 th parameter has a length of torch.Size([16])
2 th parameter has a length of torch.Size([128, 16, 5, 5])
3 th parameter has a length of torch.Size([128])
4 th parameter has a length of torch.Size([64, 3200])
5 th parameter has a length of torch.Size([64])
6 th parameter has a length of torch.Size([10, 64])
7 th parameter has a length of torch.Size([10])

```

Figure 9: Parameters