```
GPU 128
Train Epoch: 1 [0/60000 (0%)]       Loss: 2.414268
Train Epoch: 1 [6400/60000 (11%)] Loss: 0.277052
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.195978
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.243240
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.117855
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.221475
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.101384
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.094310
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.149132
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.107671

Epoch 1 time = 19.56s on device cuda

Test set: Average loss: 0.0459, Accuracy: 9848/10000 (98%)

Train Epoch: 2 [0/60000 (0%)]       Loss: 0.028224
Train Epoch: 2 [6400/60000 (11%)] Loss: 0.103099
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.043957
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.114285
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.174780
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.107917
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.148767
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.138543
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.119291
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.050153

Epoch 2 time = 12.44s on device cuda

Test set: Average loss: 0.0351, Accuracy: 9872/10000 (99%)

Train Epoch: 3 [0/60000 (0%)]       Loss: 0.065273
Train Epoch: 3 [6400/60000 (11%)] Loss: 0.075438
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.048246
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.022881
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.196851
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.021625
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.023448
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.022568
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.068177
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.048692

Epoch 3 time = 12.61s on device cuda

Test set: Average loss: 0.0327, Accuracy: 9891/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]       Loss: 0.057840
Train Epoch: 4 [6400/60000 (11%)] Loss: 0.043756
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.038450
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.096252
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.029444
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.039033
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.125818
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.106110
```

```
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.070577
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.011196

Epoch 4 time = 12.56s on device cuda

Test set: Average loss: 0.0303, Accuracy: 9901/10000 (99%)

Train Epoch: 5 [0/60000 (0%)]      Loss: 0.023543
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.018263
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.013048
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.045010
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.026486
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.018411
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.055432
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.102279
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.012677
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.033569

Epoch 5 time = 12.67s on device cuda

Test set: Average loss: 0.0287, Accuracy: 9911/10000 (99%)

GPU 256

Train Epoch: 1 [0/60000 (0%)]      Loss: 2.305827
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.339289
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.159876
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.142564
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.157582

Epoch 1 time = 18.47s on device cuda

Test set: Average loss: 0.0532, Accuracy: 9826/10000 (98%)

Train Epoch: 2 [0/60000 (0%)]      Loss: 0.101965
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.090707
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.086724
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.092786
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.092717

Epoch 2 time = 11.81s on device cuda

Test set: Average loss: 0.0361, Accuracy: 9873/10000 (99%)

Train Epoch: 3 [0/60000 (0%)]      Loss: 0.041301
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.031700
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.024200
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.057526
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.082412

Epoch 3 time = 11.97s on device cuda

Test set: Average loss: 0.0351, Accuracy: 9874/10000 (99%)
```

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.052348
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.056083
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.044855
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.049951
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.050230
```

```
Epoch 4 time = 12.05s on device cuda
```

```
Test set: Average loss: 0.0302, Accuracy: 9894/10000 (99%)
```

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.008038
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.109669
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.043095
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.033641
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.026734
```

```
Epoch 5 time = 12.02s on device cuda
```

```
Test set: Average loss: 0.0277, Accuracy: 9907/10000 (99%)
```

```
GPU 512
```

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 2.307430
Train Epoch: 1 [25600/60000 (42%)] Loss: 0.296851
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.220367
```

```
Epoch 1 time = 11.36s on device cuda
```

```
Test set: Average loss: 0.0796, Accuracy: 9761/10000 (98%)
```

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.189258
Train Epoch: 2 [25600/60000 (42%)] Loss: 0.092277
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.092924
```

```
Epoch 2 time = 11.32s on device cuda
```

```
Test set: Average loss: 0.0636, Accuracy: 9784/10000 (98%)
```

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.132860
Train Epoch: 3 [25600/60000 (42%)] Loss: 0.061310
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.122080
```

```
Epoch 3 time = 11.28s on device cuda
```

```
Test set: Average loss: 0.0390, Accuracy: 9861/10000 (99%)
```

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.057364
Train Epoch: 4 [25600/60000 (42%)] Loss: 0.051902
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.081028
```

```
Epoch 4 time = 11.27s on device cuda
```

```
Test set: Average loss: 0.0383, Accuracy: 9864/10000 (99%)
```

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.043222
Train Epoch: 5 [25600/60000 (42%)] Loss: 0.047572
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.046221
```

```
Epoch 5 time = 11.21s on device cuda
```

```
Test set: Average loss: 0.0323, Accuracy: 9882/10000 (99%)
```

```
GPU 64
```

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 2.302757
Train Epoch: 1 [3200/60000 (5%)]   Loss: 0.663841
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.314508
Train Epoch: 1 [9600/60000 (16%)]  Loss: 0.265134
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.250680
Train Epoch: 1 [16000/60000 (27%)] Loss: 0.137991
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.072663
Train Epoch: 1 [22400/60000 (37%)] Loss: 0.063475
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.148939
Train Epoch: 1 [28800/60000 (48%)] Loss: 0.073996
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.103480
Train Epoch: 1 [35200/60000 (59%)] Loss: 0.249449
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.041360
Train Epoch: 1 [41600/60000 (69%)] Loss: 0.072210
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.075656
Train Epoch: 1 [48000/60000 (80%)] Loss: 0.047348
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.079196
Train Epoch: 1 [54400/60000 (91%)] Loss: 0.083810
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.091958
```

```
Epoch 1 time = 15.87s on device cuda
```

```
Test set: Average loss: 0.0470, Accuracy: 9862/10000 (99%)
```

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.114387
Train Epoch: 2 [3200/60000 (5%)]   Loss: 0.036257
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.026804
Train Epoch: 2 [9600/60000 (16%)]  Loss: 0.088935
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.051125
Train Epoch: 2 [16000/60000 (27%)] Loss: 0.005076
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.055300
Train Epoch: 2 [22400/60000 (37%)] Loss: 0.069889
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.196763
Train Epoch: 2 [28800/60000 (48%)] Loss: 0.073178
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.009077
Train Epoch: 2 [35200/60000 (59%)] Loss: 0.090675
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.034268
Train Epoch: 2 [41600/60000 (69%)] Loss: 0.075065
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.106786
Train Epoch: 2 [48000/60000 (80%)] Loss: 0.057757
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.347936
Train Epoch: 2 [54400/60000 (91%)] Loss: 0.033055
```

```
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.001779

Epoch 2 time = 16.03s on device cuda

Test set: Average loss: 0.0335, Accuracy: 9883/10000 (99%)

Train Epoch: 3 [0/60000 (0%)]      Loss: 0.002009
Train Epoch: 3 [3200/60000 (5%)]   Loss: 0.006993
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.014722
Train Epoch: 3 [9600/60000 (16%)]  Loss: 0.020234
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.052075
Train Epoch: 3 [16000/60000 (27%)] Loss: 0.007104
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.017671
Train Epoch: 3 [22400/60000 (37%)] Loss: 0.062173
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.003626
Train Epoch: 3 [28800/60000 (48%)] Loss: 0.204829
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.010573
Train Epoch: 3 [35200/60000 (59%)] Loss: 0.033103
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.044016
Train Epoch: 3 [41600/60000 (69%)] Loss: 0.007555
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.010252
Train Epoch: 3 [48000/60000 (80%)] Loss: 0.065834
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.173782
Train Epoch: 3 [54400/60000 (91%)] Loss: 0.004389
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.117682

Epoch 3 time = 15.6s on device cuda

Test set: Average loss: 0.0325, Accuracy: 9891/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]      Loss: 0.011286
Train Epoch: 4 [3200/60000 (5%)]   Loss: 0.010590
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.001513
Train Epoch: 4 [9600/60000 (16%)]  Loss: 0.134338
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.035225
Train Epoch: 4 [16000/60000 (27%)] Loss: 0.012188
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.101562
Train Epoch: 4 [22400/60000 (37%)] Loss: 0.004078
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.056426
Train Epoch: 4 [28800/60000 (48%)] Loss: 0.098866
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.061236
Train Epoch: 4 [35200/60000 (59%)] Loss: 0.000731
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.011979
Train Epoch: 4 [41600/60000 (69%)] Loss: 0.013794
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.062367
Train Epoch: 4 [48000/60000 (80%)] Loss: 0.004326
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.092676
Train Epoch: 4 [54400/60000 (91%)] Loss: 0.006838
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.027858

Epoch 4 time = 16.09s on device cuda

Test set: Average loss: 0.0283, Accuracy: 9911/10000 (99%)
```

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.004803
Train Epoch: 5 [3200/60000 (5%)]   Loss: 0.017680
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.035292
Train Epoch: 5 [9600/60000 (16%)]  Loss: 0.002603
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.016442
Train Epoch: 5 [16000/60000 (27%)] Loss: 0.006941
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.011521
Train Epoch: 5 [22400/60000 (37%)] Loss: 0.062540
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.016536
Train Epoch: 5 [28800/60000 (48%)] Loss: 0.021021
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.014862
Train Epoch: 5 [35200/60000 (59%)] Loss: 0.031728
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.003836
Train Epoch: 5 [41600/60000 (69%)] Loss: 0.070175
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.025889
Train Epoch: 5 [48000/60000 (80%)] Loss: 0.026629
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.014489
Train Epoch: 5 [54400/60000 (91%)] Loss: 0.002793
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.009545
```

```
Epoch 5 time = 15.43s on device cuda
```

```
Test set: Average loss: 0.0260, Accuracy: 9912/10000 (99%)
```

GPU 1024

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 2.312914
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.256935
```

```
Epoch 1 time = 11.32s on device cuda
```

```
Test set: Average loss: 0.1334, Accuracy: 9604/10000 (96%)
```

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.223696
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.129380
```

```
Epoch 2 time = 11.29s on device cuda
```

```
Test set: Average loss: 0.0714, Accuracy: 9767/10000 (98%)
```

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.108353
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.137626
```

```
Epoch 3 time = 11.13s on device cuda
```

```
Test set: Average loss: 0.0515, Accuracy: 9829/10000 (98%)
```

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.100201
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.077203
```

```
Epoch 4 time = 11.25s on device cuda

Test set: Average loss: 0.0464, Accuracy: 9847/10000 (98%)

Train Epoch: 5 [0/60000 (0%)]      Loss: 0.085672
Train Epoch: 5 [51200/60000 (85%)]Loss: 0.080177

Epoch 5 time = 11.24s on device cuda

Test set: Average loss: 0.0402, Accuracy: 9861/10000 (99%)


CPU 1024
Train Epoch: 1 [0/60000 (0%)]      Loss: 2.315055
Train Epoch: 1 [51200/60000 (85%)]Loss: 0.257950

Epoch 1 time = 202.34s on device cpu

Test set: Average loss: 0.1277, Accuracy: 9609/10000 (96%)

Train Epoch: 2 [0/60000 (0%)]      Loss: 0.209479
Train Epoch: 2 [51200/60000 (85%)]Loss: 0.136769

Epoch 2 time = 208.97s on device cpu

Test set: Average loss: 0.0664, Accuracy: 9790/10000 (98%)

Train Epoch: 3 [0/60000 (0%)]      Loss: 0.132879
Train Epoch: 3 [51200/60000 (85%)]Loss: 0.107420

Epoch 3 time = 204.3s on device cpu

Test set: Average loss: 0.0558, Accuracy: 9823/10000 (98%)

Train Epoch: 4 [0/60000 (0%)]      Loss: 0.121789
Train Epoch: 4 [51200/60000 (85%)]Loss: 0.091445

Epoch 4 time = 200.68s on device cpu

Test set: Average loss: 0.0455, Accuracy: 9847/10000 (98%)

Train Epoch: 5 [0/60000 (0%)]      Loss: 0.092018
Train Epoch: 5 [51200/60000 (85%)]Loss: 0.071909

Epoch 5 time = 216.68s on device cpu

Test set: Average loss: 0.0428, Accuracy: 9851/10000 (99%)


CPU 512

Train Epoch: 1 [0/60000 (0%)]      Loss: 2.320339
Train Epoch: 1 [25600/60000 (42%)]Loss: 0.271217
Train Epoch: 1 [51200/60000 (85%)]Loss: 0.157687
```

```
Epoch 1 time = 196.21s on device cpu

Test set: Average loss: 0.1045, Accuracy: 9673/10000 (97%)

Train Epoch: 2 [0/60000 (0%)]      Loss: 0.171423
Train Epoch: 2 [25600/60000 (42%)] Loss: 0.100322
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.094170

Epoch 2 time = 175.1s on device cpu

Test set: Average loss: 0.0572, Accuracy: 9821/10000 (98%)

Train Epoch: 3 [0/60000 (0%)]      Loss: 0.111186
Train Epoch: 3 [25600/60000 (42%)] Loss: 0.078768
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.085739

Epoch 3 time = 186.65s on device cpu

Test set: Average loss: 0.0456, Accuracy: 9852/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]      Loss: 0.105068
Train Epoch: 4 [25600/60000 (42%)] Loss: 0.075485
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.062966

Epoch 4 time = 176.78s on device cpu

Test set: Average loss: 0.0389, Accuracy: 9873/10000 (99%)

Train Epoch: 5 [0/60000 (0%)]      Loss: 0.074320
Train Epoch: 5 [25600/60000 (42%)] Loss: 0.073138
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.049382

Epoch 5 time = 178.89s on device cpu

Test set: Average loss: 0.0371, Accuracy: 9874/10000 (99%)

CPU 256

Train Epoch: 1 [0/60000 (0%)]      Loss: 2.318604
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.353729
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.148729
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.166875
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.093921

Epoch 1 time = 184.68s on device cpu

Test set: Average loss: 0.0647, Accuracy: 9780/10000 (98%)

Train Epoch: 2 [0/60000 (0%)]      Loss: 0.118750
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.107628
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.118321
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.088425
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.107882
```

```
Epoch 2 time = 179.36s on device cpu

Test set: Average loss: 0.0464, Accuracy: 9849/10000 (98%)

Train Epoch: 3 [0/60000 (0%)]     Loss: 0.141737
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.083488
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.094399
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.065230
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.098067

Epoch 3 time = 180.68s on device cpu

Test set: Average loss: 0.0338, Accuracy: 9887/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]     Loss: 0.051245
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.042135
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.033028
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.038692
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.083736

Epoch 4 time = 190.11s on device cpu

Test set: Average loss: 0.0351, Accuracy: 9881/10000 (99%)

Train Epoch: 5 [0/60000 (0%)]     Loss: 0.107444
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.046703
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.047314
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.068848
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.054265

Epoch 5 time = 182.42s on device cpu

Test set: Average loss: 0.0296, Accuracy: 9895/10000 (99%)

CPU 128

Train Epoch: 1 [0/60000 (0%)]     Loss: 2.319993
Train Epoch: 1 [6400/60000 (11%)] Loss: 0.380766
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.170312
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.229061
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.199906
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.173657
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.129438
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.074074
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.277130
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.125737

Epoch 1 time = 148.37s on device cpu

Test set: Average loss: 0.0553, Accuracy: 9821/10000 (98%)

Train Epoch: 2 [0/60000 (0%)]     Loss: 0.095319
Train Epoch: 2 [6400/60000 (11%)] Loss: 0.163176
```

```
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.078314
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.080006
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.086878
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.058551
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.082847
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.066542
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.098178
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.057480

Epoch 2 time = 154.09s on device cpu

Test set: Average loss: 0.0374, Accuracy: 9880/10000 (99%)

Train Epoch: 3 [0/60000 (0%)]     Loss: 0.052881
Train Epoch: 3 [6400/60000 (11%)] Loss: 0.041696
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.023823
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.087388
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.060743
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.081374
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.052539
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.040747
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.144274
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.026689

Epoch 3 time = 155.4s on device cpu

Test set: Average loss: 0.0320, Accuracy: 9896/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]     Loss: 0.085446
Train Epoch: 4 [6400/60000 (11%)] Loss: 0.100358
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.020809
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.029298
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.073923
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.039603
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.074789
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.012961
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.154961
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.019539

Epoch 4 time = 155.37s on device cpu

Test set: Average loss: 0.0311, Accuracy: 9898/10000 (99%)

Train Epoch: 5 [0/60000 (0%)]     Loss: 0.064991
Train Epoch: 5 [6400/60000 (11%)] Loss: 0.109652
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.017424
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.034642
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.034348
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.057083
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.040457
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.091035
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.059847
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.032190
```

```
Epoch 5 time = 155.69s on device cpu

Test set: Average loss: 0.0298, Accuracy: 9894/10000 (99%)


CPU 64


Train Epoch: 1 [0/60000 (0%)]      Loss: 2.313801
Train Epoch: 1 [3200/60000 (5%)]   Loss: 0.654571
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.123058
Train Epoch: 1 [9600/60000 (16%)]  Loss: 0.104582
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.320669
Train Epoch: 1 [16000/60000 (27%)] Loss: 0.342554
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.209173
Train Epoch: 1 [22400/60000 (37%)] Loss: 0.107924
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.118724
Train Epoch: 1 [28800/60000 (48%)] Loss: 0.027149
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.110120
Train Epoch: 1 [35200/60000 (59%)] Loss: 0.200524
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.140066
Train Epoch: 1 [41600/60000 (69%)] Loss: 0.133034
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.106219
Train Epoch: 1 [48000/60000 (80%)] Loss: 0.184580
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.152842
Train Epoch: 1 [54400/60000 (91%)] Loss: 0.065063
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.171518

Epoch 1 time = 159.82s on device cpu

Test set: Average loss: 0.0476, Accuracy: 9849/10000 (98%)

Train Epoch: 2 [0/60000 (0%)]      Loss: 0.043874
Train Epoch: 2 [3200/60000 (5%)]   Loss: 0.051196
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.326375
Train Epoch: 2 [9600/60000 (16%)]  Loss: 0.054738
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.068019
Train Epoch: 2 [16000/60000 (27%)] Loss: 0.086013
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.153091
Train Epoch: 2 [22400/60000 (37%)] Loss: 0.025433
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.117877
Train Epoch: 2 [28800/60000 (48%)] Loss: 0.040364
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.088090
Train Epoch: 2 [35200/60000 (59%)] Loss: 0.068708
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.100531
Train Epoch: 2 [41600/60000 (69%)] Loss: 0.109784
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.086060
Train Epoch: 2 [48000/60000 (80%)] Loss: 0.044291
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.036860
Train Epoch: 2 [54400/60000 (91%)] Loss: 0.030410
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.063836

Epoch 2 time = 170.34s on device cpu
```

```
Test set: Average loss: 0.0414, Accuracy: 9861/10000 (99%)

Train Epoch: 3 [0/60000 (0%)]      Loss: 0.031649
Train Epoch: 3 [3200/60000 (5%)]   Loss: 0.020781
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.176763
Train Epoch: 3 [9600/60000 (16%)]  Loss: 0.075667
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.101684
Train Epoch: 3 [16000/60000 (27%)] Loss: 0.042381
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.074942
Train Epoch: 3 [22400/60000 (37%)] Loss: 0.008278
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.014084
Train Epoch: 3 [28800/60000 (48%)] Loss: 0.003656
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.078834
Train Epoch: 3 [35200/60000 (59%)] Loss: 0.185070
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.095069
Train Epoch: 3 [41600/60000 (69%)] Loss: 0.051288
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.030128
Train Epoch: 3 [48000/60000 (80%)] Loss: 0.012149
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.117786
Train Epoch: 3 [54400/60000 (91%)] Loss: 0.025953
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.131974

Epoch 3 time = 172.42s on device cpu

Test set: Average loss: 0.0339, Accuracy: 9895/10000 (99%)

Train Epoch: 4 [0/60000 (0%)]      Loss: 0.017772
Train Epoch: 4 [3200/60000 (5%)]   Loss: 0.030618
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.058960
Train Epoch: 4 [9600/60000 (16%)]  Loss: 0.012779
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.029470
Train Epoch: 4 [16000/60000 (27%)] Loss: 0.040480
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.067411
Train Epoch: 4 [22400/60000 (37%)] Loss: 0.002524
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.012967
Train Epoch: 4 [28800/60000 (48%)] Loss: 0.033713
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.056199
Train Epoch: 4 [35200/60000 (59%)] Loss: 0.125440
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.050590
Train Epoch: 4 [41600/60000 (69%)] Loss: 0.037285
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.004995
Train Epoch: 4 [48000/60000 (80%)] Loss: 0.017124
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.134530
Train Epoch: 4 [54400/60000 (91%)] Loss: 0.017571
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.057694

Epoch 4 time = 171.76s on device cpu

Test set: Average loss: 0.0335, Accuracy: 9890/10000 (99%)

Train Epoch: 5 [0/60000 (0%)]      Loss: 0.027952
Train Epoch: 5 [3200/60000 (5%)]   Loss: 0.028118
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.119115
Train Epoch: 5 [9600/60000 (16%)]  Loss: 0.001108
```

```
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.051511
Train Epoch: 5 [16000/60000 (27%)] Loss: 0.038714
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.080978
Train Epoch: 5 [22400/60000 (37%)] Loss: 0.002904
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.002440
Train Epoch: 5 [28800/60000 (48%)] Loss: 0.033051
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.095800
Train Epoch: 5 [35200/60000 (59%)] Loss: 0.075819
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.040956
Train Epoch: 5 [41600/60000 (69%)] Loss: 0.013228
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.004346
Train Epoch: 5 [48000/60000 (80%)] Loss: 0.028837
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.119950
Train Epoch: 5 [54400/60000 (91%)] Loss: 0.007861
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.021941
```

```
Epoch 5 time = 172.47s on device cpu
```

```
Test set: Average loss: 0.0304, Accuracy: 9898/10000 (99%)
```