

GPU Test

V1

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 2.316613
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.333362
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.224708
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.158794
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.150639
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.142267
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.151424
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.056569
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.157653
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.144163
```

Epoch 1 time = 20.84s on device cuda

Test set: Average loss: 0.0474, Accuracy: 9844/10000 (98%)

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.051858
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.033289
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.096785
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.093866
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.094708
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.021623
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.052466
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.048984
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.125795
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.033234
```

Epoch 2 time = 13.81s on device cuda

Test set: Average loss: 0.0316, Accuracy: 9894/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.058865
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.029533
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.022108
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.045408
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.055890
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.064253
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.054280
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.073061
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.058874
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.077584
```

Epoch 3 time = 13.49s on device cuda

Test set: Average loss: 0.0293, Accuracy: 9898/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.027116
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.027209
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.044561
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.055836
```

```
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.041011
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.023584
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.018291
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.028057
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.087238
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.076356
```

Epoch 4 time = 13.94s on device cuda

Test set: Average loss: 0.0278, Accuracy: 9907/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)] Loss: 0.018789
Train Epoch: 5 [6400/60000 (11%)] Loss: 0.018744
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.065568
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.035880
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.034893
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.030712
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.004890
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.021923
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.027875
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.054550
```

Epoch 5 time = 13.37s on device cuda

Test set: Average loss: 0.0274, Accuracy: 9914/10000 (99%)

V2

```
Train Epoch: 1 [0/60000 (0%)] Loss: 0.036703
Train Epoch: 1 [6400/60000 (11%)] Loss: 0.015126
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.013079
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.053035
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.005119
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.008064
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.032095
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.009062
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.015165
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.014049
```

Epoch 1 time = 13.43s on device cuda

Test set: Average loss: 0.0279, Accuracy: 9907/10000 (99%)

```
Train Epoch: 2 [0/60000 (0%)] Loss: 0.022398
Train Epoch: 2 [6400/60000 (11%)] Loss: 0.019554
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.070275
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.020513
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.011270
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.035425
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.020355
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.043231
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.010588
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.032195
```

Epoch 2 time = 13.44s on device cuda

Test set: Average loss: 0.0273, Accuracy: 9915/10000 (99%)

Train Epoch: 3	[0/60000 (0%)]	Loss: 0.010338
Train Epoch: 3	[6400/60000 (11%)]	Loss: 0.048879
Train Epoch: 3	[12800/60000 (21%)]	Loss: 0.021581
Train Epoch: 3	[19200/60000 (32%)]	Loss: 0.014437
Train Epoch: 3	[25600/60000 (43%)]	Loss: 0.056213
Train Epoch: 3	[32000/60000 (53%)]	Loss: 0.006734
Train Epoch: 3	[38400/60000 (64%)]	Loss: 0.007426
Train Epoch: 3	[44800/60000 (75%)]	Loss: 0.021973
Train Epoch: 3	[51200/60000 (85%)]	Loss: 0.025444
Train Epoch: 3	[57600/60000 (96%)]	Loss: 0.014130

Epoch 3 time = 13.46s on device cuda

Test set: Average loss: 0.0260, Accuracy: 9911/10000 (99%)

Train Epoch: 4	[0/60000 (0%)]	Loss: 0.036777
Train Epoch: 4	[6400/60000 (11%)]	Loss: 0.009756
Train Epoch: 4	[12800/60000 (21%)]	Loss: 0.021407
Train Epoch: 4	[19200/60000 (32%)]	Loss: 0.002505
Train Epoch: 4	[25600/60000 (43%)]	Loss: 0.019788
Train Epoch: 4	[32000/60000 (53%)]	Loss: 0.075811
Train Epoch: 4	[38400/60000 (64%)]	Loss: 0.080558
Train Epoch: 4	[44800/60000 (75%)]	Loss: 0.031123
Train Epoch: 4	[51200/60000 (85%)]	Loss: 0.044682
Train Epoch: 4	[57600/60000 (96%)]	Loss: 0.026821

Epoch 4 time = 13.63s on device cuda

Test set: Average loss: 0.0256, Accuracy: 9915/10000 (99%)

Train Epoch: 5	[0/60000 (0%)]	Loss: 0.012372
Train Epoch: 5	[6400/60000 (11%)]	Loss: 0.007016
Train Epoch: 5	[12800/60000 (21%)]	Loss: 0.018100
Train Epoch: 5	[19200/60000 (32%)]	Loss: 0.053520
Train Epoch: 5	[25600/60000 (43%)]	Loss: 0.020505
Train Epoch: 5	[32000/60000 (53%)]	Loss: 0.034437
Train Epoch: 5	[38400/60000 (64%)]	Loss: 0.011397
Train Epoch: 5	[44800/60000 (75%)]	Loss: 0.031131
Train Epoch: 5	[51200/60000 (85%)]	Loss: 0.032318
Train Epoch: 5	[57600/60000 (96%)]	Loss: 0.035948

Epoch 5 time = 13.66s on device cuda

Test set: Average loss: 0.0259, Accuracy: 9910/10000 (99%)

V3

Train Epoch: 1	[0/60000 (0%)]	Loss: 2.311979
Train Epoch: 1	[6400/60000 (11%)]	Loss: 0.253072
Train Epoch: 1	[12800/60000 (21%)]	Loss: 0.369216

```
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.214112
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.110538
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.094397
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.118044
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.106559
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.070969
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.131041
```

Epoch 1 time = 13.44s on device cuda

Test set: Average loss: 0.0509, Accuracy: 9833/10000 (98%)

```
Train Epoch: 2 [0/60000 (0%)] Loss: 0.112065
Train Epoch: 2 [6400/60000 (11%)] Loss: 0.076725
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.090555
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.027412
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.185812
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.047907
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.009863
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.145355
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.119741
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.048473
```

Epoch 2 time = 13.26s on device cuda

Test set: Average loss: 0.0353, Accuracy: 9878/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)] Loss: 0.046613
Train Epoch: 3 [6400/60000 (11%)] Loss: 0.021891
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.034701
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.056965
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.022195
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.097477
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.054128
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.061216
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.037662
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.117924
```

Epoch 3 time = 13.53s on device cuda

Test set: Average loss: 0.0286, Accuracy: 9903/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)] Loss: 0.068317
Train Epoch: 4 [6400/60000 (11%)] Loss: 0.039860
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.035334
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.021398
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.027840
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.034201
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.031955
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.004088
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.021298
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.083532
```

Epoch 4 time = 13.58s on device cuda

Test set: Average loss: 0.0271, Accuracy: 9912/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.019030
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.008654
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.056513
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.019021
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.025706
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.014591
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.023952
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.015417
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.072404
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.021001
```

Epoch 5 time = 13.47s on device cuda

Test set: Average loss: 0.0270, Accuracy: 9902/10000 (99%)

V4

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 0.011449
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.006748
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.042904
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.006027
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.041874
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.013155
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.006918
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.039102
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.017865
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.035736
```

Epoch 1 time = 13.51s on device cuda

Test set: Average loss: 0.0254, Accuracy: 9915/10000 (99%)

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.016980
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.007949
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.006239
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.021995
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.010840
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.036568
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.012346
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.022144
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.041572
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.070683
```

Epoch 2 time = 13.62s on device cuda

Test set: Average loss: 0.0260, Accuracy: 9917/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.055037
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.009451
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.011362
```

```
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.066030
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.011939
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.103408
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.004856
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.002574
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.023051
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.046737
```

Epoch 3 time = 13.57s on device cuda

Test set: Average loss: 0.0239, Accuracy: 9922/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)] Loss: 0.024442
Train Epoch: 4 [6400/60000 (11%)] Loss: 0.002714
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.001621
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.029414
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.003191
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.031805
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.034052
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.075034
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.011667
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.022785
```

Epoch 4 time = 13.36s on device cuda

Test set: Average loss: 0.0245, Accuracy: 9920/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)] Loss: 0.017187
Train Epoch: 5 [6400/60000 (11%)] Loss: 0.084324
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.044928
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.047978
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.035764
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.007908
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.005404
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.072646
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.002320
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.006000
```

Epoch 5 time = 13.46s on device cuda

Test set: Average loss: 0.0241, Accuracy: 9927/10000 (99%)

V5

```
Train Epoch: 1 [0/60000 (0%)] Loss: 0.011258
Train Epoch: 1 [6400/60000 (11%)] Loss: 0.041738
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.033911
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.037392
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.033880
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.024816
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.006558
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.047292
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.006014
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.014918
```

Epoch 1 time = 13.61s on device cuda

Test set: Average loss: 0.0242, Accuracy: 9925/10000 (99%)

Train Epoch: 2	[0/60000 (0%)]	Loss: 0.004572
Train Epoch: 2	[6400/60000 (11%)]	Loss: 0.009280
Train Epoch: 2	[12800/60000 (21%)]	Loss: 0.024960
Train Epoch: 2	[19200/60000 (32%)]	Loss: 0.009354
Train Epoch: 2	[25600/60000 (43%)]	Loss: 0.017309
Train Epoch: 2	[32000/60000 (53%)]	Loss: 0.001744
Train Epoch: 2	[38400/60000 (64%)]	Loss: 0.012341
Train Epoch: 2	[44800/60000 (75%)]	Loss: 0.009667
Train Epoch: 2	[51200/60000 (85%)]	Loss: 0.016139
Train Epoch: 2	[57600/60000 (96%)]	Loss: 0.013743

Epoch 2 time = 13.31s on device cuda

Test set: Average loss: 0.0243, Accuracy: 9925/10000 (99%)

Train Epoch: 3	[0/60000 (0%)]	Loss: 0.006973
Train Epoch: 3	[6400/60000 (11%)]	Loss: 0.076386
Train Epoch: 3	[12800/60000 (21%)]	Loss: 0.088577
Train Epoch: 3	[19200/60000 (32%)]	Loss: 0.022900
Train Epoch: 3	[25600/60000 (43%)]	Loss: 0.014779
Train Epoch: 3	[32000/60000 (53%)]	Loss: 0.007413
Train Epoch: 3	[38400/60000 (64%)]	Loss: 0.002596
Train Epoch: 3	[44800/60000 (75%)]	Loss: 0.027258
Train Epoch: 3	[51200/60000 (85%)]	Loss: 0.008347
Train Epoch: 3	[57600/60000 (96%)]	Loss: 0.003403

Epoch 3 time = 13.77s on device cuda

Test set: Average loss: 0.0241, Accuracy: 9925/10000 (99%)

Train Epoch: 4	[0/60000 (0%)]	Loss: 0.004830
Train Epoch: 4	[6400/60000 (11%)]	Loss: 0.027301
Train Epoch: 4	[12800/60000 (21%)]	Loss: 0.003255
Train Epoch: 4	[19200/60000 (32%)]	Loss: 0.009387
Train Epoch: 4	[25600/60000 (43%)]	Loss: 0.038065
Train Epoch: 4	[32000/60000 (53%)]	Loss: 0.029627
Train Epoch: 4	[38400/60000 (64%)]	Loss: 0.016042
Train Epoch: 4	[44800/60000 (75%)]	Loss: 0.003211
Train Epoch: 4	[51200/60000 (85%)]	Loss: 0.028639
Train Epoch: 4	[57600/60000 (96%)]	Loss: 0.041412

Epoch 4 time = 13.32s on device cuda

Test set: Average loss: 0.0240, Accuracy: 9923/10000 (99%)

Train Epoch: 5	[0/60000 (0%)]	Loss: 0.004751
Train Epoch: 5	[6400/60000 (11%)]	Loss: 0.013488
Train Epoch: 5	[12800/60000 (21%)]	Loss: 0.008655
Train Epoch: 5	[19200/60000 (32%)]	Loss: 0.020333

```
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.022909
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.001843
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.048590
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.006035
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.039031
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.042196
```

Epoch 5 time = 13.65s on device cuda

Test set: Average loss: 0.0241, Accuracy: 9923/10000 (99%)

CPU

V1

```
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 = 152.66s 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 = 161.34s 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 = 170.43s 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 = 174.63s 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 = 177.92s on device cpu

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

V2

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 0.050428
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.049077
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.033519
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.056136
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.055025
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.027734
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.042237
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.042441
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.068088
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.034754
```

Epoch 1 time = 183.66s on device cpu

Test set: Average loss: 0.0301, Accuracy: 9899/10000 (99%)

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.027648
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.025670
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.019075
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.031738
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.037475
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.038581
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.030122
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.049959
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.095365
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.009284
```

Epoch 2 time = 181.61s on device cpu

Test set: Average loss: 0.0278, Accuracy: 9912/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.031818
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.038197
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.010466
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.032579
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.076557
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.037130
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.041255
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.016403
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.055764
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.022111
```

Epoch 3 time = 177.66s on device cpu

Test set: Average loss: 0.0289, Accuracy: 9910/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.052673
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.054488
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.013206
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.041190
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.060060
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.038590
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.021737
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.018686
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.040700
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.005411
```

Epoch 4 time = 178.33s on device cpu

Test set: Average loss: 0.0292, Accuracy: 9902/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.091964
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.066065
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.012387
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.073285
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.053346
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.038922
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.011725
```

```
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.013165
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.020756
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.008124
```

Epoch 5 time = 180.16s on device cpu

Test set: Average loss: 0.0277, Accuracy: 9908/10000 (99%)

V3

```
Train Epoch: 1 [0/60000 (0%)] Loss: 0.035758
Train Epoch: 1 [6400/60000 (11%)] Loss: 0.034468
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.022844
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.019714
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.049297
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.031953
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.030273
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.042566
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.063037
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.011154
```

Epoch 1 time = 174.93s on device cpu

Test set: Average loss: 0.0279, Accuracy: 9907/10000 (99%)

```
Train Epoch: 2 [0/60000 (0%)] Loss: 0.067321
Train Epoch: 2 [6400/60000 (11%)] Loss: 0.085454
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.024263
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.014760
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.036748
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.011815
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.022825
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.017534
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.020539
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.021808
```

Epoch 2 time = 177.25s on device cpu

Test set: Average loss: 0.0272, Accuracy: 9908/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)] Loss: 0.036768
Train Epoch: 3 [6400/60000 (11%)] Loss: 0.060122
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.010619
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.013424
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.026557
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.020396
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.011407
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.018184
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.056976
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.018469
```

Epoch 3 time = 176.25s on device cpu

Test set: Average loss: 0.0269, Accuracy: 9913/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.030613
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.101814
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.013564
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.013157
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.041950
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.018096
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.046966
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.045143
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.064675
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.032008
```

Epoch 4 time = 174.75s on device cpu

Test set: Average loss: 0.0268, Accuracy: 9911/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.043056
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.022366
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.025556
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.029415
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.036342
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.017662
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.014544
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.015582
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.076069
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.017579
```

Epoch 5 time = 175.66s on device cpu

Test set: Average loss: 0.0268, Accuracy: 9913/10000 (99%)

V4

```
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 = 168.44s 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 = 171.48s 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 = 171.89s 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 = 180.9s 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 = 174.44s on device cpu

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

V5

```
Train Epoch: 1 [0/60000 (0%)]      Loss: 0.050428
Train Epoch: 1 [6400/60000 (11%)]  Loss: 0.049077
Train Epoch: 1 [12800/60000 (21%)] Loss: 0.033519
Train Epoch: 1 [19200/60000 (32%)] Loss: 0.056136
Train Epoch: 1 [25600/60000 (43%)] Loss: 0.055025
Train Epoch: 1 [32000/60000 (53%)] Loss: 0.027734
Train Epoch: 1 [38400/60000 (64%)] Loss: 0.042237
Train Epoch: 1 [44800/60000 (75%)] Loss: 0.042441
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.068088
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.034754
```

Epoch 1 time = 171.37s on device cpu

Test set: Average loss: 0.0301, Accuracy: 9899/10000 (99%)

```
Train Epoch: 2 [0/60000 (0%)]      Loss: 0.027648
Train Epoch: 2 [6400/60000 (11%)]  Loss: 0.025670
Train Epoch: 2 [12800/60000 (21%)] Loss: 0.019075
Train Epoch: 2 [19200/60000 (32%)] Loss: 0.031738
Train Epoch: 2 [25600/60000 (43%)] Loss: 0.037475
Train Epoch: 2 [32000/60000 (53%)] Loss: 0.038581
Train Epoch: 2 [38400/60000 (64%)] Loss: 0.030122
Train Epoch: 2 [44800/60000 (75%)] Loss: 0.049959
Train Epoch: 2 [51200/60000 (85%)] Loss: 0.095365
Train Epoch: 2 [57600/60000 (96%)] Loss: 0.009284
```

Epoch 2 time = 172.89s on device cpu

Test set: Average loss: 0.0278, Accuracy: 9912/10000 (99%)

```
Train Epoch: 3 [0/60000 (0%)]      Loss: 0.031818
Train Epoch: 3 [6400/60000 (11%)]  Loss: 0.038197
Train Epoch: 3 [12800/60000 (21%)] Loss: 0.010466
Train Epoch: 3 [19200/60000 (32%)] Loss: 0.032579
Train Epoch: 3 [25600/60000 (43%)] Loss: 0.076557
Train Epoch: 3 [32000/60000 (53%)] Loss: 0.037130
Train Epoch: 3 [38400/60000 (64%)] Loss: 0.041255
Train Epoch: 3 [44800/60000 (75%)] Loss: 0.016403
Train Epoch: 3 [51200/60000 (85%)] Loss: 0.055764
Train Epoch: 3 [57600/60000 (96%)] Loss: 0.022111
```

Epoch 3 time = 172.69s on device cpu

Test set: Average loss: 0.0289, Accuracy: 9910/10000 (99%)

```
Train Epoch: 4 [0/60000 (0%)]      Loss: 0.052673
Train Epoch: 4 [6400/60000 (11%)]  Loss: 0.054488
Train Epoch: 4 [12800/60000 (21%)] Loss: 0.013206
Train Epoch: 4 [19200/60000 (32%)] Loss: 0.041190
Train Epoch: 4 [25600/60000 (43%)] Loss: 0.060060
Train Epoch: 4 [32000/60000 (53%)] Loss: 0.038590
Train Epoch: 4 [38400/60000 (64%)] Loss: 0.021737
Train Epoch: 4 [44800/60000 (75%)] Loss: 0.018686
Train Epoch: 4 [51200/60000 (85%)] Loss: 0.040700
Train Epoch: 4 [57600/60000 (96%)] Loss: 0.005411
```

Epoch 4 time = 171.01s on device cpu

Test set: Average loss: 0.0292, Accuracy: 9902/10000 (99%)

```
Train Epoch: 5 [0/60000 (0%)]      Loss: 0.091964
Train Epoch: 5 [6400/60000 (11%)]  Loss: 0.066065
Train Epoch: 5 [12800/60000 (21%)] Loss: 0.012387
Train Epoch: 5 [19200/60000 (32%)] Loss: 0.073285
Train Epoch: 5 [25600/60000 (43%)] Loss: 0.053346
Train Epoch: 5 [32000/60000 (53%)] Loss: 0.038922
Train Epoch: 5 [38400/60000 (64%)] Loss: 0.011725
Train Epoch: 5 [44800/60000 (75%)] Loss: 0.013165
Train Epoch: 5 [51200/60000 (85%)] Loss: 0.020756
Train Epoch: 5 [57600/60000 (96%)] Loss: 0.008124
```

Epoch 5 time = 172.49s on device cpu

Test set: Average loss: 0.0277, Accuracy: 9908/10000 (99%)