**Output of single layer on MNIST:**

Train Epoch: 1 [0/60000 (0%)] Loss: 7.089541

Train Epoch: 1 [5000/60000 (8%)] Loss: 0.685971

Train Epoch: 1 [10000/60000 (17%)] Loss: 0.655903

Train Epoch: 1 [15000/60000 (25%)] Loss: 0.474321

Train Epoch: 1 [20000/60000 (33%)] Loss: 0.330206

Train Epoch: 1 [25000/60000 (42%)] Loss: 0.320224

Train Epoch: 1 [30000/60000 (50%)] Loss: 0.482131

Train Epoch: 1 [35000/60000 (58%)] Loss: 0.361452

Train Epoch: 1 [40000/60000 (67%)] Loss: 0.243129

Train Epoch: 1 [45000/60000 (75%)] Loss: 0.271554

Train Epoch: 1 [50000/60000 (83%)] Loss: 0.366397

Train Epoch: 1 [55000/60000 (92%)] Loss: 0.498909

Test set: Average loss: 0.3285, Accuracy: 9076/10000 (90%)

Train Epoch: 2 [0/60000 (0%)] Loss: 0.549675

Train Epoch: 2 [5000/60000 (8%)] Loss: 0.213172

Train Epoch: 2 [10000/60000 (17%)] Loss: 0.203139

Train Epoch: 2 [15000/60000 (25%)] Loss: 0.582523

Train Epoch: 2 [20000/60000 (33%)] Loss: 0.356306

Train Epoch: 2 [25000/60000 (42%)] Loss: 0.320346

Train Epoch: 2 [30000/60000 (50%)] Loss: 0.269673

Train Epoch: 2 [35000/60000 (58%)] Loss: 0.391097

Train Epoch: 2 [40000/60000 (67%)] Loss: 0.303784

Train Epoch: 2 [45000/60000 (75%)] Loss: 0.219469

Train Epoch: 2 [50000/60000 (83%)] Loss: 0.271125

Train Epoch: 2 [55000/60000 (92%)] Loss: 0.206784

Test set: Average loss: 0.3023, Accuracy: 9151/10000 (91%)

Train Epoch: 3 [0/60000 (0%)] Loss: 0.290725

Train Epoch: 3 [5000/60000 (8%)] Loss: 0.443386

Train Epoch: 3 [10000/60000 (17%)] Loss: 0.254507

Train Epoch: 3 [15000/60000 (25%)] Loss: 0.409476

Train Epoch: 3 [20000/60000 (33%)] Loss: 0.250239

Train Epoch: 3 [25000/60000 (42%)] Loss: 0.195785

Train Epoch: 3 [30000/60000 (50%)] Loss: 0.129806

Train Epoch: 3 [35000/60000 (58%)] Loss: 0.448743

Train Epoch: 3 [40000/60000 (67%)] Loss: 0.107554

Train Epoch: 3 [45000/60000 (75%)] Loss: 0.186254

Train Epoch: 3 [50000/60000 (83%)] Loss: 0.197138

Train Epoch: 3 [55000/60000 (92%)] Loss: 0.284312

Test set: Average loss: 0.2895, Accuracy: 9199/10000 (91%)

Train Epoch: 4 [0/60000 (0%)] Loss: 0.176699

Train Epoch: 4 [5000/60000 (8%)] Loss: 0.217077

Train Epoch: 4 [10000/60000 (17%)] Loss: 0.311399

Train Epoch: 4 [15000/60000 (25%)] Loss: 0.072951

Train Epoch: 4 [20000/60000 (33%)] Loss: 0.113952

Train Epoch: 4 [25000/60000 (42%)] Loss: 0.207557

Train Epoch: 4 [30000/60000 (50%)] Loss: 0.316108

Train Epoch: 4 [35000/60000 (58%)] Loss: 0.240777

Train Epoch: 4 [40000/60000 (67%)] Loss: 0.250785

Train Epoch: 4 [45000/60000 (75%)] Loss: 0.137382

Train Epoch: 4 [50000/60000 (83%)] Loss: 0.242569

Train Epoch: 4 [55000/60000 (92%)] Loss: 0.316030

Test set: Average loss: 0.2855, Accuracy: 9208/10000 (92%)

Train Epoch: 5 [0/60000 (0%)] Loss: 0.296732

Train Epoch: 5 [5000/60000 (8%)] Loss: 0.207898

Train Epoch: 5 [10000/60000 (17%)] Loss: 0.143977

Train Epoch: 5 [15000/60000 (25%)] Loss: 0.174512

Train Epoch: 5 [20000/60000 (33%)] Loss: 0.407390

Train Epoch: 5 [25000/60000 (42%)] Loss: 0.246729

Train Epoch: 5 [30000/60000 (50%)] Loss: 0.252243

Train Epoch: 5 [35000/60000 (58%)] Loss: 0.269473

Train Epoch: 5 [40000/60000 (67%)] Loss: 0.241514

Train Epoch: 5 [45000/60000 (75%)] Loss: 0.322532

Train Epoch: 5 [50000/60000 (83%)] Loss: 0.154211

Train Epoch: 5 [55000/60000 (92%)] Loss: 0.287052

Test set: Average loss: 0.2831, Accuracy: 9227/10000 (92%)

Train Epoch: 6 [0/60000 (0%)] Loss: 0.306102

Train Epoch: 6 [5000/60000 (8%)] Loss: 0.172272

Train Epoch: 6 [10000/60000 (17%)] Loss: 0.366518

Train Epoch: 6 [15000/60000 (25%)] Loss: 0.293182

Train Epoch: 6 [20000/60000 (33%)] Loss: 0.177749

Train Epoch: 6 [25000/60000 (42%)] Loss: 0.609034

Train Epoch: 6 [30000/60000 (50%)] Loss: 0.251356

Train Epoch: 6 [35000/60000 (58%)] Loss: 0.191507

Train Epoch: 6 [40000/60000 (67%)] Loss: 0.202745

Train Epoch: 6 [45000/60000 (75%)] Loss: 0.175938

Train Epoch: 6 [50000/60000 (83%)] Loss: 0.670988

Train Epoch: 6 [55000/60000 (92%)] Loss: 0.197317

Test set: Average loss: 0.2817, Accuracy: 9214/10000 (92%)

Train Epoch: 7 [0/60000 (0%)] Loss: 0.208793

Train Epoch: 7 [5000/60000 (8%)] Loss: 0.397933

Train Epoch: 7 [10000/60000 (17%)] Loss: 0.293195

Train Epoch: 7 [15000/60000 (25%)] Loss: 0.402731

Train Epoch: 7 [20000/60000 (33%)] Loss: 0.527438

Train Epoch: 7 [25000/60000 (42%)] Loss: 0.275572

Train Epoch: 7 [30000/60000 (50%)] Loss: 0.304021

Train Epoch: 7 [35000/60000 (58%)] Loss: 0.270642

Train Epoch: 7 [40000/60000 (67%)] Loss: 0.450811

Train Epoch: 7 [45000/60000 (75%)] Loss: 0.290301

Train Epoch: 7 [50000/60000 (83%)] Loss: 0.531794

Train Epoch: 7 [55000/60000 (92%)] Loss: 0.298194

Test set: Average loss: 0.2791, Accuracy: 9218/10000 (92%)

Train Epoch: 8 [0/60000 (0%)] Loss: 0.361164

Train Epoch: 8 [5000/60000 (8%)] Loss: 0.289070

Train Epoch: 8 [10000/60000 (17%)] Loss: 0.106304

Train Epoch: 8 [15000/60000 (25%)] Loss: 0.160262

Train Epoch: 8 [20000/60000 (33%)] Loss: 0.382239

Train Epoch: 8 [25000/60000 (42%)] Loss: 0.338557

Train Epoch: 8 [30000/60000 (50%)] Loss: 0.521079

Train Epoch: 8 [35000/60000 (58%)] Loss: 0.614125

Train Epoch: 8 [40000/60000 (67%)] Loss: 0.049240

Train Epoch: 8 [45000/60000 (75%)] Loss: 0.273943

Train Epoch: 8 [50000/60000 (83%)] Loss: 0.431000

Train Epoch: 8 [55000/60000 (92%)] Loss: 0.115235

Test set: Average loss: 0.2789, Accuracy: 9218/10000 (92%)

Train Epoch: 9 [0/60000 (0%)] Loss: 0.179356

Train Epoch: 9 [5000/60000 (8%)] Loss: 0.378331

Train Epoch: 9 [10000/60000 (17%)] Loss: 0.207762

Train Epoch: 9 [15000/60000 (25%)] Loss: 0.347898

Train Epoch: 9 [20000/60000 (33%)] Loss: 0.438586

Train Epoch: 9 [25000/60000 (42%)] Loss: 0.593306

Train Epoch: 9 [30000/60000 (50%)] Loss: 0.212837

Train Epoch: 9 [35000/60000 (58%)] Loss: 0.288125

Train Epoch: 9 [40000/60000 (67%)] Loss: 0.449102

Train Epoch: 9 [45000/60000 (75%)] Loss: 0.231296

Train Epoch: 9 [50000/60000 (83%)] Loss: 0.325284

Train Epoch: 9 [55000/60000 (92%)] Loss: 0.183072

Test set: Average loss: 0.2741, Accuracy: 9226/10000 (92%)

Train Epoch: 10 [0/60000 (0%)] Loss: 0.305058

Train Epoch: 10 [5000/60000 (8%)] Loss: 0.228628

Train Epoch: 10 [10000/60000 (17%)] Loss: 0.220839

Train Epoch: 10 [15000/60000 (25%)] Loss: 0.313329

Train Epoch: 10 [20000/60000 (33%)] Loss: 0.292246

Train Epoch: 10 [25000/60000 (42%)] Loss: 0.196667

Train Epoch: 10 [30000/60000 (50%)] Loss: 0.353260

Train Epoch: 10 [35000/60000 (58%)] Loss: 0.081142

Train Epoch: 10 [40000/60000 (67%)] Loss: 0.225102

Train Epoch: 10 [45000/60000 (75%)] Loss: 0.217183

Train Epoch: 10 [50000/60000 (83%)] Loss: 0.305108

Train Epoch: 10 [55000/60000 (92%)] Loss: 0.414368

Test set: Average loss: 0.2738, Accuracy: 9230/10000 (92%)

This is our first output of the model learning on the MNIST dataset. As shown above, we can see that the accuracy rate stagnates at 92% after Epoch 4.

**Screenshot of weight images:**

**Output of single layer on MNIST with 50 samples**

Train Epoch: 1 [0/60000 (0%)] Loss: 2.388052

Test set: Average loss: 0.7789, Accuracy: 8133/10000 (81%)

Train Epoch: 2 [0/60000 (0%)] Loss: 0.861123

Test set: Average loss: 0.5780, Accuracy: 8577/10000 (85%)

Train Epoch: 3 [0/60000 (0%)] Loss: 0.625074

Test set: Average loss: 0.5054, Accuracy: 8688/10000 (86%)

Train Epoch: 4 [0/60000 (0%)] Loss: 0.512596

Test set: Average loss: 0.4641, Accuracy: 8787/10000 (87%)

Train Epoch: 5 [0/60000 (0%)] Loss: 0.429308

Test set: Average loss: 0.4380, Accuracy: 8819/10000 (88%)

Train Epoch: 6 [0/60000 (0%)] Loss: 0.365044

Test set: Average loss: 0.4168, Accuracy: 8881/10000 (88%)

Train Epoch: 7 [0/60000 (0%)] Loss: 0.551342

Test set: Average loss: 0.4029, Accuracy: 8912/10000 (89%)

Train Epoch: 8 [0/60000 (0%)] Loss: 0.303316

Test set: Average loss: 0.3910, Accuracy: 8936/10000 (89%)

Train Epoch: 9 [0/60000 (0%)] Loss: 0.548941

Test set: Average loss: 0.3827, Accuracy: 8932/10000 (89%)

Train Epoch: 10 [0/60000 (0%)] Loss: 0.347897

Test set: Average loss: 0.3716, Accuracy: 8949/10000 (89%)

With using only 50 samples as our training set, we see that the accuracy drops off initially. The accuracy gradually increases until it plateaus at the 89% range. Instead of stopping at Epoch 4 like in our previous example, this model continues to learn past Epoch 7. With not enough data, the model is incapable of further learning and adjusting the weights.

**Output of Multi-Layer on MNIST:**

Train Epoch: 1 [0/60000 (0%)] Loss: 2.277908

Train Epoch: 1 [5000/60000 (8%)] Loss: 0.578127

Train Epoch: 1 [10000/60000 (17%)] Loss: 0.630512

Train Epoch: 1 [15000/60000 (25%)] Loss: 0.439989

Train Epoch: 1 [20000/60000 (33%)] Loss: 0.340426

Train Epoch: 1 [25000/60000 (42%)] Loss: 0.588281

Train Epoch: 1 [30000/60000 (50%)] Loss: 0.420061

Train Epoch: 1 [35000/60000 (58%)] Loss: 0.544314

Train Epoch: 1 [40000/60000 (67%)] Loss: 0.268867

Train Epoch: 1 [45000/60000 (75%)] Loss: 0.303285

Train Epoch: 1 [50000/60000 (83%)] Loss: 0.369049

Train Epoch: 1 [55000/60000 (92%)] Loss: 0.445237

Test set: Average loss: 0.3060, Accuracy: 9141/10000 (91%)

Train Epoch: 2 [0/60000 (0%)] Loss: 0.269618

Train Epoch: 2 [5000/60000 (8%)] Loss: 0.316748

Train Epoch: 2 [10000/60000 (17%)] Loss: 0.232871

Train Epoch: 2 [15000/60000 (25%)] Loss: 0.245842

Train Epoch: 2 [20000/60000 (33%)] Loss: 0.202073

Train Epoch: 2 [25000/60000 (42%)] Loss: 0.268720

Train Epoch: 2 [30000/60000 (50%)] Loss: 0.333183

Train Epoch: 2 [35000/60000 (58%)] Loss: 0.103701

Train Epoch: 2 [40000/60000 (67%)] Loss: 0.548746

Train Epoch: 2 [45000/60000 (75%)] Loss: 0.202646

Train Epoch: 2 [50000/60000 (83%)] Loss: 0.408454

Train Epoch: 2 [55000/60000 (92%)] Loss: 0.284051

Test set: Average loss: 0.2675, Accuracy: 9236/10000 (92%)

Train Epoch: 3 [0/60000 (0%)] Loss: 0.140921

Train Epoch: 3 [5000/60000 (8%)] Loss: 0.285437

Train Epoch: 3 [10000/60000 (17%)] Loss: 0.187478

Train Epoch: 3 [15000/60000 (25%)] Loss: 0.334512

Train Epoch: 3 [20000/60000 (33%)] Loss: 0.151751

Train Epoch: 3 [25000/60000 (42%)] Loss: 0.134982

Train Epoch: 3 [30000/60000 (50%)] Loss: 0.293156

Train Epoch: 3 [35000/60000 (58%)] Loss: 0.231920

Train Epoch: 3 [40000/60000 (67%)] Loss: 0.202671

Train Epoch: 3 [45000/60000 (75%)] Loss: 0.230753

Train Epoch: 3 [50000/60000 (83%)] Loss: 0.205235

Train Epoch: 3 [55000/60000 (92%)] Loss: 0.210228

Test set: Average loss: 0.2443, Accuracy: 9304/10000 (93%)

Train Epoch: 4 [0/60000 (0%)] Loss: 0.206530

Train Epoch: 4 [5000/60000 (8%)] Loss: 0.287849

Train Epoch: 4 [10000/60000 (17%)] Loss: 0.137138

Train Epoch: 4 [15000/60000 (25%)] Loss: 0.200407

Train Epoch: 4 [20000/60000 (33%)] Loss: 0.161844

Train Epoch: 4 [25000/60000 (42%)] Loss: 0.223126

Train Epoch: 4 [30000/60000 (50%)] Loss: 0.211633

Train Epoch: 4 [35000/60000 (58%)] Loss: 0.309136

Train Epoch: 4 [40000/60000 (67%)] Loss: 0.306016

Train Epoch: 4 [45000/60000 (75%)] Loss: 0.396887

Train Epoch: 4 [50000/60000 (83%)] Loss: 0.148905

Train Epoch: 4 [55000/60000 (92%)] Loss: 0.211834

Test set: Average loss: 0.2249, Accuracy: 9346/10000 (93%)

Train Epoch: 5 [0/60000 (0%)] Loss: 0.164820

Train Epoch: 5 [5000/60000 (8%)] Loss: 0.093660

Train Epoch: 5 [10000/60000 (17%)] Loss: 0.245991

Train Epoch: 5 [15000/60000 (25%)] Loss: 0.182094

Train Epoch: 5 [20000/60000 (33%)] Loss: 0.227155

Train Epoch: 5 [25000/60000 (42%)] Loss: 0.360857

Train Epoch: 5 [30000/60000 (50%)] Loss: 0.219334

Train Epoch: 5 [35000/60000 (58%)] Loss: 0.188130

Train Epoch: 5 [40000/60000 (67%)] Loss: 0.325545

Train Epoch: 5 [45000/60000 (75%)] Loss: 0.139516

Train Epoch: 5 [50000/60000 (83%)] Loss: 0.529342

Train Epoch: 5 [55000/60000 (92%)] Loss: 0.227285

Test set: Average loss: 0.2041, Accuracy: 9420/10000 (94%)

Train Epoch: 6 [0/60000 (0%)] Loss: 0.211728

Train Epoch: 6 [5000/60000 (8%)] Loss: 0.198161

Train Epoch: 6 [10000/60000 (17%)] Loss: 0.433569

Train Epoch: 6 [15000/60000 (25%)] Loss: 0.108571

Train Epoch: 6 [20000/60000 (33%)] Loss: 0.070961

Train Epoch: 6 [25000/60000 (42%)] Loss: 0.175267

Train Epoch: 6 [30000/60000 (50%)] Loss: 0.150064

Train Epoch: 6 [35000/60000 (58%)] Loss: 0.216258

Train Epoch: 6 [40000/60000 (67%)] Loss: 0.161408

Train Epoch: 6 [45000/60000 (75%)] Loss: 0.265073

Train Epoch: 6 [50000/60000 (83%)] Loss: 0.111513

Train Epoch: 6 [55000/60000 (92%)] Loss: 0.288542

Test set: Average loss: 0.1883, Accuracy: 9467/10000 (94%)

Train Epoch: 7 [0/60000 (0%)] Loss: 0.094424

Train Epoch: 7 [5000/60000 (8%)] Loss: 0.158293

Train Epoch: 7 [10000/60000 (17%)] Loss: 0.269300

Train Epoch: 7 [15000/60000 (25%)] Loss: 0.212192

Train Epoch: 7 [20000/60000 (33%)] Loss: 0.109012

Train Epoch: 7 [25000/60000 (42%)] Loss: 0.078125

Train Epoch: 7 [30000/60000 (50%)] Loss: 0.184853

Train Epoch: 7 [35000/60000 (58%)] Loss: 0.091131

Train Epoch: 7 [40000/60000 (67%)] Loss: 0.081705

Train Epoch: 7 [45000/60000 (75%)] Loss: 0.252285

Train Epoch: 7 [50000/60000 (83%)] Loss: 0.164001

Train Epoch: 7 [55000/60000 (92%)] Loss: 0.174203

Test set: Average loss: 0.1747, Accuracy: 9498/10000 (94%)

Train Epoch: 8 [0/60000 (0%)] Loss: 0.232581

Train Epoch: 8 [5000/60000 (8%)] Loss: 0.365162

Train Epoch: 8 [10000/60000 (17%)] Loss: 0.108417

Train Epoch: 8 [15000/60000 (25%)] Loss: 0.168831

Train Epoch: 8 [20000/60000 (33%)] Loss: 0.053957

Train Epoch: 8 [25000/60000 (42%)] Loss: 0.108650

Train Epoch: 8 [30000/60000 (50%)] Loss: 0.161824

Train Epoch: 8 [35000/60000 (58%)] Loss: 0.342832

Train Epoch: 8 [40000/60000 (67%)] Loss: 0.224760

Train Epoch: 8 [45000/60000 (75%)] Loss: 0.218597

Train Epoch: 8 [50000/60000 (83%)] Loss: 0.066412

Train Epoch: 8 [55000/60000 (92%)] Loss: 0.116456

Test set: Average loss: 0.1593, Accuracy: 9542/10000 (95%)

Train Epoch: 9 [0/60000 (0%)] Loss: 0.175123

Train Epoch: 9 [5000/60000 (8%)] Loss: 0.116958

Train Epoch: 9 [10000/60000 (17%)] Loss: 0.102860

Train Epoch: 9 [15000/60000 (25%)] Loss: 0.131863

Train Epoch: 9 [20000/60000 (33%)] Loss: 0.325408

Train Epoch: 9 [25000/60000 (42%)] Loss: 0.066464

Train Epoch: 9 [30000/60000 (50%)] Loss: 0.169000

Train Epoch: 9 [35000/60000 (58%)] Loss: 0.174459

Train Epoch: 9 [40000/60000 (67%)] Loss: 0.156438

Train Epoch: 9 [45000/60000 (75%)] Loss: 0.042584

Train Epoch: 9 [50000/60000 (83%)] Loss: 0.138677

Train Epoch: 9 [55000/60000 (92%)] Loss: 0.214506

Test set: Average loss: 0.1477, Accuracy: 9583/10000 (95%)

Train Epoch: 10 [0/60000 (0%)] Loss: 0.214720

Train Epoch: 10 [5000/60000 (8%)] Loss: 0.156507

Train Epoch: 10 [10000/60000 (17%)] Loss: 0.063631

Train Epoch: 10 [15000/60000 (25%)] Loss: 0.159036

Train Epoch: 10 [20000/60000 (33%)] Loss: 0.090324

Train Epoch: 10 [25000/60000 (42%)] Loss: 0.070693

Train Epoch: 10 [30000/60000 (50%)] Loss: 0.061616

Train Epoch: 10 [35000/60000 (58%)] Loss: 0.082954

Train Epoch: 10 [40000/60000 (67%)] Loss: 0.239612

Train Epoch: 10 [45000/60000 (75%)] Loss: 0.082911

Train Epoch: 10 [50000/60000 (83%)] Loss: 0.136497

Train Epoch: 10 [55000/60000 (92%)] Loss: 0.147497

Test set: Average loss: 0.1394, Accuracy: 9609/10000 (96%)

Here we see that adding more hidden layers and an activation function (tanh) improved our results from 92% up to 96%. There is still accuracy increases between epochs, even at near the end where we see increases of ~1%.

**Output of Multi-Layer on MNIST with Learning rate = 10:**

Train Epoch: 1 [0/60000 (0%)] Loss: 2.381393

Train Epoch: 1 [5000/60000 (8%)] Loss: 1012.301880

Train Epoch: 1 [10000/60000 (17%)] Loss: 663.798584

Train Epoch: 1 [15000/60000 (25%)] Loss: 946.158386

Train Epoch: 1 [20000/60000 (33%)] Loss: 693.433533

Train Epoch: 1 [25000/60000 (42%)] Loss: 332.469421

Train Epoch: 1 [30000/60000 (50%)] Loss: 305.385284

Train Epoch: 1 [35000/60000 (58%)] Loss: 235.301361

Train Epoch: 1 [40000/60000 (67%)] Loss: 441.192535

Train Epoch: 1 [45000/60000 (75%)] Loss: 334.679565

Train Epoch: 1 [50000/60000 (83%)] Loss: 405.987305

Train Epoch: 1 [55000/60000 (92%)] Loss: 115.895302

Test set: Average loss: 187.7630, Accuracy: 6453/10000 (64%)

Train Epoch: 2 [0/60000 (0%)] Loss: 281.079071

Train Epoch: 2 [5000/60000 (8%)] Loss: 262.804657

Train Epoch: 2 [10000/60000 (17%)] Loss: 405.294891

Train Epoch: 2 [15000/60000 (25%)] Loss: 360.503906

Train Epoch: 2 [20000/60000 (33%)] Loss: 275.434021

Train Epoch: 2 [25000/60000 (42%)] Loss: 319.743530

Train Epoch: 2 [30000/60000 (50%)] Loss: 204.498688

Train Epoch: 2 [35000/60000 (58%)] Loss: 527.919128

Train Epoch: 2 [40000/60000 (67%)] Loss: 333.456940

Train Epoch: 2 [45000/60000 (75%)] Loss: 196.114380

Train Epoch: 2 [50000/60000 (83%)] Loss: 459.308441

Train Epoch: 2 [55000/60000 (92%)] Loss: 434.376099

Test set: Average loss: 160.0932, Accuracy: 7044/10000 (70%)

Train Epoch: 3 [0/60000 (0%)] Loss: 201.168533

Train Epoch: 3 [5000/60000 (8%)] Loss: 288.245026

Train Epoch: 3 [10000/60000 (17%)] Loss: 205.021408

Train Epoch: 3 [15000/60000 (25%)] Loss: 250.438553

Train Epoch: 3 [20000/60000 (33%)] Loss: 154.271957

Train Epoch: 3 [25000/60000 (42%)] Loss: 288.764191

Train Epoch: 3 [30000/60000 (50%)] Loss: 255.725891

Train Epoch: 3 [35000/60000 (58%)] Loss: 278.574524

Train Epoch: 3 [40000/60000 (67%)] Loss: 676.766235

Train Epoch: 3 [45000/60000 (75%)] Loss: 118.229340

Train Epoch: 3 [50000/60000 (83%)] Loss: 274.606812

Train Epoch: 3 [55000/60000 (92%)] Loss: 320.765320

Test set: Average loss: 205.0256, Accuracy: 6720/10000 (67%)

Train Epoch: 4 [0/60000 (0%)] Loss: 220.973434

Train Epoch: 4 [5000/60000 (8%)] Loss: 128.958603

Train Epoch: 4 [10000/60000 (17%)] Loss: 149.716995

Train Epoch: 4 [15000/60000 (25%)] Loss: 294.445007

Train Epoch: 4 [20000/60000 (33%)] Loss: 927.258972

Train Epoch: 4 [25000/60000 (42%)] Loss: 194.807037

Train Epoch: 4 [30000/60000 (50%)] Loss: 154.372437

Train Epoch: 4 [35000/60000 (58%)] Loss: 105.806114

Train Epoch: 4 [40000/60000 (67%)] Loss: 46.646584

Train Epoch: 4 [45000/60000 (75%)] Loss: 234.779800

Train Epoch: 4 [50000/60000 (83%)] Loss: 330.368317

Train Epoch: 4 [55000/60000 (92%)] Loss: 152.288895

Test set: Average loss: 288.3810, Accuracy: 5838/10000 (58%)

Train Epoch: 5 [0/60000 (0%)] Loss: 252.736038

Train Epoch: 5 [5000/60000 (8%)] Loss: 249.400116

Train Epoch: 5 [10000/60000 (17%)] Loss: 115.480118

Train Epoch: 5 [15000/60000 (25%)] Loss: 45.534042

Train Epoch: 5 [20000/60000 (33%)] Loss: 217.655014

Train Epoch: 5 [25000/60000 (42%)] Loss: 271.251007

Train Epoch: 5 [30000/60000 (50%)] Loss: 143.609390

Train Epoch: 5 [35000/60000 (58%)] Loss: 144.333374

Train Epoch: 5 [40000/60000 (67%)] Loss: 96.344490

Train Epoch: 5 [45000/60000 (75%)] Loss: 158.942520

Train Epoch: 5 [50000/60000 (83%)] Loss: 193.594864

Train Epoch: 5 [55000/60000 (92%)] Loss: 185.442078

Test set: Average loss: 170.3817, Accuracy: 6830/10000 (68%)

Train Epoch: 6 [0/60000 (0%)] Loss: 65.844162

Train Epoch: 6 [5000/60000 (8%)] Loss: 323.839569

Train Epoch: 6 [10000/60000 (17%)] Loss: 310.406982

Train Epoch: 6 [15000/60000 (25%)] Loss: 115.530586

Train Epoch: 6 [20000/60000 (33%)] Loss: 208.615784

Train Epoch: 6 [25000/60000 (42%)] Loss: 136.508301

Train Epoch: 6 [30000/60000 (50%)] Loss: 151.337112

Train Epoch: 6 [35000/60000 (58%)] Loss: 281.829712

Train Epoch: 6 [40000/60000 (67%)] Loss: 188.019028

Train Epoch: 6 [45000/60000 (75%)] Loss: 209.541107

Train Epoch: 6 [50000/60000 (83%)] Loss: 68.317375

Train Epoch: 6 [55000/60000 (92%)] Loss: 156.827271

Test set: Average loss: 298.6568, Accuracy: 6228/10000 (62%)

Train Epoch: 7 [0/60000 (0%)] Loss: 334.134369

Train Epoch: 7 [5000/60000 (8%)] Loss: 518.261108

Train Epoch: 7 [10000/60000 (17%)] Loss: 189.639221

Train Epoch: 7 [15000/60000 (25%)] Loss: 51.961056

Train Epoch: 7 [20000/60000 (33%)] Loss: 159.698425

Train Epoch: 7 [25000/60000 (42%)] Loss: 173.460266

Train Epoch: 7 [30000/60000 (50%)] Loss: 250.603012

Train Epoch: 7 [35000/60000 (58%)] Loss: 135.429077

Train Epoch: 7 [40000/60000 (67%)] Loss: 107.909599

Train Epoch: 7 [45000/60000 (75%)] Loss: 206.258316

Train Epoch: 7 [50000/60000 (83%)] Loss: 223.867538

Train Epoch: 7 [55000/60000 (92%)] Loss: 193.872742

Test set: Average loss: 142.1682, Accuracy: 7527/10000 (75%)

Train Epoch: 8 [0/60000 (0%)] Loss: 58.534702

Train Epoch: 8 [5000/60000 (8%)] Loss: 253.753479

Train Epoch: 8 [10000/60000 (17%)] Loss: 141.154205

Train Epoch: 8 [15000/60000 (25%)] Loss: 78.934486

Train Epoch: 8 [20000/60000 (33%)] Loss: 79.508865

Train Epoch: 8 [25000/60000 (42%)] Loss: 168.141586

Train Epoch: 8 [30000/60000 (50%)] Loss: 121.500092

Train Epoch: 8 [35000/60000 (58%)] Loss: 207.567566

Train Epoch: 8 [40000/60000 (67%)] Loss: 57.204823

Train Epoch: 8 [45000/60000 (75%)] Loss: 328.148529

Train Epoch: 8 [50000/60000 (83%)] Loss: 128.321136

Train Epoch: 8 [55000/60000 (92%)] Loss: 110.982780

Test set: Average loss: 195.2450, Accuracy: 7265/10000 (72%)

Train Epoch: 9 [0/60000 (0%)] Loss: 215.911270

Train Epoch: 9 [5000/60000 (8%)] Loss: 101.819969

Train Epoch: 9 [10000/60000 (17%)] Loss: 186.153122

Train Epoch: 9 [15000/60000 (25%)] Loss: 303.978668

Train Epoch: 9 [20000/60000 (33%)] Loss: 368.582153

Train Epoch: 9 [25000/60000 (42%)] Loss: 234.488205

Train Epoch: 9 [30000/60000 (50%)] Loss: 80.747780

Train Epoch: 9 [35000/60000 (58%)] Loss: 143.989624

Train Epoch: 9 [40000/60000 (67%)] Loss: 195.743393

Train Epoch: 9 [45000/60000 (75%)] Loss: 109.038643

Train Epoch: 9 [50000/60000 (83%)] Loss: 106.858192

Train Epoch: 9 [55000/60000 (92%)] Loss: 92.899040

Test set: Average loss: 249.0824, Accuracy: 6965/10000 (69%)

Train Epoch: 10 [0/60000 (0%)] Loss: 229.991333

Train Epoch: 10 [5000/60000 (8%)] Loss: 145.907196

Train Epoch: 10 [10000/60000 (17%)] Loss: 189.153442

Train Epoch: 10 [15000/60000 (25%)] Loss: 109.786263

Train Epoch: 10 [20000/60000 (33%)] Loss: 166.477356

Train Epoch: 10 [25000/60000 (42%)] Loss: 211.786835

Train Epoch: 10 [30000/60000 (50%)] Loss: 104.022072

Train Epoch: 10 [35000/60000 (58%)] Loss: 57.647373

Train Epoch: 10 [40000/60000 (67%)] Loss: 170.895996

Train Epoch: 10 [45000/60000 (75%)] Loss: 101.392410

Train Epoch: 10 [50000/60000 (83%)] Loss: 119.571289

Train Epoch: 10 [55000/60000 (92%)] Loss: 409.226715

Test set: Average loss: 165.1452, Accuracy: 6869/10000 (68%)

In addition to a significantly lower accuracy rating, the model initially becomes more accurate until a certain point before losing accuracy. We see this step repeated multiple times. A reason for this is that as the model backprops the loss function to its weights, the fact that the learning rate is set so high, the model overshoots whenever it is updating the weights. This causes the true global minima of the gradients to never be approached.

**Output of Multi-Layer Convolution+Pool on CIFAR-10:**

Train Epoch: 1 [0/50000 (0%)] Loss: 2.300058

Train Epoch: 1 [6400/50000 (13%)] Loss: 2.186357

Train Epoch: 1 [12800/50000 (26%)] Loss: 2.104402

Train Epoch: 1 [19200/50000 (38%)] Loss: 1.920606

Train Epoch: 1 [25600/50000 (51%)] Loss: 1.942123

Train Epoch: 1 [32000/50000 (64%)] Loss: 1.861739

Train Epoch: 1 [38400/50000 (77%)] Loss: 1.810031

Train Epoch: 1 [44800/50000 (90%)] Loss: 1.752169

Test set: Average loss: 1.8155, Accuracy: 3680/10000 (36%)

Train Epoch: 2 [0/50000 (0%)] Loss: 1.773525

Train Epoch: 2 [6400/50000 (13%)] Loss: 1.726168

Train Epoch: 2 [12800/50000 (26%)] Loss: 1.607186

Train Epoch: 2 [19200/50000 (38%)] Loss: 1.850617

Train Epoch: 2 [25600/50000 (51%)] Loss: 1.591353

Train Epoch: 2 [32000/50000 (64%)] Loss: 1.526052

Train Epoch: 2 [38400/50000 (77%)] Loss: 1.692520

Train Epoch: 2 [44800/50000 (90%)] Loss: 1.631051

Test set: Average loss: 1.6233, Accuracy: 4229/10000 (42%)

Train Epoch: 3 [0/50000 (0%)] Loss: 1.594896

Train Epoch: 3 [6400/50000 (13%)] Loss: 1.547104

Train Epoch: 3 [12800/50000 (26%)] Loss: 1.578701

Train Epoch: 3 [19200/50000 (38%)] Loss: 1.635418

Train Epoch: 3 [25600/50000 (51%)] Loss: 1.563136

Train Epoch: 3 [32000/50000 (64%)] Loss: 1.724662

Train Epoch: 3 [38400/50000 (77%)] Loss: 1.587228

Train Epoch: 3 [44800/50000 (90%)] Loss: 1.389694

Test set: Average loss: 1.4936, Accuracy: 4628/10000 (46%)

Train Epoch: 4 [0/50000 (0%)] Loss: 1.369528

Train Epoch: 4 [6400/50000 (13%)] Loss: 1.528630

Train Epoch: 4 [12800/50000 (26%)] Loss: 1.538490

Train Epoch: 4 [19200/50000 (38%)] Loss: 1.543001

Train Epoch: 4 [25600/50000 (51%)] Loss: 1.380842

Train Epoch: 4 [32000/50000 (64%)] Loss: 1.602548

Train Epoch: 4 [38400/50000 (77%)] Loss: 1.393114

Train Epoch: 4 [44800/50000 (90%)] Loss: 1.370325

Test set: Average loss: 1.4098, Accuracy: 4968/10000 (49%)

Train Epoch: 5 [0/50000 (0%)] Loss: 1.318053

Train Epoch: 5 [6400/50000 (13%)] Loss: 1.474610

Train Epoch: 5 [12800/50000 (26%)] Loss: 1.451862

Train Epoch: 5 [19200/50000 (38%)] Loss: 1.296643

Train Epoch: 5 [25600/50000 (51%)] Loss: 1.412723

Train Epoch: 5 [32000/50000 (64%)] Loss: 1.265531

Train Epoch: 5 [38400/50000 (77%)] Loss: 1.480958

Train Epoch: 5 [44800/50000 (90%)] Loss: 1.341624

Test set: Average loss: 1.3391, Accuracy: 5217/10000 (52%)

Train Epoch: 6 [0/50000 (0%)] Loss: 1.357604

Train Epoch: 6 [6400/50000 (13%)] Loss: 1.324507

Train Epoch: 6 [12800/50000 (26%)] Loss: 1.498949

Train Epoch: 6 [19200/50000 (38%)] Loss: 1.293285

Train Epoch: 6 [25600/50000 (51%)] Loss: 1.179439

Train Epoch: 6 [32000/50000 (64%)] Loss: 1.551867

Train Epoch: 6 [38400/50000 (77%)] Loss: 1.205842

Train Epoch: 6 [44800/50000 (90%)] Loss: 1.296253

Test set: Average loss: 1.2709, Accuracy: 5539/10000 (55%)

Train Epoch: 7 [0/50000 (0%)] Loss: 1.295659

Train Epoch: 7 [6400/50000 (13%)] Loss: 1.207994

Train Epoch: 7 [12800/50000 (26%)] Loss: 1.206897

Train Epoch: 7 [19200/50000 (38%)] Loss: 1.132711

Train Epoch: 7 [25600/50000 (51%)] Loss: 1.386095

Train Epoch: 7 [32000/50000 (64%)] Loss: 1.141904

Train Epoch: 7 [38400/50000 (77%)] Loss: 1.369521

Train Epoch: 7 [44800/50000 (90%)] Loss: 1.360053

Test set: Average loss: 1.2939, Accuracy: 5402/10000 (54%)

Train Epoch: 8 [0/50000 (0%)] Loss: 1.350478

Train Epoch: 8 [6400/50000 (13%)] Loss: 1.379741

Train Epoch: 8 [12800/50000 (26%)] Loss: 1.425687

Train Epoch: 8 [19200/50000 (38%)] Loss: 0.939643

Train Epoch: 8 [25600/50000 (51%)] Loss: 1.156356

Train Epoch: 8 [32000/50000 (64%)] Loss: 1.187364

Train Epoch: 8 [38400/50000 (77%)] Loss: 1.126607

Train Epoch: 8 [44800/50000 (90%)] Loss: 1.405283

Test set: Average loss: 1.2047, Accuracy: 5695/10000 (56%)

Train Epoch: 9 [0/50000 (0%)] Loss: 1.210752

Train Epoch: 9 [6400/50000 (13%)] Loss: 1.151698

Train Epoch: 9 [12800/50000 (26%)] Loss: 1.280402

Train Epoch: 9 [19200/50000 (38%)] Loss: 1.420956

Train Epoch: 9 [25600/50000 (51%)] Loss: 1.301944

Train Epoch: 9 [32000/50000 (64%)] Loss: 0.917256

Train Epoch: 9 [38400/50000 (77%)] Loss: 1.006322

Train Epoch: 9 [44800/50000 (90%)] Loss: 1.202852

Test set: Average loss: 1.1801, Accuracy: 5812/10000 (58%)

Train Epoch: 10 [0/50000 (0%)] Loss: 1.039595

Train Epoch: 10 [6400/50000 (13%)] Loss: 1.157739

Train Epoch: 10 [12800/50000 (26%)] Loss: 1.187427

Train Epoch: 10 [19200/50000 (38%)] Loss: 0.852530

Train Epoch: 10 [25600/50000 (51%)] Loss: 1.057583

Train Epoch: 10 [32000/50000 (64%)] Loss: 1.021309

Train Epoch: 10 [38400/50000 (77%)] Loss: 1.072923

Train Epoch: 10 [44800/50000 (90%)] Loss: 1.047818

Test set: Average loss: 1.1137, Accuracy: 6111/10000 (61%)

Train Epoch: 11 [0/50000 (0%)] Loss: 1.087245

Train Epoch: 11 [6400/50000 (13%)] Loss: 0.997588

Train Epoch: 11 [12800/50000 (26%)] Loss: 0.935052

Train Epoch: 11 [19200/50000 (38%)] Loss: 1.235224

Train Epoch: 11 [25600/50000 (51%)] Loss: 1.045611

Train Epoch: 11 [32000/50000 (64%)] Loss: 1.331181

Train Epoch: 11 [38400/50000 (77%)] Loss: 1.086256

Train Epoch: 11 [44800/50000 (90%)] Loss: 1.100690

Test set: Average loss: 1.0903, Accuracy: 6154/10000 (61%)

Train Epoch: 12 [0/50000 (0%)] Loss: 1.081884

Train Epoch: 12 [6400/50000 (13%)] Loss: 0.871600

Train Epoch: 12 [12800/50000 (26%)] Loss: 0.920791

Train Epoch: 12 [19200/50000 (38%)] Loss: 0.954977

Train Epoch: 12 [25600/50000 (51%)] Loss: 1.037084

Train Epoch: 12 [32000/50000 (64%)] Loss: 0.953459

Train Epoch: 12 [38400/50000 (77%)] Loss: 0.972830

Train Epoch: 12 [44800/50000 (90%)] Loss: 0.949859

Test set: Average loss: 1.0654, Accuracy: 6204/10000 (62%)

Train Epoch: 13 [0/50000 (0%)] Loss: 0.902932

Train Epoch: 13 [6400/50000 (13%)] Loss: 0.947247

Train Epoch: 13 [12800/50000 (26%)] Loss: 0.947186

Train Epoch: 13 [19200/50000 (38%)] Loss: 0.878743

Train Epoch: 13 [25600/50000 (51%)] Loss: 1.001055

Train Epoch: 13 [32000/50000 (64%)] Loss: 0.916380

Train Epoch: 13 [38400/50000 (77%)] Loss: 1.168761

Train Epoch: 13 [44800/50000 (90%)] Loss: 0.864930

Test set: Average loss: 1.1030, Accuracy: 6102/10000 (61%)

Train Epoch: 14 [0/50000 (0%)] Loss: 1.023801

Train Epoch: 14 [6400/50000 (13%)] Loss: 0.806883

Train Epoch: 14 [12800/50000 (26%)] Loss: 0.950434

Train Epoch: 14 [19200/50000 (38%)] Loss: 0.804212

Train Epoch: 14 [25600/50000 (51%)] Loss: 1.050177

Train Epoch: 14 [32000/50000 (64%)] Loss: 0.867274

Train Epoch: 14 [38400/50000 (77%)] Loss: 0.880489

Train Epoch: 14 [44800/50000 (90%)] Loss: 0.912456

Test set: Average loss: 1.0146, Accuracy: 6459/10000 (64%)

Train Epoch: 15 [0/50000 (0%)] Loss: 0.906224

Train Epoch: 15 [6400/50000 (13%)] Loss: 0.900678

Train Epoch: 15 [12800/50000 (26%)] Loss: 0.943095

Train Epoch: 15 [19200/50000 (38%)] Loss: 1.263275

Train Epoch: 15 [25600/50000 (51%)] Loss: 1.062028

Train Epoch: 15 [32000/50000 (64%)] Loss: 1.164030

Train Epoch: 15 [38400/50000 (77%)] Loss: 0.770075

Train Epoch: 15 [44800/50000 (90%)] Loss: 0.958447

Test set: Average loss: 1.0701, Accuracy: 6211/10000 (62%)

Train Epoch: 16 [0/50000 (0%)] Loss: 1.228304

Train Epoch: 16 [6400/50000 (13%)] Loss: 0.823625

Train Epoch: 16 [12800/50000 (26%)] Loss: 0.707538

Train Epoch: 16 [19200/50000 (38%)] Loss: 1.026702

Train Epoch: 16 [25600/50000 (51%)] Loss: 0.788655

Train Epoch: 16 [32000/50000 (64%)] Loss: 0.978861

Train Epoch: 16 [38400/50000 (77%)] Loss: 1.143814

Train Epoch: 16 [44800/50000 (90%)] Loss: 1.011824

Test set: Average loss: 0.9873, Accuracy: 6540/10000 (65%)

Train Epoch: 17 [0/50000 (0%)] Loss: 0.990027

Train Epoch: 17 [6400/50000 (13%)] Loss: 0.826487

Train Epoch: 17 [12800/50000 (26%)] Loss: 0.927338

Train Epoch: 17 [19200/50000 (38%)] Loss: 0.763406

Train Epoch: 17 [25600/50000 (51%)] Loss: 0.992121

Train Epoch: 17 [32000/50000 (64%)] Loss: 0.848313

Train Epoch: 17 [38400/50000 (77%)] Loss: 0.761506

Train Epoch: 17 [44800/50000 (90%)] Loss: 0.771199

Test set: Average loss: 1.0094, Accuracy: 6437/10000 (64%)

Train Epoch: 18 [0/50000 (0%)] Loss: 0.944731

Train Epoch: 18 [6400/50000 (13%)] Loss: 0.667409

Train Epoch: 18 [12800/50000 (26%)] Loss: 0.851156

Train Epoch: 18 [19200/50000 (38%)] Loss: 0.791833

Train Epoch: 18 [25600/50000 (51%)] Loss: 0.870147

Train Epoch: 18 [32000/50000 (64%)] Loss: 1.021163

Train Epoch: 18 [38400/50000 (77%)] Loss: 0.726769

Train Epoch: 18 [44800/50000 (90%)] Loss: 0.642900

Test set: Average loss: 1.0068, Accuracy: 6474/10000 (64%)

Train Epoch: 19 [0/50000 (0%)] Loss: 0.950041

Train Epoch: 19 [6400/50000 (13%)] Loss: 0.685340

Train Epoch: 19 [12800/50000 (26%)] Loss: 0.817995

Train Epoch: 19 [19200/50000 (38%)] Loss: 0.846507

Train Epoch: 19 [25600/50000 (51%)] Loss: 0.747779

Train Epoch: 19 [32000/50000 (64%)] Loss: 0.914037

Train Epoch: 19 [38400/50000 (77%)] Loss: 0.834496

Train Epoch: 19 [44800/50000 (90%)] Loss: 1.130694

Test set: Average loss: 1.0461, Accuracy: 6385/10000 (63%)

Train Epoch: 20 [0/50000 (0%)] Loss: 0.712762

Train Epoch: 20 [6400/50000 (13%)] Loss: 0.901884

Train Epoch: 20 [12800/50000 (26%)] Loss: 0.732032

Train Epoch: 20 [19200/50000 (38%)] Loss: 0.828943

Train Epoch: 20 [25600/50000 (51%)] Loss: 0.838091

Train Epoch: 20 [32000/50000 (64%)] Loss: 0.616142

Train Epoch: 20 [38400/50000 (77%)] Loss: 0.893511

Train Epoch: 20 [44800/50000 (90%)] Loss: 0.725003

Test set: Average loss: 0.9688, Accuracy: 6591/10000 (65%)