label:tensor([[1.]])

grad:tensor([[ 1.2075e-03, -3.5275e-03, 6.1960e-04, -7.7800e-04, 6.6718e-04,

-7.2844e-03, 1.6655e-03, -2.5239e-03, -8.7274e-03, 9.0206e-03,

-4.2019e-03, -7.2530e-03, 1.6328e-05, 1.1138e-03, -1.7431e-03,

-5.2046e-03, -4.8536e-03, 5.5336e-03, -1.0208e-02, -1.2267e-02,

2.2727e-03, 1.7568e-03, -5.9776e-03, 1.1952e-03, -8.9934e-03,

3.9749e-03, 3.3249e-03, 8.7700e-03, -4.7909e-03, -3.0830e-03]])

data:tensor([[1.2470e+01, 1.8600e+01, 8.1090e+01, 4.8190e+02, 9.9650e-02, 1.0580e-01,

8.0050e-02, 3.8210e-02, 1.9250e-01, 6.3730e-02, 3.9610e-01, 1.0440e+00,

2.4970e+00, 3.0290e+01, 6.9530e-03, 1.9110e-02, 2.7010e-02, 1.0370e-02,

1.7820e-02, 3.5860e-03, 1.4970e+01, 2.4640e+01, 9.6050e+01, 6.7790e+02,

1.4260e-01, 2.3780e-01, 2.6710e-01, 1.0150e-01, 3.0140e-01, 8.7500e-02]])

loss in each character:tensor([[ 1.5058e-02, -6.5612e-02, 5.0243e-02, -3.7492e-01, 6.6484e-05,

-7.7069e-04, 1.3332e-04, -9.6437e-05, -1.6800e-03, 5.7488e-04,

-1.6644e-03, -7.5722e-03, 4.0772e-05, 3.3738e-02, -1.2120e-05,

-9.9460e-05, -1.3110e-04, 5.7384e-05, -1.8191e-04, -4.3988e-05,

3.4022e-02, 4.3288e-02, -5.7414e-01, 8.1020e-01, -1.2825e-03,

9.4523e-04, 8.8808e-04, 8.9015e-04, -1.4440e-03, -2.6976e-04]])

label:tensor([[0.]])

grad:tensor([[-0.0050, 0.0154, -0.0023, 0.0021, -0.0019, 0.0309, -0.0073, 0.0110,

0.0365, -0.0371, 0.0179, 0.0298, -0.0005, -0.0045, 0.0067, 0.0220,

0.0196, -0.0242, 0.0435, 0.0509, -0.0101, -0.0071, 0.0248, -0.0042,

0.0381, -0.0171, -0.0132, -0.0370, 0.0203, 0.0133]])

data:tensor([[1.6260e+01, 2.1880e+01, 1.0750e+02, 8.2680e+02, 1.1650e-01, 1.2830e-01,

1.7990e-01, 7.9810e-02, 1.8690e-01, 6.5320e-02, 5.7060e-01, 1.4570e+00,

2.9610e+00, 5.7720e+01, 1.0560e-02, 3.7560e-02, 5.8390e-02, 1.1860e-02,

4.0220e-02, 6.1870e-03, 1.7730e+01, 2.5210e+01, 1.1370e+02, 9.7520e+02,

1.4260e-01, 2.1160e-01, 3.3440e-01, 1.0470e-01, 2.7360e-01, 7.9530e-02]])

loss in each character:tensor([[-8.1534e-02, 3.3634e-01, -2.5130e-01, 1.7183e+00, -2.2181e-04,

3.9662e-03, -1.3178e-03, 8.8161e-04, 6.8139e-03, -2.4210e-03,

1.0217e-02, 4.3450e-02, -1.4841e-03, -2.5724e-01, 7.0318e-05,

8.2463e-04, 1.1455e-03, -2.8690e-04, 1.7483e-03, 3.1516e-04,

-1.7958e-01, -1.7905e-01, 2.8184e+00, -4.0642e+00, 5.4376e-03,

-3.6118e-03, -4.4040e-03, -3.8703e-03, 5.5404e-03, 1.0611e-03]])

label:tensor([[0.]])

grad:tensor([[-1.0114e-03, 3.1937e-03, -3.9825e-04, 2.9925e-04, -2.9662e-04,

6.3770e-03, -1.5548e-03, 2.3503e-03, 7.4807e-03, -7.5876e-03,

3.7026e-03, 6.1239e-03, -7.0948e-05, -9.6114e-04, 1.2072e-03,

4.6578e-03, 4.0512e-03, -5.0459e-03, 8.9753e-03, 1.0558e-02,

-2.1768e-03, -1.4696e-03, 5.1191e-03, -7.4757e-04, 7.8832e-03,

-3.4202e-03, -2.6768e-03, -7.6950e-03, 4.1828e-03, 2.9521e-03]])

data:tensor([[1.9100e+01, 2.6290e+01, 1.2910e+02, 1.1320e+03, 1.2150e-01, 1.7910e-01,

1.9370e-01, 1.4690e-01, 1.6340e-01, 7.2240e-02, 5.1900e-01, 2.9100e+00,

5.8010e+00, 6.7100e+01, 7.5450e-03, 6.0500e-02, 2.1340e-02, 1.8430e-02,

3.0560e-02, 1.0390e-02, 2.0330e+01, 3.2720e+01, 1.4130e+02, 1.2980e+03,

1.3920e-01, 2.8170e-01, 2.4320e-01, 1.8410e-01, 2.3110e-01, 9.2030e-02]])

loss in each character:tensor([[-1.9318e-02, 8.3963e-02, -5.1415e-02, 3.3875e-01, -3.6040e-05,

1.1421e-03, -3.0116e-04, 3.4526e-04, 1.2223e-03, -5.4813e-04,

1.9217e-03, 1.7821e-02, -4.1157e-04, -6.4492e-02, 9.1084e-06,

2.8180e-04, 8.6452e-05, -9.2996e-05, 2.7429e-04, 1.0970e-04,

-4.4253e-02, -4.8087e-02, 7.2333e-01, -9.7034e-01, 1.0973e-03,

-9.6347e-04, -6.5100e-04, -1.4167e-03, 9.6665e-04, 2.7168e-04]])

label:tensor([[1.]])

grad:tensor([[-0.0080, -0.0079, 0.0001, -0.0217, 0.0088, -0.0020, -0.0054, -0.0019,

-0.0031, 0.0244, 0.0047, -0.0177, -0.0159, -0.0034, -0.0055, 0.0024,

0.0021, -0.0081, 0.0067, -0.0064, -0.0086, -0.0010, 0.0013, 0.0174,

0.0126, 0.0020, 0.0051, -0.0028, 0.0049, -0.0099]])

data:tensor([[1.1340e+01, 2.1260e+01, 7.2480e+01, 3.9650e+02, 8.7590e-02, 6.5750e-02,

5.1330e-02, 1.8990e-02, 1.4870e-01, 6.5290e-02, 2.3440e-01, 9.8610e-01,

1.5970e+00, 1.6410e+01, 9.1130e-03, 1.5570e-02, 2.4430e-02, 6.4350e-03,

1.5680e-02, 2.4770e-03, 1.3010e+01, 2.9150e+01, 8.3990e+01, 5.1810e+02,

1.6990e-01, 2.1960e-01, 3.1200e-01, 8.2780e-02, 2.8290e-01, 8.8320e-02]])

loss in each character:tensor([[-9.1264e-02, -1.6741e-01, 1.0351e-02, -8.5945e+00, 7.7406e-04,

-1.3254e-04, -2.7484e-04, -3.5161e-05, -4.6041e-04, 1.5925e-03,

1.1050e-03, -1.7496e-02, -2.5437e-02, -5.5546e-02, -4.9695e-05,

3.7422e-05, 5.1125e-05, -5.2175e-05, 1.0497e-04, -1.5734e-05,

-1.1129e-01, -2.9502e-02, 1.1336e-01, 9.0149e+00, 2.1382e-03,

4.4684e-04, 1.5907e-03, -2.2823e-04, 1.3904e-03, -8.7588e-04]])

label:tensor([[1.]])

grad:tensor([[ 9.5184e-07, -9.3830e-04, 3.3431e-04, -2.7286e-03, 9.7167e-04,

-7.9435e-04, -2.5819e-04, -1.5107e-04, -1.7660e-03, 3.8783e-03,

4.8706e-04, -2.7458e-03, -1.8073e-03, -1.8762e-04, -1.3313e-03,

-4.2391e-04, -8.0453e-04, -9.1870e-04, -7.8053e-04, -2.5822e-03,

-6.1850e-04, 7.2510e-04, -1.0358e-03, 2.2133e-03, -2.8746e-04,

8.4935e-04, 1.2133e-03, 8.3534e-04, -6.3156e-05, -8.3267e-04]])

data:tensor([[1.1670e+01, 2.0020e+01, 7.5210e+01, 4.1620e+02, 1.0160e-01, 9.4530e-02,

4.2000e-02, 2.1570e-02, 1.8590e-01, 6.4610e-02, 2.0670e-01, 8.7450e-01,

1.3930e+00, 1.5340e+01, 5.2510e-03, 1.7270e-02, 1.8400e-02, 5.2980e-03,

1.4490e-02, 2.6710e-03, 1.3350e+01, 2.8810e+01, 8.7000e+01, 5.5060e+02,

1.5500e-01, 2.9640e-01, 2.7580e-01, 8.1200e-02, 3.2060e-01, 8.9500e-02]])

loss in each character:tensor([[ 1.1108e-05, -1.8785e-02, 2.5143e-02, -1.1357e+00, 9.8722e-05,

-7.5090e-05, -1.0844e-05, -3.2585e-06, -3.2830e-04, 2.5057e-04,

1.0068e-04, -2.4012e-03, -2.5175e-03, -2.8781e-03, -6.9909e-06,

-7.3209e-06, -1.4803e-05, -4.8673e-06, -1.1310e-05, -6.8971e-06,

-8.2570e-03, 2.0890e-02, -9.0112e-02, 1.2187e+00, -4.4557e-05,

2.5175e-04, 3.3462e-04, 6.7830e-05, -2.0248e-05, -7.4524e-05]])

label:tensor([[0.]])

grad:tensor([[-2.5893e-03, -1.9429e-04, -1.0162e-03, 2.7263e-03, -1.4103e-05,

-5.4344e-04, -4.0720e-04, -7.7389e-04, 2.4165e-03, -2.7847e-03,

-2.2359e-03, 1.5315e-03, 1.3285e-03, 2.8848e-04, 2.5855e-03,

6.8838e-04, 1.5735e-03, 2.8025e-03, 1.7081e-03, 3.1286e-03,

1.5033e-04, -2.7432e-03, 2.4060e-03, -2.1209e-03, 2.3790e-03,

-1.3463e-03, -1.3824e-03, -8.3677e-04, 2.6447e-04, -1.4360e-03]])

data:tensor([[1.3480e+01, 2.0820e+01, 8.8400e+01, 5.5920e+02, 1.0160e-01, 1.2550e-01,

1.0630e-01, 5.4390e-02, 1.7200e-01, 6.4190e-02, 2.1300e-01, 5.9140e-01,

1.5450e+00, 1.8520e+01, 5.3670e-03, 2.2390e-02, 3.0490e-02, 1.2620e-02,

1.3770e-02, 3.1870e-03, 1.5530e+01, 2.6020e+01, 1.0730e+02, 7.4040e+02,

1.6100e-01, 4.2250e-01, 5.0300e-01, 2.2580e-01, 2.8070e-01, 1.0710e-01]])

loss in each character:tensor([[-3.4904e-02, -4.0451e-03, -8.9830e-02, 1.5246e+00, -1.4329e-06,

-6.8202e-05, -4.3286e-05, -4.2092e-05, 4.1564e-04, -1.7875e-04,

-4.7626e-04, 9.0574e-04, 2.0525e-03, 5.3427e-03, 1.3877e-05,

1.5413e-05, 4.7976e-05, 3.5368e-05, 2.3521e-05, 9.9708e-06,

2.3346e-03, -7.1377e-02, 2.5816e-01, -1.5703e+00, 3.8301e-04,

-5.6881e-04, -6.9537e-04, -1.8894e-04, 7.4236e-05, -1.5379e-04]])

label:tensor([[0.]])

grad:tensor([[-1.5543e-03, 4.8252e-03, -7.1446e-04, 6.7681e-04, -6.1462e-04,

9.7077e-03, -2.2636e-03, 3.4805e-03, 1.1460e-02, -1.1724e-02,

5.6355e-03, 9.4013e-03, -8.9498e-05, -1.4137e-03, 2.0496e-03,

6.9239e-03, 6.1960e-03, -7.5603e-03, 1.3610e-02, 1.6010e-02,

-3.1676e-03, -2.2213e-03, 7.7577e-03, -1.3192e-03, 1.1911e-02,

-5.3354e-03, -4.1643e-03, -1.1613e-02, 6.3121e-03, 4.2587e-03]])

data:tensor([[1.6460e+01, 2.0110e+01, 1.0930e+02, 8.3290e+02, 9.8310e-02, 1.5560e-01,

1.7930e-01, 8.8660e-02, 1.7940e-01, 6.3230e-02, 3.0370e-01, 1.2840e+00,

2.4820e+00, 3.1590e+01, 6.6270e-03, 4.0940e-02, 5.3710e-02, 1.8130e-02,

1.6820e-02, 4.5840e-03, 1.7790e+01, 2.8450e+01, 1.2350e+02, 9.8120e+02,

1.4150e-01, 4.6670e-01, 5.8620e-01, 2.0350e-01, 3.0540e-01, 9.5190e-02]])

loss in each character:tensor([[-2.5585e-02, 9.7035e-02, -7.8090e-02, 5.6371e-01, -6.0423e-05,

1.5105e-03, -4.0587e-04, 3.0858e-04, 2.0559e-03, -7.4130e-04,

1.7115e-03, 1.2071e-02, -2.2213e-04, -4.4659e-02, 1.3583e-05,

2.8346e-04, 3.3279e-04, -1.3707e-04, 2.2893e-04, 7.3392e-05,

-5.6352e-02, -6.3195e-02, 9.5808e-01, -1.2944e+00, 1.6855e-03,

-2.4900e-03, -2.4411e-03, -2.3632e-03, 1.9277e-03, 4.0538e-04]])

label:tensor([[0.]])

grad:tensor([[-0.0006, 0.0032, -0.0005, 0.0013, -0.0008, 0.0062, -0.0012, 0.0022,

0.0073, -0.0084, 0.0034, 0.0066, 0.0006, -0.0008, 0.0016, 0.0042,

0.0039, -0.0044, 0.0082, 0.0103, -0.0016, -0.0013, 0.0048, -0.0015,

0.0069, -0.0034, -0.0029, -0.0071, 0.0038, 0.0030]])

data:tensor([[1.4480e+01, 2.1460e+01, 9.4250e+01, 6.4820e+02, 9.4440e-02, 9.9470e-02,

1.2040e-01, 4.9380e-02, 2.0750e-01, 5.6360e-02, 4.2040e-01, 2.2200e+00,

3.3010e+00, 3.8870e+01, 9.3690e-03, 2.9830e-02, 5.3710e-02, 1.7610e-02,

2.4180e-02, 3.2490e-03, 1.6210e+01, 2.9250e+01, 1.0840e+02, 8.0890e+02,

1.3060e-01, 1.9760e-01, 3.3490e-01, 1.2250e-01, 3.0200e-01, 6.8460e-02]])

loss in each character:tensor([[-9.0287e-03, 6.9689e-02, -4.4342e-02, 8.3529e-01, -7.6026e-05,

6.1459e-04, -1.4379e-04, 1.0888e-04, 1.5152e-03, -4.7069e-04,

1.4174e-03, 1.4630e-02, 1.8888e-03, -3.2321e-02, 1.4545e-05,

1.2546e-04, 2.0847e-04, -7.7026e-05, 1.9765e-04, 3.3549e-05,

-2.6210e-02, -3.8744e-02, 5.2006e-01, -1.2210e+00, 9.0394e-04,

-6.6939e-04, -9.6601e-04, -8.7413e-04, 1.1365e-03, 2.0632e-04]])

label:tensor([[1.]])

grad:tensor([[-3.0726e-04, -1.7748e-04, 1.7354e-04, -7.2505e-04, 6.6507e-04,

-1.1151e-03, 2.0100e-04, -7.4508e-05, -1.1045e-03, 1.8968e-03,

-4.2947e-04, -1.4315e-03, -4.0403e-04, 6.3094e-04, -7.9812e-04,

-3.6650e-04, -1.0420e-03, 2.5289e-04, -5.8813e-04, -2.2410e-03,

-6.1092e-05, 7.5967e-05, -8.4249e-04, 5.8076e-04, -7.1254e-04,

1.9948e-04, 9.5322e-04, 9.0569e-04, -6.8411e-04, -5.1212e-04]])

data:tensor([[1.0820e+01, 2.4210e+01, 6.8890e+01, 3.6160e+02, 8.1920e-02, 6.6020e-02,

1.5480e-02, 8.1600e-03, 1.9760e-01, 6.3280e-02, 5.1960e-01, 1.9180e+00,

3.5640e+00, 3.3000e+01, 8.2630e-03, 1.8700e-02, 1.2770e-02, 5.9170e-03,

2.4660e-02, 2.9770e-03, 1.3030e+01, 3.1450e+01, 8.3900e+01, 5.0560e+02,

1.2040e-01, 1.6330e-01, 6.1940e-02, 3.2640e-02, 3.0590e-01, 7.6260e-02]])

loss in each character:tensor([[-3.3245e-03, -4.2969e-03, 1.1955e-02, -2.6218e-01, 5.4483e-05,

-7.3622e-05, 3.1115e-06, -6.0799e-07, -2.1826e-04, 1.2003e-04,

-2.2315e-04, -2.7455e-03, -1.4400e-03, 2.0821e-02, -6.5949e-06,

-6.8535e-06, -1.3306e-05, 1.4964e-06, -1.4503e-05, -6.6715e-06,

-7.9602e-04, 2.3892e-03, -7.0685e-02, 2.9363e-01, -8.5789e-05,

3.2575e-05, 5.9042e-05, 2.9562e-05, -2.0927e-04, -3.9054e-05]])