

# MP3\_P2

November 1, 2020

```
[1]: import os
import random
os.environ["KMP_DUPLICATE_LIB_OK"]="True"
import cv2
import numpy as np

import torch
from torch.utils.data import DataLoader
from torchvision import models

from resnet_yolo import resnet50
from yolo_loss import YoloLoss
from dataset import VocDetectorDataset
from eval_voc import evaluate
from predict import predict_image
from config import VOC_CLASSES, COLORS
from kaggle_submission import output_submission_csv
import matplotlib.pyplot as plt

%matplotlib inline
%load_ext autoreload
%autoreload 2
```

## 1 Assignment3 Part2: Yolo Detection

We provide you a Yolo Detection network implementation, which is not finished. You are asked to complete the implementation by writing the loss function.

### 1.1 What to do

You are asked to implement the loss function in `yolo_loss.py`. You can use `yolo_loss_debug_tool.ipynb` to help you debug.

### 1.2 What to submit

See the submission template for what to submit.

### 1.3 Initialization

```
[2]: device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
      print(device)
```

cuda:0

```
[3]: # YOLO network hyperparameters
      B = 2 # number of bounding box predictions per cell
      S = 14 # width/height of network output grid (larger than 7x7 from paper since
      ↪we use a different network)
```

To implement Yolo we will rely on a pretrained classifier as the backbone for our detection network. PyTorch offers a variety of models which are pretrained on ImageNet in the `torchvision.models` package. In particular, we will use the ResNet50 architecture as a base for our detector. This is different from the base architecture in the Yolo paper and also results in a different output grid size (14x14 instead of 7x7).

Models are typically pretrained on ImageNet since the dataset is very large (> 1million images) and widely used. The pretrained model provides a very useful weight initialization for our detector, so that the network is able to learn quickly and effectively.

```
[4]: load_network_path = None
      pretrained = True

      # use to load a previously trained network
      if load_network_path is not None:
          print('Loading saved network from {}'.format(load_network_path))
          net = resnet50().to(device)
          net.load_state_dict(torch.load(load_network_path))
      else:
          print('Load pre-trained model')
          net = resnet50(pretrained=pretrained).to(device)
```

Load pre-trained model

```
[5]: learning_rate = 0.001
      num_epochs = 50
      batch_size = 12

      # Yolo loss component coefficients (as given in Yolo v1 paper)
      lambda_coord = 5
      lambda_noobj = 0.5
```

```
[6]: criterion = YoloLoss(S, B, lambda_coord, lambda_noobj)
      optimizer = torch.optim.SGD(net.parameters(), lr=learning_rate, momentum=0.9,
      ↪weight_decay=5e-4)
```

## 1.4 Reading Pascal Data

Since Pascal is a small dataset (5000 in train+val) we have combined the train and val splits to train our detector. This is not typically a good practice, but we will make an exception in this case to be able to get reasonable detection results with a comparatively small object detection dataset.

The train dataset loader also using a variety of data augmentation techniques including random shift, scaling, crop, and flips. Data augmentation is slightly more complicated for detection dataset since the bounding box annotations must be kept consistent through the transformations.

Since the output of the detector network we train is an  $S \times S \times (B \times 5 + C)$ , we use an encoder to convert the original bounding box coordinates into relative grid bounding box coordinates corresponding to the the expected output. We also use a decoder which allows us to convert the opposite direction into image coordinate bounding boxes.

```
[7]: file_root_train = 'VOCdevkit_2007/VOC2007/JPEGImages/'
      annotation_file_train = 'voc2007.txt'

      train_dataset =
        ↳VocDetectorDataset(root_img_dir=file_root_train,dataset_file=annotation_file_train,train=True,
        ↳S=S)
      train_loader =
        ↳DataLoader(train_dataset,batch_size=batch_size,shuffle=True,num_workers=4)
      print('Loaded %d train images' % len(train_dataset))
```

Initializing dataset

Loaded 5011 train images

```
[27]: file_root_test = 'VOCdevkit_2007/VOC2007test/JPEGImages/'
      annotation_file_test = 'voc2007test.txt'

      test_dataset =
        ↳VocDetectorDataset(root_img_dir=file_root_test,dataset_file=annotation_file_test,train=False,
        ↳S=S)
      test_loader =
        ↳DataLoader(test_dataset,batch_size=batch_size,shuffle=False,num_workers=4)
      print('Loaded %d test images' % len(test_dataset))
```

Initializing dataset

Loaded 4950 test images

## 1.5 Train detector

```
[9]: best_test_loss = np.inf

      for epoch in range(num_epochs):
          net.train()

          # Update learning rate late in training
```

```

if epoch == 30 or epoch == 40:
    learning_rate /= 10.0

for param_group in optimizer.param_groups:
    param_group['lr'] = learning_rate

print('\n\nStarting epoch %d / %d' % (epoch + 1, num_epochs))
print('Learning Rate for this epoch: {}'.format(learning_rate))

total_loss = 0.

for i, (images, target) in enumerate(train_loader):
    images, target = images.to(device), target.to(device)

    pred = net(images)
    loss = criterion(pred, target)
    total_loss += loss.item()

    optimizer.zero_grad()
    loss.backward()
    optimizer.step()
    if (i+1) % 5 == 0:
        print('Epoch [%d/%d], Iter [%d/%d] Loss: %.4f, average_loss: %.4f'
              % (epoch+1, num_epochs, i+1, len(train_loader), loss.item(),
→total_loss / (i+1)))

if epoch % 5 == 0:
    test_aprs = evaluate(net, test_dataset_file=annotation_file_test)

# evaluate the network on the test data
with torch.no_grad():
    test_loss = 0.0
    net.eval()
    for i, (images, target) in enumerate(test_loader):
        images, target = images.to(device), target.to(device)

        pred = net(images)
        loss = criterion(pred, target)
        test_loss += loss.item()
    test_loss /= len(test_loader)

if best_test_loss > test_loss:
    best_test_loss = test_loss
    print('Updating best test loss: %.5f' % best_test_loss)
    torch.save(net.state_dict(), 'best_detector.pth')

torch.save(net.state_dict(), 'detector.pth')

```

```
Starting epoch 1 / 50
Learning Rate for this epoch: 0.001
Epoch [1/50], Iter [5/418] Loss: 65.3886, average_loss: 69.6646
Epoch [1/50], Iter [10/418] Loss: 49.2433, average_loss: 61.1423
Epoch [1/50], Iter [15/418] Loss: 24.2163, average_loss: 51.7503
Epoch [1/50], Iter [20/418] Loss: 22.1707, average_loss: 44.2904
Epoch [1/50], Iter [25/418] Loss: 14.9132, average_loss: 38.3107
Epoch [1/50], Iter [30/418] Loss: 20.6197, average_loss: 34.2178
Epoch [1/50], Iter [35/418] Loss: 11.5318, average_loss: 31.0002
Epoch [1/50], Iter [40/418] Loss: 9.8201, average_loss: 28.7125
Epoch [1/50], Iter [45/418] Loss: 18.3350, average_loss: 27.1103
Epoch [1/50], Iter [50/418] Loss: 6.5053, average_loss: 25.3286
Epoch [1/50], Iter [55/418] Loss: 11.8547, average_loss: 23.9545
Epoch [1/50], Iter [60/418] Loss: 12.0990, average_loss: 22.7367
Epoch [1/50], Iter [65/418] Loss: 11.4756, average_loss: 21.8248
Epoch [1/50], Iter [70/418] Loss: 10.4918, average_loss: 20.9617
Epoch [1/50], Iter [75/418] Loss: 7.3049, average_loss: 20.1093
Epoch [1/50], Iter [80/418] Loss: 5.3298, average_loss: 19.2442
Epoch [1/50], Iter [85/418] Loss: 9.0321, average_loss: 18.5328
Epoch [1/50], Iter [90/418] Loss: 9.1787, average_loss: 17.9243
Epoch [1/50], Iter [95/418] Loss: 12.6947, average_loss: 17.4706
Epoch [1/50], Iter [100/418] Loss: 8.8721, average_loss: 16.9991
Epoch [1/50], Iter [105/418] Loss: 10.3300, average_loss: 16.5517
Epoch [1/50], Iter [110/418] Loss: 7.0819, average_loss: 16.1636
Epoch [1/50], Iter [115/418] Loss: 7.0105, average_loss: 15.7994
Epoch [1/50], Iter [120/418] Loss: 6.7109, average_loss: 15.4150
Epoch [1/50], Iter [125/418] Loss: 7.7331, average_loss: 15.1718
Epoch [1/50], Iter [130/418] Loss: 7.9856, average_loss: 14.8421
Epoch [1/50], Iter [135/418] Loss: 5.8604, average_loss: 14.5227
Epoch [1/50], Iter [140/418] Loss: 6.8238, average_loss: 14.2381
Epoch [1/50], Iter [145/418] Loss: 6.9377, average_loss: 14.0122
Epoch [1/50], Iter [150/418] Loss: 4.3260, average_loss: 13.7629
Epoch [1/50], Iter [155/418] Loss: 5.4025, average_loss: 13.5361
Epoch [1/50], Iter [160/418] Loss: 5.6752, average_loss: 13.3060
Epoch [1/50], Iter [165/418] Loss: 5.1255, average_loss: 13.1036
Epoch [1/50], Iter [170/418] Loss: 6.1087, average_loss: 12.9287
Epoch [1/50], Iter [175/418] Loss: 6.4733, average_loss: 12.7469
Epoch [1/50], Iter [180/418] Loss: 4.4048, average_loss: 12.5716
Epoch [1/50], Iter [185/418] Loss: 4.0244, average_loss: 12.3699
Epoch [1/50], Iter [190/418] Loss: 5.7435, average_loss: 12.2147
Epoch [1/50], Iter [195/418] Loss: 7.2409, average_loss: 12.0582
Epoch [1/50], Iter [200/418] Loss: 4.2728, average_loss: 11.9303
Epoch [1/50], Iter [205/418] Loss: 5.3363, average_loss: 11.7906
```

```

Epoch [1/50], Iter [210/418] Loss: 6.6876, average_loss: 11.6514
Epoch [1/50], Iter [215/418] Loss: 6.7123, average_loss: 11.5158
Epoch [1/50], Iter [220/418] Loss: 4.0968, average_loss: 11.4039
Epoch [1/50], Iter [225/418] Loss: 7.0974, average_loss: 11.2888
Epoch [1/50], Iter [230/418] Loss: 4.1523, average_loss: 11.1646
Epoch [1/50], Iter [235/418] Loss: 5.3295, average_loss: 11.0584
Epoch [1/50], Iter [240/418] Loss: 9.4687, average_loss: 10.9587
Epoch [1/50], Iter [245/418] Loss: 6.0773, average_loss: 10.8829
Epoch [1/50], Iter [250/418] Loss: 4.2032, average_loss: 10.7649
Epoch [1/50], Iter [255/418] Loss: 6.0083, average_loss: 10.6584
Epoch [1/50], Iter [260/418] Loss: 5.7676, average_loss: 10.5648
Epoch [1/50], Iter [265/418] Loss: 5.4965, average_loss: 10.4706
Epoch [1/50], Iter [270/418] Loss: 3.5563, average_loss: 10.3661
Epoch [1/50], Iter [275/418] Loss: 5.8297, average_loss: 10.2858
Epoch [1/50], Iter [280/418] Loss: 6.7755, average_loss: 10.1981
Epoch [1/50], Iter [285/418] Loss: 6.4280, average_loss: 10.1197
Epoch [1/50], Iter [290/418] Loss: 5.3167, average_loss: 10.0428
Epoch [1/50], Iter [295/418] Loss: 3.6616, average_loss: 9.9670
Epoch [1/50], Iter [300/418] Loss: 4.6364, average_loss: 9.8867
Epoch [1/50], Iter [305/418] Loss: 6.4078, average_loss: 9.8244
Epoch [1/50], Iter [310/418] Loss: 6.5280, average_loss: 9.7516
Epoch [1/50], Iter [315/418] Loss: 3.9695, average_loss: 9.6954
Epoch [1/50], Iter [320/418] Loss: 5.2071, average_loss: 9.6196
Epoch [1/50], Iter [325/418] Loss: 5.9988, average_loss: 9.5530
Epoch [1/50], Iter [330/418] Loss: 6.2824, average_loss: 9.5179
Epoch [1/50], Iter [335/418] Loss: 3.3932, average_loss: 9.4602
Epoch [1/50], Iter [340/418] Loss: 4.3472, average_loss: 9.3909
Epoch [1/50], Iter [345/418] Loss: 5.1416, average_loss: 9.3442
Epoch [1/50], Iter [350/418] Loss: 5.3631, average_loss: 9.2816
Epoch [1/50], Iter [355/418] Loss: 4.5828, average_loss: 9.2158
Epoch [1/50], Iter [360/418] Loss: 5.1077, average_loss: 9.1603
Epoch [1/50], Iter [365/418] Loss: 5.0781, average_loss: 9.1144
Epoch [1/50], Iter [370/418] Loss: 4.1438, average_loss: 9.0528
Epoch [1/50], Iter [375/418] Loss: 6.9267, average_loss: 9.0259
Epoch [1/50], Iter [380/418] Loss: 5.7580, average_loss: 8.9756
Epoch [1/50], Iter [385/418] Loss: 4.8384, average_loss: 8.9187
Epoch [1/50], Iter [390/418] Loss: 5.9994, average_loss: 8.8719
Epoch [1/50], Iter [395/418] Loss: 5.7901, average_loss: 8.8196
Epoch [1/50], Iter [400/418] Loss: 5.3300, average_loss: 8.7663
Epoch [1/50], Iter [405/418] Loss: 3.4800, average_loss: 8.7327
Epoch [1/50], Iter [410/418] Loss: 4.7400, average_loss: 8.6885
Epoch [1/50], Iter [415/418] Loss: 5.1400, average_loss: 8.6472
---Evaluate model on test samples---

```

100%|

| 4950/4950 [02:29<00:00, 33.14it/s]

---class aeroplane ap 0.0---

---class bicycle ap 0.0--- (no predictions for this class)

```
---class bird ap 0.0--- (no predictions for this class)
---class boat ap 0.0--- (no predictions for this class)
---class bottle ap 0.0--- (no predictions for this class)
---class bus ap 0.0--- (no predictions for this class)
---class car ap 0.0--- (no predictions for this class)
---class cat ap 0.0--- (no predictions for this class)
---class chair ap 0.0--- (no predictions for this class)
---class cow ap 0.0--- (no predictions for this class)
---class diningtable ap 0.0--- (no predictions for this class)
---class dog ap 0.0--- (no predictions for this class)
---class horse ap 0.0--- (no predictions for this class)
---class motorbike ap 0.0--- (no predictions for this class)
---class person ap 0.0--- (no predictions for this class)
---class pottedplant ap 0.0--- (no predictions for this class)
---class sheep ap 0.0--- (no predictions for this class)
---class sofa ap 0.0--- (no predictions for this class)
---class train ap 0.0--- (no predictions for this class)
---class tvmonitor ap 0.0--- (no predictions for this class)
---map 0.0---
Updating best test loss: 5.29597
```

Starting epoch 2 / 50

Learning Rate for this epoch: 0.001

```
Epoch [2/50], Iter [5/418] Loss: 6.2336, average_loss: 5.4039
Epoch [2/50], Iter [10/418] Loss: 5.2317, average_loss: 5.1969
Epoch [2/50], Iter [15/418] Loss: 5.7216, average_loss: 5.6443
Epoch [2/50], Iter [20/418] Loss: 5.9497, average_loss: 5.5909
Epoch [2/50], Iter [25/418] Loss: 4.9414, average_loss: 5.3517
Epoch [2/50], Iter [30/418] Loss: 3.6978, average_loss: 5.2302
Epoch [2/50], Iter [35/418] Loss: 4.1107, average_loss: 5.2826
Epoch [2/50], Iter [40/418] Loss: 3.7990, average_loss: 5.2649
Epoch [2/50], Iter [45/418] Loss: 4.6692, average_loss: 5.3098
Epoch [2/50], Iter [50/418] Loss: 7.3617, average_loss: 5.3104
Epoch [2/50], Iter [55/418] Loss: 4.4978, average_loss: 5.2600
Epoch [2/50], Iter [60/418] Loss: 3.4646, average_loss: 5.1962
Epoch [2/50], Iter [65/418] Loss: 3.4163, average_loss: 5.1622
Epoch [2/50], Iter [70/418] Loss: 3.8462, average_loss: 5.1507
Epoch [2/50], Iter [75/418] Loss: 7.6274, average_loss: 5.1538
Epoch [2/50], Iter [80/418] Loss: 4.1728, average_loss: 5.2008
Epoch [2/50], Iter [85/418] Loss: 7.3047, average_loss: 5.2091
Epoch [2/50], Iter [90/418] Loss: 5.4104, average_loss: 5.1955
Epoch [2/50], Iter [95/418] Loss: 3.8042, average_loss: 5.1569
Epoch [2/50], Iter [100/418] Loss: 4.2784, average_loss: 5.1981
Epoch [2/50], Iter [105/418] Loss: 5.9579, average_loss: 5.2324
Epoch [2/50], Iter [110/418] Loss: 5.4476, average_loss: 5.2329
Epoch [2/50], Iter [115/418] Loss: 5.7542, average_loss: 5.2269
Epoch [2/50], Iter [120/418] Loss: 6.6710, average_loss: 5.2028
```

Epoch [2/50], Iter [125/418] Loss: 4.5498, average\_loss: 5.1913  
 Epoch [2/50], Iter [130/418] Loss: 4.3386, average\_loss: 5.1580  
 Epoch [2/50], Iter [135/418] Loss: 4.0954, average\_loss: 5.1688  
 Epoch [2/50], Iter [140/418] Loss: 4.9831, average\_loss: 5.1702  
 Epoch [2/50], Iter [145/418] Loss: 3.3183, average\_loss: 5.1886  
 Epoch [2/50], Iter [150/418] Loss: 4.3732, average\_loss: 5.1516  
 Epoch [2/50], Iter [155/418] Loss: 4.4929, average\_loss: 5.1236  
 Epoch [2/50], Iter [160/418] Loss: 4.2506, average\_loss: 5.1041  
 Epoch [2/50], Iter [165/418] Loss: 5.5738, average\_loss: 5.0876  
 Epoch [2/50], Iter [170/418] Loss: 3.6758, average\_loss: 5.1036  
 Epoch [2/50], Iter [175/418] Loss: 5.2121, average\_loss: 5.0775  
 Epoch [2/50], Iter [180/418] Loss: 4.2944, average\_loss: 5.0799  
 Epoch [2/50], Iter [185/418] Loss: 3.7959, average\_loss: 5.0716  
 Epoch [2/50], Iter [190/418] Loss: 4.7457, average\_loss: 5.0584  
 Epoch [2/50], Iter [195/418] Loss: 6.0598, average\_loss: 5.0518  
 Epoch [2/50], Iter [200/418] Loss: 3.7579, average\_loss: 5.0454  
 Epoch [2/50], Iter [205/418] Loss: 4.5074, average\_loss: 5.0259  
 Epoch [2/50], Iter [210/418] Loss: 3.7466, average\_loss: 5.0082  
 Epoch [2/50], Iter [215/418] Loss: 4.7306, average\_loss: 5.0065  
 Epoch [2/50], Iter [220/418] Loss: 3.7556, average\_loss: 4.9736  
 Epoch [2/50], Iter [225/418] Loss: 5.9313, average\_loss: 4.9653  
 Epoch [2/50], Iter [230/418] Loss: 3.5268, average\_loss: 4.9630  
 Epoch [2/50], Iter [235/418] Loss: 3.6641, average\_loss: 4.9517  
 Epoch [2/50], Iter [240/418] Loss: 5.4818, average\_loss: 4.9497  
 Epoch [2/50], Iter [245/418] Loss: 4.5367, average\_loss: 4.9525  
 Epoch [2/50], Iter [250/418] Loss: 4.1378, average\_loss: 4.9436  
 Epoch [2/50], Iter [255/418] Loss: 2.7276, average\_loss: 4.9315  
 Epoch [2/50], Iter [260/418] Loss: 6.9508, average\_loss: 4.9410  
 Epoch [2/50], Iter [265/418] Loss: 5.2118, average\_loss: 4.9345  
 Epoch [2/50], Iter [270/418] Loss: 3.3776, average\_loss: 4.9244  
 Epoch [2/50], Iter [275/418] Loss: 5.9175, average\_loss: 4.9224  
 Epoch [2/50], Iter [280/418] Loss: 4.6756, average\_loss: 4.9249  
 Epoch [2/50], Iter [285/418] Loss: 4.7694, average\_loss: 4.9468  
 Epoch [2/50], Iter [290/418] Loss: 4.8624, average\_loss: 4.9407  
 Epoch [2/50], Iter [295/418] Loss: 4.6086, average\_loss: 4.9593  
 Epoch [2/50], Iter [300/418] Loss: 3.0985, average\_loss: 4.9589  
 Epoch [2/50], Iter [305/418] Loss: 4.3594, average\_loss: 4.9565  
 Epoch [2/50], Iter [310/418] Loss: 4.3160, average\_loss: 4.9422  
 Epoch [2/50], Iter [315/418] Loss: 6.2389, average\_loss: 4.9455  
 Epoch [2/50], Iter [320/418] Loss: 4.1159, average\_loss: 4.9389  
 Epoch [2/50], Iter [325/418] Loss: 4.4818, average\_loss: 4.9358  
 Epoch [2/50], Iter [330/418] Loss: 4.9736, average\_loss: 4.9371  
 Epoch [2/50], Iter [335/418] Loss: 5.2352, average\_loss: 4.9374  
 Epoch [2/50], Iter [340/418] Loss: 5.0505, average\_loss: 4.9367  
 Epoch [2/50], Iter [345/418] Loss: 4.7794, average\_loss: 4.9293  
 Epoch [2/50], Iter [350/418] Loss: 4.8319, average\_loss: 4.9287  
 Epoch [2/50], Iter [355/418] Loss: 5.2039, average\_loss: 4.9217  
 Epoch [2/50], Iter [360/418] Loss: 4.6074, average\_loss: 4.9122



Epoch [2/50], Iter [365/418] Loss: 4.2440, average\_loss: 4.9011  
Epoch [2/50], Iter [370/418] Loss: 4.1434, average\_loss: 4.8982  
Epoch [2/50], Iter [375/418] Loss: 4.6889, average\_loss: 4.9057  
Epoch [2/50], Iter [380/418] Loss: 5.7131, average\_loss: 4.9177  
Epoch [2/50], Iter [385/418] Loss: 4.7884, average\_loss: 4.9169  
Epoch [2/50], Iter [390/418] Loss: 5.2637, average\_loss: 4.9233  
Epoch [2/50], Iter [395/418] Loss: 3.4131, average\_loss: 4.9243  
Epoch [2/50], Iter [400/418] Loss: 5.5294, average\_loss: 4.9184  
Epoch [2/50], Iter [405/418] Loss: 4.9867, average\_loss: 4.9119  
Epoch [2/50], Iter [410/418] Loss: 4.0073, average\_loss: 4.9175  
Epoch [2/50], Iter [415/418] Loss: 5.0359, average\_loss: 4.9251  
Updating best test loss: 4.75686

Starting epoch 3 / 50

Learning Rate for this epoch: 0.001

Epoch [3/50], Iter [5/418] Loss: 4.8947, average\_loss: 5.0581  
Epoch [3/50], Iter [10/418] Loss: 3.3663, average\_loss: 4.5439  
Epoch [3/50], Iter [15/418] Loss: 3.8455, average\_loss: 4.6186  
Epoch [3/50], Iter [20/418] Loss: 2.7314, average\_loss: 4.5933  
Epoch [3/50], Iter [25/418] Loss: 4.2652, average\_loss: 4.6327  
Epoch [3/50], Iter [30/418] Loss: 3.6630, average\_loss: 4.4771  
Epoch [3/50], Iter [35/418] Loss: 3.4360, average\_loss: 4.3947  
Epoch [3/50], Iter [40/418] Loss: 4.9706, average\_loss: 4.3956  
Epoch [3/50], Iter [45/418] Loss: 6.4689, average\_loss: 4.4620  
Epoch [3/50], Iter [50/418] Loss: 4.0476, average\_loss: 4.4170  
Epoch [3/50], Iter [55/418] Loss: 4.8072, average\_loss: 4.4719  
Epoch [3/50], Iter [60/418] Loss: 6.6298, average\_loss: 4.5049  
Epoch [3/50], Iter [65/418] Loss: 3.9821, average\_loss: 4.5134  
Epoch [3/50], Iter [70/418] Loss: 4.0751, average\_loss: 4.5108  
Epoch [3/50], Iter [75/418] Loss: 6.4554, average\_loss: 4.5244  
Epoch [3/50], Iter [80/418] Loss: 3.8403, average\_loss: 4.5193  
Epoch [3/50], Iter [85/418] Loss: 5.5722, average\_loss: 4.5489  
Epoch [3/50], Iter [90/418] Loss: 4.5498, average\_loss: 4.5631  
Epoch [3/50], Iter [95/418] Loss: 3.5351, average\_loss: 4.5848  
Epoch [3/50], Iter [100/418] Loss: 4.2219, average\_loss: 4.5880  
Epoch [3/50], Iter [105/418] Loss: 3.6961, average\_loss: 4.6022  
Epoch [3/50], Iter [110/418] Loss: 5.1949, average\_loss: 4.6247  
Epoch [3/50], Iter [115/418] Loss: 3.4578, average\_loss: 4.5763  
Epoch [3/50], Iter [120/418] Loss: 5.2035, average\_loss: 4.5660  
Epoch [3/50], Iter [125/418] Loss: 3.0378, average\_loss: 4.5679  
Epoch [3/50], Iter [130/418] Loss: 4.3389, average\_loss: 4.5568  
Epoch [3/50], Iter [135/418] Loss: 2.8437, average\_loss: 4.5413  
Epoch [3/50], Iter [140/418] Loss: 3.2526, average\_loss: 4.5428  
Epoch [3/50], Iter [145/418] Loss: 4.8177, average\_loss: 4.5529  
Epoch [3/50], Iter [150/418] Loss: 4.3235, average\_loss: 4.5303  
Epoch [3/50], Iter [155/418] Loss: 5.9535, average\_loss: 4.5399  
Epoch [3/50], Iter [160/418] Loss: 4.3890, average\_loss: 4.5478

Epoch [3/50], Iter [165/418] Loss: 4.5106, average\_loss: 4.5338  
 Epoch [3/50], Iter [170/418] Loss: 3.5480, average\_loss: 4.5248  
 Epoch [3/50], Iter [175/418] Loss: 6.5841, average\_loss: 4.5610  
 Epoch [3/50], Iter [180/418] Loss: 2.8220, average\_loss: 4.5391  
 Epoch [3/50], Iter [185/418] Loss: 4.7179, average\_loss: 4.5535  
 Epoch [3/50], Iter [190/418] Loss: 4.6823, average\_loss: 4.5434  
 Epoch [3/50], Iter [195/418] Loss: 3.8117, average\_loss: 4.5413  
 Epoch [3/50], Iter [200/418] Loss: 2.4689, average\_loss: 4.5328  
 Epoch [3/50], Iter [205/418] Loss: 3.4904, average\_loss: 4.5388  
 Epoch [3/50], Iter [210/418] Loss: 3.1522, average\_loss: 4.5266  
 Epoch [3/50], Iter [215/418] Loss: 6.2557, average\_loss: 4.5403  
 Epoch [3/50], Iter [220/418] Loss: 5.0686, average\_loss: 4.5384  
 Epoch [3/50], Iter [225/418] Loss: 4.2380, average\_loss: 4.5335  
 Epoch [3/50], Iter [230/418] Loss: 4.9889, average\_loss: 4.5242  
 Epoch [3/50], Iter [235/418] Loss: 3.9607, average\_loss: 4.4994  
 Epoch [3/50], Iter [240/418] Loss: 4.8723, average\_loss: 4.5105  
 Epoch [3/50], Iter [245/418] Loss: 4.4774, average\_loss: 4.4991  
 Epoch [3/50], Iter [250/418] Loss: 4.4203, average\_loss: 4.5025  
 Epoch [3/50], Iter [255/418] Loss: 5.1683, average\_loss: 4.4968  
 Epoch [3/50], Iter [260/418] Loss: 3.5783, average\_loss: 4.4993  
 Epoch [3/50], Iter [265/418] Loss: 5.7031, average\_loss: 4.4975  
 Epoch [3/50], Iter [270/418] Loss: 3.7435, average\_loss: 4.4979  
 Epoch [3/50], Iter [275/418] Loss: 5.0490, average\_loss: 4.4935  
 Epoch [3/50], Iter [280/418] Loss: 6.1323, average\_loss: 4.4930  
 Epoch [3/50], Iter [285/418] Loss: 4.0084, average\_loss: 4.4824  
 Epoch [3/50], Iter [290/418] Loss: 6.1183, average\_loss: 4.4905  
 Epoch [3/50], Iter [295/418] Loss: 4.3797, average\_loss: 4.4926  
 Epoch [3/50], Iter [300/418] Loss: 6.7559, average\_loss: 4.5016  
 Epoch [3/50], Iter [305/418] Loss: 4.0044, average\_loss: 4.4949  
 Epoch [3/50], Iter [310/418] Loss: 6.2225, average\_loss: 4.4950  
 Epoch [3/50], Iter [315/418] Loss: 6.2011, average\_loss: 4.5024  
 Epoch [3/50], Iter [320/418] Loss: 4.2821, average\_loss: 4.4998  
 Epoch [3/50], Iter [325/418] Loss: 3.2028, average\_loss: 4.4869  
 Epoch [3/50], Iter [330/418] Loss: 4.0142, average\_loss: 4.4862  
 Epoch [3/50], Iter [335/418] Loss: 5.1443, average\_loss: 4.5028  
 Epoch [3/50], Iter [340/418] Loss: 2.7178, average\_loss: 4.5011  
 Epoch [3/50], Iter [345/418] Loss: 4.6059, average\_loss: 4.4970  
 Epoch [3/50], Iter [350/418] Loss: 3.5484, average\_loss: 4.4950  
 Epoch [3/50], Iter [355/418] Loss: 4.0844, average\_loss: 4.4946  
 Epoch [3/50], Iter [360/418] Loss: 4.8222, average\_loss: 4.4969  
 Epoch [3/50], Iter [365/418] Loss: 5.4019, average\_loss: 4.4945  
 Epoch [3/50], Iter [370/418] Loss: 3.5346, average\_loss: 4.4858  
 Epoch [3/50], Iter [375/418] Loss: 3.7441, average\_loss: 4.4892  
 Epoch [3/50], Iter [380/418] Loss: 4.2662, average\_loss: 4.4829  
 Epoch [3/50], Iter [385/418] Loss: 2.9597, average\_loss: 4.4843  
 Epoch [3/50], Iter [390/418] Loss: 3.9765, average\_loss: 4.4798  
 Epoch [3/50], Iter [395/418] Loss: 5.5613, average\_loss: 4.4863  
 Epoch [3/50], Iter [400/418] Loss: 5.7379, average\_loss: 4.4852

Epoch [3/50], Iter [405/418] Loss: 5.4088, average\_loss: 4.4840  
Epoch [3/50], Iter [410/418] Loss: 5.7478, average\_loss: 4.4869  
Epoch [3/50], Iter [415/418] Loss: 6.0221, average\_loss: 4.4962  
Updating best test loss: 4.43784

Starting epoch 4 / 50

Learning Rate for this epoch: 0.001

Epoch [4/50], Iter [5/418] Loss: 3.3924, average\_loss: 4.2953  
Epoch [4/50], Iter [10/418] Loss: 6.2521, average\_loss: 4.8264  
Epoch [4/50], Iter [15/418] Loss: 3.0381, average\_loss: 4.6567  
Epoch [4/50], Iter [20/418] Loss: 5.1718, average\_loss: 4.4861  
Epoch [4/50], Iter [25/418] Loss: 3.9295, average\_loss: 4.3974  
Epoch [4/50], Iter [30/418] Loss: 4.1721, average\_loss: 4.4066  
Epoch [4/50], Iter [35/418] Loss: 5.7119, average\_loss: 4.4252  
Epoch [4/50], Iter [40/418] Loss: 3.9013, average\_loss: 4.3748  
Epoch [4/50], Iter [45/418] Loss: 4.7051, average\_loss: 4.3671  
Epoch [4/50], Iter [50/418] Loss: 4.3924, average\_loss: 4.3143  
Epoch [4/50], Iter [55/418] Loss: 6.4730, average\_loss: 4.3430  
Epoch [4/50], Iter [60/418] Loss: 4.0047, average\_loss: 4.2864  
Epoch [4/50], Iter [65/418] Loss: 3.3108, average\_loss: 4.2122  
Epoch [4/50], Iter [70/418] Loss: 3.9938, average\_loss: 4.2906  
Epoch [4/50], Iter [75/418] Loss: 3.7658, average\_loss: 4.2522  
Epoch [4/50], Iter [80/418] Loss: 3.1635, average\_loss: 4.2922  
Epoch [4/50], Iter [85/418] Loss: 4.3455, average\_loss: 4.2889  
Epoch [4/50], Iter [90/418] Loss: 2.3943, average\_loss: 4.2364  
Epoch [4/50], Iter [95/418] Loss: 4.1075, average\_loss: 4.2303  
Epoch [4/50], Iter [100/418] Loss: 4.2950, average\_loss: 4.2226  
Epoch [4/50], Iter [105/418] Loss: 4.5657, average\_loss: 4.2183  
Epoch [4/50], Iter [110/418] Loss: 3.7820, average\_loss: 4.2097  
Epoch [4/50], Iter [115/418] Loss: 3.8193, average\_loss: 4.2089  
Epoch [4/50], Iter [120/418] Loss: 3.6549, average\_loss: 4.1962  
Epoch [4/50], Iter [125/418] Loss: 3.5248, average\_loss: 4.1840  
Epoch [4/50], Iter [130/418] Loss: 3.0291, average\_loss: 4.1611  
Epoch [4/50], Iter [135/418] Loss: 5.7518, average\_loss: 4.1717  
Epoch [4/50], Iter [140/418] Loss: 3.6033, average\_loss: 4.1608  
Epoch [4/50], Iter [145/418] Loss: 4.5699, average\_loss: 4.1550  
Epoch [4/50], Iter [150/418] Loss: 3.5021, average\_loss: 4.1569  
Epoch [4/50], Iter [155/418] Loss: 3.6033, average\_loss: 4.1689  
Epoch [4/50], Iter [160/418] Loss: 4.6002, average\_loss: 4.1681  
Epoch [4/50], Iter [165/418] Loss: 4.2788, average\_loss: 4.1641  
Epoch [4/50], Iter [170/418] Loss: 4.9341, average\_loss: 4.1456  
Epoch [4/50], Iter [175/418] Loss: 4.3805, average\_loss: 4.1528  
Epoch [4/50], Iter [180/418] Loss: 3.8663, average\_loss: 4.1707  
Epoch [4/50], Iter [185/418] Loss: 3.9781, average\_loss: 4.1672  
Epoch [4/50], Iter [190/418] Loss: 3.9161, average\_loss: 4.1606  
Epoch [4/50], Iter [195/418] Loss: 2.6032, average\_loss: 4.1543  
Epoch [4/50], Iter [200/418] Loss: 5.0809, average\_loss: 4.1481

Epoch [4/50], Iter [205/418] Loss: 3.8375, average\_loss: 4.1390  
Epoch [4/50], Iter [210/418] Loss: 3.7821, average\_loss: 4.1625  
Epoch [4/50], Iter [215/418] Loss: 4.3375, average\_loss: 4.1554  
Epoch [4/50], Iter [220/418] Loss: 5.0243, average\_loss: 4.1649  
Epoch [4/50], Iter [225/418] Loss: 3.9703, average\_loss: 4.1706  
Epoch [4/50], Iter [230/418] Loss: 3.6219, average\_loss: 4.1841  
Epoch [4/50], Iter [235/418] Loss: 3.9170, average\_loss: 4.1771  
Epoch [4/50], Iter [240/418] Loss: 3.3091, average\_loss: 4.1726  
Epoch [4/50], Iter [245/418] Loss: 3.3877, average\_loss: 4.1816  
Epoch [4/50], Iter [250/418] Loss: 3.6173, average\_loss: 4.1801  
Epoch [4/50], Iter [255/418] Loss: 5.8991, average\_loss: 4.1769  
Epoch [4/50], Iter [260/418] Loss: 2.4833, average\_loss: 4.1715  
Epoch [4/50], Iter [265/418] Loss: 3.0232, average\_loss: 4.1636  
Epoch [4/50], Iter [270/418] Loss: 3.0167, average\_loss: 4.1534  
Epoch [4/50], Iter [275/418] Loss: 3.8975, average\_loss: 4.1592  
Epoch [4/50], Iter [280/418] Loss: 6.0816, average\_loss: 4.1746  
Epoch [4/50], Iter [285/418] Loss: 5.0253, average\_loss: 4.2070  
Epoch [4/50], Iter [290/418] Loss: 4.8622, average\_loss: 4.2076  
Epoch [4/50], Iter [295/418] Loss: 3.3859, average\_loss: 4.2192  
Epoch [4/50], Iter [300/418] Loss: 3.4537, average\_loss: 4.2214  
Epoch [4/50], Iter [305/418] Loss: 3.7226, average\_loss: 4.2127  
Epoch [4/50], Iter [310/418] Loss: 5.2983, average\_loss: 4.2241  
Epoch [4/50], Iter [315/418] Loss: 3.2822, average\_loss: 4.2310  
Epoch [4/50], Iter [320/418] Loss: 4.2872, average\_loss: 4.2433  
Epoch [4/50], Iter [325/418] Loss: 4.0218, average\_loss: 4.2368  
Epoch [4/50], Iter [330/418] Loss: 3.1365, average\_loss: 4.2341  
Epoch [4/50], Iter [335/418] Loss: 3.5282, average\_loss: 4.2367  
Epoch [4/50], Iter [340/418] Loss: 4.4717, average\_loss: 4.2350  
Epoch [4/50], Iter [345/418] Loss: 3.7488, average\_loss: 4.2293  
Epoch [4/50], Iter [350/418] Loss: 3.1015, average\_loss: 4.2240  
Epoch [4/50], Iter [355/418] Loss: 5.8859, average\_loss: 4.2171  
Epoch [4/50], Iter [360/418] Loss: 3.5438, average\_loss: 4.2183  
Epoch [4/50], Iter [365/418] Loss: 7.1728, average\_loss: 4.2301  
Epoch [4/50], Iter [370/418] Loss: 3.7846, average\_loss: 4.2272  
Epoch [4/50], Iter [375/418] Loss: 2.9472, average\_loss: 4.2243  
Epoch [4/50], Iter [380/418] Loss: 3.5004, average\_loss: 4.2260  
Epoch [4/50], Iter [385/418] Loss: 6.8226, average\_loss: 4.2320  
Epoch [4/50], Iter [390/418] Loss: 5.5190, average\_loss: 4.2322  
Epoch [4/50], Iter [395/418] Loss: 3.4931, average\_loss: 4.2334  
Epoch [4/50], Iter [400/418] Loss: 3.3819, average\_loss: 4.2286  
Epoch [4/50], Iter [405/418] Loss: 4.9963, average\_loss: 4.2298  
Epoch [4/50], Iter [410/418] Loss: 3.2393, average\_loss: 4.2218  
Epoch [4/50], Iter [415/418] Loss: 3.7674, average\_loss: 4.2144  
Updating best test loss: 4.17308

Starting epoch 5 / 50

Learning Rate for this epoch: 0.001

Epoch [5/50], Iter [5/418] Loss: 4.7422, average\_loss: 3.9745  
 Epoch [5/50], Iter [10/418] Loss: 3.3281, average\_loss: 3.7759  
 Epoch [5/50], Iter [15/418] Loss: 3.1488, average\_loss: 3.8911  
 Epoch [5/50], Iter [20/418] Loss: 2.9334, average\_loss: 3.8955  
 Epoch [5/50], Iter [25/418] Loss: 6.6622, average\_loss: 4.0962  
 Epoch [5/50], Iter [30/418] Loss: 3.8781, average\_loss: 4.1627  
 Epoch [5/50], Iter [35/418] Loss: 3.9290, average\_loss: 4.2086  
 Epoch [5/50], Iter [40/418] Loss: 3.4286, average\_loss: 4.1891  
 Epoch [5/50], Iter [45/418] Loss: 3.6141, average\_loss: 4.1135  
 Epoch [5/50], Iter [50/418] Loss: 4.6603, average\_loss: 4.1715  
 Epoch [5/50], Iter [55/418] Loss: 2.5400, average\_loss: 4.1094  
 Epoch [5/50], Iter [60/418] Loss: 5.8002, average\_loss: 4.0971  
 Epoch [5/50], Iter [65/418] Loss: 4.1521, average\_loss: 4.0621  
 Epoch [5/50], Iter [70/418] Loss: 7.0882, average\_loss: 4.1195  
 Epoch [5/50], Iter [75/418] Loss: 3.2061, average\_loss: 4.1030  
 Epoch [5/50], Iter [80/418] Loss: 3.1955, average\_loss: 4.1918  
 Epoch [5/50], Iter [85/418] Loss: 3.7258, average\_loss: 4.1769  
 Epoch [5/50], Iter [90/418] Loss: 3.0156, average\_loss: 4.1490  
 Epoch [5/50], Iter [95/418] Loss: 2.6724, average\_loss: 4.1130  
 Epoch [5/50], Iter [100/418] Loss: 4.7242, average\_loss: 4.1000  
 Epoch [5/50], Iter [105/418] Loss: 5.4113, average\_loss: 4.0917  
 Epoch [5/50], Iter [110/418] Loss: 3.8149, average\_loss: 4.1227  
 Epoch [5/50], Iter [115/418] Loss: 5.0161, average\_loss: 4.1011  
 Epoch [5/50], Iter [120/418] Loss: 4.1843, average\_loss: 4.1048  
 Epoch [5/50], Iter [125/418] Loss: 4.0225, average\_loss: 4.1017  
 Epoch [5/50], Iter [130/418] Loss: 3.3394, average\_loss: 4.0800  
 Epoch [5/50], Iter [135/418] Loss: 4.1498, average\_loss: 4.0715  
 Epoch [5/50], Iter [140/418] Loss: 4.3600, average\_loss: 4.0769  
 Epoch [5/50], Iter [145/418] Loss: 4.4606, average\_loss: 4.0807  
 Epoch [5/50], Iter [150/418] Loss: 4.5628, average\_loss: 4.0742  
 Epoch [5/50], Iter [155/418] Loss: 3.3744, average\_loss: 4.0662  
 Epoch [5/50], Iter [160/418] Loss: 3.4271, average\_loss: 4.0698  
 Epoch [5/50], Iter [165/418] Loss: 5.8323, average\_loss: 4.0828  
 Epoch [5/50], Iter [170/418] Loss: 3.8985, average\_loss: 4.0686  
 Epoch [5/50], Iter [175/418] Loss: 5.2888, average\_loss: 4.0685  
 Epoch [5/50], Iter [180/418] Loss: 4.5248, average\_loss: 4.0578  
 Epoch [5/50], Iter [185/418] Loss: 3.5328, average\_loss: 4.0607  
 Epoch [5/50], Iter [190/418] Loss: 2.9064, average\_loss: 4.0409  
 Epoch [5/50], Iter [195/418] Loss: 4.5245, average\_loss: 4.0409  
 Epoch [5/50], Iter [200/418] Loss: 3.2935, average\_loss: 4.0257  
 Epoch [5/50], Iter [205/418] Loss: 5.2697, average\_loss: 4.0297  
 Epoch [5/50], Iter [210/418] Loss: 3.0027, average\_loss: 4.0321  
 Epoch [5/50], Iter [215/418] Loss: 3.4441, average\_loss: 4.0237  
 Epoch [5/50], Iter [220/418] Loss: 5.3664, average\_loss: 4.0195  
 Epoch [5/50], Iter [225/418] Loss: 5.1483, average\_loss: 4.0213  
 Epoch [5/50], Iter [230/418] Loss: 4.1753, average\_loss: 4.0338  
 Epoch [5/50], Iter [235/418] Loss: 3.8689, average\_loss: 4.0380  
 Epoch [5/50], Iter [240/418] Loss: 4.9598, average\_loss: 4.0359

Epoch [5/50], Iter [245/418] Loss: 3.4172, average\_loss: 4.0223  
 Epoch [5/50], Iter [250/418] Loss: 2.7510, average\_loss: 4.0216  
 Epoch [5/50], Iter [255/418] Loss: 4.4618, average\_loss: 4.0151  
 Epoch [5/50], Iter [260/418] Loss: 3.0374, average\_loss: 4.0094  
 Epoch [5/50], Iter [265/418] Loss: 3.6170, average\_loss: 3.9963  
 Epoch [5/50], Iter [270/418] Loss: 3.0121, average\_loss: 4.0001  
 Epoch [5/50], Iter [275/418] Loss: 3.1459, average\_loss: 4.0008  
 Epoch [5/50], Iter [280/418] Loss: 3.4635, average\_loss: 3.9900  
 Epoch [5/50], Iter [285/418] Loss: 4.4815, average\_loss: 3.9851  
 Epoch [5/50], Iter [290/418] Loss: 5.8299, average\_loss: 3.9965  
 Epoch [5/50], Iter [295/418] Loss: 2.2768, average\_loss: 3.9861  
 Epoch [5/50], Iter [300/418] Loss: 4.9461, average\_loss: 3.9881  
 Epoch [5/50], Iter [305/418] Loss: 3.1341, average\_loss: 3.9860  
 Epoch [5/50], Iter [310/418] Loss: 7.2617, average\_loss: 3.9936  
 Epoch [5/50], Iter [315/418] Loss: 4.5063, average\_loss: 3.9950  
 Epoch [5/50], Iter [320/418] Loss: 4.1007, average\_loss: 3.9896  
 Epoch [5/50], Iter [325/418] Loss: 4.2460, average\_loss: 3.9905  
 Epoch [5/50], Iter [330/418] Loss: 5.8714, average\_loss: 3.9975  
 Epoch [5/50], Iter [335/418] Loss: 5.0914, average\_loss: 4.0004  
 Epoch [5/50], Iter [340/418] Loss: 3.4326, average\_loss: 3.9961  
 Epoch [5/50], Iter [345/418] Loss: 3.9582, average\_loss: 3.9908  
 Epoch [5/50], Iter [350/418] Loss: 2.8096, average\_loss: 3.9850  
 Epoch [5/50], Iter [355/418] Loss: 2.5230, average\_loss: 3.9747  
 Epoch [5/50], Iter [360/418] Loss: 3.4572, average\_loss: 3.9796  
 Epoch [5/50], Iter [365/418] Loss: 2.6763, average\_loss: 3.9741  
 Epoch [5/50], Iter [370/418] Loss: 2.3680, average\_loss: 3.9807  
 Epoch [5/50], Iter [375/418] Loss: 6.0933, average\_loss: 3.9828  
 Epoch [5/50], Iter [380/418] Loss: 3.6219, average\_loss: 3.9836  
 Epoch [5/50], Iter [385/418] Loss: 5.0446, average\_loss: 3.9876  
 Epoch [5/50], Iter [390/418] Loss: 2.7486, average\_loss: 3.9799  
 Epoch [5/50], Iter [395/418] Loss: 2.5927, average\_loss: 3.9794  
 Epoch [5/50], Iter [400/418] Loss: 4.2610, average\_loss: 3.9752  
 Epoch [5/50], Iter [405/418] Loss: 5.5610, average\_loss: 3.9769  
 Epoch [5/50], Iter [410/418] Loss: 3.1277, average\_loss: 3.9802  
 Epoch [5/50], Iter [415/418] Loss: 3.4316, average\_loss: 3.9916  
 Updating best test loss: 3.96088

Starting epoch 6 / 50

Learning Rate for this epoch: 0.001

Epoch [6/50], Iter [5/418] Loss: 2.9321, average\_loss: 3.6550  
 Epoch [6/50], Iter [10/418] Loss: 2.3634, average\_loss: 3.7110  
 Epoch [6/50], Iter [15/418] Loss: 2.8976, average\_loss: 3.9155  
 Epoch [6/50], Iter [20/418] Loss: 4.5263, average\_loss: 3.8425  
 Epoch [6/50], Iter [25/418] Loss: 3.0336, average\_loss: 3.7146  
 Epoch [6/50], Iter [30/418] Loss: 5.2173, average\_loss: 3.7781  
 Epoch [6/50], Iter [35/418] Loss: 2.6839, average\_loss: 3.7052  
 Epoch [6/50], Iter [40/418] Loss: 3.9307, average\_loss: 3.7454

Epoch [6/50], Iter [45/418] Loss: 3.8495, average\_loss: 3.8104  
 Epoch [6/50], Iter [50/418] Loss: 4.0241, average\_loss: 3.7871  
 Epoch [6/50], Iter [55/418] Loss: 5.1945, average\_loss: 3.8390  
 Epoch [6/50], Iter [60/418] Loss: 2.6688, average\_loss: 3.7809  
 Epoch [6/50], Iter [65/418] Loss: 3.2902, average\_loss: 3.7393  
 Epoch [6/50], Iter [70/418] Loss: 2.9758, average\_loss: 3.7340  
 Epoch [6/50], Iter [75/418] Loss: 3.7028, average\_loss: 3.7223  
 Epoch [6/50], Iter [80/418] Loss: 2.8505, average\_loss: 3.7378  
 Epoch [6/50], Iter [85/418] Loss: 5.7766, average\_loss: 3.7520  
 Epoch [6/50], Iter [90/418] Loss: 3.3336, average\_loss: 3.8137  
 Epoch [6/50], Iter [95/418] Loss: 3.7999, average\_loss: 3.7970  
 Epoch [6/50], Iter [100/418] Loss: 3.7877, average\_loss: 3.7961  
 Epoch [6/50], Iter [105/418] Loss: 4.4689, average\_loss: 3.7921  
 Epoch [6/50], Iter [110/418] Loss: 3.9965, average\_loss: 3.7689  
 Epoch [6/50], Iter [115/418] Loss: 5.1509, average\_loss: 3.7709  
 Epoch [6/50], Iter [120/418] Loss: 5.0294, average\_loss: 3.7786  
 Epoch [6/50], Iter [125/418] Loss: 3.2128, average\_loss: 3.7602  
 Epoch [6/50], Iter [130/418] Loss: 2.9278, average\_loss: 3.7808  
 Epoch [6/50], Iter [135/418] Loss: 3.5606, average\_loss: 3.8069  
 Epoch [6/50], Iter [140/418] Loss: 3.5022, average\_loss: 3.8205  
 Epoch [6/50], Iter [145/418] Loss: 2.9824, average\_loss: 3.8017  
 Epoch [6/50], Iter [150/418] Loss: 3.4994, average\_loss: 3.8038  
 Epoch [6/50], Iter [155/418] Loss: 2.7455, average\_loss: 3.7997  
 Epoch [6/50], Iter [160/418] Loss: 3.1955, average\_loss: 3.7915  
 Epoch [6/50], Iter [165/418] Loss: 2.9819, average\_loss: 3.7745  
 Epoch [6/50], Iter [170/418] Loss: 4.9026, average\_loss: 3.7623  
 Epoch [6/50], Iter [175/418] Loss: 3.6340, average\_loss: 3.7560  
 Epoch [6/50], Iter [180/418] Loss: 5.3586, average\_loss: 3.7590  
 Epoch [6/50], Iter [185/418] Loss: 3.5363, average\_loss: 3.7494  
 Epoch [6/50], Iter [190/418] Loss: 2.7080, average\_loss: 3.7310  
 Epoch [6/50], Iter [195/418] Loss: 2.7576, average\_loss: 3.7421  
 Epoch [6/50], Iter [200/418] Loss: 3.6262, average\_loss: 3.7420  
 Epoch [6/50], Iter [205/418] Loss: 3.0959, average\_loss: 3.7191  
 Epoch [6/50], Iter [210/418] Loss: 3.7525, average\_loss: 3.7130  
 Epoch [6/50], Iter [215/418] Loss: 6.1116, average\_loss: 3.7197  
 Epoch [6/50], Iter [220/418] Loss: 4.0766, average\_loss: 3.7288  
 Epoch [6/50], Iter [225/418] Loss: 2.6831, average\_loss: 3.7229  
 Epoch [6/50], Iter [230/418] Loss: 2.7750, average\_loss: 3.7180  
 Epoch [6/50], Iter [235/418] Loss: 3.2732, average\_loss: 3.7178  
 Epoch [6/50], Iter [240/418] Loss: 2.4362, average\_loss: 3.7167  
 Epoch [6/50], Iter [245/418] Loss: 4.7043, average\_loss: 3.7128  
 Epoch [6/50], Iter [250/418] Loss: 3.7120, average\_loss: 3.7222  
 Epoch [6/50], Iter [255/418] Loss: 2.7488, average\_loss: 3.7188  
 Epoch [6/50], Iter [260/418] Loss: 3.1909, average\_loss: 3.7155  
 Epoch [6/50], Iter [265/418] Loss: 1.8832, average\_loss: 3.6951  
 Epoch [6/50], Iter [270/418] Loss: 3.6960, average\_loss: 3.7015  
 Epoch [6/50], Iter [275/418] Loss: 4.3742, average\_loss: 3.7027  
 Epoch [6/50], Iter [280/418] Loss: 4.3300, average\_loss: 3.7236

```

Epoch [6/50], Iter [285/418] Loss: 6.6676, average_loss: 3.7507
Epoch [6/50], Iter [290/418] Loss: 4.3078, average_loss: 3.7608
Epoch [6/50], Iter [295/418] Loss: 3.0237, average_loss: 3.7500
Epoch [6/50], Iter [300/418] Loss: 3.4845, average_loss: 3.7521
Epoch [6/50], Iter [305/418] Loss: 2.7085, average_loss: 3.7470
Epoch [6/50], Iter [310/418] Loss: 3.0704, average_loss: 3.7459
Epoch [6/50], Iter [315/418] Loss: 5.7006, average_loss: 3.7504
Epoch [6/50], Iter [320/418] Loss: 2.9039, average_loss: 3.7375
Epoch [6/50], Iter [325/418] Loss: 3.5037, average_loss: 3.7322
Epoch [6/50], Iter [330/418] Loss: 3.6443, average_loss: 3.7248
Epoch [6/50], Iter [335/418] Loss: 3.6813, average_loss: 3.7288
Epoch [6/50], Iter [340/418] Loss: 3.9912, average_loss: 3.7346
Epoch [6/50], Iter [345/418] Loss: 2.9188, average_loss: 3.7263
Epoch [6/50], Iter [350/418] Loss: 5.0841, average_loss: 3.7273
Epoch [6/50], Iter [355/418] Loss: 4.1703, average_loss: 3.7304
Epoch [6/50], Iter [360/418] Loss: 2.7508, average_loss: 3.7258
Epoch [6/50], Iter [365/418] Loss: 2.3774, average_loss: 3.7214
Epoch [6/50], Iter [370/418] Loss: 2.4645, average_loss: 3.7181
Epoch [6/50], Iter [375/418] Loss: 4.8040, average_loss: 3.7193
Epoch [6/50], Iter [380/418] Loss: 3.8748, average_loss: 3.7321
Epoch [6/50], Iter [385/418] Loss: 4.1830, average_loss: 3.7316
Epoch [6/50], Iter [390/418] Loss: 3.7409, average_loss: 3.7293
Epoch [6/50], Iter [395/418] Loss: 3.1879, average_loss: 3.7329
Epoch [6/50], Iter [400/418] Loss: 2.9941, average_loss: 3.7266
Epoch [6/50], Iter [405/418] Loss: 3.1734, average_loss: 3.7306
Epoch [6/50], Iter [410/418] Loss: 3.9176, average_loss: 3.7249
Epoch [6/50], Iter [415/418] Loss: 2.9521, average_loss: 3.7259

```

---Evaluate model on test samples---

100%|

| 4950/4950 [02:26<00:00, 33.68it/s]

```

---class aeroplane ap 0.15677445321987085---
---class bicycle ap 0.13549403809197724---
---class bird ap 0.023484656250711153---
---class boat ap 0.0--- (no predictions for this class)
---class bottle ap 0.010457710171351269---
---class bus ap 0.06287979593809999---
---class car ap 0.3196125223822618---
---class cat ap 0.10284463089006092---
---class chair ap 0.061336566931985095---
---class cow ap 0.011270491803278689---
---class diningtable ap 0.0--- (no predictions for this class)
---class dog ap 0.005112474437627812---
---class horse ap 0.11084328502102064---
---class motorbike ap 0.020384615384615383---
---class person ap 0.21305105875684155---
---class pottedplant ap 0.015939401066282267---
---class sheep ap 0.03926462040240236---

```



---class sofa ap 0.0--- (no predictions for this class)  
---class train ap 0.2684100173468424---  
---class tvmonitor ap 0.10459685076236552---  
---map 0.08308785944287975---  
Updating best test loss: 3.81385

Starting epoch 7 / 50

Learning Rate for this epoch: 0.001

Epoch [7/50], Iter [5/418] Loss: 3.9523, average\_loss: 3.5317  
Epoch [7/50], Iter [10/418] Loss: 2.3123, average\_loss: 3.5288  
Epoch [7/50], Iter [15/418] Loss: 2.5208, average\_loss: 3.5099  
Epoch [7/50], Iter [20/418] Loss: 3.7546, average\_loss: 3.4626  
Epoch [7/50], Iter [25/418] Loss: 3.8122, average\_loss: 3.4527  
Epoch [7/50], Iter [30/418] Loss: 3.6851, average\_loss: 3.4113  
Epoch [7/50], Iter [35/418] Loss: 3.7090, average\_loss: 3.4932  
Epoch [7/50], Iter [40/418] Loss: 3.8397, average\_loss: 3.5767  
Epoch [7/50], Iter [45/418] Loss: 3.6480, average\_loss: 3.5509  
Epoch [7/50], Iter [50/418] Loss: 2.9084, average\_loss: 3.5170  
Epoch [7/50], Iter [55/418] Loss: 2.9744, average\_loss: 3.5058  
Epoch [7/50], Iter [60/418] Loss: 4.8594, average\_loss: 3.5641  
Epoch [7/50], Iter [65/418] Loss: 3.8989, average\_loss: 3.5841  
Epoch [7/50], Iter [70/418] Loss: 4.4708, average\_loss: 3.6182  
Epoch [7/50], Iter [75/418] Loss: 3.2388, average\_loss: 3.5880  
Epoch [7/50], Iter [80/418] Loss: 2.9615, average\_loss: 3.5405  
Epoch [7/50], Iter [85/418] Loss: 3.5567, average\_loss: 3.5482  
Epoch [7/50], Iter [90/418] Loss: 5.5973, average\_loss: 3.5534  
Epoch [7/50], Iter [95/418] Loss: 3.3743, average\_loss: 3.5546  
Epoch [7/50], Iter [100/418] Loss: 5.8001, average\_loss: 3.5796  
Epoch [7/50], Iter [105/418] Loss: 3.5560, average\_loss: 3.5769  
Epoch [7/50], Iter [110/418] Loss: 4.2260, average\_loss: 3.5822  
Epoch [7/50], Iter [115/418] Loss: 2.8897, average\_loss: 3.5894  
Epoch [7/50], Iter [120/418] Loss: 2.6502, average\_loss: 3.5714  
Epoch [7/50], Iter [125/418] Loss: 3.3721, average\_loss: 3.5647  
Epoch [7/50], Iter [130/418] Loss: 4.0179, average\_loss: 3.5547  
Epoch [7/50], Iter [135/418] Loss: 4.2352, average\_loss: 3.5870  
Epoch [7/50], Iter [140/418] Loss: 6.2313, average\_loss: 3.5940  
Epoch [7/50], Iter [145/418] Loss: 2.6485, average\_loss: 3.5847  
Epoch [7/50], Iter [150/418] Loss: 4.2781, average\_loss: 3.5757  
Epoch [7/50], Iter [155/418] Loss: 3.0615, average\_loss: 3.5699  
Epoch [7/50], Iter [160/418] Loss: 5.6434, average\_loss: 3.5692  
Epoch [7/50], Iter [165/418] Loss: 3.2592, average\_loss: 3.5685  
Epoch [7/50], Iter [170/418] Loss: 3.4785, average\_loss: 3.5747  
Epoch [7/50], Iter [175/418] Loss: 3.4235, average\_loss: 3.5764  
Epoch [7/50], Iter [180/418] Loss: 7.3947, average\_loss: 3.5887  
Epoch [7/50], Iter [185/418] Loss: 3.2257, average\_loss: 3.5919  
Epoch [7/50], Iter [190/418] Loss: 4.0447, average\_loss: 3.5912  
Epoch [7/50], Iter [195/418] Loss: 3.2784, average\_loss: 3.5799

Epoch [7/50], Iter [200/418] Loss: 2.2044, average\_loss: 3.5673  
Epoch [7/50], Iter [205/418] Loss: 2.9139, average\_loss: 3.5586  
Epoch [7/50], Iter [210/418] Loss: 3.3142, average\_loss: 3.5644  
Epoch [7/50], Iter [215/418] Loss: 3.7091, average\_loss: 3.5577  
Epoch [7/50], Iter [220/418] Loss: 3.1889, average\_loss: 3.5567  
Epoch [7/50], Iter [225/418] Loss: 2.9927, average\_loss: 3.5652  
Epoch [7/50], Iter [230/418] Loss: 3.7423, average\_loss: 3.5630  
Epoch [7/50], Iter [235/418] Loss: 4.0079, average\_loss: 3.5666  
Epoch [7/50], Iter [240/418] Loss: 3.9825, average\_loss: 3.5766  
Epoch [7/50], Iter [245/418] Loss: 4.4558, average\_loss: 3.5799  
Epoch [7/50], Iter [250/418] Loss: 4.1440, average\_loss: 3.5860  
Epoch [7/50], Iter [255/418] Loss: 3.3066, average\_loss: 3.5839  
Epoch [7/50], Iter [260/418] Loss: 2.9551, average\_loss: 3.5821  
Epoch [7/50], Iter [265/418] Loss: 2.7438, average\_loss: 3.5837  
Epoch [7/50], Iter [270/418] Loss: 2.7198, average\_loss: 3.5807  
Epoch [7/50], Iter [275/418] Loss: 3.9297, average\_loss: 3.5971  
Epoch [7/50], Iter [280/418] Loss: 2.7841, average\_loss: 3.6001  
Epoch [7/50], Iter [285/418] Loss: 4.1227, average\_loss: 3.6022  
Epoch [7/50], Iter [290/418] Loss: 3.1810, average\_loss: 3.5987  
Epoch [7/50], Iter [295/418] Loss: 2.9860, average\_loss: 3.5953  
Epoch [7/50], Iter [300/418] Loss: 4.0874, average\_loss: 3.5927  
Epoch [7/50], Iter [305/418] Loss: 2.9914, average\_loss: 3.5857  
Epoch [7/50], Iter [310/418] Loss: 3.2779, average\_loss: 3.5788  
Epoch [7/50], Iter [315/418] Loss: 2.1585, average\_loss: 3.5745  
Epoch [7/50], Iter [320/418] Loss: 3.1141, average\_loss: 3.5736  
Epoch [7/50], Iter [325/418] Loss: 2.9726, average\_loss: 3.5764  
Epoch [7/50], Iter [330/418] Loss: 3.8117, average\_loss: 3.5769  
Epoch [7/50], Iter [335/418] Loss: 2.1542, average\_loss: 3.5676  
Epoch [7/50], Iter [340/418] Loss: 2.8802, average\_loss: 3.5586  
Epoch [7/50], Iter [345/418] Loss: 2.0742, average\_loss: 3.5621  
Epoch [7/50], Iter [350/418] Loss: 3.7069, average\_loss: 3.5580  
Epoch [7/50], Iter [355/418] Loss: 3.6472, average\_loss: 3.5565  
Epoch [7/50], Iter [360/418] Loss: 4.1252, average\_loss: 3.5607  
Epoch [7/50], Iter [365/418] Loss: 4.0683, average\_loss: 3.5647  
Epoch [7/50], Iter [370/418] Loss: 3.7148, average\_loss: 3.5664  
Epoch [7/50], Iter [375/418] Loss: 3.1577, average\_loss: 3.5652  
Epoch [7/50], Iter [380/418] Loss: 3.4290, average\_loss: 3.5639  
Epoch [7/50], Iter [385/418] Loss: 3.1490, average\_loss: 3.5588  
Epoch [7/50], Iter [390/418] Loss: 3.3023, average\_loss: 3.5566  
Epoch [7/50], Iter [395/418] Loss: 3.2771, average\_loss: 3.5509  
Epoch [7/50], Iter [400/418] Loss: 3.3339, average\_loss: 3.5512  
Epoch [7/50], Iter [405/418] Loss: 4.2080, average\_loss: 3.5579  
Epoch [7/50], Iter [410/418] Loss: 2.6700, average\_loss: 3.5587  
Epoch [7/50], Iter [415/418] Loss: 4.1498, average\_loss: 3.5598  
Updating best test loss: 3.64272

Starting epoch 8 / 50

Learning Rate for this epoch: 0.001

Epoch [8/50], Iter [5/418] Loss: 3.3756, average\_loss: 3.4890  
Epoch [8/50], Iter [10/418] Loss: 6.3712, average\_loss: 3.7994  
Epoch [8/50], Iter [15/418] Loss: 2.7276, average\_loss: 3.5128  
Epoch [8/50], Iter [20/418] Loss: 3.0743, average\_loss: 3.4508  
Epoch [8/50], Iter [25/418] Loss: 2.5762, average\_loss: 3.3890  
Epoch [8/50], Iter [30/418] Loss: 2.0981, average\_loss: 3.3478  
Epoch [8/50], Iter [35/418] Loss: 5.3108, average\_loss: 3.3659  
Epoch [8/50], Iter [40/418] Loss: 2.4509, average\_loss: 3.4176  
Epoch [8/50], Iter [45/418] Loss: 3.3995, average\_loss: 3.4275  
Epoch [8/50], Iter [50/418] Loss: 4.4262, average\_loss: 3.4983  
Epoch [8/50], Iter [55/418] Loss: 3.7498, average\_loss: 3.4817  
Epoch [8/50], Iter [60/418] Loss: 2.3334, average\_loss: 3.4951  
Epoch [8/50], Iter [65/418] Loss: 3.1874, average\_loss: 3.4791  
Epoch [8/50], Iter [70/418] Loss: 4.6106, average\_loss: 3.4886  
Epoch [8/50], Iter [75/418] Loss: 4.7945, average\_loss: 3.5888  
Epoch [8/50], Iter [80/418] Loss: 3.0432, average\_loss: 3.5644  
Epoch [8/50], Iter [85/418] Loss: 4.2934, average\_loss: 3.5780  
Epoch [8/50], Iter [90/418] Loss: 3.9482, average\_loss: 3.5757  
Epoch [8/50], Iter [95/418] Loss: 3.2802, average\_loss: 3.5490  
Epoch [8/50], Iter [100/418] Loss: 2.7950, average\_loss: 3.5333  
Epoch [8/50], Iter [105/418] Loss: 3.7068, average\_loss: 3.5398  
Epoch [8/50], Iter [110/418] Loss: 4.8146, average\_loss: 3.5510  
Epoch [8/50], Iter [115/418] Loss: 2.4868, average\_loss: 3.5355  
Epoch [8/50], Iter [120/418] Loss: 3.2179, average\_loss: 3.5233  
Epoch [8/50], Iter [125/418] Loss: 4.0257, average\_loss: 3.5089  
Epoch [8/50], Iter [130/418] Loss: 3.9451, average\_loss: 3.5092  
Epoch [8/50], Iter [135/418] Loss: 3.1485, average\_loss: 3.4996  
Epoch [8/50], Iter [140/418] Loss: 2.7243, average\_loss: 3.5027  
Epoch [8/50], Iter [145/418] Loss: 4.5699, average\_loss: 3.5080  
Epoch [8/50], Iter [150/418] Loss: 2.8376, average\_loss: 3.4904  
Epoch [8/50], Iter [155/418] Loss: 2.2811, average\_loss: 3.4815  
Epoch [8/50], Iter [160/418] Loss: 2.7052, average\_loss: 3.4733  
Epoch [8/50], Iter [165/418] Loss: 2.5953, average\_loss: 3.4838  
Epoch [8/50], Iter [170/418] Loss: 2.5935, average\_loss: 3.4812  
Epoch [8/50], Iter [175/418] Loss: 4.2614, average\_loss: 3.4815  
Epoch [8/50], Iter [180/418] Loss: 3.2196, average\_loss: 3.4802  
Epoch [8/50], Iter [185/418] Loss: 2.5211, average\_loss: 3.4808  
Epoch [8/50], Iter [190/418] Loss: 3.7939, average\_loss: 3.4826  
Epoch [8/50], Iter [195/418] Loss: 2.9917, average\_loss: 3.4801  
Epoch [8/50], Iter [200/418] Loss: 2.8497, average\_loss: 3.4950  
Epoch [8/50], Iter [205/418] Loss: 2.6941, average\_loss: 3.5055  
Epoch [8/50], Iter [210/418] Loss: 3.9694, average\_loss: 3.5041  
Epoch [8/50], Iter [215/418] Loss: 2.7054, average\_loss: 3.4866  
Epoch [8/50], Iter [220/418] Loss: 2.7483, average\_loss: 3.4807  
Epoch [8/50], Iter [225/418] Loss: 3.2868, average\_loss: 3.4858  
Epoch [8/50], Iter [230/418] Loss: 4.0235, average\_loss: 3.4846  
Epoch [8/50], Iter [235/418] Loss: 2.8223, average\_loss: 3.4819

Epoch [8/50], Iter [240/418] Loss: 3.7350, average\_loss: 3.4727  
Epoch [8/50], Iter [245/418] Loss: 2.6250, average\_loss: 3.4608  
Epoch [8/50], Iter [250/418] Loss: 2.5566, average\_loss: 3.4432  
Epoch [8/50], Iter [255/418] Loss: 3.0105, average\_loss: 3.4497  
Epoch [8/50], Iter [260/418] Loss: 4.0426, average\_loss: 3.4484  
Epoch [8/50], Iter [265/418] Loss: 3.0097, average\_loss: 3.4451  
Epoch [8/50], Iter [270/418] Loss: 1.8753, average\_loss: 3.4305  
Epoch [8/50], Iter [275/418] Loss: 2.5030, average\_loss: 3.4248  
Epoch [8/50], Iter [280/418] Loss: 2.4943, average\_loss: 3.4297  
Epoch [8/50], Iter [285/418] Loss: 3.9644, average\_loss: 3.4270  
Epoch [8/50], Iter [290/418] Loss: 3.6660, average\_loss: 3.4275  
Epoch [8/50], Iter [295/418] Loss: 3.4106, average\_loss: 3.4277  
Epoch [8/50], Iter [300/418] Loss: 6.4314, average\_loss: 3.4399  
Epoch [8/50], Iter [305/418] Loss: 3.3446, average\_loss: 3.4389  
Epoch [8/50], Iter [310/418] Loss: 3.6226, average\_loss: 3.4338  
Epoch [8/50], Iter [315/418] Loss: 2.1356, average\_loss: 3.4358  
Epoch [8/50], Iter [320/418] Loss: 5.9214, average\_loss: 3.4481  
Epoch [8/50], Iter [325/418] Loss: 3.4505, average\_loss: 3.4454  
Epoch [8/50], Iter [330/418] Loss: 3.5724, average\_loss: 3.4479  
Epoch [8/50], Iter [335/418] Loss: 3.7826, average\_loss: 3.4491  
Epoch [8/50], Iter [340/418] Loss: 3.3244, average\_loss: 3.4438  
Epoch [8/50], Iter [345/418] Loss: 2.8568, average\_loss: 3.4400  
Epoch [8/50], Iter [350/418] Loss: 2.7158, average\_loss: 3.4329  
Epoch [8/50], Iter [355/418] Loss: 2.6089, average\_loss: 3.4348  
Epoch [8/50], Iter [360/418] Loss: 3.8504, average\_loss: 3.4387  
Epoch [8/50], Iter [365/418] Loss: 3.1281, average\_loss: 3.4375  
Epoch [8/50], Iter [370/418] Loss: 3.7744, average\_loss: 3.4352  
Epoch [8/50], Iter [375/418] Loss: 4.0397, average\_loss: 3.4324  
Epoch [8/50], Iter [380/418] Loss: 3.3511, average\_loss: 3.4359  
Epoch [8/50], Iter [385/418] Loss: 3.1949, average\_loss: 3.4288  
Epoch [8/50], Iter [390/418] Loss: 4.2881, average\_loss: 3.4343  
Epoch [8/50], Iter [395/418] Loss: 3.5182, average\_loss: 3.4314  
Epoch [8/50], Iter [400/418] Loss: 3.0160, average\_loss: 3.4260  
Epoch [8/50], Iter [405/418] Loss: 4.1715, average\_loss: 3.4195  
Epoch [8/50], Iter [410/418] Loss: 2.2574, average\_loss: 3.4178  
Epoch [8/50], Iter [415/418] Loss: 3.3027, average\_loss: 3.4187  
Updating best test loss: 3.47357

Starting epoch 9 / 50

Learning Rate for this epoch: 0.001

Epoch [9/50], Iter [5/418] Loss: 3.5313, average\_loss: 3.8136  
Epoch [9/50], Iter [10/418] Loss: 2.8568, average\_loss: 3.3037  
Epoch [9/50], Iter [15/418] Loss: 3.5909, average\_loss: 3.2834  
Epoch [9/50], Iter [20/418] Loss: 3.4293, average\_loss: 3.1696  
Epoch [9/50], Iter [25/418] Loss: 2.4937, average\_loss: 3.1284  
Epoch [9/50], Iter [30/418] Loss: 2.3968, average\_loss: 3.1141  
Epoch [9/50], Iter [35/418] Loss: 3.6305, average\_loss: 3.1549

Epoch [9/50], Iter [40/418] Loss: 3.1875, average\_loss: 3.2147  
 Epoch [9/50], Iter [45/418] Loss: 4.9232, average\_loss: 3.2362  
 Epoch [9/50], Iter [50/418] Loss: 2.6900, average\_loss: 3.1627  
 Epoch [9/50], Iter [55/418] Loss: 3.2126, average\_loss: 3.2408  
 Epoch [9/50], Iter [60/418] Loss: 2.5694, average\_loss: 3.2360  
 Epoch [9/50], Iter [65/418] Loss: 4.7879, average\_loss: 3.2578  
 Epoch [9/50], Iter [70/418] Loss: 2.8793, average\_loss: 3.2796  
 Epoch [9/50], Iter [75/418] Loss: 3.5932, average\_loss: 3.2828  
 Epoch [9/50], Iter [80/418] Loss: 3.5948, average\_loss: 3.2748  
 Epoch [9/50], Iter [85/418] Loss: 2.9152, average\_loss: 3.2705  
 Epoch [9/50], Iter [90/418] Loss: 2.9794, average\_loss: 3.2896  
 Epoch [9/50], Iter [95/418] Loss: 3.3918, average\_loss: 3.2791  
 Epoch [9/50], Iter [100/418] Loss: 3.8889, average\_loss: 3.2689  
 Epoch [9/50], Iter [105/418] Loss: 2.9072, average\_loss: 3.2580  
 Epoch [9/50], Iter [110/418] Loss: 4.5328, average\_loss: 3.2859  
 Epoch [9/50], Iter [115/418] Loss: 3.4953, average\_loss: 3.2837  
 Epoch [9/50], Iter [120/418] Loss: 4.8858, average\_loss: 3.2809  
 Epoch [9/50], Iter [125/418] Loss: 2.3155, average\_loss: 3.2768  
 Epoch [9/50], Iter [130/418] Loss: 3.8947, average\_loss: 3.3003  
 Epoch [9/50], Iter [135/418] Loss: 3.1358, average\_loss: 3.2943  
 Epoch [9/50], Iter [140/418] Loss: 3.2039, average\_loss: 3.3048  
 Epoch [9/50], Iter [145/418] Loss: 3.4456, average\_loss: 3.2938  
 Epoch [9/50], Iter [150/418] Loss: 3.3011, average\_loss: 3.3038  
 Epoch [9/50], Iter [155/418] Loss: 2.7597, average\_loss: 3.2793  
 Epoch [9/50], Iter [160/418] Loss: 3.7336, average\_loss: 3.2999  
 Epoch [9/50], Iter [165/418] Loss: 2.4306, average\_loss: 3.3027  
 Epoch [9/50], Iter [170/418] Loss: 4.0181, average\_loss: 3.2997  
 Epoch [9/50], Iter [175/418] Loss: 4.0274, average\_loss: 3.3169  
 Epoch [9/50], Iter [180/418] Loss: 3.9843, average\_loss: 3.3140  
 Epoch [9/50], Iter [185/418] Loss: 2.7465, average\_loss: 3.2991  
 Epoch [9/50], Iter [190/418] Loss: 3.9931, average\_loss: 3.3141  
 Epoch [9/50], Iter [195/418] Loss: 2.9861, average\_loss: 3.2999  
 Epoch [9/50], Iter [200/418] Loss: 3.1613, average\_loss: 3.2993  
 Epoch [9/50], Iter [205/418] Loss: 3.3986, average\_loss: 3.2907  
 Epoch [9/50], Iter [210/418] Loss: 3.1566, average\_loss: 3.2824  
 Epoch [9/50], Iter [215/418] Loss: 5.0389, average\_loss: 3.2850  
 Epoch [9/50], Iter [220/418] Loss: 3.7978, average\_loss: 3.2822  
 Epoch [9/50], Iter [225/418] Loss: 2.8393, average\_loss: 3.2786  
 Epoch [9/50], Iter [230/418] Loss: 2.1511, average\_loss: 3.2739  
 Epoch [9/50], Iter [235/418] Loss: 2.4931, average\_loss: 3.2805  
 Epoch [9/50], Iter [240/418] Loss: 4.5653, average\_loss: 3.2835  
 Epoch [9/50], Iter [245/418] Loss: 2.2255, average\_loss: 3.2719  
 Epoch [9/50], Iter [250/418] Loss: 2.7767, average\_loss: 3.2742  
 Epoch [9/50], Iter [255/418] Loss: 3.0365, average\_loss: 3.2696  
 Epoch [9/50], Iter [260/418] Loss: 3.6884, average\_loss: 3.2696  
 Epoch [9/50], Iter [265/418] Loss: 4.0017, average\_loss: 3.2769  
 Epoch [9/50], Iter [270/418] Loss: 2.3643, average\_loss: 3.2690  
 Epoch [9/50], Iter [275/418] Loss: 2.1558, average\_loss: 3.2618

Epoch [9/50], Iter [280/418] Loss: 3.9595, average\_loss: 3.2567  
Epoch [9/50], Iter [285/418] Loss: 2.4725, average\_loss: 3.2552  
Epoch [9/50], Iter [290/418] Loss: 3.1401, average\_loss: 3.2472  
Epoch [9/50], Iter [295/418] Loss: 3.2959, average\_loss: 3.2479  
Epoch [9/50], Iter [300/418] Loss: 3.5985, average\_loss: 3.2530  
Epoch [9/50], Iter [305/418] Loss: 2.7322, average\_loss: 3.2571  
Epoch [9/50], Iter [310/418] Loss: 3.2527, average\_loss: 3.2582  
Epoch [9/50], Iter [315/418] Loss: 2.1119, average\_loss: 3.2544  
Epoch [9/50], Iter [320/418] Loss: 4.1324, average\_loss: 3.2582  
Epoch [9/50], Iter [325/418] Loss: 3.9615, average\_loss: 3.2689  
Epoch [9/50], Iter [330/418] Loss: 2.8977, average\_loss: 3.2659  
Epoch [9/50], Iter [335/418] Loss: 2.7875, average\_loss: 3.2672  
Epoch [9/50], Iter [340/418] Loss: 4.3660, average\_loss: 3.2682  
Epoch [9/50], Iter [345/418] Loss: 3.6155, average\_loss: 3.2743  
Epoch [9/50], Iter [350/418] Loss: 2.1156, average\_loss: 3.2728  
Epoch [9/50], Iter [355/418] Loss: 3.6681, average\_loss: 3.2718  
Epoch [9/50], Iter [360/418] Loss: 3.8229, average\_loss: 3.2719  
Epoch [9/50], Iter [365/418] Loss: 3.0965, average\_loss: 3.2714  
Epoch [9/50], Iter [370/418] Loss: 2.3765, average\_loss: 3.2701  
Epoch [9/50], Iter [375/418] Loss: 4.1497, average\_loss: 3.2726  
Epoch [9/50], Iter [380/418] Loss: 2.9217, average\_loss: 3.2672  
Epoch [9/50], Iter [385/418] Loss: 2.6831, average\_loss: 3.2656  
Epoch [9/50], Iter [390/418] Loss: 4.6947, average\_loss: 3.2679  
Epoch [9/50], Iter [395/418] Loss: 3.2749, average\_loss: 3.2632  
Epoch [9/50], Iter [400/418] Loss: 6.6673, average\_loss: 3.2664  
Epoch [9/50], Iter [405/418] Loss: 2.1805, average\_loss: 3.2643  
Epoch [9/50], Iter [410/418] Loss: 1.7253, average\_loss: 3.2611  
Epoch [9/50], Iter [415/418] Loss: 5.5722, average\_loss: 3.2685  
Updating best test loss: 3.36789

Starting epoch 10 / 50

Learning Rate for this epoch: 0.001

Epoch [10/50], Iter [5/418] Loss: 4.9394, average\_loss: 3.9492  
Epoch [10/50], Iter [10/418] Loss: 3.5021, average\_loss: 3.3124  
Epoch [10/50], Iter [15/418] Loss: 4.0345, average\_loss: 3.3328  
Epoch [10/50], Iter [20/418] Loss: 2.7531, average\_loss: 3.1943  
Epoch [10/50], Iter [25/418] Loss: 5.4400, average\_loss: 3.3810  
Epoch [10/50], Iter [30/418] Loss: 2.4333, average\_loss: 3.4553  
Epoch [10/50], Iter [35/418] Loss: 4.1052, average\_loss: 3.4090  
Epoch [10/50], Iter [40/418] Loss: 4.2894, average\_loss: 3.3855  
Epoch [10/50], Iter [45/418] Loss: 3.1348, average\_loss: 3.3340  
Epoch [10/50], Iter [50/418] Loss: 2.8005, average\_loss: 3.2971  
Epoch [10/50], Iter [55/418] Loss: 2.4608, average\_loss: 3.2541  
Epoch [10/50], Iter [60/418] Loss: 2.8338, average\_loss: 3.3020  
Epoch [10/50], Iter [65/418] Loss: 2.4501, average\_loss: 3.2583  
Epoch [10/50], Iter [70/418] Loss: 3.4581, average\_loss: 3.2737  
Epoch [10/50], Iter [75/418] Loss: 3.5584, average\_loss: 3.2499

Epoch [10/50], Iter [80/418] Loss: 3.3870, average\_loss: 3.2666  
 Epoch [10/50], Iter [85/418] Loss: 2.5062, average\_loss: 3.2466  
 Epoch [10/50], Iter [90/418] Loss: 2.1193, average\_loss: 3.2543  
 Epoch [10/50], Iter [95/418] Loss: 2.8779, average\_loss: 3.2339  
 Epoch [10/50], Iter [100/418] Loss: 2.6920, average\_loss: 3.2075  
 Epoch [10/50], Iter [105/418] Loss: 4.3039, average\_loss: 3.2087  
 Epoch [10/50], Iter [110/418] Loss: 2.2402, average\_loss: 3.2167  
 Epoch [10/50], Iter [115/418] Loss: 3.8726, average\_loss: 3.2008  
 Epoch [10/50], Iter [120/418] Loss: 3.2157, average\_loss: 3.1944  
 Epoch [10/50], Iter [125/418] Loss: 2.3742, average\_loss: 3.1825  
 Epoch [10/50], Iter [130/418] Loss: 3.9683, average\_loss: 3.1855  
 Epoch [10/50], Iter [135/418] Loss: 3.4345, average\_loss: 3.1828  
 Epoch [10/50], Iter [140/418] Loss: 2.8118, average\_loss: 3.1959  
 Epoch [10/50], Iter [145/418] Loss: 2.0887, average\_loss: 3.1846  
 Epoch [10/50], Iter [150/418] Loss: 2.6967, average\_loss: 3.1827  
 Epoch [10/50], Iter [155/418] Loss: 2.9477, average\_loss: 3.1738  
 Epoch [10/50], Iter [160/418] Loss: 4.4581, average\_loss: 3.1817  
 Epoch [10/50], Iter [165/418] Loss: 4.9596, average\_loss: 3.1830  
 Epoch [10/50], Iter [170/418] Loss: 3.3627, average\_loss: 3.1966  
 Epoch [10/50], Iter [175/418] Loss: 3.9260, average\_loss: 3.1890  
 Epoch [10/50], Iter [180/418] Loss: 2.9883, average\_loss: 3.1968  
 Epoch [10/50], Iter [185/418] Loss: 2.9983, average\_loss: 3.1924  
 Epoch [10/50], Iter [190/418] Loss: 2.7240, average\_loss: 3.1896  
 Epoch [10/50], Iter [195/418] Loss: 2.0310, average\_loss: 3.1720  
 Epoch [10/50], Iter [200/418] Loss: 3.0654, average\_loss: 3.1738  
 Epoch [10/50], Iter [205/418] Loss: 6.4030, average\_loss: 3.1831  
 Epoch [10/50], Iter [210/418] Loss: 4.6452, average\_loss: 3.1780  
 Epoch [10/50], Iter [215/418] Loss: 2.4273, average\_loss: 3.1597  
 Epoch [10/50], Iter [220/418] Loss: 2.2998, average\_loss: 3.1533  
 Epoch [10/50], Iter [225/418] Loss: 3.7474, average\_loss: 3.1440  
 Epoch [10/50], Iter [230/418] Loss: 2.9821, average\_loss: 3.1478  
 Epoch [10/50], Iter [235/418] Loss: 4.4805, average\_loss: 3.1407  
 Epoch [10/50], Iter [240/418] Loss: 2.0950, average\_loss: 3.1395  
 Epoch [10/50], Iter [245/418] Loss: 2.7706, average\_loss: 3.1389  
 Epoch [10/50], Iter [250/418] Loss: 2.1609, average\_loss: 3.1315  
 Epoch [10/50], Iter [255/418] Loss: 1.9308, average\_loss: 3.1193  
 Epoch [10/50], Iter [260/418] Loss: 2.8821, average\_loss: 3.1146  
 Epoch [10/50], Iter [265/418] Loss: 2.8491, average\_loss: 3.1121  
 Epoch [10/50], Iter [270/418] Loss: 5.3104, average\_loss: 3.1289  
 Epoch [10/50], Iter [275/418] Loss: 4.0329, average\_loss: 3.1369  
 Epoch [10/50], Iter [280/418] Loss: 3.9446, average\_loss: 3.1424  
 Epoch [10/50], Iter [285/418] Loss: 2.0349, average\_loss: 3.1351  
 Epoch [10/50], Iter [290/418] Loss: 3.6472, average\_loss: 3.1351  
 Epoch [10/50], Iter [295/418] Loss: 2.5549, average\_loss: 3.1398  
 Epoch [10/50], Iter [300/418] Loss: 1.8388, average\_loss: 3.1267  
 Epoch [10/50], Iter [305/418] Loss: 2.8866, average\_loss: 3.1281  
 Epoch [10/50], Iter [310/418] Loss: 2.6215, average\_loss: 3.1281  
 Epoch [10/50], Iter [315/418] Loss: 3.0972, average\_loss: 3.1229

Epoch [10/50], Iter [320/418] Loss: 4.1242, average\_loss: 3.1271  
 Epoch [10/50], Iter [325/418] Loss: 4.1929, average\_loss: 3.1295  
 Epoch [10/50], Iter [330/418] Loss: 4.0604, average\_loss: 3.1265  
 Epoch [10/50], Iter [335/418] Loss: 3.8889, average\_loss: 3.1237  
 Epoch [10/50], Iter [340/418] Loss: 3.7433, average\_loss: 3.1216  
 Epoch [10/50], Iter [345/418] Loss: 2.8796, average\_loss: 3.1212  
 Epoch [10/50], Iter [350/418] Loss: 3.1059, average\_loss: 3.1206  
 Epoch [10/50], Iter [355/418] Loss: 2.3207, average\_loss: 3.1216  
 Epoch [10/50], Iter [360/418] Loss: 3.5693, average\_loss: 3.1222  
 Epoch [10/50], Iter [365/418] Loss: 3.8384, average\_loss: 3.1203  
 Epoch [10/50], Iter [370/418] Loss: 3.0252, average\_loss: 3.1220  
 Epoch [10/50], Iter [375/418] Loss: 2.7633, average\_loss: 3.1190  
 Epoch [10/50], Iter [380/418] Loss: 2.3634, average\_loss: 3.1193  
 Epoch [10/50], Iter [385/418] Loss: 4.2718, average\_loss: 3.1237  
 Epoch [10/50], Iter [390/418] Loss: 2.5143, average\_loss: 3.1200  
 Epoch [10/50], Iter [395/418] Loss: 2.9236, average\_loss: 3.1151  
 Epoch [10/50], Iter [400/418] Loss: 3.7272, average\_loss: 3.1199  
 Epoch [10/50], Iter [405/418] Loss: 3.5076, average\_loss: 3.1202  
 Epoch [10/50], Iter [410/418] Loss: 3.6045, average\_loss: 3.1207  
 Epoch [10/50], Iter [415/418] Loss: 2.3611, average\_loss: 3.1182  
 Updating best test loss: 3.27245

Starting epoch 11 / 50

Learning Rate for this epoch: 0.001

Epoch [11/50], Iter [5/418] Loss: 1.9756, average\_loss: 3.2094  
 Epoch [11/50], Iter [10/418] Loss: 2.7428, average\_loss: 3.0410  
 Epoch [11/50], Iter [15/418] Loss: 2.7726, average\_loss: 3.2049  
 Epoch [11/50], Iter [20/418] Loss: 3.4221, average\_loss: 3.3533  
 Epoch [11/50], Iter [25/418] Loss: 4.2018, average\_loss: 3.4575  
 Epoch [11/50], Iter [30/418] Loss: 2.5221, average\_loss: 3.3837  
 Epoch [11/50], Iter [35/418] Loss: 3.8963, average\_loss: 3.4274  
 Epoch [11/50], Iter [40/418] Loss: 3.0606, average\_loss: 3.3413  
 Epoch [11/50], Iter [45/418] Loss: 2.4486, average\_loss: 3.3250  
 Epoch [11/50], Iter [50/418] Loss: 2.1899, average\_loss: 3.2639  
 Epoch [11/50], Iter [55/418] Loss: 1.8735, average\_loss: 3.2207  
 Epoch [11/50], Iter [60/418] Loss: 2.9395, average\_loss: 3.1907  
 Epoch [11/50], Iter [65/418] Loss: 1.9000, average\_loss: 3.1428  
 Epoch [11/50], Iter [70/418] Loss: 2.8176, average\_loss: 3.1636  
 Epoch [11/50], Iter [75/418] Loss: 3.9633, average\_loss: 3.1514  
 Epoch [11/50], Iter [80/418] Loss: 4.0434, average\_loss: 3.1318  
 Epoch [11/50], Iter [85/418] Loss: 2.9738, average\_loss: 3.1099  
 Epoch [11/50], Iter [90/418] Loss: 2.7933, average\_loss: 3.1380  
 Epoch [11/50], Iter [95/418] Loss: 4.2334, average\_loss: 3.1555  
 Epoch [11/50], Iter [100/418] Loss: 3.6646, average\_loss: 3.1843  
 Epoch [11/50], Iter [105/418] Loss: 1.6669, average\_loss: 3.1747  
 Epoch [11/50], Iter [110/418] Loss: 2.9200, average\_loss: 3.1600  
 Epoch [11/50], Iter [115/418] Loss: 3.4966, average\_loss: 3.1519



Epoch [11/50], Iter [120/418] Loss: 2.7366, average\_loss: 3.1399  
 Epoch [11/50], Iter [125/418] Loss: 2.6582, average\_loss: 3.1337  
 Epoch [11/50], Iter [130/418] Loss: 5.6061, average\_loss: 3.1397  
 Epoch [11/50], Iter [135/418] Loss: 2.5293, average\_loss: 3.1370  
 Epoch [11/50], Iter [140/418] Loss: 1.7315, average\_loss: 3.1338  
 Epoch [11/50], Iter [145/418] Loss: 1.9468, average\_loss: 3.1156  
 Epoch [11/50], Iter [150/418] Loss: 4.2563, average\_loss: 3.1266  
 Epoch [11/50], Iter [155/418] Loss: 2.3983, average\_loss: 3.1360  
 Epoch [11/50], Iter [160/418] Loss: 3.7208, average\_loss: 3.1389  
 Epoch [11/50], Iter [165/418] Loss: 4.2522, average\_loss: 3.1619  
 Epoch [11/50], Iter [170/418] Loss: 1.9443, average\_loss: 3.1653  
 Epoch [11/50], Iter [175/418] Loss: 2.7555, average\_loss: 3.1639  
 Epoch [11/50], Iter [180/418] Loss: 1.6121, average\_loss: 3.1653  
 Epoch [11/50], Iter [185/418] Loss: 3.6224, average\_loss: 3.1624  
 Epoch [11/50], Iter [190/418] Loss: 2.0980, average\_loss: 3.1532  
 Epoch [11/50], Iter [195/418] Loss: 3.1885, average\_loss: 3.1441  
 Epoch [11/50], Iter [200/418] Loss: 4.1407, average\_loss: 3.1389  
 Epoch [11/50], Iter [205/418] Loss: 2.1039, average\_loss: 3.1321  
 Epoch [11/50], Iter [210/418] Loss: 2.6335, average\_loss: 3.1408  
 Epoch [11/50], Iter [215/418] Loss: 1.8697, average\_loss: 3.1226  
 Epoch [11/50], Iter [220/418] Loss: 3.0301, average\_loss: 3.1237  
 Epoch [11/50], Iter [225/418] Loss: 2.4267, average\_loss: 3.1124  
 Epoch [11/50], Iter [230/418] Loss: 2.7600, average\_loss: 3.1088  
 Epoch [11/50], Iter [235/418] Loss: 2.1640, average\_loss: 3.0941  
 Epoch [11/50], Iter [240/418] Loss: 4.4099, average\_loss: 3.0943  
 Epoch [11/50], Iter [245/418] Loss: 2.6908, average\_loss: 3.0992  
 Epoch [11/50], Iter [250/418] Loss: 2.8045, average\_loss: 3.0976  
 Epoch [11/50], Iter [255/418] Loss: 2.4606, average\_loss: 3.0836  
 Epoch [11/50], Iter [260/418] Loss: 1.6490, average\_loss: 3.0723  
 Epoch [11/50], Iter [265/418] Loss: 2.5730, average\_loss: 3.0706  
 Epoch [11/50], Iter [270/418] Loss: 2.6490, average\_loss: 3.0593  
 Epoch [11/50], Iter [275/418] Loss: 3.0635, average\_loss: 3.0616  
 Epoch [11/50], Iter [280/418] Loss: 2.5389, average\_loss: 3.0564  
 Epoch [11/50], Iter [285/418] Loss: 2.9877, average\_loss: 3.0572  
 Epoch [11/50], Iter [290/418] Loss: 3.0561, average\_loss: 3.0522  
 Epoch [11/50], Iter [295/418] Loss: 2.7211, average\_loss: 3.0550  
 Epoch [11/50], Iter [300/418] Loss: 2.4814, average\_loss: 3.0466  
 Epoch [11/50], Iter [305/418] Loss: 3.0198, average\_loss: 3.0447  
 Epoch [11/50], Iter [310/418] Loss: 4.0097, average\_loss: 3.0565  
 Epoch [11/50], Iter [315/418] Loss: 2.1642, average\_loss: 3.0592  
 Epoch [11/50], Iter [320/418] Loss: 2.4505, average\_loss: 3.0508  
 Epoch [11/50], Iter [325/418] Loss: 3.5487, average\_loss: 3.0620  
 Epoch [11/50], Iter [330/418] Loss: 2.5535, average\_loss: 3.0631  
 Epoch [11/50], Iter [335/418] Loss: 2.8629, average\_loss: 3.0677  
 Epoch [11/50], Iter [340/418] Loss: 2.8691, average\_loss: 3.0674  
 Epoch [11/50], Iter [345/418] Loss: 2.9085, average\_loss: 3.0630  
 Epoch [11/50], Iter [350/418] Loss: 3.9544, average\_loss: 3.0600  
 Epoch [11/50], Iter [355/418] Loss: 2.6986, average\_loss: 3.0539

Epoch [11/50], Iter [360/418] Loss: 2.8438, average\_loss: 3.0502  
Epoch [11/50], Iter [365/418] Loss: 2.0500, average\_loss: 3.0444  
Epoch [11/50], Iter [370/418] Loss: 2.3525, average\_loss: 3.0449  
Epoch [11/50], Iter [375/418] Loss: 2.8799, average\_loss: 3.0457  
Epoch [11/50], Iter [380/418] Loss: 3.1935, average\_loss: 3.0466  
Epoch [11/50], Iter [385/418] Loss: 1.8889, average\_loss: 3.0372  
Epoch [11/50], Iter [390/418] Loss: 2.3690, average\_loss: 3.0297  
Epoch [11/50], Iter [395/418] Loss: 1.9611, average\_loss: 3.0297  
Epoch [11/50], Iter [400/418] Loss: 2.9513, average\_loss: 3.0303  
Epoch [11/50], Iter [405/418] Loss: 2.0810, average\_loss: 3.0249  
Epoch [11/50], Iter [410/418] Loss: 2.9051, average\_loss: 3.0178  
Epoch [11/50], Iter [415/418] Loss: 2.1883, average\_loss: 3.0145

---Evaluate model on test samples---

100%|

| 4950/4950 [02:29<00:00, 33.08it/s]

---class aeroplane ap 0.295768783763334---  
---class bicycle ap 0.379346323328453---  
---class bird ap 0.21685817174478597---  
---class boat ap 0.08592000761755941---  
---class bottle ap 0.09926367245828573---  
---class bus ap 0.4143307026801051---  
---class car ap 0.495194614373971---  
---class cat ap 0.4316385932822256---  
---class chair ap 0.16143781328898357---  
---class cow ap 0.1863496011073207---  
---class diningtable ap 0.07069603805040697---  
---class dog ap 0.3623213598647317---  
---class horse ap 0.4263373915114903---  
---class motorbike ap 0.4032274274228236---  
---class person ap 0.3413324331753324---  
---class pottedplant ap 0.05729249903321587---  
---class sheep ap 0.20048851302312976---  
---class sofa ap 0.2091141166722329---  
---class train ap 0.4335083517167603---  
---class tvmonitor ap 0.34211508592680906---  
---map 0.2806270750020979---

Updating best test loss: 3.15290

Starting epoch 12 / 50

Learning Rate for this epoch: 0.001

Epoch [12/50], Iter [5/418] Loss: 2.8690, average\_loss: 3.0858  
Epoch [12/50], Iter [10/418] Loss: 2.5524, average\_loss: 2.7536  
Epoch [12/50], Iter [15/418] Loss: 1.9531, average\_loss: 2.6506  
Epoch [12/50], Iter [20/418] Loss: 2.3618, average\_loss: 2.6861  
Epoch [12/50], Iter [25/418] Loss: 3.0633, average\_loss: 2.7246  
Epoch [12/50], Iter [30/418] Loss: 2.4473, average\_loss: 2.6290

Epoch [12/50], Iter [35/418] Loss: 2.8262, average\_loss: 2.7756  
 Epoch [12/50], Iter [40/418] Loss: 3.7866, average\_loss: 2.8320  
 Epoch [12/50], Iter [45/418] Loss: 2.1335, average\_loss: 2.8085  
 Epoch [12/50], Iter [50/418] Loss: 4.3067, average\_loss: 2.8057  
 Epoch [12/50], Iter [55/418] Loss: 2.6905, average\_loss: 2.7746  
 Epoch [12/50], Iter [60/418] Loss: 2.4645, average\_loss: 2.7848  
 Epoch [12/50], Iter [65/418] Loss: 1.3515, average\_loss: 2.7734  
 Epoch [12/50], Iter [70/418] Loss: 3.8389, average\_loss: 2.8113  
 Epoch [12/50], Iter [75/418] Loss: 2.2106, average\_loss: 2.8123  
 Epoch [12/50], Iter [80/418] Loss: 2.9084, average\_loss: 2.8102  
 Epoch [12/50], Iter [85/418] Loss: 2.7454, average\_loss: 2.8273  
 Epoch [12/50], Iter [90/418] Loss: 2.7129, average\_loss: 2.8120  
 Epoch [12/50], Iter [95/418] Loss: 2.7066, average\_loss: 2.8013  
 Epoch [12/50], Iter [100/418] Loss: 2.4382, average\_loss: 2.7920  
 Epoch [12/50], Iter [105/418] Loss: 1.9699, average\_loss: 2.8102  
 Epoch [12/50], Iter [110/418] Loss: 4.0102, average\_loss: 2.7937  
 Epoch [12/50], Iter [115/418] Loss: 2.4841, average\_loss: 2.7868  
 Epoch [12/50], Iter [120/418] Loss: 2.7361, average\_loss: 2.7900  
 Epoch [12/50], Iter [125/418] Loss: 2.8665, average\_loss: 2.7947  
 Epoch [12/50], Iter [130/418] Loss: 3.2213, average\_loss: 2.8128  
 Epoch [12/50], Iter [135/418] Loss: 3.5093, average\_loss: 2.8143  
 Epoch [12/50], Iter [140/418] Loss: 2.9457, average\_loss: 2.8143  
 Epoch [12/50], Iter [145/418] Loss: 1.8630, average\_loss: 2.8172  
 Epoch [12/50], Iter [150/418] Loss: 2.9836, average\_loss: 2.8267  
 Epoch [12/50], Iter [155/418] Loss: 2.6883, average\_loss: 2.8228  
 Epoch [12/50], Iter [160/418] Loss: 3.0058, average\_loss: 2.8311  
 Epoch [12/50], Iter [165/418] Loss: 2.7019, average\_loss: 2.8423  
 Epoch [12/50], Iter [170/418] Loss: 3.1618, average\_loss: 2.8329  
 Epoch [12/50], Iter [175/418] Loss: 3.1253, average\_loss: 2.8326  
 Epoch [12/50], Iter [180/418] Loss: 4.4206, average\_loss: 2.8352  
 Epoch [12/50], Iter [185/418] Loss: 4.5389, average\_loss: 2.8367  
 Epoch [12/50], Iter [190/418] Loss: 1.9545, average\_loss: 2.8293  
 Epoch [12/50], Iter [195/418] Loss: 3.7955, average\_loss: 2.8331  
 Epoch [12/50], Iter [200/418] Loss: 2.3352, average\_loss: 2.8328  
 Epoch [12/50], Iter [205/418] Loss: 3.2255, average\_loss: 2.8322  
 Epoch [12/50], Iter [210/418] Loss: 3.4922, average\_loss: 2.8441  
 Epoch [12/50], Iter [215/418] Loss: 2.8770, average\_loss: 2.8511  
 Epoch [12/50], Iter [220/418] Loss: 2.5348, average\_loss: 2.8479  
 Epoch [12/50], Iter [225/418] Loss: 3.0829, average\_loss: 2.8406  
 Epoch [12/50], Iter [230/418] Loss: 2.4932, average\_loss: 2.8449  
 Epoch [12/50], Iter [235/418] Loss: 2.1778, average\_loss: 2.8425  
 Epoch [12/50], Iter [240/418] Loss: 3.2585, average\_loss: 2.8532  
 Epoch [12/50], Iter [245/418] Loss: 2.3341, average\_loss: 2.8496  
 Epoch [12/50], Iter [250/418] Loss: 3.2269, average\_loss: 2.8466  
 Epoch [12/50], Iter [255/418] Loss: 3.0470, average\_loss: 2.8473  
 Epoch [12/50], Iter [260/418] Loss: 2.1785, average\_loss: 2.8483  
 Epoch [12/50], Iter [265/418] Loss: 2.2195, average\_loss: 2.8541  
 Epoch [12/50], Iter [270/418] Loss: 3.6224, average\_loss: 2.8653

Epoch [12/50], Iter [275/418] Loss: 3.0358, average\_loss: 2.8685  
 Epoch [12/50], Iter [280/418] Loss: 4.0622, average\_loss: 2.8704  
 Epoch [12/50], Iter [285/418] Loss: 2.2012, average\_loss: 2.8684  
 Epoch [12/50], Iter [290/418] Loss: 2.6154, average\_loss: 2.8663  
 Epoch [12/50], Iter [295/418] Loss: 2.4172, average\_loss: 2.8744  
 Epoch [12/50], Iter [300/418] Loss: 2.2097, average\_loss: 2.8782  
 Epoch [12/50], Iter [305/418] Loss: 3.3236, average\_loss: 2.8875  
 Epoch [12/50], Iter [310/418] Loss: 2.7034, average\_loss: 2.8816  
 Epoch [12/50], Iter [315/418] Loss: 3.3023, average\_loss: 2.8889  
 Epoch [12/50], Iter [320/418] Loss: 3.7571, average\_loss: 2.8927  
 Epoch [12/50], Iter [325/418] Loss: 3.0431, average\_loss: 2.8923  
 Epoch [12/50], Iter [330/418] Loss: 3.1200, average\_loss: 2.8939  
 Epoch [12/50], Iter [335/418] Loss: 2.8494, average\_loss: 2.8908  
 Epoch [12/50], Iter [340/418] Loss: 1.7679, average\_loss: 2.8897  
 Epoch [12/50], Iter [345/418] Loss: 2.6128, average\_loss: 2.8871  
 Epoch [12/50], Iter [350/418] Loss: 3.9639, average\_loss: 2.8879  
 Epoch [12/50], Iter [355/418] Loss: 2.2522, average\_loss: 2.8976  
 Epoch [12/50], Iter [360/418] Loss: 3.4593, average\_loss: 2.9006  
 Epoch [12/50], Iter [365/418] Loss: 3.2639, average\_loss: 2.8961  
 Epoch [12/50], Iter [370/418] Loss: 2.8561, average\_loss: 2.8949  
 Epoch [12/50], Iter [375/418] Loss: 4.5679, average\_loss: 2.8947  
 Epoch [12/50], Iter [380/418] Loss: 2.5516, average\_loss: 2.8938  
 Epoch [12/50], Iter [385/418] Loss: 2.1019, average\_loss: 2.8925  
 Epoch [12/50], Iter [390/418] Loss: 2.1300, average\_loss: 2.8862  
 Epoch [12/50], Iter [395/418] Loss: 2.6556, average\_loss: 2.8893  
 Epoch [12/50], Iter [400/418] Loss: 4.5259, average\_loss: 2.9012  
 Epoch [12/50], Iter [405/418] Loss: 3.0616, average\_loss: 2.8973  
 Epoch [12/50], Iter [410/418] Loss: 2.2462, average\_loss: 2.8963  
 Epoch [12/50], Iter [415/418] Loss: 3.2384, average\_loss: 2.8953  
 Updating best test loss: 3.11097

Starting epoch 13 / 50

Learning Rate for this epoch: 0.001

Epoch [13/50], Iter [5/418] Loss: 1.6391, average\_loss: 2.8849  
 Epoch [13/50], Iter [10/418] Loss: 4.3303, average\_loss: 2.8197  
 Epoch [13/50], Iter [15/418] Loss: 2.2257, average\_loss: 2.7179  
 Epoch [13/50], Iter [20/418] Loss: 2.3318, average\_loss: 2.6932  
 Epoch [13/50], Iter [25/418] Loss: 2.9675, average\_loss: 2.7895  
 Epoch [13/50], Iter [30/418] Loss: 3.7091, average\_loss: 2.8110  
 Epoch [13/50], Iter [35/418] Loss: 1.9699, average\_loss: 2.8151  
 Epoch [13/50], Iter [40/418] Loss: 3.1186, average\_loss: 2.8839  
 Epoch [13/50], Iter [45/418] Loss: 2.4646, average\_loss: 2.8355  
 Epoch [13/50], Iter [50/418] Loss: 3.6118, average\_loss: 2.8417  
 Epoch [13/50], Iter [55/418] Loss: 2.0020, average\_loss: 2.8202  
 Epoch [13/50], Iter [60/418] Loss: 2.3615, average\_loss: 2.8205  
 Epoch [13/50], Iter [65/418] Loss: 3.3336, average\_loss: 2.8435  
 Epoch [13/50], Iter [70/418] Loss: 2.9233, average\_loss: 2.8540

Epoch [13/50], Iter [75/418] Loss: 2.4024, average\_loss: 2.8607  
Epoch [13/50], Iter [80/418] Loss: 2.8327, average\_loss: 2.8365  
Epoch [13/50], Iter [85/418] Loss: 2.3510, average\_loss: 2.8322  
Epoch [13/50], Iter [90/418] Loss: 1.9619, average\_loss: 2.8074  
Epoch [13/50], Iter [95/418] Loss: 3.0009, average\_loss: 2.8074  
Epoch [13/50], Iter [100/418] Loss: 2.2037, average\_loss: 2.7880  
Epoch [13/50], Iter [105/418] Loss: 3.9826, average\_loss: 2.7874  
Epoch [13/50], Iter [110/418] Loss: 3.2162, average\_loss: 2.7816  
Epoch [13/50], Iter [115/418] Loss: 2.6129, average\_loss: 2.7603  
Epoch [13/50], Iter [120/418] Loss: 3.6741, average\_loss: 2.7819  
Epoch [13/50], Iter [125/418] Loss: 2.3127, average\_loss: 2.7802  
Epoch [13/50], Iter [130/418] Loss: 4.1557, average\_loss: 2.7826  
Epoch [13/50], Iter [135/418] Loss: 4.0315, average\_loss: 2.8247  
Epoch [13/50], Iter [140/418] Loss: 2.6048, average\_loss: 2.8096  
Epoch [13/50], Iter [145/418] Loss: 2.3493, average\_loss: 2.8175  
Epoch [13/50], Iter [150/418] Loss: 2.9947, average\_loss: 2.8196  
Epoch [13/50], Iter [155/418] Loss: 2.4415, average\_loss: 2.8141  
Epoch [13/50], Iter [160/418] Loss: 2.5704, average\_loss: 2.7949  
Epoch [13/50], Iter [165/418] Loss: 2.8933, average\_loss: 2.7850  
Epoch [13/50], Iter [170/418] Loss: 4.3753, average\_loss: 2.8004  
Epoch [13/50], Iter [175/418] Loss: 2.3534, average\_loss: 2.8016  
Epoch [13/50], Iter [180/418] Loss: 2.0969, average\_loss: 2.8029  
Epoch [13/50], Iter [185/418] Loss: 3.3142, average\_loss: 2.7973  
Epoch [13/50], Iter [190/418] Loss: 2.2393, average\_loss: 2.8043  
Epoch [13/50], Iter [195/418] Loss: 2.3645, average\_loss: 2.7965  
Epoch [13/50], Iter [200/418] Loss: 2.0853, average\_loss: 2.7975  
Epoch [13/50], Iter [205/418] Loss: 2.0896, average\_loss: 2.7921  
Epoch [13/50], Iter [210/418] Loss: 3.1458, average\_loss: 2.8010  
Epoch [13/50], Iter [215/418] Loss: 2.0046, average\_loss: 2.8078  
Epoch [13/50], Iter [220/418] Loss: 2.4372, average\_loss: 2.8022  
Epoch [13/50], Iter [225/418] Loss: 3.4860, average\_loss: 2.8125  
Epoch [13/50], Iter [230/418] Loss: 2.4541, average\_loss: 2.8034  
Epoch [13/50], Iter [235/418] Loss: 3.1965, average\_loss: 2.8056  
Epoch [13/50], Iter [240/418] Loss: 3.3011, average\_loss: 2.8093  
Epoch [13/50], Iter [245/418] Loss: 3.1928, average\_loss: 2.8148  
Epoch [13/50], Iter [250/418] Loss: 5.5361, average\_loss: 2.8219  
Epoch [13/50], Iter [255/418] Loss: 2.0752, average\_loss: 2.8349  
Epoch [13/50], Iter [260/418] Loss: 2.6341, average\_loss: 2.8277  
Epoch [13/50], Iter [265/418] Loss: 2.7237, average\_loss: 2.8220  
Epoch [13/50], Iter [270/418] Loss: 3.2078, average\_loss: 2.8253  
Epoch [13/50], Iter [275/418] Loss: 3.3136, average\_loss: 2.8295  
Epoch [13/50], Iter [280/418] Loss: 2.1125, average\_loss: 2.8189  
Epoch [13/50], Iter [285/418] Loss: 2.7300, average\_loss: 2.8176  
Epoch [13/50], Iter [290/418] Loss: 2.4081, average\_loss: 2.8235  
Epoch [13/50], Iter [295/418] Loss: 2.7912, average\_loss: 2.8237  
Epoch [13/50], Iter [300/418] Loss: 2.9099, average\_loss: 2.8213  
Epoch [13/50], Iter [305/418] Loss: 3.2908, average\_loss: 2.8246  
Epoch [13/50], Iter [310/418] Loss: 3.3792, average\_loss: 2.8310

Epoch [13/50], Iter [315/418] Loss: 2.4620, average\_loss: 2.8248  
 Epoch [13/50], Iter [320/418] Loss: 2.8430, average\_loss: 2.8226  
 Epoch [13/50], Iter [325/418] Loss: 2.4572, average\_loss: 2.8166  
 Epoch [13/50], Iter [330/418] Loss: 1.8378, average\_loss: 2.8131  
 Epoch [13/50], Iter [335/418] Loss: 3.1339, average\_loss: 2.8128  
 Epoch [13/50], Iter [340/418] Loss: 3.2451, average\_loss: 2.8088  
 Epoch [13/50], Iter [345/418] Loss: 2.5193, average\_loss: 2.8149  
 Epoch [13/50], Iter [350/418] Loss: 2.7052, average\_loss: 2.8116  
 Epoch [13/50], Iter [355/418] Loss: 2.4074, average\_loss: 2.8127  
 Epoch [13/50], Iter [360/418] Loss: 2.2192, average\_loss: 2.8006  
 Epoch [13/50], Iter [365/418] Loss: 3.4183, average\_loss: 2.7941  
 Epoch [13/50], Iter [370/418] Loss: 1.9174, average\_loss: 2.7925  
 Epoch [13/50], Iter [375/418] Loss: 2.7460, average\_loss: 2.7971  
 Epoch [13/50], Iter [380/418] Loss: 3.0734, average\_loss: 2.8024  
 Epoch [13/50], Iter [385/418] Loss: 2.5573, average\_loss: 2.8030  
 Epoch [13/50], Iter [390/418] Loss: 2.2552, average\_loss: 2.8067  
 Epoch [13/50], Iter [395/418] Loss: 4.3835, average\_loss: 2.8048  
 Epoch [13/50], Iter [400/418] Loss: 2.1993, average\_loss: 2.8039  
 Epoch [13/50], Iter [405/418] Loss: 2.4488, average\_loss: 2.8059  
 Epoch [13/50], Iter [410/418] Loss: 2.2608, average\_loss: 2.8095  
 Epoch [13/50], Iter [415/418] Loss: 3.8806, average\_loss: 2.8131  
 Updating best test loss: 3.03012

Starting epoch 14 / 50

Learning Rate for this epoch: 0.001

Epoch [14/50], Iter [5/418] Loss: 3.4633, average\_loss: 2.7062  
 Epoch [14/50], Iter [10/418] Loss: 2.7888, average\_loss: 2.8300  
 Epoch [14/50], Iter [15/418] Loss: 2.7783, average\_loss: 2.9896  
 Epoch [14/50], Iter [20/418] Loss: 3.0265, average\_loss: 2.9516  
 Epoch [14/50], Iter [25/418] Loss: 2.3189, average\_loss: 2.9243  
 Epoch [14/50], Iter [30/418] Loss: 2.7391, average\_loss: 2.9722  
 Epoch [14/50], Iter [35/418] Loss: 2.4492, average\_loss: 2.9294  
 Epoch [14/50], Iter [40/418] Loss: 2.6503, average\_loss: 2.9155  
 Epoch [14/50], Iter [45/418] Loss: 3.9450, average\_loss: 2.9529  
 Epoch [14/50], Iter [50/418] Loss: 3.2003, average\_loss: 2.9450  
 Epoch [14/50], Iter [55/418] Loss: 2.9144, average\_loss: 2.9193  
 Epoch [14/50], Iter [60/418] Loss: 1.8065, average\_loss: 2.8874  
 Epoch [14/50], Iter [65/418] Loss: 2.5982, average\_loss: 2.8766  
 Epoch [14/50], Iter [70/418] Loss: 2.6138, average\_loss: 2.8367  
 Epoch [14/50], Iter [75/418] Loss: 2.8251, average\_loss: 2.8443  
 Epoch [14/50], Iter [80/418] Loss: 2.8860, average\_loss: 2.8331  
 Epoch [14/50], Iter [85/418] Loss: 2.4068, average\_loss: 2.8245  
 Epoch [14/50], Iter [90/418] Loss: 2.2678, average\_loss: 2.8254  
 Epoch [14/50], Iter [95/418] Loss: 5.5910, average\_loss: 2.8654  
 Epoch [14/50], Iter [100/418] Loss: 1.9827, average\_loss: 2.8494  
 Epoch [14/50], Iter [105/418] Loss: 2.5321, average\_loss: 2.8397  
 Epoch [14/50], Iter [110/418] Loss: 2.5111, average\_loss: 2.8286

Epoch [14/50], Iter [115/418] Loss: 2.5369, average\_loss: 2.8476  
 Epoch [14/50], Iter [120/418] Loss: 3.5136, average\_loss: 2.8458  
 Epoch [14/50], Iter [125/418] Loss: 2.1380, average\_loss: 2.8403  
 Epoch [14/50], Iter [130/418] Loss: 5.2203, average\_loss: 2.8387  
 Epoch [14/50], Iter [135/418] Loss: 2.5668, average\_loss: 2.8380  
 Epoch [14/50], Iter [140/418] Loss: 1.9077, average\_loss: 2.8439  
 Epoch [14/50], Iter [145/418] Loss: 2.1541, average\_loss: 2.8383  
 Epoch [14/50], Iter [150/418] Loss: 2.5495, average\_loss: 2.8310  
 Epoch [14/50], Iter [155/418] Loss: 2.7064, average\_loss: 2.8335  
 Epoch [14/50], Iter [160/418] Loss: 3.3721, average\_loss: 2.8352  
 Epoch [14/50], Iter [165/418] Loss: 2.8104, average\_loss: 2.8285  
 Epoch [14/50], Iter [170/418] Loss: 2.3448, average\_loss: 2.8280  
 Epoch [14/50], Iter [175/418] Loss: 2.3908, average\_loss: 2.8235  
 Epoch [14/50], Iter [180/418] Loss: 2.1084, average\_loss: 2.8280  
 Epoch [14/50], Iter [185/418] Loss: 2.5950, average\_loss: 2.8239  
 Epoch [14/50], Iter [190/418] Loss: 3.6189, average\_loss: 2.8171  
 Epoch [14/50], Iter [195/418] Loss: 3.6445, average\_loss: 2.8202  
 Epoch [14/50], Iter [200/418] Loss: 3.1686, average\_loss: 2.8077  
 Epoch [14/50], Iter [205/418] Loss: 3.1053, average\_loss: 2.8128  
 Epoch [14/50], Iter [210/418] Loss: 2.3833, average\_loss: 2.8082  
 Epoch [14/50], Iter [215/418] Loss: 2.9572, average\_loss: 2.7965  
 Epoch [14/50], Iter [220/418] Loss: 4.5599, average\_loss: 2.8044  
 Epoch [14/50], Iter [225/418] Loss: 2.2026, average\_loss: 2.7939  
 Epoch [14/50], Iter [230/418] Loss: 2.7919, average\_loss: 2.7872  
 Epoch [14/50], Iter [235/418] Loss: 2.5241, average\_loss: 2.7814  
 Epoch [14/50], Iter [240/418] Loss: 2.6544, average\_loss: 2.7781  
 Epoch [14/50], Iter [245/418] Loss: 2.1864, average\_loss: 2.7914  
 Epoch [14/50], Iter [250/418] Loss: 2.3758, average\_loss: 2.7989  
 Epoch [14/50], Iter [255/418] Loss: 3.1676, average\_loss: 2.8045  
 Epoch [14/50], Iter [260/418] Loss: 2.7739, average\_loss: 2.8008  
 Epoch [14/50], Iter [265/418] Loss: 3.6488, average\_loss: 2.8044  
 Epoch [14/50], Iter [270/418] Loss: 2.2164, average\_loss: 2.7982  
 Epoch [14/50], Iter [275/418] Loss: 2.1324, average\_loss: 2.7933  
 Epoch [14/50], Iter [280/418] Loss: 2.2735, average\_loss: 2.7894  
 Epoch [14/50], Iter [285/418] Loss: 4.6463, average\_loss: 2.7879  
 Epoch [14/50], Iter [290/418] Loss: 2.8907, average\_loss: 2.7892  
 Epoch [14/50], Iter [295/418] Loss: 4.0649, average\_loss: 2.8023  
 Epoch [14/50], Iter [300/418] Loss: 2.1223, average\_loss: 2.8045  
 Epoch [14/50], Iter [305/418] Loss: 2.4923, average\_loss: 2.7981  
 Epoch [14/50], Iter [310/418] Loss: 1.8756, average\_loss: 2.7954  
 Epoch [14/50], Iter [315/418] Loss: 2.2476, average\_loss: 2.7928  
 Epoch [14/50], Iter [320/418] Loss: 3.1495, average\_loss: 2.7934  
 Epoch [14/50], Iter [325/418] Loss: 1.7500, average\_loss: 2.7858  
 Epoch [14/50], Iter [330/418] Loss: 2.6856, average\_loss: 2.7936  
 Epoch [14/50], Iter [335/418] Loss: 1.8451, average\_loss: 2.7921  
 Epoch [14/50], Iter [340/418] Loss: 3.5707, average\_loss: 2.7943  
 Epoch [14/50], Iter [345/418] Loss: 5.3328, average\_loss: 2.7954  
 Epoch [14/50], Iter [350/418] Loss: 3.8429, average\_loss: 2.7979

Epoch [14/50], Iter [355/418] Loss: 3.2459, average\_loss: 2.7984  
 Epoch [14/50], Iter [360/418] Loss: 2.6892, average\_loss: 2.7949  
 Epoch [14/50], Iter [365/418] Loss: 3.3953, average\_loss: 2.7892  
 Epoch [14/50], Iter [370/418] Loss: 5.1694, average\_loss: 2.7929  
 Epoch [14/50], Iter [375/418] Loss: 2.0468, average\_loss: 2.7881  
 Epoch [14/50], Iter [380/418] Loss: 2.4253, average\_loss: 2.7922  
 Epoch [14/50], Iter [385/418] Loss: 1.9663, average\_loss: 2.7934  
 Epoch [14/50], Iter [390/418] Loss: 2.4924, average\_loss: 2.7939  
 Epoch [14/50], Iter [395/418] Loss: 2.9653, average\_loss: 2.7918  
 Epoch [14/50], Iter [400/418] Loss: 3.7740, average\_loss: 2.7946  
 Epoch [14/50], Iter [405/418] Loss: 2.0498, average\_loss: 2.7924  
 Epoch [14/50], Iter [410/418] Loss: 1.7532, average\_loss: 2.7865  
 Epoch [14/50], Iter [415/418] Loss: 2.0069, average\_loss: 2.7826  
 Updating best test loss: 3.02143

Starting epoch 15 / 50

Learning Rate for this epoch: 0.001

Epoch [15/50], Iter [5/418] Loss: 2.9903, average\_loss: 2.9179  
 Epoch [15/50], Iter [10/418] Loss: 2.1998, average\_loss: 2.8498  
 Epoch [15/50], Iter [15/418] Loss: 2.5347, average\_loss: 2.7443  
 Epoch [15/50], Iter [20/418] Loss: 2.5203, average\_loss: 2.6500  
 Epoch [15/50], Iter [25/418] Loss: 2.2794, average\_loss: 2.6091  
 Epoch [15/50], Iter [30/418] Loss: 3.2444, average\_loss: 2.6368  
 Epoch [15/50], Iter [35/418] Loss: 2.0029, average\_loss: 2.5846  
 Epoch [15/50], Iter [40/418] Loss: 2.7693, average\_loss: 2.6827  
 Epoch [15/50], Iter [45/418] Loss: 2.4292, average\_loss: 2.6569  
 Epoch [15/50], Iter [50/418] Loss: 1.6695, average\_loss: 2.6558  
 Epoch [15/50], Iter [55/418] Loss: 2.4928, average\_loss: 2.6595  
 Epoch [15/50], Iter [60/418] Loss: 2.4973, average\_loss: 2.6635  
 Epoch [15/50], Iter [65/418] Loss: 2.3093, average\_loss: 2.6370  
 Epoch [15/50], Iter [70/418] Loss: 3.6276, average\_loss: 2.6724  
 Epoch [15/50], Iter [75/418] Loss: 1.9375, average\_loss: 2.6541  
 Epoch [15/50], Iter [80/418] Loss: 2.7237, average\_loss: 2.6725  
 Epoch [15/50], Iter [85/418] Loss: 2.5848, average\_loss: 2.6745  
 Epoch [15/50], Iter [90/418] Loss: 2.4341, average\_loss: 2.6772  
 Epoch [15/50], Iter [95/418] Loss: 3.0072, average\_loss: 2.6757  
 Epoch [15/50], Iter [100/418] Loss: 1.9146, average\_loss: 2.6717  
 Epoch [15/50], Iter [105/418] Loss: 2.7048, average\_loss: 2.6738  
 Epoch [15/50], Iter [110/418] Loss: 3.1288, average\_loss: 2.6738  
 Epoch [15/50], Iter [115/418] Loss: 2.6286, average\_loss: 2.6647  
 Epoch [15/50], Iter [120/418] Loss: 2.9604, average\_loss: 2.6651  
 Epoch [15/50], Iter [125/418] Loss: 2.5621, average\_loss: 2.6533  
 Epoch [15/50], Iter [130/418] Loss: 2.3640, average\_loss: 2.6524  
 Epoch [15/50], Iter [135/418] Loss: 3.4843, average\_loss: 2.6464  
 Epoch [15/50], Iter [140/418] Loss: 1.7821, average\_loss: 2.6412  
 Epoch [15/50], Iter [145/418] Loss: 3.0540, average\_loss: 2.6470  
 Epoch [15/50], Iter [150/418] Loss: 1.6701, average\_loss: 2.6455



Epoch [15/50], Iter [155/418] Loss: 4.3128, average\_loss: 2.6502  
 Epoch [15/50], Iter [160/418] Loss: 2.1506, average\_loss: 2.6416  
 Epoch [15/50], Iter [165/418] Loss: 2.8709, average\_loss: 2.6406  
 Epoch [15/50], Iter [170/418] Loss: 3.6310, average\_loss: 2.6676  
 Epoch [15/50], Iter [175/418] Loss: 1.9218, average\_loss: 2.6627  
 Epoch [15/50], Iter [180/418] Loss: 2.3755, average\_loss: 2.6674  
 Epoch [15/50], Iter [185/418] Loss: 3.0594, average\_loss: 2.6568  
 Epoch [15/50], Iter [190/418] Loss: 2.6637, average\_loss: 2.6545  
 Epoch [15/50], Iter [195/418] Loss: 2.6048, average\_loss: 2.6624  
 Epoch [15/50], Iter [200/418] Loss: 2.0739, average\_loss: 2.6654  
 Epoch [15/50], Iter [205/418] Loss: 3.2533, average\_loss: 2.6614  
 Epoch [15/50], Iter [210/418] Loss: 1.8523, average\_loss: 2.6715  
 Epoch [15/50], Iter [215/418] Loss: 2.8519, average\_loss: 2.6674  
 Epoch [15/50], Iter [220/418] Loss: 3.4982, average\_loss: 2.6749  
 Epoch [15/50], Iter [225/418] Loss: 3.1978, average\_loss: 2.6730  
 Epoch [15/50], Iter [230/418] Loss: 3.3064, average\_loss: 2.6779  
 Epoch [15/50], Iter [235/418] Loss: 4.0198, average\_loss: 2.6754  
 Epoch [15/50], Iter [240/418] Loss: 2.7656, average\_loss: 2.6694  
 Epoch [15/50], Iter [245/418] Loss: 1.8808, average\_loss: 2.6656  
 Epoch [15/50], Iter [250/418] Loss: 1.7195, average\_loss: 2.6654  
 Epoch [15/50], Iter [255/418] Loss: 3.0620, average\_loss: 2.6623  
 Epoch [15/50], Iter [260/418] Loss: 2.9026, average\_loss: 2.6672  
 Epoch [15/50], Iter [265/418] Loss: 2.3262, average\_loss: 2.6625  
 Epoch [15/50], Iter [270/418] Loss: 2.6661, average\_loss: 2.6577  
 Epoch [15/50], Iter [275/418] Loss: 3.4505, average\_loss: 2.6646  
 Epoch [15/50], Iter [280/418] Loss: 4.9758, average\_loss: 2.6750  
 Epoch [15/50], Iter [285/418] Loss: 2.4734, average\_loss: 2.6742  
 Epoch [15/50], Iter [290/418] Loss: 3.0238, average\_loss: 2.6787  
 Epoch [15/50], Iter [295/418] Loss: 2.5458, average\_loss: 2.6750  
 Epoch [15/50], Iter [300/418] Loss: 2.9014, average\_loss: 2.6720  
 Epoch [15/50], Iter [305/418] Loss: 3.2794, average\_loss: 2.6774  
 Epoch [15/50], Iter [310/418] Loss: 2.5224, average\_loss: 2.6729  
 Epoch [15/50], Iter [315/418] Loss: 1.7032, average\_loss: 2.6698  
 Epoch [15/50], Iter [320/418] Loss: 2.9162, average\_loss: 2.6671  
 Epoch [15/50], Iter [325/418] Loss: 1.1658, average\_loss: 2.6602  
 Epoch [15/50], Iter [330/418] Loss: 1.8947, average\_loss: 2.6596  
 Epoch [15/50], Iter [335/418] Loss: 1.5833, average\_loss: 2.6588  
 Epoch [15/50], Iter [340/418] Loss: 2.1985, average\_loss: 2.6597  
 Epoch [15/50], Iter [345/418] Loss: 3.1961, average\_loss: 2.6559  
 Epoch [15/50], Iter [350/418] Loss: 2.3333, average\_loss: 2.6566  
 Epoch [15/50], Iter [355/418] Loss: 2.5961, average\_loss: 2.6589  
 Epoch [15/50], Iter [360/418] Loss: 2.5182, average\_loss: 2.6573  
 Epoch [15/50], Iter [365/418] Loss: 2.5668, average\_loss: 2.6587  
 Epoch [15/50], Iter [370/418] Loss: 3.0421, average\_loss: 2.6644  
 Epoch [15/50], Iter [375/418] Loss: 3.3871, average\_loss: 2.6760  
 Epoch [15/50], Iter [380/418] Loss: 1.8377, average\_loss: 2.6784  
 Epoch [15/50], Iter [385/418] Loss: 3.3863, average\_loss: 2.6808  
 Epoch [15/50], Iter [390/418] Loss: 2.7314, average\_loss: 2.6852

Epoch [15/50], Iter [395/418] Loss: 1.4375, average\_loss: 2.6874  
Epoch [15/50], Iter [400/418] Loss: 3.3686, average\_loss: 2.6868  
Epoch [15/50], Iter [405/418] Loss: 1.8185, average\_loss: 2.6867  
Epoch [15/50], Iter [410/418] Loss: 2.5019, average\_loss: 2.6828  
Epoch [15/50], Iter [415/418] Loss: 2.3895, average\_loss: 2.6805  
Updating best test loss: 2.99538

Starting epoch 16 / 50

Learning Rate for this epoch: 0.001

Epoch [16/50], Iter [5/418] Loss: 1.5844, average\_loss: 2.4308  
Epoch [16/50], Iter [10/418] Loss: 2.5345, average\_loss: 2.6363  
Epoch [16/50], Iter [15/418] Loss: 3.0253, average\_loss: 2.6513  
Epoch [16/50], Iter [20/418] Loss: 2.7795, average\_loss: 2.5989  
Epoch [16/50], Iter [25/418] Loss: 1.8818, average\_loss: 2.6298  
Epoch [16/50], Iter [30/418] Loss: 2.1925, average\_loss: 2.5427  
Epoch [16/50], Iter [35/418] Loss: 2.2895, average\_loss: 2.5199  
Epoch [16/50], Iter [40/418] Loss: 2.3546, average\_loss: 2.5024  
Epoch [16/50], Iter [45/418] Loss: 2.8777, average\_loss: 2.4998  
Epoch [16/50], Iter [50/418] Loss: 2.5116, average\_loss: 2.4975  
Epoch [16/50], Iter [55/418] Loss: 2.3673, average\_loss: 2.5031  
Epoch [16/50], Iter [60/418] Loss: 3.1089, average\_loss: 2.4888  
Epoch [16/50], Iter [65/418] Loss: 2.1940, average\_loss: 2.5085  
Epoch [16/50], Iter [70/418] Loss: 1.5719, average\_loss: 2.5253  
Epoch [16/50], Iter [75/418] Loss: 1.9155, average\_loss: 2.5365  
Epoch [16/50], Iter [80/418] Loss: 4.7089, average\_loss: 2.5745  
Epoch [16/50], Iter [85/418] Loss: 2.0631, average\_loss: 2.5953  
Epoch [16/50], Iter [90/418] Loss: 1.5742, average\_loss: 2.5759  
Epoch [16/50], Iter [95/418] Loss: 2.7026, average\_loss: 2.5630  
Epoch [16/50], Iter [100/418] Loss: 2.0873, average\_loss: 2.5975  
Epoch [16/50], Iter [105/418] Loss: 1.9913, average\_loss: 2.5961  
Epoch [16/50], Iter [110/418] Loss: 2.6189, average\_loss: 2.6220  
Epoch [16/50], Iter [115/418] Loss: 2.0923, average\_loss: 2.6089  
Epoch [16/50], Iter [120/418] Loss: 2.6308, average\_loss: 2.6138  
Epoch [16/50], Iter [125/418] Loss: 2.2106, average\_loss: 2.6248  
Epoch [16/50], Iter [130/418] Loss: 2.4900, average\_loss: 2.6201  
Epoch [16/50], Iter [135/418] Loss: 3.9962, average\_loss: 2.6182  
Epoch [16/50], Iter [140/418] Loss: 3.1213, average\_loss: 2.6229  
Epoch [16/50], Iter [145/418] Loss: 1.9086, average\_loss: 2.6313  
Epoch [16/50], Iter [150/418] Loss: 2.1584, average\_loss: 2.6452  
Epoch [16/50], Iter [155/418] Loss: 1.5105, average\_loss: 2.6330  
Epoch [16/50], Iter [160/418] Loss: 2.0077, average\_loss: 2.6134  
Epoch [16/50], Iter [165/418] Loss: 1.0132, average\_loss: 2.6105  
Epoch [16/50], Iter [170/418] Loss: 1.5500, average\_loss: 2.5943  
Epoch [16/50], Iter [175/418] Loss: 2.2440, average\_loss: 2.5987  
Epoch [16/50], Iter [180/418] Loss: 2.7324, average\_loss: 2.5987  
Epoch [16/50], Iter [185/418] Loss: 2.3496, average\_loss: 2.6027  
Epoch [16/50], Iter [190/418] Loss: 2.4032, average\_loss: 2.6041

```

Epoch [16/50], Iter [195/418] Loss: 2.5855, average_loss: 2.6070
Epoch [16/50], Iter [200/418] Loss: 3.6567, average_loss: 2.6094
Epoch [16/50], Iter [205/418] Loss: 2.5860, average_loss: 2.6056
Epoch [16/50], Iter [210/418] Loss: 2.4169, average_loss: 2.6137
Epoch [16/50], Iter [215/418] Loss: 3.9758, average_loss: 2.6078
Epoch [16/50], Iter [220/418] Loss: 2.5228, average_loss: 2.6114
Epoch [16/50], Iter [225/418] Loss: 1.8323, average_loss: 2.6169
Epoch [16/50], Iter [230/418] Loss: 3.5167, average_loss: 2.6217
Epoch [16/50], Iter [235/418] Loss: 1.9479, average_loss: 2.6232
Epoch [16/50], Iter [240/418] Loss: 2.5396, average_loss: 2.6231
Epoch [16/50], Iter [245/418] Loss: 1.9538, average_loss: 2.6189
Epoch [16/50], Iter [250/418] Loss: 2.8569, average_loss: 2.6164
Epoch [16/50], Iter [255/418] Loss: 1.6888, average_loss: 2.6063
Epoch [16/50], Iter [260/418] Loss: 1.6370, average_loss: 2.5986
Epoch [16/50], Iter [265/418] Loss: 2.7376, average_loss: 2.5997
Epoch [16/50], Iter [270/418] Loss: 3.9901, average_loss: 2.5947
Epoch [16/50], Iter [275/418] Loss: 2.8663, average_loss: 2.5927
Epoch [16/50], Iter [280/418] Loss: 3.3426, average_loss: 2.5927
Epoch [16/50], Iter [285/418] Loss: 1.9148, average_loss: 2.5841
Epoch [16/50], Iter [290/418] Loss: 3.2239, average_loss: 2.5829
Epoch [16/50], Iter [295/418] Loss: 2.1444, average_loss: 2.5917
Epoch [16/50], Iter [300/418] Loss: 2.4399, average_loss: 2.5871
Epoch [16/50], Iter [305/418] Loss: 4.2137, average_loss: 2.5899
Epoch [16/50], Iter [310/418] Loss: 2.1394, average_loss: 2.5874
Epoch [16/50], Iter [315/418] Loss: 2.5305, average_loss: 2.5936
Epoch [16/50], Iter [320/418] Loss: 2.6605, average_loss: 2.5972
Epoch [16/50], Iter [325/418] Loss: 3.5224, average_loss: 2.6061
Epoch [16/50], Iter [330/418] Loss: 3.0641, average_loss: 2.6027
Epoch [16/50], Iter [335/418] Loss: 3.0243, average_loss: 2.6073
Epoch [16/50], Iter [340/418] Loss: 3.1290, average_loss: 2.6075
Epoch [16/50], Iter [345/418] Loss: 2.2111, average_loss: 2.6087
Epoch [16/50], Iter [350/418] Loss: 2.7636, average_loss: 2.6068
Epoch [16/50], Iter [355/418] Loss: 2.9779, average_loss: 2.6070
Epoch [16/50], Iter [360/418] Loss: 2.6765, average_loss: 2.6080
Epoch [16/50], Iter [365/418] Loss: 2.9521, average_loss: 2.6104
Epoch [16/50], Iter [370/418] Loss: 2.0968, average_loss: 2.6116
Epoch [16/50], Iter [375/418] Loss: 2.4787, average_loss: 2.6082
Epoch [16/50], Iter [380/418] Loss: 3.2447, average_loss: 2.6116
Epoch [16/50], Iter [385/418] Loss: 2.3042, average_loss: 2.6129
Epoch [16/50], Iter [390/418] Loss: 3.1116, average_loss: 2.6148
Epoch [16/50], Iter [395/418] Loss: 3.5006, average_loss: 2.6142
Epoch [16/50], Iter [400/418] Loss: 3.9883, average_loss: 2.6205
Epoch [16/50], Iter [405/418] Loss: 3.4286, average_loss: 2.6288
Epoch [16/50], Iter [410/418] Loss: 1.8302, average_loss: 2.6238
Epoch [16/50], Iter [415/418] Loss: 1.8936, average_loss: 2.6209
---Evaluate model on test samples---

```

100%|

| 4950/4950 [02:30<00:00, 32.88it/s]

---class aeroplane ap 0.37285237585051584---  
---class bicycle ap 0.4676476861537768---  
---class bird ap 0.3666061847331922---  
---class boat ap 0.1720380810682079---  
---class bottle ap 0.08349009627502596---  
---class bus ap 0.4795882875416946---  
---class car ap 0.5241351875104908---  
---class cat ap 0.5858728484348155---  
---class chair ap 0.25888455849634606---  
---class cow ap 0.30679627675630644---  
---class diningtable ap 0.2560731423059066---  
---class dog ap 0.4517628216903178---  
---class horse ap 0.5170806115703241---  
---class motorbike ap 0.465489037802489---  
---class person ap 0.4373157071527014---  
---class pottedplant ap 0.09873670510741604---  
---class sheep ap 0.31648684428543783---  
---class sofa ap 0.3730579164937942---  
---class train ap 0.5454691090488633---  
---class tvmonitor ap 0.39109766651241157---  
---map 0.37352405723950166---

Updating best test loss: 2.93706

Starting epoch 17 / 50

Learning Rate for this epoch: 0.001

Epoch [17/50], Iter [5/418] Loss: 2.0921, average\_loss: 2.4741  
Epoch [17/50], Iter [10/418] Loss: 2.8573, average\_loss: 2.8303  
Epoch [17/50], Iter [15/418] Loss: 2.5762, average\_loss: 2.7388  
Epoch [17/50], Iter [20/418] Loss: 2.8573, average\_loss: 2.7520  
Epoch [17/50], Iter [25/418] Loss: 1.6296, average\_loss: 2.6467  
Epoch [17/50], Iter [30/418] Loss: 2.0204, average\_loss: 2.6215  
Epoch [17/50], Iter [35/418] Loss: 2.5763, average\_loss: 2.6031  
Epoch [17/50], Iter [40/418] Loss: 1.3771, average\_loss: 2.5860  
Epoch [17/50], Iter [45/418] Loss: 4.2348, average\_loss: 2.6901  
Epoch [17/50], Iter [50/418] Loss: 2.8086, average\_loss: 2.6960  
Epoch [17/50], Iter [55/418] Loss: 1.9981, average\_loss: 2.6780  
Epoch [17/50], Iter [60/418] Loss: 3.9565, average\_loss: 2.7212  
Epoch [17/50], Iter [65/418] Loss: 2.7510, average\_loss: 2.6898  
Epoch [17/50], Iter [70/418] Loss: 2.3195, average\_loss: 2.6836  
Epoch [17/50], Iter [75/418] Loss: 1.4008, average\_loss: 2.6666  
Epoch [17/50], Iter [80/418] Loss: 2.1534, average\_loss: 2.6465  
Epoch [17/50], Iter [85/418] Loss: 1.7191, average\_loss: 2.6260  
Epoch [17/50], Iter [90/418] Loss: 1.4560, average\_loss: 2.6208  
Epoch [17/50], Iter [95/418] Loss: 3.1183, average\_loss: 2.6134  
Epoch [17/50], Iter [100/418] Loss: 2.2779, average\_loss: 2.5938

Epoch [17/50], Iter [105/418] Loss: 3.3030, average\_loss: 2.5873  
 Epoch [17/50], Iter [110/418] Loss: 2.5072, average\_loss: 2.5700  
 Epoch [17/50], Iter [115/418] Loss: 2.8695, average\_loss: 2.5807  
 Epoch [17/50], Iter [120/418] Loss: 3.8263, average\_loss: 2.5883  
 Epoch [17/50], Iter [125/418] Loss: 1.9824, average\_loss: 2.5856  
 Epoch [17/50], Iter [130/418] Loss: 2.3573, average\_loss: 2.5785  
 Epoch [17/50], Iter [135/418] Loss: 2.9498, average\_loss: 2.5814  
 Epoch [17/50], Iter [140/418] Loss: 3.8557, average\_loss: 2.5993  
 Epoch [17/50], Iter [145/418] Loss: 1.7005, average\_loss: 2.5834  
 Epoch [17/50], Iter [150/418] Loss: 2.0221, average\_loss: 2.6012  
 Epoch [17/50], Iter [155/418] Loss: 3.4918, average\_loss: 2.6079  
 Epoch [17/50], Iter [160/418] Loss: 2.1350, average\_loss: 2.6006  
 Epoch [17/50], Iter [165/418] Loss: 3.4610, average\_loss: 2.6123  
 Epoch [17/50], Iter [170/418] Loss: 1.6251, average\_loss: 2.5974  
 Epoch [17/50], Iter [175/418] Loss: 1.9530, average\_loss: 2.5871  
 Epoch [17/50], Iter [180/418] Loss: 2.1309, average\_loss: 2.5847  
 Epoch [17/50], Iter [185/418] Loss: 2.9332, average\_loss: 2.5769  
 Epoch [17/50], Iter [190/418] Loss: 1.4167, average\_loss: 2.5796  
 Epoch [17/50], Iter [195/418] Loss: 2.5298, average\_loss: 2.5983  
 Epoch [17/50], Iter [200/418] Loss: 1.3989, average\_loss: 2.5835  
 Epoch [17/50], Iter [205/418] Loss: 1.9427, average\_loss: 2.5705  
 Epoch [17/50], Iter [210/418] Loss: 2.0697, average\_loss: 2.5589  
 Epoch [17/50], Iter [215/418] Loss: 2.4380, average\_loss: 2.5690  
 Epoch [17/50], Iter [220/418] Loss: 2.4245, average\_loss: 2.5626  
 Epoch [17/50], Iter [225/418] Loss: 2.4572, average\_loss: 2.5532  
 Epoch [17/50], Iter [230/418] Loss: 2.0440, average\_loss: 2.5666  
 Epoch [17/50], Iter [235/418] Loss: 2.6237, average\_loss: 2.5645  
 Epoch [17/50], Iter [240/418] Loss: 3.0622, average\_loss: 2.5632  
 Epoch [17/50], Iter [245/418] Loss: 1.7469, average\_loss: 2.5647  
 Epoch [17/50], Iter [250/418] Loss: 2.5335, average\_loss: 2.5557  
 Epoch [17/50], Iter [255/418] Loss: 2.1332, average\_loss: 2.5554  
 Epoch [17/50], Iter [260/418] Loss: 2.0675, average\_loss: 2.5553  
 Epoch [17/50], Iter [265/418] Loss: 1.8079, average\_loss: 2.5535  
 Epoch [17/50], Iter [270/418] Loss: 1.5169, average\_loss: 2.5479  
 Epoch [17/50], Iter [275/418] Loss: 2.8573, average\_loss: 2.5576  
 Epoch [17/50], Iter [280/418] Loss: 2.7068, average\_loss: 2.5610  
 Epoch [17/50], Iter [285/418] Loss: 2.1030, average\_loss: 2.5544  
 Epoch [17/50], Iter [290/418] Loss: 2.2978, average\_loss: 2.5460  
 Epoch [17/50], Iter [295/418] Loss: 2.8462, average\_loss: 2.5473  
 Epoch [17/50], Iter [300/418] Loss: 3.1398, average\_loss: 2.5486  
 Epoch [17/50], Iter [305/418] Loss: 1.9417, average\_loss: 2.5435  
 Epoch [17/50], Iter [310/418] Loss: 3.1429, average\_loss: 2.5566  
 Epoch [17/50], Iter [315/418] Loss: 2.4001, average\_loss: 2.5524  
 Epoch [17/50], Iter [320/418] Loss: 2.9842, average\_loss: 2.5517  
 Epoch [17/50], Iter [325/418] Loss: 4.2073, average\_loss: 2.5567  
 Epoch [17/50], Iter [330/418] Loss: 1.8434, average\_loss: 2.5574  
 Epoch [17/50], Iter [335/418] Loss: 2.3487, average\_loss: 2.5583  
 Epoch [17/50], Iter [340/418] Loss: 2.7264, average\_loss: 2.5606

Epoch [17/50], Iter [345/418] Loss: 2.1716, average\_loss: 2.5674  
 Epoch [17/50], Iter [350/418] Loss: 3.4457, average\_loss: 2.5719  
 Epoch [17/50], Iter [355/418] Loss: 3.7197, average\_loss: 2.5805  
 Epoch [17/50], Iter [360/418] Loss: 1.9915, average\_loss: 2.5794  
 Epoch [17/50], Iter [365/418] Loss: 2.2192, average\_loss: 2.5746  
 Epoch [17/50], Iter [370/418] Loss: 1.6026, average\_loss: 2.5711  
 Epoch [17/50], Iter [375/418] Loss: 3.7670, average\_loss: 2.5764  
 Epoch [17/50], Iter [380/418] Loss: 3.0137, average\_loss: 2.5775  
 Epoch [17/50], Iter [385/418] Loss: 2.2261, average\_loss: 2.5773  
 Epoch [17/50], Iter [390/418] Loss: 2.7243, average\_loss: 2.5811  
 Epoch [17/50], Iter [395/418] Loss: 3.2194, average\_loss: 2.5861  
 Epoch [17/50], Iter [400/418] Loss: 1.9714, average\_loss: 2.5885  
 Epoch [17/50], Iter [405/418] Loss: 1.9493, average\_loss: 2.5810  
 Epoch [17/50], Iter [410/418] Loss: 2.7654, average\_loss: 2.5761  
 Epoch [17/50], Iter [415/418] Loss: 3.8404, average\_loss: 2.5733  
 Updating best test loss: 2.90203

Starting epoch 18 / 50

Learning Rate for this epoch: 0.001

Epoch [18/50], Iter [5/418] Loss: 2.4903, average\_loss: 2.0634  
 Epoch [18/50], Iter [10/418] Loss: 2.8222, average\_loss: 2.5463  
 Epoch [18/50], Iter [15/418] Loss: 1.7415, average\_loss: 2.5294  
 Epoch [18/50], Iter [20/418] Loss: 1.9393, average\_loss: 2.4423  
 Epoch [18/50], Iter [25/418] Loss: 2.1203, average\_loss: 2.4406  
 Epoch [18/50], Iter [30/418] Loss: 2.2769, average\_loss: 2.4413  
 Epoch [18/50], Iter [35/418] Loss: 2.6310, average\_loss: 2.5242  
 Epoch [18/50], Iter [40/418] Loss: 1.3604, average\_loss: 2.4842  
 Epoch [18/50], Iter [45/418] Loss: 2.0737, average\_loss: 2.4641  
 Epoch [18/50], Iter [50/418] Loss: 2.4065, average\_loss: 2.4419  
 Epoch [18/50], Iter [55/418] Loss: 1.6487, average\_loss: 2.4164  
 Epoch [18/50], Iter [60/418] Loss: 3.6292, average\_loss: 2.4315  
 Epoch [18/50], Iter [65/418] Loss: 3.0444, average\_loss: 2.4946  
 Epoch [18/50], Iter [70/418] Loss: 1.6520, average\_loss: 2.4970  
 Epoch [18/50], Iter [75/418] Loss: 2.1290, average\_loss: 2.4866  
 Epoch [18/50], Iter [80/418] Loss: 1.8136, average\_loss: 2.4915  
 Epoch [18/50], Iter [85/418] Loss: 2.4590, average\_loss: 2.4753  
 Epoch [18/50], Iter [90/418] Loss: 2.8822, average\_loss: 2.4912  
 Epoch [18/50], Iter [95/418] Loss: 2.0381, average\_loss: 2.4930  
 Epoch [18/50], Iter [100/418] Loss: 3.1720, average\_loss: 2.4891  
 Epoch [18/50], Iter [105/418] Loss: 2.2790, average\_loss: 2.4871  
 Epoch [18/50], Iter [110/418] Loss: 4.4561, average\_loss: 2.4875  
 Epoch [18/50], Iter [115/418] Loss: 1.8488, average\_loss: 2.4847  
 Epoch [18/50], Iter [120/418] Loss: 3.3024, average\_loss: 2.4878  
 Epoch [18/50], Iter [125/418] Loss: 3.8795, average\_loss: 2.5069  
 Epoch [18/50], Iter [130/418] Loss: 2.5410, average\_loss: 2.5068  
 Epoch [18/50], Iter [135/418] Loss: 3.4707, average\_loss: 2.5205  
 Epoch [18/50], Iter [140/418] Loss: 1.3537, average\_loss: 2.5033

Epoch [18/50], Iter [145/418] Loss: 2.0425, average\_loss: 2.5008  
 Epoch [18/50], Iter [150/418] Loss: 2.2185, average\_loss: 2.5109  
 Epoch [18/50], Iter [155/418] Loss: 2.5697, average\_loss: 2.5058  
 Epoch [18/50], Iter [160/418] Loss: 2.5902, average\_loss: 2.5065  
 Epoch [18/50], Iter [165/418] Loss: 1.9834, average\_loss: 2.5034  
 Epoch [18/50], Iter [170/418] Loss: 2.3721, average\_loss: 2.4970  
 Epoch [18/50], Iter [175/418] Loss: 2.8074, average\_loss: 2.4945  
 Epoch [18/50], Iter [180/418] Loss: 2.7699, average\_loss: 2.5008  
 Epoch [18/50], Iter [185/418] Loss: 1.9149, average\_loss: 2.4998  
 Epoch [18/50], Iter [190/418] Loss: 2.5321, average\_loss: 2.4931  
 Epoch [18/50], Iter [195/418] Loss: 4.2122, average\_loss: 2.5030  
 Epoch [18/50], Iter [200/418] Loss: 2.3172, average\_loss: 2.5090  
 Epoch [18/50], Iter [205/418] Loss: 2.1862, average\_loss: 2.4928  
 Epoch [18/50], Iter [210/418] Loss: 2.5889, average\_loss: 2.5123  
 Epoch [18/50], Iter [215/418] Loss: 2.3813, average\_loss: 2.5254  
 Epoch [18/50], Iter [220/418] Loss: 3.2377, average\_loss: 2.5263  
 Epoch [18/50], Iter [225/418] Loss: 2.0341, average\_loss: 2.5240  
 Epoch [18/50], Iter [230/418] Loss: 3.7777, average\_loss: 2.5226  
 Epoch [18/50], Iter [235/418] Loss: 1.6341, average\_loss: 2.5177  
 Epoch [18/50], Iter [240/418] Loss: 2.0522, average\_loss: 2.5153  
 Epoch [18/50], Iter [245/418] Loss: 2.4781, average\_loss: 2.5118  
 Epoch [18/50], Iter [250/418] Loss: 2.4303, average\_loss: 2.5182  
 Epoch [18/50], Iter [255/418] Loss: 1.8408, average\_loss: 2.5191  
 Epoch [18/50], Iter [260/418] Loss: 1.9669, average\_loss: 2.5141  
 Epoch [18/50], Iter [265/418] Loss: 1.8889, average\_loss: 2.5162  
 Epoch [18/50], Iter [270/418] Loss: 3.9539, average\_loss: 2.5282  
 Epoch [18/50], Iter [275/418] Loss: 2.8446, average\_loss: 2.5275  
 Epoch [18/50], Iter [280/418] Loss: 3.2818, average\_loss: 2.5361  
 Epoch [18/50], Iter [285/418] Loss: 1.6437, average\_loss: 2.5414  
 Epoch [18/50], Iter [290/418] Loss: 2.7429, average\_loss: 2.5385  
 Epoch [18/50], Iter [295/418] Loss: 2.1624, average\_loss: 2.5328  
 Epoch [18/50], Iter [300/418] Loss: 3.2334, average\_loss: 2.5371  
 Epoch [18/50], Iter [305/418] Loss: 2.3408, average\_loss: 2.5363  
 Epoch [18/50], Iter [310/418] Loss: 2.3869, average\_loss: 2.5415  
 Epoch [18/50], Iter [315/418] Loss: 2.6510, average\_loss: 2.5390  
 Epoch [18/50], Iter [320/418] Loss: 1.9843, average\_loss: 2.5358  
 Epoch [18/50], Iter [325/418] Loss: 2.2194, average\_loss: 2.5281  
 Epoch [18/50], Iter [330/418] Loss: 2.1694, average\_loss: 2.5276  
 Epoch [18/50], Iter [335/418] Loss: 1.6669, average\_loss: 2.5201  
 Epoch [18/50], Iter [340/418] Loss: 2.0950, average\_loss: 2.5168  
 Epoch [18/50], Iter [345/418] Loss: 2.3597, average\_loss: 2.5124  
 Epoch [18/50], Iter [350/418] Loss: 1.9115, average\_loss: 2.5121  
 Epoch [18/50], Iter [355/418] Loss: 2.8217, average\_loss: 2.5164  
 Epoch [18/50], Iter [360/418] Loss: 1.7096, average\_loss: 2.5087  
 Epoch [18/50], Iter [365/418] Loss: 2.2981, average\_loss: 2.5049  
 Epoch [18/50], Iter [370/418] Loss: 2.1767, average\_loss: 2.5046  
 Epoch [18/50], Iter [375/418] Loss: 2.8641, average\_loss: 2.5045  
 Epoch [18/50], Iter [380/418] Loss: 2.2922, average\_loss: 2.5039

Epoch [18/50], Iter [385/418] Loss: 3.1392, average\_loss: 2.5076  
Epoch [18/50], Iter [390/418] Loss: 1.9969, average\_loss: 2.5052  
Epoch [18/50], Iter [395/418] Loss: 2.5488, average\_loss: 2.5055  
Epoch [18/50], Iter [400/418] Loss: 2.7426, average\_loss: 2.5128  
Epoch [18/50], Iter [405/418] Loss: 4.0667, average\_loss: 2.5140  
Epoch [18/50], Iter [410/418] Loss: 2.6550, average\_loss: 2.5136  
Epoch [18/50], Iter [415/418] Loss: 1.8604, average\_loss: 2.5143  
Updating best test loss: 2.88991

Starting epoch 19 / 50

Learning Rate for this epoch: 0.001

Epoch [19/50], Iter [5/418] Loss: 3.1672, average\_loss: 2.3272  
Epoch [19/50], Iter [10/418] Loss: 1.6327, average\_loss: 2.0746  
Epoch [19/50], Iter [15/418] Loss: 2.4612, average\_loss: 2.2304  
Epoch [19/50], Iter [20/418] Loss: 1.7546, average\_loss: 2.2278  
Epoch [19/50], Iter [25/418] Loss: 1.9202, average\_loss: 2.2083  
Epoch [19/50], Iter [30/418] Loss: 2.8967, average\_loss: 2.2688  
Epoch [19/50], Iter [35/418] Loss: 3.4363, average\_loss: 2.3087  
Epoch [19/50], Iter [40/418] Loss: 1.7639, average\_loss: 2.3036  
Epoch [19/50], Iter [45/418] Loss: 2.3249, average\_loss: 2.3468  
Epoch [19/50], Iter [50/418] Loss: 1.8202, average\_loss: 2.3656  
Epoch [19/50], Iter [55/418] Loss: 2.2361, average\_loss: 2.3742  
Epoch [19/50], Iter [60/418] Loss: 1.6857, average\_loss: 2.3947  
Epoch [19/50], Iter [65/418] Loss: 3.8789, average\_loss: 2.3944  
Epoch [19/50], Iter [70/418] Loss: 2.4561, average\_loss: 2.3933  
Epoch [19/50], Iter [75/418] Loss: 2.2228, average\_loss: 2.4383  
Epoch [19/50], Iter [80/418] Loss: 2.5868, average\_loss: 2.4405  
Epoch [19/50], Iter [85/418] Loss: 2.2058, average\_loss: 2.4283  
Epoch [19/50], Iter [90/418] Loss: 3.7433, average\_loss: 2.4501  
Epoch [19/50], Iter [95/418] Loss: 2.5320, average\_loss: 2.4833  
Epoch [19/50], Iter [100/418] Loss: 1.8767, average\_loss: 2.4644  
Epoch [19/50], Iter [105/418] Loss: 3.6664, average\_loss: 2.4692  
Epoch [19/50], Iter [110/418] Loss: 2.4485, average\_loss: 2.4905  
Epoch [19/50], Iter [115/418] Loss: 2.3279, average\_loss: 2.4918  
Epoch [19/50], Iter [120/418] Loss: 2.9731, average\_loss: 2.5024  
Epoch [19/50], Iter [125/418] Loss: 2.1141, average\_loss: 2.4775  
Epoch [19/50], Iter [130/418] Loss: 3.0804, average\_loss: 2.4796  
Epoch [19/50], Iter [135/418] Loss: 2.1259, average\_loss: 2.4780  
Epoch [19/50], Iter [140/418] Loss: 2.1390, average\_loss: 2.4622  
Epoch [19/50], Iter [145/418] Loss: 3.2340, average\_loss: 2.4705  
Epoch [19/50], Iter [150/418] Loss: 4.1149, average\_loss: 2.4758  
Epoch [19/50], Iter [155/418] Loss: 2.1794, average\_loss: 2.4707  
Epoch [19/50], Iter [160/418] Loss: 3.7212, average\_loss: 2.4825  
Epoch [19/50], Iter [165/418] Loss: 2.0644, average\_loss: 2.4675  
Epoch [19/50], Iter [170/418] Loss: 2.0292, average\_loss: 2.4688  
Epoch [19/50], Iter [175/418] Loss: 2.5831, average\_loss: 2.4678  
Epoch [19/50], Iter [180/418] Loss: 2.5521, average\_loss: 2.4683



Epoch [19/50], Iter [185/418] Loss: 2.0368, average\_loss: 2.4683  
Epoch [19/50], Iter [190/418] Loss: 3.5900, average\_loss: 2.4658  
Epoch [19/50], Iter [195/418] Loss: 1.5326, average\_loss: 2.4618  
Epoch [19/50], Iter [200/418] Loss: 2.6996, average\_loss: 2.4552  
Epoch [19/50], Iter [205/418] Loss: 2.2689, average\_loss: 2.4539  
Epoch [19/50], Iter [210/418] Loss: 2.0608, average\_loss: 2.4516  
Epoch [19/50], Iter [215/418] Loss: 2.3917, average\_loss: 2.4436  
Epoch [19/50], Iter [220/418] Loss: 2.4782, average\_loss: 2.4467  
Epoch [19/50], Iter [225/418] Loss: 2.1376, average\_loss: 2.4453  
Epoch [19/50], Iter [230/418] Loss: 2.3149, average\_loss: 2.4611  
Epoch [19/50], Iter [235/418] Loss: 3.6353, average\_loss: 2.4643  
Epoch [19/50], Iter [240/418] Loss: 2.0991, average\_loss: 2.4602  
Epoch [19/50], Iter [245/418] Loss: 2.8051, average\_loss: 2.4560  
Epoch [19/50], Iter [250/418] Loss: 3.0367, average\_loss: 2.4598  
Epoch [19/50], Iter [255/418] Loss: 1.9752, average\_loss: 2.4545  
Epoch [19/50], Iter [260/418] Loss: 2.3196, average\_loss: 2.4490  
Epoch [19/50], Iter [265/418] Loss: 2.5416, average\_loss: 2.4562  
Epoch [19/50], Iter [270/418] Loss: 2.2343, average\_loss: 2.4585  
Epoch [19/50], Iter [275/418] Loss: 1.8218, average\_loss: 2.4544  
Epoch [19/50], Iter [280/418] Loss: 2.3802, average\_loss: 2.4531  
Epoch [19/50], Iter [285/418] Loss: 1.7069, average\_loss: 2.4478  
Epoch [19/50], Iter [290/418] Loss: 3.0032, average\_loss: 2.4524  
Epoch [19/50], Iter [295/418] Loss: 1.5628, average\_loss: 2.4448  
Epoch [19/50], Iter [300/418] Loss: 2.3522, average\_loss: 2.4460  
Epoch [19/50], Iter [305/418] Loss: 1.9514, average\_loss: 2.4375  
Epoch [19/50], Iter [310/418] Loss: 2.3529, average\_loss: 2.4378  
Epoch [19/50], Iter [315/418] Loss: 1.5209, average\_loss: 2.4351  
Epoch [19/50], Iter [320/418] Loss: 2.3112, average\_loss: 2.4332  
Epoch [19/50], Iter [325/418] Loss: 1.5918, average\_loss: 2.4282  
Epoch [19/50], Iter [330/418] Loss: 1.8559, average\_loss: 2.4313  
Epoch [19/50], Iter [335/418] Loss: 1.2344, average\_loss: 2.4261  
Epoch [19/50], Iter [340/418] Loss: 1.7193, average\_loss: 2.4245  
Epoch [19/50], Iter [345/418] Loss: 1.4713, average\_loss: 2.4278  
Epoch [19/50], Iter [350/418] Loss: 1.9177, average\_loss: 2.4263  
Epoch [19/50], Iter [355/418] Loss: 2.5822, average\_loss: 2.4321  
Epoch [19/50], Iter [360/418] Loss: 2.0580, average\_loss: 2.4302  
Epoch [19/50], Iter [365/418] Loss: 2.6328, average\_loss: 2.4335  
Epoch [19/50], Iter [370/418] Loss: 2.6210, average\_loss: 2.4297  
Epoch [19/50], Iter [375/418] Loss: 3.0566, average\_loss: 2.4332  
Epoch [19/50], Iter [380/418] Loss: 2.4982, average\_loss: 2.4315  
Epoch [19/50], Iter [385/418] Loss: 2.5095, average\_loss: 2.4308  
Epoch [19/50], Iter [390/418] Loss: 1.8211, average\_loss: 2.4296  
Epoch [19/50], Iter [395/418] Loss: 2.4225, average\_loss: 2.4309  
Epoch [19/50], Iter [400/418] Loss: 2.3501, average\_loss: 2.4321  
Epoch [19/50], Iter [405/418] Loss: 2.2268, average\_loss: 2.4291  
Epoch [19/50], Iter [410/418] Loss: 2.0651, average\_loss: 2.4292  
Epoch [19/50], Iter [415/418] Loss: 2.0902, average\_loss: 2.4264  
Updating best test loss: 2.87671

Starting epoch 20 / 50

Learning Rate for this epoch: 0.001

Epoch [20/50],	Iter [5/418]	Loss: 1.5387,	average_loss: 2.2430
Epoch [20/50],	Iter [10/418]	Loss: 1.9424,	average_loss: 2.1332
Epoch [20/50],	Iter [15/418]	Loss: 2.2422,	average_loss: 2.1013
Epoch [20/50],	Iter [20/418]	Loss: 3.1774,	average_loss: 2.1915
Epoch [20/50],	Iter [25/418]	Loss: 3.8811,	average_loss: 2.2944
Epoch [20/50],	Iter [30/418]	Loss: 2.8206,	average_loss: 2.3778
Epoch [20/50],	Iter [35/418]	Loss: 2.6084,	average_loss: 2.3997
Epoch [20/50],	Iter [40/418]	Loss: 4.0064,	average_loss: 2.4314
Epoch [20/50],	Iter [45/418]	Loss: 2.4026,	average_loss: 2.3809
Epoch [20/50],	Iter [50/418]	Loss: 1.9199,	average_loss: 2.4525
Epoch [20/50],	Iter [55/418]	Loss: 2.1946,	average_loss: 2.4290
Epoch [20/50],	Iter [60/418]	Loss: 2.5875,	average_loss: 2.4217
Epoch [20/50],	Iter [65/418]	Loss: 3.0883,	average_loss: 2.4820
Epoch [20/50],	Iter [70/418]	Loss: 1.9362,	average_loss: 2.4805
Epoch [20/50],	Iter [75/418]	Loss: 2.1763,	average_loss: 2.4740
Epoch [20/50],	Iter [80/418]	Loss: 2.2272,	average_loss: 2.4697
Epoch [20/50],	Iter [85/418]	Loss: 3.1537,	average_loss: 2.4834
Epoch [20/50],	Iter [90/418]	Loss: 2.0290,	average_loss: 2.4661
Epoch [20/50],	Iter [95/418]	Loss: 3.6521,	average_loss: 2.4715
Epoch [20/50],	Iter [100/418]	Loss: 1.8855,	average_loss: 2.4761
Epoch [20/50],	Iter [105/418]	Loss: 1.8031,	average_loss: 2.4590
Epoch [20/50],	Iter [110/418]	Loss: 3.3371,	average_loss: 2.4570
Epoch [20/50],	Iter [115/418]	Loss: 2.5795,	average_loss: 2.4610
Epoch [20/50],	Iter [120/418]	Loss: 3.0435,	average_loss: 2.4500
Epoch [20/50],	Iter [125/418]	Loss: 1.4460,	average_loss: 2.4505
Epoch [20/50],	Iter [130/418]	Loss: 2.5939,	average_loss: 2.4415
Epoch [20/50],	Iter [135/418]	Loss: 2.8190,	average_loss: 2.4450
Epoch [20/50],	Iter [140/418]	Loss: 2.3647,	average_loss: 2.4480
Epoch [20/50],	Iter [145/418]	Loss: 1.8641,	average_loss: 2.4329
Epoch [20/50],	Iter [150/418]	Loss: 2.8765,	average_loss: 2.4533
Epoch [20/50],	Iter [155/418]	Loss: 3.3220,	average_loss: 2.4571
Epoch [20/50],	Iter [160/418]	Loss: 2.0721,	average_loss: 2.4528
Epoch [20/50],	Iter [165/418]	Loss: 1.9828,	average_loss: 2.4660
Epoch [20/50],	Iter [170/418]	Loss: 2.3034,	average_loss: 2.4631
Epoch [20/50],	Iter [175/418]	Loss: 3.4591,	average_loss: 2.4580
Epoch [20/50],	Iter [180/418]	Loss: 4.6331,	average_loss: 2.4741
Epoch [20/50],	Iter [185/418]	Loss: 2.5599,	average_loss: 2.4808
Epoch [20/50],	Iter [190/418]	Loss: 2.5686,	average_loss: 2.4759
Epoch [20/50],	Iter [195/418]	Loss: 2.5863,	average_loss: 2.4777
Epoch [20/50],	Iter [200/418]	Loss: 2.7243,	average_loss: 2.4834
Epoch [20/50],	Iter [205/418]	Loss: 2.4096,	average_loss: 2.4823
Epoch [20/50],	Iter [210/418]	Loss: 2.7293,	average_loss: 2.4733
Epoch [20/50],	Iter [215/418]	Loss: 2.3982,	average_loss: 2.4635
Epoch [20/50],	Iter [220/418]	Loss: 2.1510,	average_loss: 2.4473

Epoch [20/50], Iter [225/418] Loss: 2.5210, average\_loss: 2.4409  
 Epoch [20/50], Iter [230/418] Loss: 2.3613, average\_loss: 2.4359  
 Epoch [20/50], Iter [235/418] Loss: 2.1906, average\_loss: 2.4306  
 Epoch [20/50], Iter [240/418] Loss: 2.4447, average\_loss: 2.4330  
 Epoch [20/50], Iter [245/418] Loss: 4.1360, average\_loss: 2.4302  
 Epoch [20/50], Iter [250/418] Loss: 2.5118, average\_loss: 2.4353  
 Epoch [20/50], Iter [255/418] Loss: 2.5900, average\_loss: 2.4291  
 Epoch [20/50], Iter [260/418] Loss: 1.4445, average\_loss: 2.4299  
 Epoch [20/50], Iter [265/418] Loss: 3.2014, average\_loss: 2.4363  
 Epoch [20/50], Iter [270/418] Loss: 3.1513, average\_loss: 2.4363  
 Epoch [20/50], Iter [275/418] Loss: 3.4920, average\_loss: 2.4410  
 Epoch [20/50], Iter [280/418] Loss: 3.1172, average\_loss: 2.4417  
 Epoch [20/50], Iter [285/418] Loss: 1.4512, average\_loss: 2.4330  
 Epoch [20/50], Iter [290/418] Loss: 2.4862, average\_loss: 2.4318  
 Epoch [20/50], Iter [295/418] Loss: 2.6568, average\_loss: 2.4325  
 Epoch [20/50], Iter [300/418] Loss: 2.5798, average\_loss: 2.4254  
 Epoch [20/50], Iter [305/418] Loss: 2.3335, average\_loss: 2.4238  
 Epoch [20/50], Iter [310/418] Loss: 2.6521, average\_loss: 2.4221  
 Epoch [20/50], Iter [315/418] Loss: 2.9795, average\_loss: 2.4235  
 Epoch [20/50], Iter [320/418] Loss: 3.5093, average\_loss: 2.4304  
 Epoch [20/50], Iter [325/418] Loss: 3.0953, average\_loss: 2.4294  
 Epoch [20/50], Iter [330/418] Loss: 2.2328, average\_loss: 2.4233  
 Epoch [20/50], Iter [335/418] Loss: 3.7208, average\_loss: 2.4215  
 Epoch [20/50], Iter [340/418] Loss: 1.7701, average\_loss: 2.4225  
 Epoch [20/50], Iter [345/418] Loss: 1.9049, average\_loss: 2.4229  
 Epoch [20/50], Iter [350/418] Loss: 3.1869, average\_loss: 2.4190  
 Epoch [20/50], Iter [355/418] Loss: 2.0659, average\_loss: 2.4175  
 Epoch [20/50], Iter [360/418] Loss: 1.4867, average\_loss: 2.4143  
 Epoch [20/50], Iter [365/418] Loss: 3.1308, average\_loss: 2.4118  
 Epoch [20/50], Iter [370/418] Loss: 1.6767, average\_loss: 2.4097  
 Epoch [20/50], Iter [375/418] Loss: 1.5818, average\_loss: 2.4093  
 Epoch [20/50], Iter [380/418] Loss: 1.9335, average\_loss: 2.4084  
 Epoch [20/50], Iter [385/418] Loss: 1.6106, average\_loss: 2.4074  
 Epoch [20/50], Iter [390/418] Loss: 2.9847, average\_loss: 2.4051  
 Epoch [20/50], Iter [395/418] Loss: 2.4008, average\_loss: 2.4029  
 Epoch [20/50], Iter [400/418] Loss: 2.6036, average\_loss: 2.4075  
 Epoch [20/50], Iter [405/418] Loss: 3.0547, average\_loss: 2.4095  
 Epoch [20/50], Iter [410/418] Loss: 3.2532, average\_loss: 2.4102  
 Epoch [20/50], Iter [415/418] Loss: 1.8151, average\_loss: 2.4048  
 Updating best test loss: 2.85807

Starting epoch 21 / 50

Learning Rate for this epoch: 0.001

Epoch [21/50], Iter [5/418] Loss: 1.6963, average\_loss: 1.9350  
 Epoch [21/50], Iter [10/418] Loss: 3.3299, average\_loss: 2.0889  
 Epoch [21/50], Iter [15/418] Loss: 2.7756, average\_loss: 2.2018  
 Epoch [21/50], Iter [20/418] Loss: 2.0056, average\_loss: 2.2241

Epoch [21/50], Iter [25/418] Loss: 1.5427, average\_loss: 2.1979  
 Epoch [21/50], Iter [30/418] Loss: 2.6637, average\_loss: 2.2291  
 Epoch [21/50], Iter [35/418] Loss: 2.0920, average\_loss: 2.2708  
 Epoch [21/50], Iter [40/418] Loss: 2.0494, average\_loss: 2.2748  
 Epoch [21/50], Iter [45/418] Loss: 2.0148, average\_loss: 2.2717  
 Epoch [21/50], Iter [50/418] Loss: 1.5175, average\_loss: 2.2521  
 Epoch [21/50], Iter [55/418] Loss: 1.9352, average\_loss: 2.2370  
 Epoch [21/50], Iter [60/418] Loss: 2.1155, average\_loss: 2.2463  
 Epoch [21/50], Iter [65/418] Loss: 2.4402, average\_loss: 2.2406  
 Epoch [21/50], Iter [70/418] Loss: 1.5650, average\_loss: 2.2252  
 Epoch [21/50], Iter [75/418] Loss: 3.4904, average\_loss: 2.2238  
 Epoch [21/50], Iter [80/418] Loss: 1.7475, average\_loss: 2.2204  
 Epoch [21/50], Iter [85/418] Loss: 2.0563, average\_loss: 2.2302  
 Epoch [21/50], Iter [90/418] Loss: 1.7490, average\_loss: 2.2283  
 Epoch [21/50], Iter [95/418] Loss: 1.9043, average\_loss: 2.2154  
 Epoch [21/50], Iter [100/418] Loss: 3.2556, average\_loss: 2.2213  
 Epoch [21/50], Iter [105/418] Loss: 1.8278, average\_loss: 2.2232  
 Epoch [21/50], Iter [110/418] Loss: 3.0382, average\_loss: 2.2239  
 Epoch [21/50], Iter [115/418] Loss: 1.5003, average\_loss: 2.2139  
 Epoch [21/50], Iter [120/418] Loss: 2.0965, average\_loss: 2.2139  
 Epoch [21/50], Iter [125/418] Loss: 3.0682, average\_loss: 2.2246  
 Epoch [21/50], Iter [130/418] Loss: 2.9958, average\_loss: 2.2427  
 Epoch [21/50], Iter [135/418] Loss: 2.5618, average\_loss: 2.2469  
 Epoch [21/50], Iter [140/418] Loss: 3.5109, average\_loss: 2.2580  
 Epoch [21/50], Iter [145/418] Loss: 2.4830, average\_loss: 2.2511  
 Epoch [21/50], Iter [150/418] Loss: 2.0430, average\_loss: 2.2451  
 Epoch [21/50], Iter [155/418] Loss: 3.0151, average\_loss: 2.2555  
 Epoch [21/50], Iter [160/418] Loss: 2.7854, average\_loss: 2.2785  
 Epoch [21/50], Iter [165/418] Loss: 2.1272, average\_loss: 2.2846  
 Epoch [21/50], Iter [170/418] Loss: 2.9167, average\_loss: 2.3000  
 Epoch [21/50], Iter [175/418] Loss: 1.5975, average\_loss: 2.2992  
 Epoch [21/50], Iter [180/418] Loss: 2.7753, average\_loss: 2.3193  
 Epoch [21/50], Iter [185/418] Loss: 1.9413, average\_loss: 2.3159  
 Epoch [21/50], Iter [190/418] Loss: 2.2019, average\_loss: 2.3103  
 Epoch [21/50], Iter [195/418] Loss: 2.8550, average\_loss: 2.3097  
 Epoch [21/50], Iter [200/418] Loss: 2.8200, average\_loss: 2.3180  
 Epoch [21/50], Iter [205/418] Loss: 2.5447, average\_loss: 2.3155  
 Epoch [21/50], Iter [210/418] Loss: 2.1417, average\_loss: 2.3138  
 Epoch [21/50], Iter [215/418] Loss: 2.2687, average\_loss: 2.3274  
 Epoch [21/50], Iter [220/418] Loss: 1.9661, average\_loss: 2.3238  
 Epoch [21/50], Iter [225/418] Loss: 2.1658, average\_loss: 2.3187  
 Epoch [21/50], Iter [230/418] Loss: 3.2242, average\_loss: 2.3214  
 Epoch [21/50], Iter [235/418] Loss: 2.4348, average\_loss: 2.3268  
 Epoch [21/50], Iter [240/418] Loss: 2.8110, average\_loss: 2.3303  
 Epoch [21/50], Iter [245/418] Loss: 1.8455, average\_loss: 2.3331  
 Epoch [21/50], Iter [250/418] Loss: 2.3809, average\_loss: 2.3405  
 Epoch [21/50], Iter [255/418] Loss: 3.0281, average\_loss: 2.3420  
 Epoch [21/50], Iter [260/418] Loss: 2.1409, average\_loss: 2.3347

Epoch [21/50], Iter [265/418] Loss: 1.5789, average\_loss: 2.3300  
Epoch [21/50], Iter [270/418] Loss: 2.7481, average\_loss: 2.3411  
Epoch [21/50], Iter [275/418] Loss: 1.6032, average\_loss: 2.3436  
Epoch [21/50], Iter [280/418] Loss: 2.7144, average\_loss: 2.3426  
Epoch [21/50], Iter [285/418] Loss: 1.9138, average\_loss: 2.3374  
Epoch [21/50], Iter [290/418] Loss: 2.0150, average\_loss: 2.3394  
Epoch [21/50], Iter [295/418] Loss: 2.5326, average\_loss: 2.3383  
Epoch [21/50], Iter [300/418] Loss: 1.6404, average\_loss: 2.3345  
Epoch [21/50], Iter [305/418] Loss: 2.4375, average\_loss: 2.3398  
Epoch [21/50], Iter [310/418] Loss: 3.3204, average\_loss: 2.3368  
Epoch [21/50], Iter [315/418] Loss: 2.4568, average\_loss: 2.3406  
Epoch [21/50], Iter [320/418] Loss: 4.4161, average\_loss: 2.3414  
Epoch [21/50], Iter [325/418] Loss: 2.4298, average\_loss: 2.3379  
Epoch [21/50], Iter [330/418] Loss: 3.9142, average\_loss: 2.3452  
Epoch [21/50], Iter [335/418] Loss: 1.5861, average\_loss: 2.3400  
Epoch [21/50], Iter [340/418] Loss: 3.7192, average\_loss: 2.3444  
Epoch [21/50], Iter [345/418] Loss: 1.5124, average\_loss: 2.3386  
Epoch [21/50], Iter [350/418] Loss: 2.6771, average\_loss: 2.3428  
Epoch [21/50], Iter [355/418] Loss: 2.9320, average\_loss: 2.3466  
Epoch [21/50], Iter [360/418] Loss: 1.9146, average\_loss: 2.3504  
Epoch [21/50], Iter [365/418] Loss: 1.3814, average\_loss: 2.3515  
Epoch [21/50], Iter [370/418] Loss: 3.0504, average\_loss: 2.3528  
Epoch [21/50], Iter [375/418] Loss: 3.2729, average\_loss: 2.3536  
Epoch [21/50], Iter [380/418] Loss: 2.7984, average\_loss: 2.3531  
Epoch [21/50], Iter [385/418] Loss: 2.6846, average\_loss: 2.3525  
Epoch [21/50], Iter [390/418] Loss: 3.1493, average\_loss: 2.3599  
Epoch [21/50], Iter [395/418] Loss: 3.4295, average\_loss: 2.3653  
Epoch [21/50], Iter [400/418] Loss: 1.9972, average\_loss: 2.3638  
Epoch [21/50], Iter [405/418] Loss: 2.4431, average\_loss: 2.3686  
Epoch [21/50], Iter [410/418] Loss: 1.7781, average\_loss: 2.3656  
Epoch [21/50], Iter [415/418] Loss: 1.6594, average\_loss: 2.3686  
---Evaluate model on test samples---

100%|

| 4950/4950 [02:29<00:00, 33.00it/s]

---class aeroplane ap 0.49527723916354427---  
---class bicycle ap 0.554469262967542---  
---class bird ap 0.4088712338637629---  
---class boat ap 0.22445096250553298---  
---class bottle ap 0.2001294240391368---  
---class bus ap 0.5644049611547959---  
---class car ap 0.591911072442602---  
---class cat ap 0.5595001508798307---  
---class chair ap 0.23199019097062606---  
---class cow ap 0.3882388480738114---  
---class diningtable ap 0.23400138175311888---  
---class dog ap 0.5236995882063804---  
---class horse ap 0.5779228278500603---

```
---class motorbike ap 0.5070554144826592---
---class person ap 0.44300151371755936---
---class pottedplant ap 0.1455522655229368---
---class sheep ap 0.3716933873709764---
---class sofa ap 0.41868268911529977---
---class train ap 0.5991360083357484---
---class tvmonitor ap 0.4115268155031486---
---map 0.4225757618959536---
Updating best test loss: 2.81465
```

Starting epoch 22 / 50

Learning Rate for this epoch: 0.001

```
Epoch [22/50], Iter [5/418] Loss: 1.9292, average_loss: 2.4362
Epoch [22/50], Iter [10/418] Loss: 4.0163, average_loss: 2.4002
Epoch [22/50], Iter [15/418] Loss: 2.1477, average_loss: 2.3125
Epoch [22/50], Iter [20/418] Loss: 1.5509, average_loss: 2.3062
Epoch [22/50], Iter [25/418] Loss: 1.9871, average_loss: 2.2722
Epoch [22/50], Iter [30/418] Loss: 1.3599, average_loss: 2.3455
Epoch [22/50], Iter [35/418] Loss: 2.0885, average_loss: 2.3377
Epoch [22/50], Iter [40/418] Loss: 1.7107, average_loss: 2.3317
Epoch [22/50], Iter [45/418] Loss: 1.8735, average_loss: 2.2836
Epoch [22/50], Iter [50/418] Loss: 2.4128, average_loss: 2.2721
Epoch [22/50], Iter [55/418] Loss: 2.4756, average_loss: 2.2537
Epoch [22/50], Iter [60/418] Loss: 2.2307, average_loss: 2.2252
Epoch [22/50], Iter [65/418] Loss: 1.5347, average_loss: 2.2207
Epoch [22/50], Iter [70/418] Loss: 2.0429, average_loss: 2.2229
Epoch [22/50], Iter [75/418] Loss: 2.8220, average_loss: 2.2059
Epoch [22/50], Iter [80/418] Loss: 3.0729, average_loss: 2.2477
Epoch [22/50], Iter [85/418] Loss: 3.4640, average_loss: 2.2838
Epoch [22/50], Iter [90/418] Loss: 2.5329, average_loss: 2.2869
Epoch [22/50], Iter [95/418] Loss: 2.2868, average_loss: 2.3008
Epoch [22/50], Iter [100/418] Loss: 5.0010, average_loss: 2.3119
Epoch [22/50], Iter [105/418] Loss: 1.7045, average_loss: 2.2870
Epoch [22/50], Iter [110/418] Loss: 1.4035, average_loss: 2.2735
Epoch [22/50], Iter [115/418] Loss: 2.0548, average_loss: 2.2803
Epoch [22/50], Iter [120/418] Loss: 2.3003, average_loss: 2.2737
Epoch [22/50], Iter [125/418] Loss: 2.2885, average_loss: 2.2726
Epoch [22/50], Iter [130/418] Loss: 2.4475, average_loss: 2.2855
Epoch [22/50], Iter [135/418] Loss: 3.3136, average_loss: 2.2841
Epoch [22/50], Iter [140/418] Loss: 2.9286, average_loss: 2.2879
Epoch [22/50], Iter [145/418] Loss: 2.4279, average_loss: 2.3022
Epoch [22/50], Iter [150/418] Loss: 1.9544, average_loss: 2.2941
Epoch [22/50], Iter [155/418] Loss: 3.1909, average_loss: 2.2967
Epoch [22/50], Iter [160/418] Loss: 2.1228, average_loss: 2.3075
Epoch [22/50], Iter [165/418] Loss: 1.6336, average_loss: 2.3128
Epoch [22/50], Iter [170/418] Loss: 1.9285, average_loss: 2.2997
Epoch [22/50], Iter [175/418] Loss: 2.5895, average_loss: 2.3003
```

Epoch [22/50], Iter [180/418] Loss: 2.4965, average\_loss: 2.2940  
 Epoch [22/50], Iter [185/418] Loss: 1.8370, average\_loss: 2.2841  
 Epoch [22/50], Iter [190/418] Loss: 2.2124, average\_loss: 2.2838  
 Epoch [22/50], Iter [195/418] Loss: 2.1346, average\_loss: 2.2899  
 Epoch [22/50], Iter [200/418] Loss: 2.3746, average\_loss: 2.2956  
 Epoch [22/50], Iter [205/418] Loss: 1.8654, average\_loss: 2.2991  
 Epoch [22/50], Iter [210/418] Loss: 1.5811, average\_loss: 2.3039  
 Epoch [22/50], Iter [215/418] Loss: 2.8758, average\_loss: 2.3052  
 Epoch [22/50], Iter [220/418] Loss: 2.6829, average\_loss: 2.3015  
 Epoch [22/50], Iter [225/418] Loss: 2.8318, average\_loss: 2.3022  
 Epoch [22/50], Iter [230/418] Loss: 1.7507, average\_loss: 2.3140  
 Epoch [22/50], Iter [235/418] Loss: 2.0505, average\_loss: 2.3074  
 Epoch [22/50], Iter [240/418] Loss: 2.2618, average\_loss: 2.3001  
 Epoch [22/50], Iter [245/418] Loss: 3.0203, average\_loss: 2.3166  
 Epoch [22/50], Iter [250/418] Loss: 1.8141, average\_loss: 2.3144  
 Epoch [22/50], Iter [255/418] Loss: 2.3527, average\_loss: 2.3071  
 Epoch [22/50], Iter [260/418] Loss: 2.9460, average\_loss: 2.3114  
 Epoch [22/50], Iter [265/418] Loss: 1.8154, average\_loss: 2.3052  
 Epoch [22/50], Iter [270/418] Loss: 2.4112, average\_loss: 2.3099  
 Epoch [22/50], Iter [275/418] Loss: 2.5952, average\_loss: 2.3058  
 Epoch [22/50], Iter [280/418] Loss: 1.6572, average\_loss: 2.2985  
 Epoch [22/50], Iter [285/418] Loss: 2.3513, average\_loss: 2.2965  
 Epoch [22/50], Iter [290/418] Loss: 2.3672, average\_loss: 2.2890  
 Epoch [22/50], Iter [295/418] Loss: 1.9896, average\_loss: 2.2936  
 Epoch [22/50], Iter [300/418] Loss: 4.4271, average\_loss: 2.3015  
 Epoch [22/50], Iter [305/418] Loss: 2.0754, average\_loss: 2.3041  
 Epoch [22/50], Iter [310/418] Loss: 1.8432, average\_loss: 2.3046  
 Epoch [22/50], Iter [315/418] Loss: 2.3872, average\_loss: 2.3077  
 Epoch [22/50], Iter [320/418] Loss: 3.5933, average\_loss: 2.3142  
 Epoch [22/50], Iter [325/418] Loss: 1.7180, average\_loss: 2.3088  
 Epoch [22/50], Iter [330/418] Loss: 1.5484, average\_loss: 2.3072  
 Epoch [22/50], Iter [335/418] Loss: 2.3140, average\_loss: 2.3087  
 Epoch [22/50], Iter [340/418] Loss: 1.7013, average\_loss: 2.3100  
 Epoch [22/50], Iter [345/418] Loss: 2.7809, average\_loss: 2.3102  
 Epoch [22/50], Iter [350/418] Loss: 1.8697, average\_loss: 2.3107  
 Epoch [22/50], Iter [355/418] Loss: 2.5518, average\_loss: 2.3134  
 Epoch [22/50], Iter [360/418] Loss: 2.5101, average\_loss: 2.3098  
 Epoch [22/50], Iter [365/418] Loss: 2.3148, average\_loss: 2.3116  
 Epoch [22/50], Iter [370/418] Loss: 3.0049, average\_loss: 2.3144  
 Epoch [22/50], Iter [375/418] Loss: 1.8826, average\_loss: 2.3122  
 Epoch [22/50], Iter [380/418] Loss: 2.0398, average\_loss: 2.3138  
 Epoch [22/50], Iter [385/418] Loss: 1.7810, average\_loss: 2.3127  
 Epoch [22/50], Iter [390/418] Loss: 3.1781, average\_loss: 2.3113  
 Epoch [22/50], Iter [395/418] Loss: 3.2106, average\_loss: 2.3159  
 Epoch [22/50], Iter [400/418] Loss: 2.1737, average\_loss: 2.3136  
 Epoch [22/50], Iter [405/418] Loss: 1.7491, average\_loss: 2.3123  
 Epoch [22/50], Iter [410/418] Loss: 1.9782, average\_loss: 2.3112  
 Epoch [22/50], Iter [415/418] Loss: 2.3376, average\_loss: 2.3121

Starting epoch 23 / 50

Learning Rate for this epoch: 0.001

Epoch [23/50],	Iter [5/418]	Loss: 1.8943,	average_loss: 2.1528
Epoch [23/50],	Iter [10/418]	Loss: 1.8409,	average_loss: 2.1089
Epoch [23/50],	Iter [15/418]	Loss: 2.1676,	average_loss: 2.1480
Epoch [23/50],	Iter [20/418]	Loss: 2.0927,	average_loss: 2.1522
Epoch [23/50],	Iter [25/418]	Loss: 1.7794,	average_loss: 2.1162
Epoch [23/50],	Iter [30/418]	Loss: 2.7014,	average_loss: 2.1494
Epoch [23/50],	Iter [35/418]	Loss: 1.9159,	average_loss: 2.0966
Epoch [23/50],	Iter [40/418]	Loss: 2.3402,	average_loss: 2.1303
Epoch [23/50],	Iter [45/418]	Loss: 2.8800,	average_loss: 2.1779
Epoch [23/50],	Iter [50/418]	Loss: 2.3083,	average_loss: 2.2402
Epoch [23/50],	Iter [55/418]	Loss: 2.9062,	average_loss: 2.2686
Epoch [23/50],	Iter [60/418]	Loss: 2.8421,	average_loss: 2.2702
Epoch [23/50],	Iter [65/418]	Loss: 1.4604,	average_loss: 2.2243
Epoch [23/50],	Iter [70/418]	Loss: 1.5766,	average_loss: 2.2010
Epoch [23/50],	Iter [75/418]	Loss: 1.7820,	average_loss: 2.1907
Epoch [23/50],	Iter [80/418]	Loss: 1.7673,	average_loss: 2.1948
Epoch [23/50],	Iter [85/418]	Loss: 1.8665,	average_loss: 2.2026
Epoch [23/50],	Iter [90/418]	Loss: 2.4338,	average_loss: 2.2323
Epoch [23/50],	Iter [95/418]	Loss: 2.2520,	average_loss: 2.2404
Epoch [23/50],	Iter [100/418]	Loss: 2.4143,	average_loss: 2.2277
Epoch [23/50],	Iter [105/418]	Loss: 1.8236,	average_loss: 2.2229
Epoch [23/50],	Iter [110/418]	Loss: 3.4785,	average_loss: 2.2634
Epoch [23/50],	Iter [115/418]	Loss: 2.2590,	average_loss: 2.2748
Epoch [23/50],	Iter [120/418]	Loss: 2.5715,	average_loss: 2.2647
Epoch [23/50],	Iter [125/418]	Loss: 2.8133,	average_loss: 2.2803
Epoch [23/50],	Iter [130/418]	Loss: 2.1051,	average_loss: 2.2831
Epoch [23/50],	Iter [135/418]	Loss: 1.9652,	average_loss: 2.2917
Epoch [23/50],	Iter [140/418]	Loss: 2.8984,	average_loss: 2.2899
Epoch [23/50],	Iter [145/418]	Loss: 1.7024,	average_loss: 2.2905
Epoch [23/50],	Iter [150/418]	Loss: 2.4672,	average_loss: 2.2968
Epoch [23/50],	Iter [155/418]	Loss: 2.3756,	average_loss: 2.3074
Epoch [23/50],	Iter [160/418]	Loss: 2.1116,	average_loss: 2.3026
Epoch [23/50],	Iter [165/418]	Loss: 2.6197,	average_loss: 2.3150
Epoch [23/50],	Iter [170/418]	Loss: 1.8670,	average_loss: 2.3051
Epoch [23/50],	Iter [175/418]	Loss: 1.6712,	average_loss: 2.3007
Epoch [23/50],	Iter [180/418]	Loss: 2.0089,	average_loss: 2.3012
Epoch [23/50],	Iter [185/418]	Loss: 1.7749,	average_loss: 2.2980
Epoch [23/50],	Iter [190/418]	Loss: 1.9438,	average_loss: 2.3045
Epoch [23/50],	Iter [195/418]	Loss: 2.5775,	average_loss: 2.2877
Epoch [23/50],	Iter [200/418]	Loss: 2.0360,	average_loss: 2.3000
Epoch [23/50],	Iter [205/418]	Loss: 1.5873,	average_loss: 2.2894
Epoch [23/50],	Iter [210/418]	Loss: 3.0565,	average_loss: 2.2957
Epoch [23/50],	Iter [215/418]	Loss: 1.8618,	average_loss: 2.2969
Epoch [23/50],	Iter [220/418]	Loss: 2.1546,	average_loss: 2.3019



Epoch [23/50], Iter [225/418] Loss: 1.2226, average\_loss: 2.3000  
 Epoch [23/50], Iter [230/418] Loss: 1.9546, average\_loss: 2.3001  
 Epoch [23/50], Iter [235/418] Loss: 2.9492, average\_loss: 2.3068  
 Epoch [23/50], Iter [240/418] Loss: 2.1766, average\_loss: 2.3060  
 Epoch [23/50], Iter [245/418] Loss: 2.1938, average\_loss: 2.3064  
 Epoch [23/50], Iter [250/418] Loss: 2.4161, average\_loss: 2.3138  
 Epoch [23/50], Iter [255/418] Loss: 1.6301, average\_loss: 2.3059  
 Epoch [23/50], Iter [260/418] Loss: 2.6500, average\_loss: 2.3007  
 Epoch [23/50], Iter [265/418] Loss: 1.5834, average\_loss: 2.3036  
 Epoch [23/50], Iter [270/418] Loss: 1.7613, average\_loss: 2.3011  
 Epoch [23/50], Iter [275/418] Loss: 2.8002, average\_loss: 2.3069  
 Epoch [23/50], Iter [280/418] Loss: 2.6539, average\_loss: 2.3047  
 Epoch [23/50], Iter [285/418] Loss: 1.7320, average\_loss: 2.2991  
 Epoch [23/50], Iter [290/418] Loss: 2.2667, average\_loss: 2.3028  
 Epoch [23/50], Iter [295/418] Loss: 2.9227, average\_loss: 2.3105  
 Epoch [23/50], Iter [300/418] Loss: 1.8009, average\_loss: 2.3109  
 Epoch [23/50], Iter [305/418] Loss: 2.1507, average\_loss: 2.3123  
 Epoch [23/50], Iter [310/418] Loss: 1.9264, average\_loss: 2.3067  
 Epoch [23/50], Iter [315/418] Loss: 2.1503, average\_loss: 2.3033  
 Epoch [23/50], Iter [320/418] Loss: 1.8508, average\_loss: 2.3019  
 Epoch [23/50], Iter [325/418] Loss: 1.9472, average\_loss: 2.3032  
 Epoch [23/50], Iter [330/418] Loss: 1.4648, average\_loss: 2.2974  
 Epoch [23/50], Iter [335/418] Loss: 3.1935, average\_loss: 2.3020  
 Epoch [23/50], Iter [340/418] Loss: 2.7013, average\_loss: 2.2972  
 Epoch [23/50], Iter [345/418] Loss: 2.7099, average\_loss: 2.2991  
 Epoch [23/50], Iter [350/418] Loss: 2.9857, average\_loss: 2.3089  
 Epoch [23/50], Iter [355/418] Loss: 2.3646, average\_loss: 2.3078  
 Epoch [23/50], Iter [360/418] Loss: 1.9707, average\_loss: 2.3042  
 Epoch [23/50], Iter [365/418] Loss: 3.6848, average\_loss: 2.3098  
 Epoch [23/50], Iter [370/418] Loss: 2.5113, average\_loss: 2.3101  
 Epoch [23/50], Iter [375/418] Loss: 2.0995, average\_loss: 2.3087  
 Epoch [23/50], Iter [380/418] Loss: 3.0585, average\_loss: 2.3059  
 Epoch [23/50], Iter [385/418] Loss: 2.1090, average\_loss: 2.3057  
 Epoch [23/50], Iter [390/418] Loss: 2.1438, average\_loss: 2.3049  
 Epoch [23/50], Iter [395/418] Loss: 2.2078, average\_loss: 2.3039  
 Epoch [23/50], Iter [400/418] Loss: 2.5360, average\_loss: 2.3068  
 Epoch [23/50], Iter [405/418] Loss: 2.3013, average\_loss: 2.3015  
 Epoch [23/50], Iter [410/418] Loss: 3.2974, average\_loss: 2.3058  
 Epoch [23/50], Iter [415/418] Loss: 2.0667, average\_loss: 2.3067  
 Updating best test loss: 2.79041

Starting epoch 24 / 50

Learning Rate for this epoch: 0.001

Epoch [24/50], Iter [5/418] Loss: 1.8098, average\_loss: 2.1340  
 Epoch [24/50], Iter [10/418] Loss: 1.5815, average\_loss: 2.1542  
 Epoch [24/50], Iter [15/418] Loss: 1.5535, average\_loss: 2.1301  
 Epoch [24/50], Iter [20/418] Loss: 1.9255, average\_loss: 2.1774

Epoch [24/50], Iter [25/418] Loss: 1.0927, average\_loss: 2.1677  
Epoch [24/50], Iter [30/418] Loss: 1.6373, average\_loss: 2.1859  
Epoch [24/50], Iter [35/418] Loss: 2.0519, average\_loss: 2.1763  
Epoch [24/50], Iter [40/418] Loss: 1.8719, average\_loss: 2.1540  
Epoch [24/50], Iter [45/418] Loss: 2.7331, average\_loss: 2.1919  
Epoch [24/50], Iter [50/418] Loss: 2.0993, average\_loss: 2.1784  
Epoch [24/50], Iter [55/418] Loss: 2.1258, average\_loss: 2.1654  
Epoch [24/50], Iter [60/418] Loss: 2.7374, average\_loss: 2.2093  
Epoch [24/50], Iter [65/418] Loss: 2.8881, average\_loss: 2.2096  
Epoch [24/50], Iter [70/418] Loss: 3.5805, average\_loss: 2.2413  
Epoch [24/50], Iter [75/418] Loss: 1.7507, average\_loss: 2.2258  
Epoch [24/50], Iter [80/418] Loss: 2.2085, average\_loss: 2.2071  
Epoch [24/50], Iter [85/418] Loss: 2.9113, average\_loss: 2.2163  
Epoch [24/50], Iter [90/418] Loss: 2.3918, average\_loss: 2.2058  
Epoch [24/50], Iter [95/418] Loss: 2.5381, average\_loss: 2.1946  
Epoch [24/50], Iter [100/418] Loss: 2.7569, average\_loss: 2.1815  
Epoch [24/50], Iter [105/418] Loss: 2.9102, average\_loss: 2.1857  
Epoch [24/50], Iter [110/418] Loss: 3.8000, average\_loss: 2.2063  
Epoch [24/50], Iter [115/418] Loss: 2.3191, average\_loss: 2.1924  
Epoch [24/50], Iter [120/418] Loss: 2.0420, average\_loss: 2.1853  
Epoch [24/50], Iter [125/418] Loss: 3.2730, average\_loss: 2.1987  
Epoch [24/50], Iter [130/418] Loss: 2.8248, average\_loss: 2.1872  
Epoch [24/50], Iter [135/418] Loss: 2.9539, average\_loss: 2.2007  
Epoch [24/50], Iter [140/418] Loss: 2.3126, average\_loss: 2.1961  
Epoch [24/50], Iter [145/418] Loss: 1.4304, average\_loss: 2.2085  
Epoch [24/50], Iter [150/418] Loss: 2.3129, average\_loss: 2.2053  
Epoch [24/50], Iter [155/418] Loss: 2.5840, average\_loss: 2.1949  
Epoch [24/50], Iter [160/418] Loss: 1.7415, average\_loss: 2.1905  
Epoch [24/50], Iter [165/418] Loss: 2.6038, average\_loss: 2.1960  
Epoch [24/50], Iter [170/418] Loss: 2.9959, average\_loss: 2.1964  
Epoch [24/50], Iter [175/418] Loss: 1.5592, average\_loss: 2.2039  
Epoch [24/50], Iter [180/418] Loss: 2.0784, average\_loss: 2.1997  
Epoch [24/50], Iter [185/418] Loss: 1.4845, average\_loss: 2.2030  
Epoch [24/50], Iter [190/418] Loss: 2.4683, average\_loss: 2.2164  
Epoch [24/50], Iter [195/418] Loss: 1.8509, average\_loss: 2.2152  
Epoch [24/50], Iter [200/418] Loss: 2.6982, average\_loss: 2.2279  
Epoch [24/50], Iter [205/418] Loss: 1.3694, average\_loss: 2.2302  
Epoch [24/50], Iter [210/418] Loss: 1.9400, average\_loss: 2.2344  
Epoch [24/50], Iter [215/418] Loss: 2.8328, average\_loss: 2.2485  
Epoch [24/50], Iter [220/418] Loss: 1.6743, average\_loss: 2.2452  
Epoch [24/50], Iter [225/418] Loss: 2.3506, average\_loss: 2.2465  
Epoch [24/50], Iter [230/418] Loss: 2.1180, average\_loss: 2.2433  
Epoch [24/50], Iter [235/418] Loss: 2.9344, average\_loss: 2.2500  
Epoch [24/50], Iter [240/418] Loss: 2.1163, average\_loss: 2.2520  
Epoch [24/50], Iter [245/418] Loss: 1.8528, average\_loss: 2.2458  
Epoch [24/50], Iter [250/418] Loss: 1.0803, average\_loss: 2.2398  
Epoch [24/50], Iter [255/418] Loss: 2.4412, average\_loss: 2.2366  
Epoch [24/50], Iter [260/418] Loss: 2.1186, average\_loss: 2.2342

Epoch [24/50], Iter [265/418] Loss: 2.7919, average\_loss: 2.2356  
 Epoch [24/50], Iter [270/418] Loss: 1.9810, average\_loss: 2.2368  
 Epoch [24/50], Iter [275/418] Loss: 1.0758, average\_loss: 2.2345  
 Epoch [24/50], Iter [280/418] Loss: 2.5122, average\_loss: 2.2454  
 Epoch [24/50], Iter [285/418] Loss: 2.5170, average\_loss: 2.2480  
 Epoch [24/50], Iter [290/418] Loss: 1.0135, average\_loss: 2.2469  
 Epoch [24/50], Iter [295/418] Loss: 4.0792, average\_loss: 2.2573  
 Epoch [24/50], Iter [300/418] Loss: 1.4393, average\_loss: 2.2513  
 Epoch [24/50], Iter [305/418] Loss: 2.3974, average\_loss: 2.2433  
 Epoch [24/50], Iter [310/418] Loss: 1.3128, average\_loss: 2.2469  
 Epoch [24/50], Iter [315/418] Loss: 2.3032, average\_loss: 2.2463  
 Epoch [24/50], Iter [320/418] Loss: 1.8880, average\_loss: 2.2425  
 Epoch [24/50], Iter [325/418] Loss: 2.2775, average\_loss: 2.2423  
 Epoch [24/50], Iter [330/418] Loss: 2.0772, average\_loss: 2.2388  
 Epoch [24/50], Iter [335/418] Loss: 1.1774, average\_loss: 2.2449  
 Epoch [24/50], Iter [340/418] Loss: 1.9925, average\_loss: 2.2486  
 Epoch [24/50], Iter [345/418] Loss: 3.5058, average\_loss: 2.2497  
 Epoch [24/50], Iter [350/418] Loss: 2.6448, average\_loss: 2.2532  
 Epoch [24/50], Iter [355/418] Loss: 1.9453, average\_loss: 2.2529  
 Epoch [24/50], Iter [360/418] Loss: 2.6496, average\_loss: 2.2487  
 Epoch [24/50], Iter [365/418] Loss: 2.8846, average\_loss: 2.2495  
 Epoch [24/50], Iter [370/418] Loss: 1.9950, average\_loss: 2.2480  
 Epoch [24/50], Iter [375/418] Loss: 2.4582, average\_loss: 2.2536  
 Epoch [24/50], Iter [380/418] Loss: 2.7444, average\_loss: 2.2566  
 Epoch [24/50], Iter [385/418] Loss: 2.6421, average\_loss: 2.2548  
 Epoch [24/50], Iter [390/418] Loss: 2.3332, average\_loss: 2.2500  
 Epoch [24/50], Iter [395/418] Loss: 1.8523, average\_loss: 2.2533  
 Epoch [24/50], Iter [400/418] Loss: 1.9688, average\_loss: 2.2524  
 Epoch [24/50], Iter [405/418] Loss: 2.1050, average\_loss: 2.2519  
 Epoch [24/50], Iter [410/418] Loss: 3.3869, average\_loss: 2.2548  
 Epoch [24/50], Iter [415/418] Loss: 3.4004, average\_loss: 2.2578  
 Updating best test loss: 2.77750

Starting epoch 25 / 50

Learning Rate for this epoch: 0.001

Epoch [25/50], Iter [5/418] Loss: 2.4302, average\_loss: 2.5265  
 Epoch [25/50], Iter [10/418] Loss: 1.5543, average\_loss: 2.2862  
 Epoch [25/50], Iter [15/418] Loss: 1.2813, average\_loss: 2.3705  
 Epoch [25/50], Iter [20/418] Loss: 3.5269, average\_loss: 2.3261  
 Epoch [25/50], Iter [25/418] Loss: 1.7516, average\_loss: 2.2879  
 Epoch [25/50], Iter [30/418] Loss: 3.1795, average\_loss: 2.2671  
 Epoch [25/50], Iter [35/418] Loss: 1.7491, average\_loss: 2.2068  
 Epoch [25/50], Iter [40/418] Loss: 2.2466, average\_loss: 2.1601  
 Epoch [25/50], Iter [45/418] Loss: 3.7402, average\_loss: 2.2175  
 Epoch [25/50], Iter [50/418] Loss: 2.3120, average\_loss: 2.2410  
 Epoch [25/50], Iter [55/418] Loss: 2.6792, average\_loss: 2.3133  
 Epoch [25/50], Iter [60/418] Loss: 2.4097, average\_loss: 2.2864

Epoch [25/50], Iter [65/418] Loss: 2.2886, average\_loss: 2.2735  
 Epoch [25/50], Iter [70/418] Loss: 2.6010, average\_loss: 2.2603  
 Epoch [25/50], Iter [75/418] Loss: 2.7853, average\_loss: 2.2549  
 Epoch [25/50], Iter [80/418] Loss: 2.0422, average\_loss: 2.2433  
 Epoch [25/50], Iter [85/418] Loss: 2.0349, average\_loss: 2.2365  
 Epoch [25/50], Iter [90/418] Loss: 2.9172, average\_loss: 2.2552  
 Epoch [25/50], Iter [95/418] Loss: 1.7666, average\_loss: 2.2600  
 Epoch [25/50], Iter [100/418] Loss: 2.2672, average\_loss: 2.2433  
 Epoch [25/50], Iter [105/418] Loss: 3.1669, average\_loss: 2.2451  
 Epoch [25/50], Iter [110/418] Loss: 1.8977, average\_loss: 2.2272  
 Epoch [25/50], Iter [115/418] Loss: 1.5499, average\_loss: 2.2272  
 Epoch [25/50], Iter [120/418] Loss: 2.5144, average\_loss: 2.2358  
 Epoch [25/50], Iter [125/418] Loss: 1.6840, average\_loss: 2.2308  
 Epoch [25/50], Iter [130/418] Loss: 2.4337, average\_loss: 2.2208  
 Epoch [25/50], Iter [135/418] Loss: 2.1846, average\_loss: 2.2213  
 Epoch [25/50], Iter [140/418] Loss: 1.6617, average\_loss: 2.2019  
 Epoch [25/50], Iter [145/418] Loss: 2.5305, average\_loss: 2.1895  
 Epoch [25/50], Iter [150/418] Loss: 2.8907, average\_loss: 2.1968  
 Epoch [25/50], Iter [155/418] Loss: 1.8123, average\_loss: 2.1939  
 Epoch [25/50], Iter [160/418] Loss: 1.8749, average\_loss: 2.1998  
 Epoch [25/50], Iter [165/418] Loss: 2.0575, average\_loss: 2.2014  
 Epoch [25/50], Iter [170/418] Loss: 2.3786, average\_loss: 2.2025  
 Epoch [25/50], Iter [175/418] Loss: 1.5738, average\_loss: 2.1871  
 Epoch [25/50], Iter [180/418] Loss: 1.8093, average\_loss: 2.1770  
 Epoch [25/50], Iter [185/418] Loss: 1.7248, average\_loss: 2.1762  
 Epoch [25/50], Iter [190/418] Loss: 2.6693, average\_loss: 2.1780  
 Epoch [25/50], Iter [195/418] Loss: 2.5876, average\_loss: 2.1866  
 Epoch [25/50], Iter [200/418] Loss: 4.1379, average\_loss: 2.1893  
 Epoch [25/50], Iter [205/418] Loss: 1.7530, average\_loss: 2.1889  
 Epoch [25/50], Iter [210/418] Loss: 2.1915, average\_loss: 2.1989  
 Epoch [25/50], Iter [215/418] Loss: 2.1505, average\_loss: 2.2009  
 Epoch [25/50], Iter [220/418] Loss: 2.1233, average\_loss: 2.2000  
 Epoch [25/50], Iter [225/418] Loss: 1.7313, average\_loss: 2.1993  
 Epoch [25/50], Iter [230/418] Loss: 2.0952, average\_loss: 2.2020  
 Epoch [25/50], Iter [235/418] Loss: 1.2398, average\_loss: 2.1935  
 Epoch [25/50], Iter [240/418] Loss: 1.5229, average\_loss: 2.1956  
 Epoch [25/50], Iter [245/418] Loss: 3.0625, average\_loss: 2.1947  
 Epoch [25/50], Iter [250/418] Loss: 2.1115, average\_loss: 2.1892  
 Epoch [25/50], Iter [255/418] Loss: 1.7852, average\_loss: 2.1951  
 Epoch [25/50], Iter [260/418] Loss: 2.0128, average\_loss: 2.1927  
 Epoch [25/50], Iter [265/418] Loss: 1.9961, average\_loss: 2.1889  
 Epoch [25/50], Iter [270/418] Loss: 2.4345, average\_loss: 2.1941  
 Epoch [25/50], Iter [275/418] Loss: 2.6004, average\_loss: 2.2016  
 Epoch [25/50], Iter [280/418] Loss: 1.7143, average\_loss: 2.1950  
 Epoch [25/50], Iter [285/418] Loss: 1.9733, average\_loss: 2.1919  
 Epoch [25/50], Iter [290/418] Loss: 2.1052, average\_loss: 2.1882  
 Epoch [25/50], Iter [295/418] Loss: 1.7945, average\_loss: 2.1878  
 Epoch [25/50], Iter [300/418] Loss: 2.1807, average\_loss: 2.1866

Epoch [25/50], Iter [305/418] Loss: 1.3066, average\_loss: 2.1864  
 Epoch [25/50], Iter [310/418] Loss: 2.1164, average\_loss: 2.1967  
 Epoch [25/50], Iter [315/418] Loss: 2.0225, average\_loss: 2.2054  
 Epoch [25/50], Iter [320/418] Loss: 3.0874, average\_loss: 2.2077  
 Epoch [25/50], Iter [325/418] Loss: 1.7307, average\_loss: 2.1981  
 Epoch [25/50], Iter [330/418] Loss: 2.0514, average\_loss: 2.2006  
 Epoch [25/50], Iter [335/418] Loss: 2.2587, average\_loss: 2.1990  
 Epoch [25/50], Iter [340/418] Loss: 1.7470, average\_loss: 2.1973  
 Epoch [25/50], Iter [345/418] Loss: 1.9131, average\_loss: 2.1952  
 Epoch [25/50], Iter [350/418] Loss: 1.5478, average\_loss: 2.1898  
 Epoch [25/50], Iter [355/418] Loss: 2.3750, average\_loss: 2.1894  
 Epoch [25/50], Iter [360/418] Loss: 1.1060, average\_loss: 2.1858  
 Epoch [25/50], Iter [365/418] Loss: 1.9184, average\_loss: 2.1840  
 Epoch [25/50], Iter [370/418] Loss: 1.7996, average\_loss: 2.1825  
 Epoch [25/50], Iter [375/418] Loss: 2.4353, average\_loss: 2.1825  
 Epoch [25/50], Iter [380/418] Loss: 1.9996, average\_loss: 2.1819  
 Epoch [25/50], Iter [385/418] Loss: 1.7943, average\_loss: 2.1787  
 Epoch [25/50], Iter [390/418] Loss: 2.2431, average\_loss: 2.1785  
 Epoch [25/50], Iter [395/418] Loss: 2.8968, average\_loss: 2.1800  
 Epoch [25/50], Iter [400/418] Loss: 1.5263, average\_loss: 2.1807  
 Epoch [25/50], Iter [405/418] Loss: 2.4423, average\_loss: 2.1859  
 Epoch [25/50], Iter [410/418] Loss: 2.0946, average\_loss: 2.1847  
 Epoch [25/50], Iter [415/418] Loss: 2.7283, average\_loss: 2.1843  
 Updating best test loss: 2.76853

Starting epoch 26 / 50

Learning Rate for this epoch: 0.001

Epoch [26/50], Iter [5/418] Loss: 1.7417, average\_loss: 2.0609  
 Epoch [26/50], Iter [10/418] Loss: 1.6547, average\_loss: 1.9202  
 Epoch [26/50], Iter [15/418] Loss: 2.1787, average\_loss: 1.9247  
 Epoch [26/50], Iter [20/418] Loss: 2.8821, average\_loss: 2.0693  
 Epoch [26/50], Iter [25/418] Loss: 2.4167, average\_loss: 2.1138  
 Epoch [26/50], Iter [30/418] Loss: 2.6850, average\_loss: 2.1398  
 Epoch [26/50], Iter [35/418] Loss: 1.2200, average\_loss: 2.1422  
 Epoch [26/50], Iter [40/418] Loss: 2.2496, average\_loss: 2.0988  
 Epoch [26/50], Iter [45/418] Loss: 2.5068, average\_loss: 2.1236  
 Epoch [26/50], Iter [50/418] Loss: 1.8745, average\_loss: 2.1085  
 Epoch [26/50], Iter [55/418] Loss: 2.5490, average\_loss: 2.0913  
 Epoch [26/50], Iter [60/418] Loss: 1.7092, average\_loss: 2.0801  
 Epoch [26/50], Iter [65/418] Loss: 1.7302, average\_loss: 2.0596  
 Epoch [26/50], Iter [70/418] Loss: 2.3515, average\_loss: 2.0877  
 Epoch [26/50], Iter [75/418] Loss: 1.3904, average\_loss: 2.0689  
 Epoch [26/50], Iter [80/418] Loss: 3.2982, average\_loss: 2.1042  
 Epoch [26/50], Iter [85/418] Loss: 2.2473, average\_loss: 2.1267  
 Epoch [26/50], Iter [90/418] Loss: 1.4654, average\_loss: 2.1112  
 Epoch [26/50], Iter [95/418] Loss: 2.1146, average\_loss: 2.1108  
 Epoch [26/50], Iter [100/418] Loss: 1.6462, average\_loss: 2.1095

Epoch [26/50], Iter [105/418] Loss: 1.8309, average\_loss: 2.0992  
 Epoch [26/50], Iter [110/418] Loss: 1.6775, average\_loss: 2.1131  
 Epoch [26/50], Iter [115/418] Loss: 1.7966, average\_loss: 2.1363  
 Epoch [26/50], Iter [120/418] Loss: 1.5722, average\_loss: 2.1315  
 Epoch [26/50], Iter [125/418] Loss: 1.1889, average\_loss: 2.1301  
 Epoch [26/50], Iter [130/418] Loss: 1.5043, average\_loss: 2.1299  
 Epoch [26/50], Iter [135/418] Loss: 2.0239, average\_loss: 2.1360  
 Epoch [26/50], Iter [140/418] Loss: 2.8704, average\_loss: 2.1478  
 Epoch [26/50], Iter [145/418] Loss: 1.8610, average\_loss: 2.1502  
 Epoch [26/50], Iter [150/418] Loss: 1.7492, average\_loss: 2.1509  
 Epoch [26/50], Iter [155/418] Loss: 2.0997, average\_loss: 2.1485  
 Epoch [26/50], Iter [160/418] Loss: 2.7193, average\_loss: 2.1492  
 Epoch [26/50], Iter [165/418] Loss: 1.5145, average\_loss: 2.1455  
 Epoch [26/50], Iter [170/418] Loss: 1.7269, average\_loss: 2.1458  
 Epoch [26/50], Iter [175/418] Loss: 1.6760, average\_loss: 2.1381  
 Epoch [26/50], Iter [180/418] Loss: 1.4274, average\_loss: 2.1430  
 Epoch [26/50], Iter [185/418] Loss: 1.4452, average\_loss: 2.1375  
 Epoch [26/50], Iter [190/418] Loss: 1.7558, average\_loss: 2.1435  
 Epoch [26/50], Iter [195/418] Loss: 1.8621, average\_loss: 2.1415  
 Epoch [26/50], Iter [200/418] Loss: 2.4962, average\_loss: 2.1548  
 Epoch [26/50], Iter [205/418] Loss: 2.9575, average\_loss: 2.1754  
 Epoch [26/50], Iter [210/418] Loss: 2.2645, average\_loss: 2.1776  
 Epoch [26/50], Iter [215/418] Loss: 1.7427, average\_loss: 2.1830  
 Epoch [26/50], Iter [220/418] Loss: 1.3812, average\_loss: 2.1784  
 Epoch [26/50], Iter [225/418] Loss: 1.5118, average\_loss: 2.1771  
 Epoch [26/50], Iter [230/418] Loss: 2.6220, average\_loss: 2.1779  
 Epoch [26/50], Iter [235/418] Loss: 2.2329, average\_loss: 2.1806  
 Epoch [26/50], Iter [240/418] Loss: 2.1030, average\_loss: 2.1884  
 Epoch [26/50], Iter [245/418] Loss: 2.4165, average\_loss: 2.1886  
 Epoch [26/50], Iter [250/418] Loss: 1.8489, average\_loss: 2.1816  
 Epoch [26/50], Iter [255/418] Loss: 1.8875, average\_loss: 2.1797  
 Epoch [26/50], Iter [260/418] Loss: 2.5248, average\_loss: 2.1727  
 Epoch [26/50], Iter [265/418] Loss: 2.0588, average\_loss: 2.1677  
 Epoch [26/50], Iter [270/418] Loss: 1.4627, average\_loss: 2.1636  
 Epoch [26/50], Iter [275/418] Loss: 2.3928, average\_loss: 2.1630  
 Epoch [26/50], Iter [280/418] Loss: 2.0058, average\_loss: 2.1613  
 Epoch [26/50], Iter [285/418] Loss: 1.5625, average\_loss: 2.1569  
 Epoch [26/50], Iter [290/418] Loss: 1.4886, average\_loss: 2.1638  
 Epoch [26/50], Iter [295/418] Loss: 3.3272, average\_loss: 2.1666  
 Epoch [26/50], Iter [300/418] Loss: 1.5664, average\_loss: 2.1631  
 Epoch [26/50], Iter [305/418] Loss: 1.2455, average\_loss: 2.1627  
 Epoch [26/50], Iter [310/418] Loss: 2.0860, average\_loss: 2.1644  
 Epoch [26/50], Iter [315/418] Loss: 2.8896, average\_loss: 2.1709  
 Epoch [26/50], Iter [320/418] Loss: 1.8197, average\_loss: 2.1720  
 Epoch [26/50], Iter [325/418] Loss: 3.0325, average\_loss: 2.1703  
 Epoch [26/50], Iter [330/418] Loss: 2.4496, average\_loss: 2.1711  
 Epoch [26/50], Iter [335/418] Loss: 2.0565, average\_loss: 2.1703  
 Epoch [26/50], Iter [340/418] Loss: 1.7147, average\_loss: 2.1716

Epoch [26/50], Iter [345/418] Loss: 3.6117, average\_loss: 2.1771  
Epoch [26/50], Iter [350/418] Loss: 2.4716, average\_loss: 2.1773  
Epoch [26/50], Iter [355/418] Loss: 2.3050, average\_loss: 2.1776  
Epoch [26/50], Iter [360/418] Loss: 2.4708, average\_loss: 2.1767  
Epoch [26/50], Iter [365/418] Loss: 2.6090, average\_loss: 2.1755  
Epoch [26/50], Iter [370/418] Loss: 2.0978, average\_loss: 2.1786  
Epoch [26/50], Iter [375/418] Loss: 2.6227, average\_loss: 2.1792  
Epoch [26/50], Iter [380/418] Loss: 1.8935, average\_loss: 2.1773  
Epoch [26/50], Iter [385/418] Loss: 1.7658, average\_loss: 2.1778  
Epoch [26/50], Iter [390/418] Loss: 2.9299, average\_loss: 2.1780  
Epoch [26/50], Iter [395/418] Loss: 2.1351, average\_loss: 2.1708  
Epoch [26/50], Iter [400/418] Loss: 1.8925, average\_loss: 2.1700  
Epoch [26/50], Iter [405/418] Loss: 2.7393, average\_loss: 2.1708  
Epoch [26/50], Iter [410/418] Loss: 1.8996, average\_loss: 2.1689  
Epoch [26/50], Iter [415/418] Loss: 2.0320, average\_loss: 2.1806  
---Evaluate model on test samples---

100%|

| 4950/4950 [02:30<00:00, 32.88it/s]

---class aeroplane ap 0.4114940259725325---  
---class bicycle ap 0.4926805419220275---  
---class bird ap 0.44816190017699---  
---class boat ap 0.22395884432813185---  
---class bottle ap 0.20258872317566234---  
---class bus ap 0.49812768746677494---  
---class car ap 0.6246878486931444---  
---class cat ap 0.6520388169151785---  
---class chair ap 0.2558906736454172---  
---class cow ap 0.45855420708828465---  
---class diningtable ap 0.2117889197286326---  
---class dog ap 0.594121433373076---  
---class horse ap 0.6059119025649435---  
---class motorbike ap 0.48962408246379624---  
---class person ap 0.526683287007056---  
---class pottedplant ap 0.15723934105504644---  
---class sheep ap 0.4102061576996643---  
---class sofa ap 0.36423257693748085---  
---class train ap 0.5984478142756032---  
---class tvmonitor ap 0.3788770988383344---  
---map 0.4302657941663889---

Starting epoch 27 / 50

Learning Rate for this epoch: 0.001

Epoch [27/50], Iter [5/418] Loss: 1.6646, average\_loss: 2.0954  
Epoch [27/50], Iter [10/418] Loss: 2.1113, average\_loss: 2.2226  
Epoch [27/50], Iter [15/418] Loss: 1.3117, average\_loss: 2.2713  
Epoch [27/50], Iter [20/418] Loss: 1.8166, average\_loss: 2.1569

Epoch [27/50], Iter [25/418] Loss: 1.6640, average\_loss: 2.2011  
 Epoch [27/50], Iter [30/418] Loss: 3.1871, average\_loss: 2.1952  
 Epoch [27/50], Iter [35/418] Loss: 1.8791, average\_loss: 2.1818  
 Epoch [27/50], Iter [40/418] Loss: 1.7467, average\_loss: 2.1374  
 Epoch [27/50], Iter [45/418] Loss: 2.1512, average\_loss: 2.1072  
 Epoch [27/50], Iter [50/418] Loss: 1.4022, average\_loss: 2.1133  
 Epoch [27/50], Iter [55/418] Loss: 2.4915, average\_loss: 2.1325  
 Epoch [27/50], Iter [60/418] Loss: 1.9964, average\_loss: 2.1307  
 Epoch [27/50], Iter [65/418] Loss: 1.6525, average\_loss: 2.1120  
 Epoch [27/50], Iter [70/418] Loss: 1.8727, average\_loss: 2.1037  
 Epoch [27/50], Iter [75/418] Loss: 1.9612, average\_loss: 2.1120  
 Epoch [27/50], Iter [80/418] Loss: 2.4198, average\_loss: 2.1386  
 Epoch [27/50], Iter [85/418] Loss: 2.1282, average\_loss: 2.1464  
 Epoch [27/50], Iter [90/418] Loss: 1.7778, average\_loss: 2.1368  
 Epoch [27/50], Iter [95/418] Loss: 2.6270, average\_loss: 2.1603  
 Epoch [27/50], Iter [100/418] Loss: 1.8635, average\_loss: 2.1549  
 Epoch [27/50], Iter [105/418] Loss: 2.3369, average\_loss: 2.1400  
 Epoch [27/50], Iter [110/418] Loss: 1.3191, average\_loss: 2.1568  
 Epoch [27/50], Iter [115/418] Loss: 1.7557, average\_loss: 2.1461  
 Epoch [27/50], Iter [120/418] Loss: 1.9571, average\_loss: 2.1458  
 Epoch [27/50], Iter [125/418] Loss: 4.1045, average\_loss: 2.1573  
 Epoch [27/50], Iter [130/418] Loss: 1.6584, average\_loss: 2.1591  
 Epoch [27/50], Iter [135/418] Loss: 3.7424, average\_loss: 2.1746  
 Epoch [27/50], Iter [140/418] Loss: 1.9868, average\_loss: 2.1556  
 Epoch [27/50], Iter [145/418] Loss: 2.5412, average\_loss: 2.1520  
 Epoch [27/50], Iter [150/418] Loss: 1.2496, average\_loss: 2.1263  
 Epoch [27/50], Iter [155/418] Loss: 2.8495, average\_loss: 2.1341  
 Epoch [27/50], Iter [160/418] Loss: 2.1908, average\_loss: 2.1360  
 Epoch [27/50], Iter [165/418] Loss: 2.3216, average\_loss: 2.1442  
 Epoch [27/50], Iter [170/418] Loss: 4.0050, average\_loss: 2.1599  
 Epoch [27/50], Iter [175/418] Loss: 1.5581, average\_loss: 2.1573  
 Epoch [27/50], Iter [180/418] Loss: 1.7553, average\_loss: 2.1463  
 Epoch [27/50], Iter [185/418] Loss: 1.8816, average\_loss: 2.1398  
 Epoch [27/50], Iter [190/418] Loss: 1.4029, average\_loss: 2.1386  
 Epoch [27/50], Iter [195/418] Loss: 2.8350, average\_loss: 2.1513  
 Epoch [27/50], Iter [200/418] Loss: 2.6137, average\_loss: 2.1640  
 Epoch [27/50], Iter [205/418] Loss: 2.0265, average\_loss: 2.1607  
 Epoch [27/50], Iter [210/418] Loss: 3.0357, average\_loss: 2.1602  
 Epoch [27/50], Iter [215/418] Loss: 0.9633, average\_loss: 2.1580  
 Epoch [27/50], Iter [220/418] Loss: 2.0382, average\_loss: 2.1567  
 Epoch [27/50], Iter [225/418] Loss: 2.0606, average\_loss: 2.1599  
 Epoch [27/50], Iter [230/418] Loss: 1.8763, average\_loss: 2.1532  
 Epoch [27/50], Iter [235/418] Loss: 2.2666, average\_loss: 2.1537  
 Epoch [27/50], Iter [240/418] Loss: 2.0940, average\_loss: 2.1544  
 Epoch [27/50], Iter [245/418] Loss: 2.0939, average\_loss: 2.1575  
 Epoch [27/50], Iter [250/418] Loss: 3.7446, average\_loss: 2.1664  
 Epoch [27/50], Iter [255/418] Loss: 2.8451, average\_loss: 2.1718  
 Epoch [27/50], Iter [260/418] Loss: 2.0901, average\_loss: 2.1656



Epoch [27/50], Iter [265/418] Loss: 1.9267, average\_loss: 2.1645  
 Epoch [27/50], Iter [270/418] Loss: 2.8824, average\_loss: 2.1644  
 Epoch [27/50], Iter [275/418] Loss: 2.3207, average\_loss: 2.1615  
 Epoch [27/50], Iter [280/418] Loss: 1.8958, average\_loss: 2.1573  
 Epoch [27/50], Iter [285/418] Loss: 3.2197, average\_loss: 2.1561  
 Epoch [27/50], Iter [290/418] Loss: 4.3262, average\_loss: 2.1584  
 Epoch [27/50], Iter [295/418] Loss: 1.5179, average\_loss: 2.1621  
 Epoch [27/50], Iter [300/418] Loss: 2.3084, average\_loss: 2.1664  
 Epoch [27/50], Iter [305/418] Loss: 2.9751, average\_loss: 2.1684  
 Epoch [27/50], Iter [310/418] Loss: 2.5492, average\_loss: 2.1669  
 Epoch [27/50], Iter [315/418] Loss: 1.4325, average\_loss: 2.1637  
 Epoch [27/50], Iter [320/418] Loss: 2.0909, average\_loss: 2.1642  
 Epoch [27/50], Iter [325/418] Loss: 2.0177, average\_loss: 2.1619  
 Epoch [27/50], Iter [330/418] Loss: 1.1945, average\_loss: 2.1532  
 Epoch [27/50], Iter [335/418] Loss: 1.0744, average\_loss: 2.1502  
 Epoch [27/50], Iter [340/418] Loss: 2.1528, average\_loss: 2.1576  
 Epoch [27/50], Iter [345/418] Loss: 1.1446, average\_loss: 2.1597  
 Epoch [27/50], Iter [350/418] Loss: 2.1665, average\_loss: 2.1578  
 Epoch [27/50], Iter [355/418] Loss: 1.8789, average\_loss: 2.1604  
 Epoch [27/50], Iter [360/418] Loss: 2.3670, average\_loss: 2.1592  
 Epoch [27/50], Iter [365/418] Loss: 2.9575, average\_loss: 2.1604  
 Epoch [27/50], Iter [370/418] Loss: 1.8765, average\_loss: 2.1556  
 Epoch [27/50], Iter [375/418] Loss: 1.3463, average\_loss: 2.1512  
 Epoch [27/50], Iter [380/418] Loss: 3.0620, average\_loss: 2.1497  
 Epoch [27/50], Iter [385/418] Loss: 2.8823, average\_loss: 2.1525  
 Epoch [27/50], Iter [390/418] Loss: 1.3703, average\_loss: 2.1535  
 Epoch [27/50], Iter [395/418] Loss: 1.9362, average\_loss: 2.1517  
 Epoch [27/50], Iter [400/418] Loss: 1.4003, average\_loss: 2.1512  
 Epoch [27/50], Iter [405/418] Loss: 1.7628, average\_loss: 2.1546  
 Epoch [27/50], Iter [410/418] Loss: 1.7717, average\_loss: 2.1542  
 Epoch [27/50], Iter [415/418] Loss: 2.1035, average\_loss: 2.1510  
 Updating best test loss: 2.72013

Starting epoch 28 / 50

Learning Rate for this epoch: 0.001

Epoch [28/50], Iter [5/418] Loss: 1.7128, average\_loss: 1.6312  
 Epoch [28/50], Iter [10/418] Loss: 1.3901, average\_loss: 1.9066  
 Epoch [28/50], Iter [15/418] Loss: 2.5809, average\_loss: 1.8721  
 Epoch [28/50], Iter [20/418] Loss: 2.1079, average\_loss: 1.8774  
 Epoch [28/50], Iter [25/418] Loss: 1.8883, average\_loss: 1.9911  
 Epoch [28/50], Iter [30/418] Loss: 1.6042, average\_loss: 1.9919  
 Epoch [28/50], Iter [35/418] Loss: 1.9467, average\_loss: 2.0170  
 Epoch [28/50], Iter [40/418] Loss: 1.5007, average\_loss: 2.0433  
 Epoch [28/50], Iter [45/418] Loss: 1.4155, average\_loss: 2.0245  
 Epoch [28/50], Iter [50/418] Loss: 1.6979, average\_loss: 2.0005  
 Epoch [28/50], Iter [55/418] Loss: 1.4660, average\_loss: 2.0170  
 Epoch [28/50], Iter [60/418] Loss: 1.6645, average\_loss: 2.0124

Epoch [28/50], Iter [65/418] Loss: 1.7504, average\_loss: 2.0265  
 Epoch [28/50], Iter [70/418] Loss: 4.5120, average\_loss: 2.0912  
 Epoch [28/50], Iter [75/418] Loss: 2.6784, average\_loss: 2.1051  
 Epoch [28/50], Iter [80/418] Loss: 3.1701, average\_loss: 2.1273  
 Epoch [28/50], Iter [85/418] Loss: 2.3557, average\_loss: 2.1116  
 Epoch [28/50], Iter [90/418] Loss: 2.4866, average\_loss: 2.1198  
 Epoch [28/50], Iter [95/418] Loss: 1.5646, average\_loss: 2.1265  
 Epoch [28/50], Iter [100/418] Loss: 2.5518, average\_loss: 2.1411  
 Epoch [28/50], Iter [105/418] Loss: 2.2605, average\_loss: 2.1315  
 Epoch [28/50], Iter [110/418] Loss: 2.1635, average\_loss: 2.1298  
 Epoch [28/50], Iter [115/418] Loss: 2.3974, average\_loss: 2.1217  
 Epoch [28/50], Iter [120/418] Loss: 1.9602, average\_loss: 2.1168  
 Epoch [28/50], Iter [125/418] Loss: 2.4410, average\_loss: 2.1156  
 Epoch [28/50], Iter [130/418] Loss: 1.8767, average\_loss: 2.1407  
 Epoch [28/50], Iter [135/418] Loss: 1.8134, average\_loss: 2.1461  
 Epoch [28/50], Iter [140/418] Loss: 1.2458, average\_loss: 2.1414  
 Epoch [28/50], Iter [145/418] Loss: 1.9026, average\_loss: 2.1263  
 Epoch [28/50], Iter [150/418] Loss: 2.5937, average\_loss: 2.1238  
 Epoch [28/50], Iter [155/418] Loss: 2.5002, average\_loss: 2.1153  
 Epoch [28/50], Iter [160/418] Loss: 2.4381, average\_loss: 2.1207  
 Epoch [28/50], Iter [165/418] Loss: 1.5539, average\_loss: 2.1209  
 Epoch [28/50], Iter [170/418] Loss: 2.7772, average\_loss: 2.1177  
 Epoch [28/50], Iter [175/418] Loss: 1.8003, average\_loss: 2.1084  
 Epoch [28/50], Iter [180/418] Loss: 1.2239, average\_loss: 2.1024  
 Epoch [28/50], Iter [185/418] Loss: 2.4869, average\_loss: 2.0969  
 Epoch [28/50], Iter [190/418] Loss: 1.9078, average\_loss: 2.1041  
 Epoch [28/50], Iter [195/418] Loss: 1.1233, average\_loss: 2.1068  
 Epoch [28/50], Iter [200/418] Loss: 3.1791, average\_loss: 2.1187  
 Epoch [28/50], Iter [205/418] Loss: 2.3713, average\_loss: 2.1117  
 Epoch [28/50], Iter [210/418] Loss: 1.4245, average\_loss: 2.1084  
 Epoch [28/50], Iter [215/418] Loss: 1.8634, average\_loss: 2.1046  
 Epoch [28/50], Iter [220/418] Loss: 2.5642, average\_loss: 2.1036  
 Epoch [28/50], Iter [225/418] Loss: 2.5123, average\_loss: 2.1120  
 Epoch [28/50], Iter [230/418] Loss: 2.9795, average\_loss: 2.1138  
 Epoch [28/50], Iter [235/418] Loss: 2.3598, average\_loss: 2.1248  
 Epoch [28/50], Iter [240/418] Loss: 2.1666, average\_loss: 2.1222  
 Epoch [28/50], Iter [245/418] Loss: 2.1129, average\_loss: 2.1215  
 Epoch [28/50], Iter [250/418] Loss: 1.6972, average\_loss: 2.1189  
 Epoch [28/50], Iter [255/418] Loss: 1.9781, average\_loss: 2.1142  
 Epoch [28/50], Iter [260/418] Loss: 2.2522, average\_loss: 2.1194  
 Epoch [28/50], Iter [265/418] Loss: 3.0667, average\_loss: 2.1204  
 Epoch [28/50], Iter [270/418] Loss: 1.8262, average\_loss: 2.1155  
 Epoch [28/50], Iter [275/418] Loss: 1.8587, average\_loss: 2.1111  
 Epoch [28/50], Iter [280/418] Loss: 2.2809, average\_loss: 2.1093  
 Epoch [28/50], Iter [285/418] Loss: 1.8990, average\_loss: 2.1055  
 Epoch [28/50], Iter [290/418] Loss: 1.4320, average\_loss: 2.1084  
 Epoch [28/50], Iter [295/418] Loss: 2.6357, average\_loss: 2.1040  
 Epoch [28/50], Iter [300/418] Loss: 1.6675, average\_loss: 2.1101

Epoch [28/50], Iter [305/418] Loss: 1.6990, average\_loss: 2.1101  
Epoch [28/50], Iter [310/418] Loss: 1.8469, average\_loss: 2.1105  
Epoch [28/50], Iter [315/418] Loss: 1.0752, average\_loss: 2.1062  
Epoch [28/50], Iter [320/418] Loss: 1.6295, average\_loss: 2.1053  
Epoch [28/50], Iter [325/418] Loss: 2.0545, average\_loss: 2.1098  
Epoch [28/50], Iter [330/418] Loss: 2.9756, average\_loss: 2.1122  
Epoch [28/50], Iter [335/418] Loss: 2.1738, average\_loss: 2.1101  
Epoch [28/50], Iter [340/418] Loss: 1.5154, average\_loss: 2.1086  
Epoch [28/50], Iter [345/418] Loss: 1.5951, average\_loss: 2.1103  
Epoch [28/50], Iter [350/418] Loss: 2.8745, average\_loss: 2.1119  
Epoch [28/50], Iter [355/418] Loss: 2.8742, average\_loss: 2.1158  
Epoch [28/50], Iter [360/418] Loss: 2.1673, average\_loss: 2.1144  
Epoch [28/50], Iter [365/418] Loss: 1.4803, average\_loss: 2.1103  
Epoch [28/50], Iter [370/418] Loss: 3.9997, average\_loss: 2.1144  
Epoch [28/50], Iter [375/418] Loss: 2.7191, average\_loss: 2.1182  
Epoch [28/50], Iter [380/418] Loss: 2.7878, average\_loss: 2.1209  
Epoch [28/50], Iter [385/418] Loss: 2.1922, average\_loss: 2.1191  
Epoch [28/50], Iter [390/418] Loss: 2.5057, average\_loss: 2.1212  
Epoch [28/50], Iter [395/418] Loss: 2.0857, average\_loss: 2.1230  
Epoch [28/50], Iter [400/418] Loss: 2.1482, average\_loss: 2.1245  
Epoch [28/50], Iter [405/418] Loss: 1.9874, average\_loss: 2.1262  
Epoch [28/50], Iter [410/418] Loss: 2.7037, average\_loss: 2.1248  
Epoch [28/50], Iter [415/418] Loss: 2.0955, average\_loss: 2.1276

Starting epoch 29 / 50

Learning Rate for this epoch: 0.001

Epoch [29/50], Iter [5/418] Loss: 1.9412, average\_loss: 2.3551  
Epoch [29/50], Iter [10/418] Loss: 2.6277, average\_loss: 2.4480  
Epoch [29/50], Iter [15/418] Loss: 2.3997, average\_loss: 2.3309  
Epoch [29/50], Iter [20/418] Loss: 2.1265, average\_loss: 2.1994  
Epoch [29/50], Iter [25/418] Loss: 2.0746, average\_loss: 2.1191  
Epoch [29/50], Iter [30/418] Loss: 3.1495, average\_loss: 2.0933  
Epoch [29/50], Iter [35/418] Loss: 2.4777, average\_loss: 2.0833  
Epoch [29/50], Iter [40/418] Loss: 1.6803, average\_loss: 2.0552  
Epoch [29/50], Iter [45/418] Loss: 1.8967, average\_loss: 2.0901  
Epoch [29/50], Iter [50/418] Loss: 1.7753, average\_loss: 2.1003  
Epoch [29/50], Iter [55/418] Loss: 1.9512, average\_loss: 2.0892  
Epoch [29/50], Iter [60/418] Loss: 1.6200, average\_loss: 2.1212  
Epoch [29/50], Iter [65/418] Loss: 1.7762, average\_loss: 2.1547  
Epoch [29/50], Iter [70/418] Loss: 1.1334, average\_loss: 2.1115  
Epoch [29/50], Iter [75/418] Loss: 1.4423, average\_loss: 2.1076  
Epoch [29/50], Iter [80/418] Loss: 1.6454, average\_loss: 2.1261  
Epoch [29/50], Iter [85/418] Loss: 2.0882, average\_loss: 2.1262  
Epoch [29/50], Iter [90/418] Loss: 1.7482, average\_loss: 2.1069  
Epoch [29/50], Iter [95/418] Loss: 2.6235, average\_loss: 2.1005  
Epoch [29/50], Iter [100/418] Loss: 1.8471, average\_loss: 2.1249  
Epoch [29/50], Iter [105/418] Loss: 1.7617, average\_loss: 2.1295

Epoch [29/50], Iter [110/418] Loss: 2.0321, average\_loss: 2.1378  
 Epoch [29/50], Iter [115/418] Loss: 1.8909, average\_loss: 2.1311  
 Epoch [29/50], Iter [120/418] Loss: 1.8835, average\_loss: 2.1270  
 Epoch [29/50], Iter [125/418] Loss: 1.8199, average\_loss: 2.1228  
 Epoch [29/50], Iter [130/418] Loss: 3.1225, average\_loss: 2.1390  
 Epoch [29/50], Iter [135/418] Loss: 2.2365, average\_loss: 2.1319  
 Epoch [29/50], Iter [140/418] Loss: 1.2708, average\_loss: 2.1224  
 Epoch [29/50], Iter [145/418] Loss: 2.6241, average\_loss: 2.1294  
 Epoch [29/50], Iter [150/418] Loss: 2.6946, average\_loss: 2.1168  
 Epoch [29/50], Iter [155/418] Loss: 2.0250, average\_loss: 2.1160  
 Epoch [29/50], Iter [160/418] Loss: 1.6849, average\_loss: 2.1196  
 Epoch [29/50], Iter [165/418] Loss: 2.0489, average\_loss: 2.1218  
 Epoch [29/50], Iter [170/418] Loss: 1.9308, average\_loss: 2.1248  
 Epoch [29/50], Iter [175/418] Loss: 1.8487, average\_loss: 2.1279  
 Epoch [29/50], Iter [180/418] Loss: 2.5579, average\_loss: 2.1195  
 Epoch [29/50], Iter [185/418] Loss: 1.3409, average\_loss: 2.1125  
 Epoch [29/50], Iter [190/418] Loss: 2.9760, average\_loss: 2.1195  
 Epoch [29/50], Iter [195/418] Loss: 1.5607, average\_loss: 2.1179  
 Epoch [29/50], Iter [200/418] Loss: 1.8161, average\_loss: 2.1193  
 Epoch [29/50], Iter [205/418] Loss: 2.8506, average\_loss: 2.1213  
 Epoch [29/50], Iter [210/418] Loss: 1.7932, average\_loss: 2.1169  
 Epoch [29/50], Iter [215/418] Loss: 3.4070, average\_loss: 2.1270  
 Epoch [29/50], Iter [220/418] Loss: 2.2650, average\_loss: 2.1177  
 Epoch [29/50], Iter [225/418] Loss: 3.0633, average\_loss: 2.1186  
 Epoch [29/50], Iter [230/418] Loss: 2.0122, average\_loss: 2.1180  
 Epoch [29/50], Iter [235/418] Loss: 1.7807, average\_loss: 2.1186  
 Epoch [29/50], Iter [240/418] Loss: 2.9008, average\_loss: 2.1160  
 Epoch [29/50], Iter [245/418] Loss: 2.1975, average\_loss: 2.1175  
 Epoch [29/50], Iter [250/418] Loss: 1.7417, average\_loss: 2.1224  
 Epoch [29/50], Iter [255/418] Loss: 2.0252, average\_loss: 2.1263  
 Epoch [29/50], Iter [260/418] Loss: 1.6596, average\_loss: 2.1319  
 Epoch [29/50], Iter [265/418] Loss: 1.6071, average\_loss: 2.1253  
 Epoch [29/50], Iter [270/418] Loss: 1.5726, average\_loss: 2.1248  
 Epoch [29/50], Iter [275/418] Loss: 0.9664, average\_loss: 2.1255  
 Epoch [29/50], Iter [280/418] Loss: 2.3136, average\_loss: 2.1235  
 Epoch [29/50], Iter [285/418] Loss: 1.7591, average\_loss: 2.1169  
 Epoch [29/50], Iter [290/418] Loss: 2.8336, average\_loss: 2.1173  
 Epoch [29/50], Iter [295/418] Loss: 1.9223, average\_loss: 2.1127  
 Epoch [29/50], Iter [300/418] Loss: 1.7944, average\_loss: 2.1046  
 Epoch [29/50], Iter [305/418] Loss: 1.9235, average\_loss: 2.1069  
 Epoch [29/50], Iter [310/418] Loss: 1.5119, average\_loss: 2.1003  
 Epoch [29/50], Iter [315/418] Loss: 1.4559, average\_loss: 2.0934  
 Epoch [29/50], Iter [320/418] Loss: 1.7656, average\_loss: 2.0946  
 Epoch [29/50], Iter [325/418] Loss: 1.5657, average\_loss: 2.0982  
 Epoch [29/50], Iter [330/418] Loss: 1.4780, average\_loss: 2.0981  
 Epoch [29/50], Iter [335/418] Loss: 1.9434, average\_loss: 2.0985  
 Epoch [29/50], Iter [340/418] Loss: 2.8061, average\_loss: 2.1016  
 Epoch [29/50], Iter [345/418] Loss: 1.6825, average\_loss: 2.1004

Epoch [29/50], Iter [350/418] Loss: 1.5956, average\_loss: 2.1047  
Epoch [29/50], Iter [355/418] Loss: 2.6712, average\_loss: 2.1071  
Epoch [29/50], Iter [360/418] Loss: 1.7821, average\_loss: 2.1082  
Epoch [29/50], Iter [365/418] Loss: 1.6098, average\_loss: 2.1106  
Epoch [29/50], Iter [370/418] Loss: 1.2433, average\_loss: 2.1079  
Epoch [29/50], Iter [375/418] Loss: 2.9325, average\_loss: 2.1071  
Epoch [29/50], Iter [380/418] Loss: 2.4513, average\_loss: 2.1101  
Epoch [29/50], Iter [385/418] Loss: 1.9747, average\_loss: 2.1080  
Epoch [29/50], Iter [390/418] Loss: 2.2975, average\_loss: 2.1109  
Epoch [29/50], Iter [395/418] Loss: 1.5937, average\_loss: 2.1101  
Epoch [29/50], Iter [400/418] Loss: 2.7988, average\_loss: 2.1080  
Epoch [29/50], Iter [405/418] Loss: 1.7095, average\_loss: 2.1051  
Epoch [29/50], Iter [410/418] Loss: 3.1387, average\_loss: 2.1089  
Epoch [29/50], Iter [415/418] Loss: 1.9981, average\_loss: 2.1165

Starting epoch 30 / 50

Learning Rate for this epoch: 0.001

Epoch [30/50], Iter [5/418] Loss: 1.8762, average\_loss: 2.3770  
Epoch [30/50], Iter [10/418] Loss: 1.9221, average\_loss: 2.1539  
Epoch [30/50], Iter [15/418] Loss: 1.5434, average\_loss: 2.1897  
Epoch [30/50], Iter [20/418] Loss: 2.0226, average\_loss: 2.1496  
Epoch [30/50], Iter [25/418] Loss: 2.4218, average\_loss: 2.1505  
Epoch [30/50], Iter [30/418] Loss: 2.4820, average\_loss: 2.1013  
Epoch [30/50], Iter [35/418] Loss: 2.2464, average\_loss: 2.1460  
Epoch [30/50], Iter [40/418] Loss: 1.4994, average\_loss: 2.0732  
Epoch [30/50], Iter [45/418] Loss: 1.7967, average\_loss: 2.0487  
Epoch [30/50], Iter [50/418] Loss: 2.5873, average\_loss: 2.0907  
Epoch [30/50], Iter [55/418] Loss: 1.6826, average\_loss: 2.1037  
Epoch [30/50], Iter [60/418] Loss: 1.7476, average\_loss: 2.0949  
Epoch [30/50], Iter [65/418] Loss: 1.8951, average\_loss: 2.0980  
Epoch [30/50], Iter [70/418] Loss: 3.0339, average\_loss: 2.0986  
Epoch [30/50], Iter [75/418] Loss: 2.6793, average\_loss: 2.0988  
Epoch [30/50], Iter [80/418] Loss: 2.5199, average\_loss: 2.1000  
Epoch [30/50], Iter [85/418] Loss: 2.9060, average\_loss: 2.1071  
Epoch [30/50], Iter [90/418] Loss: 1.5788, average\_loss: 2.1400  
Epoch [30/50], Iter [95/418] Loss: 1.8486, average\_loss: 2.1225  
Epoch [30/50], Iter [100/418] Loss: 2.3986, average\_loss: 2.1263  
Epoch [30/50], Iter [105/418] Loss: 3.3719, average\_loss: 2.1343  
Epoch [30/50], Iter [110/418] Loss: 2.5529, average\_loss: 2.1366  
Epoch [30/50], Iter [115/418] Loss: 2.2137, average\_loss: 2.1285  
Epoch [30/50], Iter [120/418] Loss: 1.7843, average\_loss: 2.1186  
Epoch [30/50], Iter [125/418] Loss: 2.2942, average\_loss: 2.1290  
Epoch [30/50], Iter [130/418] Loss: 2.1959, average\_loss: 2.1096  
Epoch [30/50], Iter [135/418] Loss: 1.7324, average\_loss: 2.1011  
Epoch [30/50], Iter [140/418] Loss: 2.5559, average\_loss: 2.1078  
Epoch [30/50], Iter [145/418] Loss: 2.3964, average\_loss: 2.1072  
Epoch [30/50], Iter [150/418] Loss: 3.5408, average\_loss: 2.1065

Epoch [30/50], Iter [155/418] Loss: 1.6245, average\_loss: 2.0989  
 Epoch [30/50], Iter [160/418] Loss: 2.0289, average\_loss: 2.0913  
 Epoch [30/50], Iter [165/418] Loss: 1.8365, average\_loss: 2.0935  
 Epoch [30/50], Iter [170/418] Loss: 1.9869, average\_loss: 2.0898  
 Epoch [30/50], Iter [175/418] Loss: 2.8584, average\_loss: 2.0962  
 Epoch [30/50], Iter [180/418] Loss: 2.6068, average\_loss: 2.0973  
 Epoch [30/50], Iter [185/418] Loss: 1.5622, average\_loss: 2.0916  
 Epoch [30/50], Iter [190/418] Loss: 1.7914, average\_loss: 2.0913  
 Epoch [30/50], Iter [195/418] Loss: 1.8423, average\_loss: 2.0881  
 Epoch [30/50], Iter [200/418] Loss: 1.0794, average\_loss: 2.0779  
 Epoch [30/50], Iter [205/418] Loss: 2.5313, average\_loss: 2.0810  
 Epoch [30/50], Iter [210/418] Loss: 1.5603, average\_loss: 2.0785  
 Epoch [30/50], Iter [215/418] Loss: 2.2767, average\_loss: 2.0774  
 Epoch [30/50], Iter [220/418] Loss: 2.8736, average\_loss: 2.0837  
 Epoch [30/50], Iter [225/418] Loss: 1.6185, average\_loss: 2.0782  
 Epoch [30/50], Iter [230/418] Loss: 2.8166, average\_loss: 2.0908  
 Epoch [30/50], Iter [235/418] Loss: 2.3792, average\_loss: 2.0919  
 Epoch [30/50], Iter [240/418] Loss: 1.5179, average\_loss: 2.0823  
 Epoch [30/50], Iter [245/418] Loss: 2.3755, average\_loss: 2.0779  
 Epoch [30/50], Iter [250/418] Loss: 1.4119, average\_loss: 2.0714  
 Epoch [30/50], Iter [255/418] Loss: 2.1813, average\_loss: 2.0732  
 Epoch [30/50], Iter [260/418] Loss: 2.3273, average\_loss: 2.0823  
 Epoch [30/50], Iter [265/418] Loss: 2.2171, average\_loss: 2.0853  
 Epoch [30/50], Iter [270/418] Loss: 1.7407, average\_loss: 2.0836  
 Epoch [30/50], Iter [275/418] Loss: 2.6137, average\_loss: 2.0864  
 Epoch [30/50], Iter [280/418] Loss: 1.6643, average\_loss: 2.0891  
 Epoch [30/50], Iter [285/418] Loss: 2.1189, average\_loss: 2.0852  
 Epoch [30/50], Iter [290/418] Loss: 2.4998, average\_loss: 2.0830  
 Epoch [30/50], Iter [295/418] Loss: 2.2029, average\_loss: 2.0788  
 Epoch [30/50], Iter [300/418] Loss: 1.4151, average\_loss: 2.0822  
 Epoch [30/50], Iter [305/418] Loss: 2.4315, average\_loss: 2.0844  
 Epoch [30/50], Iter [310/418] Loss: 2.4517, average\_loss: 2.0867  
 Epoch [30/50], Iter [315/418] Loss: 2.5704, average\_loss: 2.0835  
 Epoch [30/50], Iter [320/418] Loss: 2.3983, average\_loss: 2.0833  
 Epoch [30/50], Iter [325/418] Loss: 2.0590, average\_loss: 2.0807  
 Epoch [30/50], Iter [330/418] Loss: 1.2754, average\_loss: 2.0761  
 Epoch [30/50], Iter [335/418] Loss: 1.6491, average\_loss: 2.0687  
 Epoch [30/50], Iter [340/418] Loss: 1.8774, average\_loss: 2.0652  
 Epoch [30/50], Iter [345/418] Loss: 2.3246, average\_loss: 2.0649  
 Epoch [30/50], Iter [350/418] Loss: 1.6096, average\_loss: 2.0615  
 Epoch [30/50], Iter [355/418] Loss: 3.7858, average\_loss: 2.0664  
 Epoch [30/50], Iter [360/418] Loss: 1.2530, average\_loss: 2.0630  
 Epoch [30/50], Iter [365/418] Loss: 1.8485, average\_loss: 2.0640  
 Epoch [30/50], Iter [370/418] Loss: 1.7052, average\_loss: 2.0633  
 Epoch [30/50], Iter [375/418] Loss: 2.5057, average\_loss: 2.0684  
 Epoch [30/50], Iter [380/418] Loss: 1.4933, average\_loss: 2.0624  
 Epoch [30/50], Iter [385/418] Loss: 1.2323, average\_loss: 2.0562  
 Epoch [30/50], Iter [390/418] Loss: 2.5785, average\_loss: 2.0567

Epoch [30/50], Iter [395/418] Loss: 1.6352, average\_loss: 2.0539  
Epoch [30/50], Iter [400/418] Loss: 2.3889, average\_loss: 2.0574  
Epoch [30/50], Iter [405/418] Loss: 2.1708, average\_loss: 2.0565  
Epoch [30/50], Iter [410/418] Loss: 2.4743, average\_loss: 2.0612  
Epoch [30/50], Iter [415/418] Loss: 2.1898, average\_loss: 2.0622  
Updating best test loss: 2.71970

Starting epoch 31 / 50

Learning Rate for this epoch: 0.0001

Epoch [31/50], Iter [5/418] Loss: 1.9131, average\_loss: 2.2866  
Epoch [31/50], Iter [10/418] Loss: 2.0257, average\_loss: 2.1999  
Epoch [31/50], Iter [15/418] Loss: 3.1152, average\_loss: 2.2434  
Epoch [31/50], Iter [20/418] Loss: 1.7389, average\_loss: 2.2709  
Epoch [31/50], Iter [25/418] Loss: 1.8442, average\_loss: 2.2282  
Epoch [31/50], Iter [30/418] Loss: 2.6825, average\_loss: 2.2813  
Epoch [31/50], Iter [35/418] Loss: 1.4907, average\_loss: 2.2190  
Epoch [31/50], Iter [40/418] Loss: 2.3189, average\_loss: 2.1871  
Epoch [31/50], Iter [45/418] Loss: 3.0538, average\_loss: 2.1878  
Epoch [31/50], Iter [50/418] Loss: 1.7435, average\_loss: 2.1640  
Epoch [31/50], Iter [55/418] Loss: 2.2400, average\_loss: 2.1460  
Epoch [31/50], Iter [60/418] Loss: 1.9828, average\_loss: 2.1380  
Epoch [31/50], Iter [65/418] Loss: 1.7842, average\_loss: 2.1378  
Epoch [31/50], Iter [70/418] Loss: 1.7937, average\_loss: 2.1156  
Epoch [31/50], Iter [75/418] Loss: 2.0647, average\_loss: 2.1159  
Epoch [31/50], Iter [80/418] Loss: 1.8026, average\_loss: 2.0999  
Epoch [31/50], Iter [85/418] Loss: 1.4180, average\_loss: 2.0712  
Epoch [31/50], Iter [90/418] Loss: 1.9499, average\_loss: 2.0634  
Epoch [31/50], Iter [95/418] Loss: 2.3331, average\_loss: 2.0587  
Epoch [31/50], Iter [100/418] Loss: 1.8821, average\_loss: 2.0446  
Epoch [31/50], Iter [105/418] Loss: 1.8675, average\_loss: 2.0344  
Epoch [31/50], Iter [110/418] Loss: 1.8156, average\_loss: 2.0220  
Epoch [31/50], Iter [115/418] Loss: 1.2569, average\_loss: 2.0080  
Epoch [31/50], Iter [120/418] Loss: 2.0325, average\_loss: 1.9961  
Epoch [31/50], Iter [125/418] Loss: 1.3668, average\_loss: 2.0000  
Epoch [31/50], Iter [130/418] Loss: 1.6217, average\_loss: 1.9891  
Epoch [31/50], Iter [135/418] Loss: 1.1568, average\_loss: 1.9798  
Epoch [31/50], Iter [140/418] Loss: 1.7564, average\_loss: 1.9838  
Epoch [31/50], Iter [145/418] Loss: 2.8376, average\_loss: 1.9979  
Epoch [31/50], Iter [150/418] Loss: 3.5752, average\_loss: 2.0040  
Epoch [31/50], Iter [155/418] Loss: 1.9262, average\_loss: 1.9872  
Epoch [31/50], Iter [160/418] Loss: 1.5510, average\_loss: 1.9817  
Epoch [31/50], Iter [165/418] Loss: 2.4581, average\_loss: 1.9866  
Epoch [31/50], Iter [170/418] Loss: 1.4758, average\_loss: 1.9832  
Epoch [31/50], Iter [175/418] Loss: 2.1749, average\_loss: 1.9791  
Epoch [31/50], Iter [180/418] Loss: 1.6289, average\_loss: 1.9801  
Epoch [31/50], Iter [185/418] Loss: 2.5118, average\_loss: 1.9821  
Epoch [31/50], Iter [190/418] Loss: 2.6853, average\_loss: 1.9921

```

Epoch [31/50], Iter [195/418] Loss: 2.5547, average_loss: 1.9939
Epoch [31/50], Iter [200/418] Loss: 1.8358, average_loss: 1.9983
Epoch [31/50], Iter [205/418] Loss: 1.7983, average_loss: 1.9911
Epoch [31/50], Iter [210/418] Loss: 1.1749, average_loss: 1.9848
Epoch [31/50], Iter [215/418] Loss: 1.5968, average_loss: 1.9921
Epoch [31/50], Iter [220/418] Loss: 1.7151, average_loss: 1.9853
Epoch [31/50], Iter [225/418] Loss: 1.6928, average_loss: 1.9802
Epoch [31/50], Iter [230/418] Loss: 2.6335, average_loss: 1.9787
Epoch [31/50], Iter [235/418] Loss: 3.3199, average_loss: 1.9742
Epoch [31/50], Iter [240/418] Loss: 2.1089, average_loss: 1.9719
Epoch [31/50], Iter [245/418] Loss: 1.8596, average_loss: 1.9669
Epoch [31/50], Iter [250/418] Loss: 1.6319, average_loss: 1.9649
Epoch [31/50], Iter [255/418] Loss: 1.3225, average_loss: 1.9695
Epoch [31/50], Iter [260/418] Loss: 1.5521, average_loss: 1.9723
Epoch [31/50], Iter [265/418] Loss: 1.6221, average_loss: 1.9686
Epoch [31/50], Iter [270/418] Loss: 1.4720, average_loss: 1.9646
Epoch [31/50], Iter [275/418] Loss: 3.6316, average_loss: 1.9683
Epoch [31/50], Iter [280/418] Loss: 1.4439, average_loss: 1.9604
Epoch [31/50], Iter [285/418] Loss: 1.6460, average_loss: 1.9546
Epoch [31/50], Iter [290/418] Loss: 2.0382, average_loss: 1.9521
Epoch [31/50], Iter [295/418] Loss: 1.7247, average_loss: 1.9459
Epoch [31/50], Iter [300/418] Loss: 1.6191, average_loss: 1.9474
Epoch [31/50], Iter [305/418] Loss: 1.7846, average_loss: 1.9408
Epoch [31/50], Iter [310/418] Loss: 2.8196, average_loss: 1.9458
Epoch [31/50], Iter [315/418] Loss: 2.3788, average_loss: 1.9465
Epoch [31/50], Iter [320/418] Loss: 2.9267, average_loss: 1.9478
Epoch [31/50], Iter [325/418] Loss: 1.4961, average_loss: 1.9479
Epoch [31/50], Iter [330/418] Loss: 1.8005, average_loss: 1.9491
Epoch [31/50], Iter [335/418] Loss: 2.1854, average_loss: 1.9482
Epoch [31/50], Iter [340/418] Loss: 2.5305, average_loss: 1.9490
Epoch [31/50], Iter [345/418] Loss: 1.2883, average_loss: 1.9490
Epoch [31/50], Iter [350/418] Loss: 1.3849, average_loss: 1.9445
Epoch [31/50], Iter [355/418] Loss: 2.0934, average_loss: 1.9429
Epoch [31/50], Iter [360/418] Loss: 2.8308, average_loss: 1.9426
Epoch [31/50], Iter [365/418] Loss: 1.1119, average_loss: 1.9378
Epoch [31/50], Iter [370/418] Loss: 2.1478, average_loss: 1.9387
Epoch [31/50], Iter [375/418] Loss: 1.4742, average_loss: 1.9385
Epoch [31/50], Iter [380/418] Loss: 3.4166, average_loss: 1.9422
Epoch [31/50], Iter [385/418] Loss: 1.9446, average_loss: 1.9411
Epoch [31/50], Iter [390/418] Loss: 2.0713, average_loss: 1.9433
Epoch [31/50], Iter [395/418] Loss: 1.7021, average_loss: 1.9416
Epoch [31/50], Iter [400/418] Loss: 2.1395, average_loss: 1.9422
Epoch [31/50], Iter [405/418] Loss: 2.0872, average_loss: 1.9438
Epoch [31/50], Iter [410/418] Loss: 1.7293, average_loss: 1.9413
Epoch [31/50], Iter [415/418] Loss: 1.6839, average_loss: 1.9438
---Evaluate model on test samples---

```

100%|



```
| 4950/4950 [02:29<00:00, 33.08it/s]

---class aeroplane ap 0.528877119957077---
---class bicycle ap 0.6093749675795747---
---class bird ap 0.48809334194145354---
---class boat ap 0.2822376478102236---
---class bottle ap 0.24697835153628822---
---class bus ap 0.6268886812043865---
---class car ap 0.6789248420966967---
---class cat ap 0.6692465124544416---
---class chair ap 0.2908206544751993---
---class cow ap 0.5294304359807506---
---class diningtable ap 0.3786984806442993---
---class dog ap 0.5954118139118849---
---class horse ap 0.6707539895793995---
---class motorbike ap 0.5654385256775505---
---class person ap 0.5757478616853977---
---class pottedplant ap 0.19310195212746625---
---class sheep ap 0.49180412242647353---
---class sofa ap 0.4757817901402315---
---class train ap 0.6320479546170865---
---class tvmonitor ap 0.49742760911372874---
---map 0.5013543327479806---
Updating best test loss: 2.62129
```

Starting epoch 32 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [32/50], Iter [5/418] Loss: 1.9898, average_loss: 2.1979
Epoch [32/50], Iter [10/418] Loss: 2.2365, average_loss: 2.0682
Epoch [32/50], Iter [15/418] Loss: 1.4839, average_loss: 1.9296
Epoch [32/50], Iter [20/418] Loss: 1.7368, average_loss: 1.8239
Epoch [32/50], Iter [25/418] Loss: 1.7719, average_loss: 1.8325
Epoch [32/50], Iter [30/418] Loss: 1.5645, average_loss: 1.8394
Epoch [32/50], Iter [35/418] Loss: 3.5577, average_loss: 1.8690
Epoch [32/50], Iter [40/418] Loss: 0.8694, average_loss: 1.8376
Epoch [32/50], Iter [45/418] Loss: 2.3896, average_loss: 1.8501
Epoch [32/50], Iter [50/418] Loss: 1.3663, average_loss: 1.8157
Epoch [32/50], Iter [55/418] Loss: 1.8489, average_loss: 1.8025
Epoch [32/50], Iter [60/418] Loss: 2.1302, average_loss: 1.8169
Epoch [32/50], Iter [65/418] Loss: 2.0944, average_loss: 1.8286
Epoch [32/50], Iter [70/418] Loss: 1.4038, average_loss: 1.8269
Epoch [32/50], Iter [75/418] Loss: 1.7176, average_loss: 1.8078
Epoch [32/50], Iter [80/418] Loss: 2.1238, average_loss: 1.8319
Epoch [32/50], Iter [85/418] Loss: 1.6610, average_loss: 1.8257
Epoch [32/50], Iter [90/418] Loss: 3.4949, average_loss: 1.8467
Epoch [32/50], Iter [95/418] Loss: 1.9647, average_loss: 1.8455
Epoch [32/50], Iter [100/418] Loss: 2.2646, average_loss: 1.8533
```

Epoch [32/50], Iter [105/418] Loss: 2.0566, average\_loss: 1.8556  
 Epoch [32/50], Iter [110/418] Loss: 2.6401, average\_loss: 1.8532  
 Epoch [32/50], Iter [115/418] Loss: 1.3519, average\_loss: 1.8384  
 Epoch [32/50], Iter [120/418] Loss: 2.7856, average\_loss: 1.8476  
 Epoch [32/50], Iter [125/418] Loss: 2.2161, average\_loss: 1.8533  
 Epoch [32/50], Iter [130/418] Loss: 2.0415, average\_loss: 1.8691  
 Epoch [32/50], Iter [135/418] Loss: 2.0244, average\_loss: 1.8671  
 Epoch [32/50], Iter [140/418] Loss: 1.7686, average\_loss: 1.8663  
 Epoch [32/50], Iter [145/418] Loss: 1.9402, average\_loss: 1.8674  
 Epoch [32/50], Iter [150/418] Loss: 1.8954, average\_loss: 1.8595  
 Epoch [32/50], Iter [155/418] Loss: 1.1878, average\_loss: 1.8523  
 Epoch [32/50], Iter [160/418] Loss: 1.5313, average\_loss: 1.8462  
 Epoch [32/50], Iter [165/418] Loss: 1.1264, average\_loss: 1.8472  
 Epoch [32/50], Iter [170/418] Loss: 1.9900, average\_loss: 1.8447  
 Epoch [32/50], Iter [175/418] Loss: 1.3773, average\_loss: 1.8523  
 Epoch [32/50], Iter [180/418] Loss: 1.3408, average\_loss: 1.8600  
 Epoch [32/50], Iter [185/418] Loss: 2.1724, average\_loss: 1.8567  
 Epoch [32/50], Iter [190/418] Loss: 1.1855, average\_loss: 1.8520  
 Epoch [32/50], Iter [195/418] Loss: 1.3921, average\_loss: 1.8570  
 Epoch [32/50], Iter [200/418] Loss: 1.9657, average\_loss: 1.8582  
 Epoch [32/50], Iter [205/418] Loss: 1.8920, average\_loss: 1.8569  
 Epoch [32/50], Iter [210/418] Loss: 1.3997, average\_loss: 1.8538  
 Epoch [32/50], Iter [215/418] Loss: 2.0440, average\_loss: 1.8544  
 Epoch [32/50], Iter [220/418] Loss: 1.9303, average\_loss: 1.8580  
 Epoch [32/50], Iter [225/418] Loss: 1.2979, average\_loss: 1.8595  
 Epoch [32/50], Iter [230/418] Loss: 1.7815, average\_loss: 1.8540  
 Epoch [32/50], Iter [235/418] Loss: 2.1174, average\_loss: 1.8603  
 Epoch [32/50], Iter [240/418] Loss: 1.3453, average\_loss: 1.8595  
 Epoch [32/50], Iter [245/418] Loss: 1.8235, average\_loss: 1.8690  
 Epoch [32/50], Iter [250/418] Loss: 1.3831, average\_loss: 1.8642  
 Epoch [32/50], Iter [255/418] Loss: 2.6133, average\_loss: 1.8662  
 Epoch [32/50], Iter [260/418] Loss: 1.9098, average\_loss: 1.8630  
 Epoch [32/50], Iter [265/418] Loss: 2.9830, average\_loss: 1.8762  
 Epoch [32/50], Iter [270/418] Loss: 1.9505, average\_loss: 1.8851  
 Epoch [32/50], Iter [275/418] Loss: 2.4008, average\_loss: 1.8817  
 Epoch [32/50], Iter [280/418] Loss: 1.9933, average\_loss: 1.8910  
 Epoch [32/50], Iter [285/418] Loss: 1.2346, average\_loss: 1.8858  
 Epoch [32/50], Iter [290/418] Loss: 1.6524, average\_loss: 1.8872  
 Epoch [32/50], Iter [295/418] Loss: 2.4555, average\_loss: 1.8829  
 Epoch [32/50], Iter [300/418] Loss: 2.4099, average\_loss: 1.8873  
 Epoch [32/50], Iter [305/418] Loss: 1.9611, average\_loss: 1.8887  
 Epoch [32/50], Iter [310/418] Loss: 2.1574, average\_loss: 1.8923  
 Epoch [32/50], Iter [315/418] Loss: 2.3731, average\_loss: 1.8914  
 Epoch [32/50], Iter [320/418] Loss: 2.1361, average\_loss: 1.8973  
 Epoch [32/50], Iter [325/418] Loss: 2.2709, average\_loss: 1.8937  
 Epoch [32/50], Iter [330/418] Loss: 2.2062, average\_loss: 1.8918  
 Epoch [32/50], Iter [335/418] Loss: 1.1772, average\_loss: 1.8862  
 Epoch [32/50], Iter [340/418] Loss: 2.2363, average\_loss: 1.8827

Epoch [32/50], Iter [345/418] Loss: 1.4256, average\_loss: 1.8825  
 Epoch [32/50], Iter [350/418] Loss: 1.1081, average\_loss: 1.8854  
 Epoch [32/50], Iter [355/418] Loss: 1.6555, average\_loss: 1.8839  
 Epoch [32/50], Iter [360/418] Loss: 2.1085, average\_loss: 1.8895  
 Epoch [32/50], Iter [365/418] Loss: 1.7265, average\_loss: 1.8931  
 Epoch [32/50], Iter [370/418] Loss: 2.1190, average\_loss: 1.8921  
 Epoch [32/50], Iter [375/418] Loss: 2.0342, average\_loss: 1.8945  
 Epoch [32/50], Iter [380/418] Loss: 1.5524, average\_loss: 1.8932  
 Epoch [32/50], Iter [385/418] Loss: 3.0996, average\_loss: 1.8948  
 Epoch [32/50], Iter [390/418] Loss: 2.4209, average\_loss: 1.8971  
 Epoch [32/50], Iter [395/418] Loss: 2.5875, average\_loss: 1.9001  
 Epoch [32/50], Iter [400/418] Loss: 1.6120, average\_loss: 1.8957  
 Epoch [32/50], Iter [405/418] Loss: 1.6913, average\_loss: 1.8945  
 Epoch [32/50], Iter [410/418] Loss: 1.4630, average\_loss: 1.8922  
 Epoch [32/50], Iter [415/418] Loss: 2.1381, average\_loss: 1.8969  
 Updating best test loss: 2.59806

Starting epoch 33 / 50

Learning Rate for this epoch: 0.0001

Epoch [33/50], Iter [5/418] Loss: 2.1268, average\_loss: 1.9183  
 Epoch [33/50], Iter [10/418] Loss: 1.9959, average\_loss: 1.8474  
 Epoch [33/50], Iter [15/418] Loss: 1.2567, average\_loss: 1.8021  
 Epoch [33/50], Iter [20/418] Loss: 1.7624, average\_loss: 1.9058  
 Epoch [33/50], Iter [25/418] Loss: 1.6730, average\_loss: 1.8412  
 Epoch [33/50], Iter [30/418] Loss: 1.7188, average\_loss: 1.7747  
 Epoch [33/50], Iter [35/418] Loss: 1.1850, average\_loss: 1.7792  
 Epoch [33/50], Iter [40/418] Loss: 2.2701, average\_loss: 1.7851  
 Epoch [33/50], Iter [45/418] Loss: 2.2419, average\_loss: 1.7902  
 Epoch [33/50], Iter [50/418] Loss: 1.0597, average\_loss: 1.7850  
 Epoch [33/50], Iter [55/418] Loss: 1.8912, average\_loss: 1.7579  
 Epoch [33/50], Iter [60/418] Loss: 1.2952, average\_loss: 1.7254  
 Epoch [33/50], Iter [65/418] Loss: 2.5217, average\_loss: 1.7517  
 Epoch [33/50], Iter [70/418] Loss: 1.3027, average\_loss: 1.7350  
 Epoch [33/50], Iter [75/418] Loss: 1.8773, average\_loss: 1.7767  
 Epoch [33/50], Iter [80/418] Loss: 1.0886, average\_loss: 1.7646  
 Epoch [33/50], Iter [85/418] Loss: 1.5784, average\_loss: 1.7453  
 Epoch [33/50], Iter [90/418] Loss: 1.1835, average\_loss: 1.7752  
 Epoch [33/50], Iter [95/418] Loss: 1.5344, average\_loss: 1.7722  
 Epoch [33/50], Iter [100/418] Loss: 1.1669, average\_loss: 1.7764  
 Epoch [33/50], Iter [105/418] Loss: 2.9542, average\_loss: 1.7989  
 Epoch [33/50], Iter [110/418] Loss: 1.8551, average\_loss: 1.7968  
 Epoch [33/50], Iter [115/418] Loss: 2.0008, average\_loss: 1.8273  
 Epoch [33/50], Iter [120/418] Loss: 2.0640, average\_loss: 1.8326  
 Epoch [33/50], Iter [125/418] Loss: 2.2496, average\_loss: 1.8474  
 Epoch [33/50], Iter [130/418] Loss: 2.1906, average\_loss: 1.8477  
 Epoch [33/50], Iter [135/418] Loss: 1.5616, average\_loss: 1.8483  
 Epoch [33/50], Iter [140/418] Loss: 1.2689, average\_loss: 1.8497

Epoch [33/50], Iter [145/418] Loss: 1.8696, average\_loss: 1.8445  
 Epoch [33/50], Iter [150/418] Loss: 2.3740, average\_loss: 1.8430  
 Epoch [33/50], Iter [155/418] Loss: 1.2164, average\_loss: 1.8307  
 Epoch [33/50], Iter [160/418] Loss: 1.6271, average\_loss: 1.8328  
 Epoch [33/50], Iter [165/418] Loss: 2.3249, average\_loss: 1.8386  
 Epoch [33/50], Iter [170/418] Loss: 2.6298, average\_loss: 1.8450  
 Epoch [33/50], Iter [175/418] Loss: 1.8818, average\_loss: 1.8480  
 Epoch [33/50], Iter [180/418] Loss: 3.6978, average\_loss: 1.8520  
 Epoch [33/50], Iter [185/418] Loss: 1.5760, average\_loss: 1.8520  
 Epoch [33/50], Iter [190/418] Loss: 1.6974, average\_loss: 1.8467  
 Epoch [33/50], Iter [195/418] Loss: 1.2989, average\_loss: 1.8427  
 Epoch [33/50], Iter [200/418] Loss: 1.1043, average\_loss: 1.8453  
 Epoch [33/50], Iter [205/418] Loss: 1.1917, average\_loss: 1.8450  
 Epoch [33/50], Iter [210/418] Loss: 1.8651, average\_loss: 1.8464  
 Epoch [33/50], Iter [215/418] Loss: 1.9489, average\_loss: 1.8457  
 Epoch [33/50], Iter [220/418] Loss: 1.7892, average\_loss: 1.8455  
 Epoch [33/50], Iter [225/418] Loss: 1.0078, average\_loss: 1.8370  
 Epoch [33/50], Iter [230/418] Loss: 2.2336, average\_loss: 1.8450  
 Epoch [33/50], Iter [235/418] Loss: 2.1652, average\_loss: 1.8440  
 Epoch [33/50], Iter [240/418] Loss: 1.8707, average\_loss: 1.8420  
 Epoch [33/50], Iter [245/418] Loss: 1.6767, average\_loss: 1.8465  
 Epoch [33/50], Iter [250/418] Loss: 2.8077, average\_loss: 1.8454  
 Epoch [33/50], Iter [255/418] Loss: 1.4632, average\_loss: 1.8398  
 Epoch [33/50], Iter [260/418] Loss: 1.8841, average\_loss: 1.8342  
 Epoch [33/50], Iter [265/418] Loss: 1.5602, average\_loss: 1.8416  
 Epoch [33/50], Iter [270/418] Loss: 1.6583, average\_loss: 1.8407  
 Epoch [33/50], Iter [275/418] Loss: 1.9819, average\_loss: 1.8387  
 Epoch [33/50], Iter [280/418] Loss: 1.9740, average\_loss: 1.8377  
 Epoch [33/50], Iter [285/418] Loss: 1.8198, average\_loss: 1.8391  
 Epoch [33/50], Iter [290/418] Loss: 2.5648, average\_loss: 1.8438  
 Epoch [33/50], Iter [295/418] Loss: 2.1437, average\_loss: 1.8514  
 Epoch [33/50], Iter [300/418] Loss: 1.6733, average\_loss: 1.8471  
 Epoch [33/50], Iter [305/418] Loss: 1.5636, average\_loss: 1.8459  
 Epoch [33/50], Iter [310/418] Loss: 2.0581, average\_loss: 1.8442  
 Epoch [33/50], Iter [315/418] Loss: 1.6662, average\_loss: 1.8422  
 Epoch [33/50], Iter [320/418] Loss: 1.4605, average\_loss: 1.8385  
 Epoch [33/50], Iter [325/418] Loss: 1.3668, average\_loss: 1.8395  
 Epoch [33/50], Iter [330/418] Loss: 1.7928, average\_loss: 1.8407  
 Epoch [33/50], Iter [335/418] Loss: 1.6407, average\_loss: 1.8394  
 Epoch [33/50], Iter [340/418] Loss: 1.0413, average\_loss: 1.8428  
 Epoch [33/50], Iter [345/418] Loss: 2.2812, average\_loss: 1.8413  
 Epoch [33/50], Iter [350/418] Loss: 1.9194, average\_loss: 1.8454  
 Epoch [33/50], Iter [355/418] Loss: 1.1511, average\_loss: 1.8451  
 Epoch [33/50], Iter [360/418] Loss: 2.8474, average\_loss: 1.8460  
 Epoch [33/50], Iter [365/418] Loss: 3.0983, average\_loss: 1.8522  
 Epoch [33/50], Iter [370/418] Loss: 1.8209, average\_loss: 1.8519  
 Epoch [33/50], Iter [375/418] Loss: 2.8575, average\_loss: 1.8505  
 Epoch [33/50], Iter [380/418] Loss: 1.7525, average\_loss: 1.8533

Epoch [33/50], Iter [385/418] Loss: 1.5424, average\_loss: 1.8523  
Epoch [33/50], Iter [390/418] Loss: 1.5190, average\_loss: 1.8538  
Epoch [33/50], Iter [395/418] Loss: 1.6914, average\_loss: 1.8545  
Epoch [33/50], Iter [400/418] Loss: 2.5153, average\_loss: 1.8571  
Epoch [33/50], Iter [405/418] Loss: 1.0380, average\_loss: 1.8560  
Epoch [33/50], Iter [410/418] Loss: 1.5072, average\_loss: 1.8572  
Epoch [33/50], Iter [415/418] Loss: 1.9555, average\_loss: 1.8586  
Updating best test loss: 2.58935

Starting epoch 34 / 50

Learning Rate for this epoch: 0.0001

Epoch [34/50], Iter [5/418] Loss: 1.8494, average\_loss: 1.8446  
Epoch [34/50], Iter [10/418] Loss: 1.6975, average\_loss: 1.9758  
Epoch [34/50], Iter [15/418] Loss: 2.2841, average\_loss: 1.9837  
Epoch [34/50], Iter [20/418] Loss: 1.4789, average\_loss: 1.8737  
Epoch [34/50], Iter [25/418] Loss: 1.1444, average\_loss: 1.8743  
Epoch [34/50], Iter [30/418] Loss: 2.4607, average\_loss: 2.0029  
Epoch [34/50], Iter [35/418] Loss: 1.9763, average\_loss: 1.9628  
Epoch [34/50], Iter [40/418] Loss: 1.8819, average\_loss: 1.9561  
Epoch [34/50], Iter [45/418] Loss: 1.5687, average\_loss: 1.9449  
Epoch [34/50], Iter [50/418] Loss: 1.6331, average\_loss: 1.9214  
Epoch [34/50], Iter [55/418] Loss: 2.1618, average\_loss: 1.9277  
Epoch [34/50], Iter [60/418] Loss: 1.4946, average\_loss: 1.9321  
Epoch [34/50], Iter [65/418] Loss: 1.6419, average\_loss: 1.9163  
Epoch [34/50], Iter [70/418] Loss: 1.5279, average\_loss: 1.8828  
Epoch [34/50], Iter [75/418] Loss: 1.1068, average\_loss: 1.8708  
Epoch [34/50], Iter [80/418] Loss: 2.3155, average\_loss: 1.8790  
Epoch [34/50], Iter [85/418] Loss: 1.6188, average\_loss: 1.8569  
Epoch [34/50], Iter [90/418] Loss: 1.9038, average\_loss: 1.8533  
Epoch [34/50], Iter [95/418] Loss: 0.9951, average\_loss: 1.8369  
Epoch [34/50], Iter [100/418] Loss: 1.6123, average\_loss: 1.8125  
Epoch [34/50], Iter [105/418] Loss: 1.9894, average\_loss: 1.8256  
Epoch [34/50], Iter [110/418] Loss: 1.9538, average\_loss: 1.8242  
Epoch [34/50], Iter [115/418] Loss: 1.6428, average\_loss: 1.8155  
Epoch [34/50], Iter [120/418] Loss: 2.1481, average\_loss: 1.8106  
Epoch [34/50], Iter [125/418] Loss: 1.0204, average\_loss: 1.7995  
Epoch [34/50], Iter [130/418] Loss: 2.0182, average\_loss: 1.8185  
Epoch [34/50], Iter [135/418] Loss: 1.7457, average\_loss: 1.8163  
Epoch [34/50], Iter [140/418] Loss: 2.1292, average\_loss: 1.8106  
Epoch [34/50], Iter [145/418] Loss: 1.2069, average\_loss: 1.7951  
Epoch [34/50], Iter [150/418] Loss: 1.0147, average\_loss: 1.7818  
Epoch [34/50], Iter [155/418] Loss: 1.5036, average\_loss: 1.7764  
Epoch [34/50], Iter [160/418] Loss: 2.3073, average\_loss: 1.7863  
Epoch [34/50], Iter [165/418] Loss: 1.8868, average\_loss: 1.7846  
Epoch [34/50], Iter [170/418] Loss: 1.0701, average\_loss: 1.7850  
Epoch [34/50], Iter [175/418] Loss: 1.5014, average\_loss: 1.7820  
Epoch [34/50], Iter [180/418] Loss: 2.6623, average\_loss: 1.7849

Epoch [34/50], Iter [185/418] Loss: 2.2213, average\_loss: 1.7834  
 Epoch [34/50], Iter [190/418] Loss: 1.1874, average\_loss: 1.7784  
 Epoch [34/50], Iter [195/418] Loss: 2.0474, average\_loss: 1.7801  
 Epoch [34/50], Iter [200/418] Loss: 2.5868, average\_loss: 1.7837  
 Epoch [34/50], Iter [205/418] Loss: 1.7908, average\_loss: 1.7803  
 Epoch [34/50], Iter [210/418] Loss: 1.7949, average\_loss: 1.7835  
 Epoch [34/50], Iter [215/418] Loss: 2.1714, average\_loss: 1.7806  
 Epoch [34/50], Iter [220/418] Loss: 2.6886, average\_loss: 1.7803  
 Epoch [34/50], Iter [225/418] Loss: 1.2790, average\_loss: 1.7765  
 Epoch [34/50], Iter [230/418] Loss: 2.1407, average\_loss: 1.7827  
 Epoch [34/50], Iter [235/418] Loss: 1.3943, average\_loss: 1.7791  
 Epoch [34/50], Iter [240/418] Loss: 2.5599, average\_loss: 1.7809  
 Epoch [34/50], Iter [245/418] Loss: 1.9893, average\_loss: 1.7816  
 Epoch [34/50], Iter [250/418] Loss: 1.5980, average\_loss: 1.7786  
 Epoch [34/50], Iter [255/418] Loss: 1.8881, average\_loss: 1.7832  
 Epoch [34/50], Iter [260/418] Loss: 1.3279, average\_loss: 1.7826  
 Epoch [34/50], Iter [265/418] Loss: 2.1122, average\_loss: 1.7938  
 Epoch [34/50], Iter [270/418] Loss: 1.2886, average\_loss: 1.7893  
 Epoch [34/50], Iter [275/418] Loss: 1.9465, average\_loss: 1.7899  
 Epoch [34/50], Iter [280/418] Loss: 1.4481, average\_loss: 1.7843  
 Epoch [34/50], Iter [285/418] Loss: 1.9218, average\_loss: 1.7839  
 Epoch [34/50], Iter [290/418] Loss: 1.2766, average\_loss: 1.7832  
 Epoch [34/50], Iter [295/418] Loss: 1.1360, average\_loss: 1.7840  
 Epoch [34/50], Iter [300/418] Loss: 1.4663, average\_loss: 1.7872  
 Epoch [34/50], Iter [305/418] Loss: 1.8940, average\_loss: 1.7904  
 Epoch [34/50], Iter [310/418] Loss: 2.0601, average\_loss: 1.7915  
 Epoch [34/50], Iter [315/418] Loss: 1.2785, average\_loss: 1.7846  
 Epoch [34/50], Iter [320/418] Loss: 1.5060, average\_loss: 1.7827  
 Epoch [34/50], Iter [325/418] Loss: 1.1852, average\_loss: 1.7853  
 Epoch [34/50], Iter [330/418] Loss: 3.1568, average\_loss: 1.7886  
 Epoch [34/50], Iter [335/418] Loss: 1.6888, average\_loss: 1.7904  
 Epoch [34/50], Iter [340/418] Loss: 1.5842, average\_loss: 1.7962  
 Epoch [34/50], Iter [345/418] Loss: 1.6133, average\_loss: 1.7948  
 Epoch [34/50], Iter [350/418] Loss: 2.5042, average\_loss: 1.7996  
 Epoch [34/50], Iter [355/418] Loss: 1.6787, average\_loss: 1.8008  
 Epoch [34/50], Iter [360/418] Loss: 1.6989, average\_loss: 1.8059  
 Epoch [34/50], Iter [365/418] Loss: 1.5188, average\_loss: 1.8020  
 Epoch [34/50], Iter [370/418] Loss: 2.3544, average\_loss: 1.8066  
 Epoch [34/50], Iter [375/418] Loss: 1.7326, average\_loss: 1.8068  
 Epoch [34/50], Iter [380/418] Loss: 2.1471, average\_loss: 1.8105  
 Epoch [34/50], Iter [385/418] Loss: 1.7915, average\_loss: 1.8069  
 Epoch [34/50], Iter [390/418] Loss: 1.4622, average\_loss: 1.8036  
 Epoch [34/50], Iter [395/418] Loss: 1.7155, average\_loss: 1.8006  
 Epoch [34/50], Iter [400/418] Loss: 1.5893, average\_loss: 1.7962  
 Epoch [34/50], Iter [405/418] Loss: 1.9948, average\_loss: 1.7973  
 Epoch [34/50], Iter [410/418] Loss: 2.2094, average\_loss: 1.7990  
 Epoch [34/50], Iter [415/418] Loss: 2.0345, average\_loss: 1.8049

Starting epoch 35 / 50

Learning Rate for this epoch: 0.0001

Epoch [35/50],	Iter [5/418]	Loss: 1.3695,	average_loss: 1.7315
Epoch [35/50],	Iter [10/418]	Loss: 2.1959,	average_loss: 1.9099
Epoch [35/50],	Iter [15/418]	Loss: 1.6290,	average_loss: 1.7695
Epoch [35/50],	Iter [20/418]	Loss: 3.2349,	average_loss: 1.8888
Epoch [35/50],	Iter [25/418]	Loss: 1.6592,	average_loss: 1.7883
Epoch [35/50],	Iter [30/418]	Loss: 2.3835,	average_loss: 1.8519
Epoch [35/50],	Iter [35/418]	Loss: 1.7214,	average_loss: 1.8451
Epoch [35/50],	Iter [40/418]	Loss: 1.5421,	average_loss: 1.8433
Epoch [35/50],	Iter [45/418]	Loss: 1.6340,	average_loss: 1.8392
Epoch [35/50],	Iter [50/418]	Loss: 1.0742,	average_loss: 1.8373
Epoch [35/50],	Iter [55/418]	Loss: 1.6269,	average_loss: 1.8456
Epoch [35/50],	Iter [60/418]	Loss: 2.2268,	average_loss: 1.8664
Epoch [35/50],	Iter [65/418]	Loss: 1.3411,	average_loss: 1.8567
Epoch [35/50],	Iter [70/418]	Loss: 1.9647,	average_loss: 1.8665
Epoch [35/50],	Iter [75/418]	Loss: 1.2985,	average_loss: 1.8642
Epoch [35/50],	Iter [80/418]	Loss: 1.9749,	average_loss: 1.8433
Epoch [35/50],	Iter [85/418]	Loss: 3.1990,	average_loss: 1.8595
Epoch [35/50],	Iter [90/418]	Loss: 1.3831,	average_loss: 1.8470
Epoch [35/50],	Iter [95/418]	Loss: 1.5040,	average_loss: 1.8606
Epoch [35/50],	Iter [100/418]	Loss: 1.2955,	average_loss: 1.8725
Epoch [35/50],	Iter [105/418]	Loss: 1.6586,	average_loss: 1.8546
Epoch [35/50],	Iter [110/418]	Loss: 2.6695,	average_loss: 1.8504
Epoch [35/50],	Iter [115/418]	Loss: 1.6767,	average_loss: 1.8512
Epoch [35/50],	Iter [120/418]	Loss: 2.5402,	average_loss: 1.8512
Epoch [35/50],	Iter [125/418]	Loss: 1.9843,	average_loss: 1.8471
Epoch [35/50],	Iter [130/418]	Loss: 1.9412,	average_loss: 1.8598
Epoch [35/50],	Iter [135/418]	Loss: 1.4033,	average_loss: 1.8499
Epoch [35/50],	Iter [140/418]	Loss: 1.3917,	average_loss: 1.8470
Epoch [35/50],	Iter [145/418]	Loss: 1.4533,	average_loss: 1.8439
Epoch [35/50],	Iter [150/418]	Loss: 1.8799,	average_loss: 1.8502
Epoch [35/50],	Iter [155/418]	Loss: 3.7192,	average_loss: 1.8712
Epoch [35/50],	Iter [160/418]	Loss: 1.0814,	average_loss: 1.8613
Epoch [35/50],	Iter [165/418]	Loss: 1.3609,	average_loss: 1.8572
Epoch [35/50],	Iter [170/418]	Loss: 1.6327,	average_loss: 1.8547
Epoch [35/50],	Iter [175/418]	Loss: 2.1008,	average_loss: 1.8516
Epoch [35/50],	Iter [180/418]	Loss: 1.2017,	average_loss: 1.8430
Epoch [35/50],	Iter [185/418]	Loss: 2.9132,	average_loss: 1.8420
Epoch [35/50],	Iter [190/418]	Loss: 1.8517,	average_loss: 1.8422
Epoch [35/50],	Iter [195/418]	Loss: 2.4718,	average_loss: 1.8389
Epoch [35/50],	Iter [200/418]	Loss: 1.4482,	average_loss: 1.8320
Epoch [35/50],	Iter [205/418]	Loss: 2.4438,	average_loss: 1.8383
Epoch [35/50],	Iter [210/418]	Loss: 1.7016,	average_loss: 1.8395
Epoch [35/50],	Iter [215/418]	Loss: 1.5360,	average_loss: 1.8392
Epoch [35/50],	Iter [220/418]	Loss: 1.2734,	average_loss: 1.8445
Epoch [35/50],	Iter [225/418]	Loss: 1.7324,	average_loss: 1.8417

Epoch [35/50], Iter [230/418] Loss: 2.2797, average\_loss: 1.8437  
 Epoch [35/50], Iter [235/418] Loss: 1.5649, average\_loss: 1.8405  
 Epoch [35/50], Iter [240/418] Loss: 1.6241, average\_loss: 1.8413  
 Epoch [35/50], Iter [245/418] Loss: 1.8339, average\_loss: 1.8457  
 Epoch [35/50], Iter [250/418] Loss: 1.5984, average\_loss: 1.8368  
 Epoch [35/50], Iter [255/418] Loss: 1.6636, average\_loss: 1.8363  
 Epoch [35/50], Iter [260/418] Loss: 1.6437, average\_loss: 1.8325  
 Epoch [35/50], Iter [265/418] Loss: 1.6463, average\_loss: 1.8187  
 Epoch [35/50], Iter [270/418] Loss: 1.4284, average\_loss: 1.8149  
 Epoch [35/50], Iter [275/418] Loss: 1.1025, average\_loss: 1.8137  
 Epoch [35/50], Iter [280/418] Loss: 1.9734, average\_loss: 1.8161  
 Epoch [35/50], Iter [285/418] Loss: 2.2116, average\_loss: 1.8238  
 Epoch [35/50], Iter [290/418] Loss: 1.4062, average\_loss: 1.8181  
 Epoch [35/50], Iter [295/418] Loss: 1.9990, average\_loss: 1.8171  
 Epoch [35/50], Iter [300/418] Loss: 2.1679, average\_loss: 1.8184  
 Epoch [35/50], Iter [305/418] Loss: 2.1973, average\_loss: 1.8164  
 Epoch [35/50], Iter [310/418] Loss: 1.4783, average\_loss: 1.8122  
 Epoch [35/50], Iter [315/418] Loss: 1.4421, average\_loss: 1.8066  
 Epoch [35/50], Iter [320/418] Loss: 1.3884, average\_loss: 1.8064  
 Epoch [35/50], Iter [325/418] Loss: 1.9118, average\_loss: 1.8055  
 Epoch [35/50], Iter [330/418] Loss: 1.1571, average\_loss: 1.7990  
 Epoch [35/50], Iter [335/418] Loss: 1.3553, average\_loss: 1.8030  
 Epoch [35/50], Iter [340/418] Loss: 1.9334, average\_loss: 1.8069  
 Epoch [35/50], Iter [345/418] Loss: 1.7732, average\_loss: 1.8059  
 Epoch [35/50], Iter [350/418] Loss: 1.3879, average\_loss: 1.8002  
 Epoch [35/50], Iter [355/418] Loss: 1.5972, average\_loss: 1.8012  
 Epoch [35/50], Iter [360/418] Loss: 1.9431, average\_loss: 1.8047  
 Epoch [35/50], Iter [365/418] Loss: 2.1387, average\_loss: 1.8042  
 Epoch [35/50], Iter [370/418] Loss: 1.8695, average\_loss: 1.8088  
 Epoch [35/50], Iter [375/418] Loss: 2.0609, average\_loss: 1.8109  
 Epoch [35/50], Iter [380/418] Loss: 1.5639, average\_loss: 1.8093  
 Epoch [35/50], Iter [385/418] Loss: 1.4586, average\_loss: 1.8151  
 Epoch [35/50], Iter [390/418] Loss: 1.6166, average\_loss: 1.8139  
 Epoch [35/50], Iter [395/418] Loss: 2.4809, average\_loss: 1.8150  
 Epoch [35/50], Iter [400/418] Loss: 0.9592, average\_loss: 1.8135  
 Epoch [35/50], Iter [405/418] Loss: 2.0424, average\_loss: 1.8137  
 Epoch [35/50], Iter [410/418] Loss: 1.7977, average\_loss: 1.8150  
 Epoch [35/50], Iter [415/418] Loss: 2.0837, average\_loss: 1.8179

Starting epoch 36 / 50

Learning Rate for this epoch: 0.0001

Epoch [36/50], Iter [5/418] Loss: 1.4369, average\_loss: 1.7314  
 Epoch [36/50], Iter [10/418] Loss: 1.6834, average\_loss: 1.6853  
 Epoch [36/50], Iter [15/418] Loss: 1.3600, average\_loss: 1.6113  
 Epoch [36/50], Iter [20/418] Loss: 1.4676, average\_loss: 1.7174  
 Epoch [36/50], Iter [25/418] Loss: 1.5345, average\_loss: 1.7417  
 Epoch [36/50], Iter [30/418] Loss: 1.1953, average\_loss: 1.7403



Epoch [36/50], Iter [35/418] Loss: 3.2637, average\_loss: 1.7684  
 Epoch [36/50], Iter [40/418] Loss: 1.6933, average\_loss: 1.7620  
 Epoch [36/50], Iter [45/418] Loss: 1.4838, average\_loss: 1.7940  
 Epoch [36/50], Iter [50/418] Loss: 1.8879, average\_loss: 1.7934  
 Epoch [36/50], Iter [55/418] Loss: 2.1242, average\_loss: 1.8029  
 Epoch [36/50], Iter [60/418] Loss: 2.6149, average\_loss: 1.7958  
 Epoch [36/50], Iter [65/418] Loss: 1.4202, average\_loss: 1.7700  
 Epoch [36/50], Iter [70/418] Loss: 1.7927, average\_loss: 1.7688  
 Epoch [36/50], Iter [75/418] Loss: 1.8945, average\_loss: 1.7457  
 Epoch [36/50], Iter [80/418] Loss: 1.9843, average\_loss: 1.7467  
 Epoch [36/50], Iter [85/418] Loss: 1.3294, average\_loss: 1.7368  
 Epoch [36/50], Iter [90/418] Loss: 1.8636, average\_loss: 1.7338  
 Epoch [36/50], Iter [95/418] Loss: 1.3103, average\_loss: 1.7395  
 Epoch [36/50], Iter [100/418] Loss: 1.0539, average\_loss: 1.7564  
 Epoch [36/50], Iter [105/418] Loss: 1.8098, average\_loss: 1.7722  
 Epoch [36/50], Iter [110/418] Loss: 2.2647, average\_loss: 1.7890  
 Epoch [36/50], Iter [115/418] Loss: 1.3940, average\_loss: 1.7860  
 Epoch [36/50], Iter [120/418] Loss: 3.2152, average\_loss: 1.7936  
 Epoch [36/50], Iter [125/418] Loss: 1.2696, average\_loss: 1.7906  
 Epoch [36/50], Iter [130/418] Loss: 1.6693, average\_loss: 1.8236  
 Epoch [36/50], Iter [135/418] Loss: 1.6100, average\_loss: 1.8120  
 Epoch [36/50], Iter [140/418] Loss: 2.5882, average\_loss: 1.8217  
 Epoch [36/50], Iter [145/418] Loss: 1.9613, average\_loss: 1.8220  
 Epoch [36/50], Iter [150/418] Loss: 1.7414, average\_loss: 1.8182  
 Epoch [36/50], Iter [155/418] Loss: 1.5609, average\_loss: 1.8234  
 Epoch [36/50], Iter [160/418] Loss: 1.9315, average\_loss: 1.8284  
 Epoch [36/50], Iter [165/418] Loss: 1.7043, average\_loss: 1.8198  
 Epoch [36/50], Iter [170/418] Loss: 1.3183, average\_loss: 1.8182  
 Epoch [36/50], Iter [175/418] Loss: 1.7580, average\_loss: 1.8230  
 Epoch [36/50], Iter [180/418] Loss: 2.8193, average\_loss: 1.8344  
 Epoch [36/50], Iter [185/418] Loss: 2.0447, average\_loss: 1.8465  
 Epoch [36/50], Iter [190/418] Loss: 3.2461, average\_loss: 1.8523  
 Epoch [36/50], Iter [195/418] Loss: 1.3025, average\_loss: 1.8420  
 Epoch [36/50], Iter [200/418] Loss: 1.8963, average\_loss: 1.8438  
 Epoch [36/50], Iter [205/418] Loss: 1.4238, average\_loss: 1.8477  
 Epoch [36/50], Iter [210/418] Loss: 1.3111, average\_loss: 1.8482  
 Epoch [36/50], Iter [215/418] Loss: 1.7831, average\_loss: 1.8457  
 Epoch [36/50], Iter [220/418] Loss: 1.8752, average\_loss: 1.8381  
 Epoch [36/50], Iter [225/418] Loss: 0.9231, average\_loss: 1.8353  
 Epoch [36/50], Iter [230/418] Loss: 1.9360, average\_loss: 1.8247  
 Epoch [36/50], Iter [235/418] Loss: 1.8543, average\_loss: 1.8277  
 Epoch [36/50], Iter [240/418] Loss: 1.4941, average\_loss: 1.8308  
 Epoch [36/50], Iter [245/418] Loss: 2.1077, average\_loss: 1.8305  
 Epoch [36/50], Iter [250/418] Loss: 2.5073, average\_loss: 1.8262  
 Epoch [36/50], Iter [255/418] Loss: 1.4149, average\_loss: 1.8204  
 Epoch [36/50], Iter [260/418] Loss: 2.0264, average\_loss: 1.8223  
 Epoch [36/50], Iter [265/418] Loss: 1.6702, average\_loss: 1.8243  
 Epoch [36/50], Iter [270/418] Loss: 2.8766, average\_loss: 1.8251

```

Epoch [36/50], Iter [275/418] Loss: 0.9101, average_loss: 1.8222
Epoch [36/50], Iter [280/418] Loss: 1.5076, average_loss: 1.8215
Epoch [36/50], Iter [285/418] Loss: 2.0610, average_loss: 1.8221
Epoch [36/50], Iter [290/418] Loss: 1.7682, average_loss: 1.8255
Epoch [36/50], Iter [295/418] Loss: 3.4777, average_loss: 1.8295
Epoch [36/50], Iter [300/418] Loss: 2.3744, average_loss: 1.8349
Epoch [36/50], Iter [305/418] Loss: 1.6314, average_loss: 1.8379
Epoch [36/50], Iter [310/418] Loss: 2.6921, average_loss: 1.8408
Epoch [36/50], Iter [315/418] Loss: 2.0750, average_loss: 1.8411
Epoch [36/50], Iter [320/418] Loss: 1.9505, average_loss: 1.8410
Epoch [36/50], Iter [325/418] Loss: 1.7697, average_loss: 1.8403
Epoch [36/50], Iter [330/418] Loss: 1.6871, average_loss: 1.8425
Epoch [36/50], Iter [335/418] Loss: 1.2483, average_loss: 1.8388
Epoch [36/50], Iter [340/418] Loss: 1.1946, average_loss: 1.8321
Epoch [36/50], Iter [345/418] Loss: 2.0102, average_loss: 1.8352
Epoch [36/50], Iter [350/418] Loss: 2.0235, average_loss: 1.8326
Epoch [36/50], Iter [355/418] Loss: 1.4973, average_loss: 1.8315
Epoch [36/50], Iter [360/418] Loss: 0.9827, average_loss: 1.8267
Epoch [36/50], Iter [365/418] Loss: 1.5895, average_loss: 1.8206
Epoch [36/50], Iter [370/418] Loss: 1.2383, average_loss: 1.8144
Epoch [36/50], Iter [375/418] Loss: 1.5354, average_loss: 1.8146
Epoch [36/50], Iter [380/418] Loss: 1.8381, average_loss: 1.8137
Epoch [36/50], Iter [385/418] Loss: 1.5793, average_loss: 1.8144
Epoch [36/50], Iter [390/418] Loss: 1.9963, average_loss: 1.8171
Epoch [36/50], Iter [395/418] Loss: 2.0164, average_loss: 1.8138
Epoch [36/50], Iter [400/418] Loss: 1.1490, average_loss: 1.8097
Epoch [36/50], Iter [405/418] Loss: 2.1038, average_loss: 1.8106
Epoch [36/50], Iter [410/418] Loss: 2.1503, average_loss: 1.8128
Epoch [36/50], Iter [415/418] Loss: 1.8474, average_loss: 1.8172

```

---Evaluate model on test samples---

100%|

| 4950/4950 [02:30<00:00, 32.98it/s]

```

---class aeroplane ap 0.5583491757125381---
---class bicycle ap 0.6277661078454237---
---class bird ap 0.5150708857769584---
---class boat ap 0.24345235047541103---
---class bottle ap 0.2273179996610508---
---class bus ap 0.6139253736538909---
---class car ap 0.6783177800514015---
---class cat ap 0.7100213384273657---
---class chair ap 0.2981406432748636---
---class cow ap 0.55388234508627---
---class diningtable ap 0.3626264528476675---
---class dog ap 0.6138911189890806---
---class horse ap 0.6735226957340688---
---class motorbike ap 0.5978788657593546---
---class person ap 0.5657900210983533---

```

---class pottedplant ap 0.22592417491578617---  
---class sheep ap 0.48383267144564657---  
---class sofa ap 0.4983407518046529---  
---class train ap 0.6495399977394608---  
---class tvmonitor ap 0.49620995286638614---  
---map 0.5096900351582816---

Starting epoch 37 / 50

Learning Rate for this epoch: 0.0001

Epoch [37/50], Iter [5/418] Loss: 2.3082, average\_loss: 1.9010  
Epoch [37/50], Iter [10/418] Loss: 1.6883, average\_loss: 1.7503  
Epoch [37/50], Iter [15/418] Loss: 1.3643, average\_loss: 1.6640  
Epoch [37/50], Iter [20/418] Loss: 2.1290, average\_loss: 1.6777  
Epoch [37/50], Iter [25/418] Loss: 1.6238, average\_loss: 1.7142  
Epoch [37/50], Iter [30/418] Loss: 2.1990, average\_loss: 1.7161  
Epoch [37/50], Iter [35/418] Loss: 2.5959, average\_loss: 1.8109  
Epoch [37/50], Iter [40/418] Loss: 1.7608, average\_loss: 1.8358  
Epoch [37/50], Iter [45/418] Loss: 1.6135, average\_loss: 1.8167  
Epoch [37/50], Iter [50/418] Loss: 2.3727, average\_loss: 1.8684  
Epoch [37/50], Iter [55/418] Loss: 1.5250, average\_loss: 1.8782  
Epoch [37/50], Iter [60/418] Loss: 1.3343, average\_loss: 1.8499  
Epoch [37/50], Iter [65/418] Loss: 1.8796, average\_loss: 1.8727  
Epoch [37/50], Iter [70/418] Loss: 2.8114, average\_loss: 1.8810  
Epoch [37/50], Iter [75/418] Loss: 0.8695, average\_loss: 1.8333  
Epoch [37/50], Iter [80/418] Loss: 1.6569, average\_loss: 1.8151  
Epoch [37/50], Iter [85/418] Loss: 1.0555, average\_loss: 1.7978  
Epoch [37/50], Iter [90/418] Loss: 1.3830, average\_loss: 1.8036  
Epoch [37/50], Iter [95/418] Loss: 3.2504, average\_loss: 1.7962  
Epoch [37/50], Iter [100/418] Loss: 1.5899, average\_loss: 1.7944  
Epoch [37/50], Iter [105/418] Loss: 1.8032, average\_loss: 1.7747  
Epoch [37/50], Iter [110/418] Loss: 1.3656, average\_loss: 1.7805  
Epoch [37/50], Iter [115/418] Loss: 2.1634, average\_loss: 1.8015  
Epoch [37/50], Iter [120/418] Loss: 1.3037, average\_loss: 1.7969  
Epoch [37/50], Iter [125/418] Loss: 1.0641, average\_loss: 1.7879  
Epoch [37/50], Iter [130/418] Loss: 3.0843, average\_loss: 1.8095  
Epoch [37/50], Iter [135/418] Loss: 2.5596, average\_loss: 1.8240  
Epoch [37/50], Iter [140/418] Loss: 1.3323, average\_loss: 1.8186  
Epoch [37/50], Iter [145/418] Loss: 1.5548, average\_loss: 1.8170  
Epoch [37/50], Iter [150/418] Loss: 1.8902, average\_loss: 1.8090  
Epoch [37/50], Iter [155/418] Loss: 1.6204, average\_loss: 1.8070  
Epoch [37/50], Iter [160/418] Loss: 2.4455, average\_loss: 1.8164  
Epoch [37/50], Iter [165/418] Loss: 1.6277, average\_loss: 1.8216  
Epoch [37/50], Iter [170/418] Loss: 1.4500, average\_loss: 1.8067  
Epoch [37/50], Iter [175/418] Loss: 2.2978, average\_loss: 1.8130  
Epoch [37/50], Iter [180/418] Loss: 1.3370, average\_loss: 1.8029  
Epoch [37/50], Iter [185/418] Loss: 1.2388, average\_loss: 1.8074  
Epoch [37/50], Iter [190/418] Loss: 1.9268, average\_loss: 1.8076

Epoch [37/50], Iter [195/418]	Loss: 1.1470, average_loss: 1.8169
Epoch [37/50], Iter [200/418]	Loss: 1.5711, average_loss: 1.8168
Epoch [37/50], Iter [205/418]	Loss: 1.8741, average_loss: 1.8102
Epoch [37/50], Iter [210/418]	Loss: 0.9916, average_loss: 1.8036
Epoch [37/50], Iter [215/418]	Loss: 2.7549, average_loss: 1.8044
Epoch [37/50], Iter [220/418]	Loss: 1.3120, average_loss: 1.7996
Epoch [37/50], Iter [225/418]	Loss: 1.3308, average_loss: 1.7971
Epoch [37/50], Iter [230/418]	Loss: 2.1012, average_loss: 1.8016
Epoch [37/50], Iter [235/418]	Loss: 1.7008, average_loss: 1.7985
Epoch [37/50], Iter [240/418]	Loss: 2.0059, average_loss: 1.7972
Epoch [37/50], Iter [245/418]	Loss: 1.6520, average_loss: 1.8040
Epoch [37/50], Iter [250/418]	Loss: 0.9868, average_loss: 1.7981
Epoch [37/50], Iter [255/418]	Loss: 1.4773, average_loss: 1.7979
Epoch [37/50], Iter [260/418]	Loss: 1.3378, average_loss: 1.7904
Epoch [37/50], Iter [265/418]	Loss: 1.8680, average_loss: 1.7900
Epoch [37/50], Iter [270/418]	Loss: 1.1726, average_loss: 1.7844
Epoch [37/50], Iter [275/418]	Loss: 1.6592, average_loss: 1.7869
Epoch [37/50], Iter [280/418]	Loss: 1.4864, average_loss: 1.7837
Epoch [37/50], Iter [285/418]	Loss: 1.1742, average_loss: 1.7881
Epoch [37/50], Iter [290/418]	Loss: 1.2037, average_loss: 1.7975
Epoch [37/50], Iter [295/418]	Loss: 1.5545, average_loss: 1.7920
Epoch [37/50], Iter [300/418]	Loss: 1.0195, average_loss: 1.7950
Epoch [37/50], Iter [305/418]	Loss: 1.9435, average_loss: 1.7950
Epoch [37/50], Iter [310/418]	Loss: 2.2664, average_loss: 1.7942
Epoch [37/50], Iter [315/418]	Loss: 0.9881, average_loss: 1.7905
Epoch [37/50], Iter [320/418]	Loss: 1.3361, average_loss: 1.7874
Epoch [37/50], Iter [325/418]	Loss: 1.9513, average_loss: 1.7851
Epoch [37/50], Iter [330/418]	Loss: 2.2880, average_loss: 1.7878
Epoch [37/50], Iter [335/418]	Loss: 1.2502, average_loss: 1.7893
Epoch [37/50], Iter [340/418]	Loss: 2.6287, average_loss: 1.7870
Epoch [37/50], Iter [345/418]	Loss: 2.0354, average_loss: 1.7847
Epoch [37/50], Iter [350/418]	Loss: 1.9715, average_loss: 1.7834
Epoch [37/50], Iter [355/418]	Loss: 1.9746, average_loss: 1.7816
Epoch [37/50], Iter [360/418]	Loss: 0.9264, average_loss: 1.7740
Epoch [37/50], Iter [365/418]	Loss: 1.7543, average_loss: 1.7768
Epoch [37/50], Iter [370/418]	Loss: 2.2864, average_loss: 1.7786
Epoch [37/50], Iter [375/418]	Loss: 2.7273, average_loss: 1.7884
Epoch [37/50], Iter [380/418]	Loss: 1.4780, average_loss: 1.7873
Epoch [37/50], Iter [385/418]	Loss: 2.0712, average_loss: 1.7881
Epoch [37/50], Iter [390/418]	Loss: 1.5615, average_loss: 1.7938
Epoch [37/50], Iter [395/418]	Loss: 1.2900, average_loss: 1.7932
Epoch [37/50], Iter [400/418]	Loss: 2.2473, average_loss: 1.7916
Epoch [37/50], Iter [405/418]	Loss: 1.2795, average_loss: 1.7930
Epoch [37/50], Iter [410/418]	Loss: 1.4810, average_loss: 1.7871
Epoch [37/50], Iter [415/418]	Loss: 3.0116, average_loss: 1.7887

Starting epoch 38 / 50

Learning Rate for this epoch: 0.0001

Epoch [38/50], Iter [5/418] Loss: 1.3904, average\_loss: 1.9244  
Epoch [38/50], Iter [10/418] Loss: 1.2726, average\_loss: 1.8251  
Epoch [38/50], Iter [15/418] Loss: 2.0120, average\_loss: 1.8081  
Epoch [38/50], Iter [20/418] Loss: 2.1645, average\_loss: 1.8176  
Epoch [38/50], Iter [25/418] Loss: 1.5986, average\_loss: 1.8044  
Epoch [38/50], Iter [30/418] Loss: 1.9681, average\_loss: 1.8232  
Epoch [38/50], Iter [35/418] Loss: 1.4613, average\_loss: 1.8774  
Epoch [38/50], Iter [40/418] Loss: 2.0935, average\_loss: 1.8281  
Epoch [38/50], Iter [45/418] Loss: 1.8991, average\_loss: 1.8017  
Epoch [38/50], Iter [50/418] Loss: 2.1226, average\_loss: 1.7666  
Epoch [38/50], Iter [55/418] Loss: 1.1885, average\_loss: 1.7621  
Epoch [38/50], Iter [60/418] Loss: 1.9739, average\_loss: 1.7768  
Epoch [38/50], Iter [65/418] Loss: 1.3453, average\_loss: 1.7700  
Epoch [38/50], Iter [70/418] Loss: 2.1336, average\_loss: 1.7568  
Epoch [38/50], Iter [75/418] Loss: 1.4994, average\_loss: 1.7420  
Epoch [38/50], Iter [80/418] Loss: 1.5098, average\_loss: 1.7267  
Epoch [38/50], Iter [85/418] Loss: 1.3345, average\_loss: 1.7241  
Epoch [38/50], Iter [90/418] Loss: 1.1893, average\_loss: 1.7092  
Epoch [38/50], Iter [95/418] Loss: 1.5833, average\_loss: 1.7253  
Epoch [38/50], Iter [100/418] Loss: 1.6557, average\_loss: 1.7181  
Epoch [38/50], Iter [105/418] Loss: 1.1949, average\_loss: 1.7319  
Epoch [38/50], Iter [110/418] Loss: 1.5992, average\_loss: 1.7393  
Epoch [38/50], Iter [115/418] Loss: 1.6521, average\_loss: 1.7448  
Epoch [38/50], Iter [120/418] Loss: 2.6859, average\_loss: 1.7482  
Epoch [38/50], Iter [125/418] Loss: 1.1469, average\_loss: 1.7374  
Epoch [38/50], Iter [130/418] Loss: 1.6617, average\_loss: 1.7529  
Epoch [38/50], Iter [135/418] Loss: 1.8706, average\_loss: 1.7579  
Epoch [38/50], Iter [140/418] Loss: 1.2012, average\_loss: 1.7556  
Epoch [38/50], Iter [145/418] Loss: 1.6334, average\_loss: 1.7646  
Epoch [38/50], Iter [150/418] Loss: 1.6156, average\_loss: 1.7516  
Epoch [38/50], Iter [155/418] Loss: 1.4696, average\_loss: 1.7547  
Epoch [38/50], Iter [160/418] Loss: 1.3885, average\_loss: 1.7457  
Epoch [38/50], Iter [165/418] Loss: 1.8096, average\_loss: 1.7469  
Epoch [38/50], Iter [170/418] Loss: 1.5426, average\_loss: 1.7436  
Epoch [38/50], Iter [175/418] Loss: 1.2926, average\_loss: 1.7395  
Epoch [38/50], Iter [180/418] Loss: 1.6793, average\_loss: 1.7324  
Epoch [38/50], Iter [185/418] Loss: 1.9386, average\_loss: 1.7313  
Epoch [38/50], Iter [190/418] Loss: 2.0204, average\_loss: 1.7290  
Epoch [38/50], Iter [195/418] Loss: 2.4166, average\_loss: 1.7329  
Epoch [38/50], Iter [200/418] Loss: 1.4274, average\_loss: 1.7280  
Epoch [38/50], Iter [205/418] Loss: 1.3843, average\_loss: 1.7372  
Epoch [38/50], Iter [210/418] Loss: 1.4051, average\_loss: 1.7459  
Epoch [38/50], Iter [215/418] Loss: 1.1738, average\_loss: 1.7563  
Epoch [38/50], Iter [220/418] Loss: 1.9622, average\_loss: 1.7605  
Epoch [38/50], Iter [225/418] Loss: 1.8034, average\_loss: 1.7592  
Epoch [38/50], Iter [230/418] Loss: 1.5070, average\_loss: 1.7530  
Epoch [38/50], Iter [235/418] Loss: 2.2046, average\_loss: 1.7566

Epoch [38/50], Iter [240/418] Loss: 1.5801, average\_loss: 1.7538  
 Epoch [38/50], Iter [245/418] Loss: 2.5070, average\_loss: 1.7504  
 Epoch [38/50], Iter [250/418] Loss: 1.3982, average\_loss: 1.7468  
 Epoch [38/50], Iter [255/418] Loss: 2.5053, average\_loss: 1.7550  
 Epoch [38/50], Iter [260/418] Loss: 2.0386, average\_loss: 1.7548  
 Epoch [38/50], Iter [265/418] Loss: 1.6013, average\_loss: 1.7522  
 Epoch [38/50], Iter [270/418] Loss: 1.8127, average\_loss: 1.7514  
 Epoch [38/50], Iter [275/418] Loss: 1.2930, average\_loss: 1.7486  
 Epoch [38/50], Iter [280/418] Loss: 1.3146, average\_loss: 1.7470  
 Epoch [38/50], Iter [285/418] Loss: 1.7454, average\_loss: 1.7473  
 Epoch [38/50], Iter [290/418] Loss: 2.1970, average\_loss: 1.7559  
 Epoch [38/50], Iter [295/418] Loss: 1.9697, average\_loss: 1.7572  
 Epoch [38/50], Iter [300/418] Loss: 2.1410, average\_loss: 1.7639  
 Epoch [38/50], Iter [305/418] Loss: 1.6950, average\_loss: 1.7666  
 Epoch [38/50], Iter [310/418] Loss: 3.4076, average\_loss: 1.7702  
 Epoch [38/50], Iter [315/418] Loss: 2.0994, average\_loss: 1.7689  
 Epoch [38/50], Iter [320/418] Loss: 1.3696, average\_loss: 1.7722  
 Epoch [38/50], Iter [325/418] Loss: 1.1677, average\_loss: 1.7729  
 Epoch [38/50], Iter [330/418] Loss: 2.3492, average\_loss: 1.7738  
 Epoch [38/50], Iter [335/418] Loss: 1.9043, average\_loss: 1.7721  
 Epoch [38/50], Iter [340/418] Loss: 1.3400, average\_loss: 1.7712  
 Epoch [38/50], Iter [345/418] Loss: 3.5256, average\_loss: 1.7741  
 Epoch [38/50], Iter [350/418] Loss: 2.0231, average\_loss: 1.7770  
 Epoch [38/50], Iter [355/418] Loss: 2.5936, average\_loss: 1.7811  
 Epoch [38/50], Iter [360/418] Loss: 1.0518, average\_loss: 1.7774  
 Epoch [38/50], Iter [365/418] Loss: 1.3231, average\_loss: 1.7721  
 Epoch [38/50], Iter [370/418] Loss: 1.3821, average\_loss: 1.7715  
 Epoch [38/50], Iter [375/418] Loss: 1.6696, average\_loss: 1.7695  
 Epoch [38/50], Iter [380/418] Loss: 2.2994, average\_loss: 1.7696  
 Epoch [38/50], Iter [385/418] Loss: 1.3872, average\_loss: 1.7647  
 Epoch [38/50], Iter [390/418] Loss: 1.1520, average\_loss: 1.7687  
 Epoch [38/50], Iter [395/418] Loss: 2.8865, average\_loss: 1.7681  
 Epoch [38/50], Iter [400/418] Loss: 2.0300, average\_loss: 1.7655  
 Epoch [38/50], Iter [405/418] Loss: 2.3953, average\_loss: 1.7648  
 Epoch [38/50], Iter [410/418] Loss: 1.5934, average\_loss: 1.7647  
 Epoch [38/50], Iter [415/418] Loss: 2.0817, average\_loss: 1.7666

Starting epoch 39 / 50

Learning Rate for this epoch: 0.0001

Epoch [39/50], Iter [5/418] Loss: 1.3994, average\_loss: 1.7242  
 Epoch [39/50], Iter [10/418] Loss: 1.3475, average\_loss: 1.8052  
 Epoch [39/50], Iter [15/418] Loss: 1.3246, average\_loss: 1.7665  
 Epoch [39/50], Iter [20/418] Loss: 1.2818, average\_loss: 1.7039  
 Epoch [39/50], Iter [25/418] Loss: 1.5162, average\_loss: 1.6789  
 Epoch [39/50], Iter [30/418] Loss: 2.1971, average\_loss: 1.6962  
 Epoch [39/50], Iter [35/418] Loss: 1.8946, average\_loss: 1.6929  
 Epoch [39/50], Iter [40/418] Loss: 1.5149, average\_loss: 1.6812

Epoch [39/50], Iter [45/418] Loss: 2.1273, average\_loss: 1.7039  
 Epoch [39/50], Iter [50/418] Loss: 1.7789, average\_loss: 1.7098  
 Epoch [39/50], Iter [55/418] Loss: 1.5748, average\_loss: 1.6946  
 Epoch [39/50], Iter [60/418] Loss: 1.0743, average\_loss: 1.6966  
 Epoch [39/50], Iter [65/418] Loss: 2.6934, average\_loss: 1.7028  
 Epoch [39/50], Iter [70/418] Loss: 2.1308, average\_loss: 1.6820  
 Epoch [39/50], Iter [75/418] Loss: 2.5680, average\_loss: 1.6912  
 Epoch [39/50], Iter [80/418] Loss: 1.3248, average\_loss: 1.6874  
 Epoch [39/50], Iter [85/418] Loss: 2.0682, average\_loss: 1.6964  
 Epoch [39/50], Iter [90/418] Loss: 1.2984, average\_loss: 1.6763  
 Epoch [39/50], Iter [95/418] Loss: 1.8117, average\_loss: 1.7105  
 Epoch [39/50], Iter [100/418] Loss: 2.1068, average\_loss: 1.7020  
 Epoch [39/50], Iter [105/418] Loss: 1.9997, average\_loss: 1.7035  
 Epoch [39/50], Iter [110/418] Loss: 2.1606, average\_loss: 1.7017  
 Epoch [39/50], Iter [115/418] Loss: 2.0116, average\_loss: 1.7034  
 Epoch [39/50], Iter [120/418] Loss: 2.1277, average\_loss: 1.6997  
 Epoch [39/50], Iter [125/418] Loss: 2.1007, average\_loss: 1.6923  
 Epoch [39/50], Iter [130/418] Loss: 2.1962, average\_loss: 1.6927  
 Epoch [39/50], Iter [135/418] Loss: 1.6721, average\_loss: 1.7015  
 Epoch [39/50], Iter [140/418] Loss: 1.9648, average\_loss: 1.6996  
 Epoch [39/50], Iter [145/418] Loss: 1.8943, average\_loss: 1.7123  
 Epoch [39/50], Iter [150/418] Loss: 2.0467, average\_loss: 1.7250  
 Epoch [39/50], Iter [155/418] Loss: 1.6768, average\_loss: 1.7281  
 Epoch [39/50], Iter [160/418] Loss: 1.9437, average\_loss: 1.7206  
 Epoch [39/50], Iter [165/418] Loss: 1.5511, average\_loss: 1.7276  
 Epoch [39/50], Iter [170/418] Loss: 1.5227, average\_loss: 1.7171  
 Epoch [39/50], Iter [175/418] Loss: 1.4673, average\_loss: 1.7078  
 Epoch [39/50], Iter [180/418] Loss: 1.7650, average\_loss: 1.7059  
 Epoch [39/50], Iter [185/418] Loss: 2.0574, average\_loss: 1.7089  
 Epoch [39/50], Iter [190/418] Loss: 1.4851, average\_loss: 1.7156  
 Epoch [39/50], Iter [195/418] Loss: 2.3479, average\_loss: 1.7300  
 Epoch [39/50], Iter [200/418] Loss: 3.2391, average\_loss: 1.7385  
 Epoch [39/50], Iter [205/418] Loss: 1.2722, average\_loss: 1.7297  
 Epoch [39/50], Iter [210/418] Loss: 2.0618, average\_loss: 1.7303  
 Epoch [39/50], Iter [215/418] Loss: 0.5938, average\_loss: 1.7262  
 Epoch [39/50], Iter [220/418] Loss: 1.3847, average\_loss: 1.7236  
 Epoch [39/50], Iter [225/418] Loss: 2.1148, average\_loss: 1.7278  
 Epoch [39/50], Iter [230/418] Loss: 1.5553, average\_loss: 1.7273  
 Epoch [39/50], Iter [235/418] Loss: 1.3732, average\_loss: 1.7375  
 Epoch [39/50], Iter [240/418] Loss: 2.1905, average\_loss: 1.7378  
 Epoch [39/50], Iter [245/418] Loss: 1.8387, average\_loss: 1.7393  
 Epoch [39/50], Iter [250/418] Loss: 1.0377, average\_loss: 1.7296  
 Epoch [39/50], Iter [255/418] Loss: 1.5265, average\_loss: 1.7258  
 Epoch [39/50], Iter [260/418] Loss: 0.9032, average\_loss: 1.7172  
 Epoch [39/50], Iter [265/418] Loss: 1.4659, average\_loss: 1.7251  
 Epoch [39/50], Iter [270/418] Loss: 2.0571, average\_loss: 1.7259  
 Epoch [39/50], Iter [275/418] Loss: 2.4083, average\_loss: 1.7273  
 Epoch [39/50], Iter [280/418] Loss: 1.4594, average\_loss: 1.7344

Epoch [39/50], Iter [285/418] Loss: 1.4475, average\_loss: 1.7365  
 Epoch [39/50], Iter [290/418] Loss: 1.6924, average\_loss: 1.7388  
 Epoch [39/50], Iter [295/418] Loss: 1.2949, average\_loss: 1.7405  
 Epoch [39/50], Iter [300/418] Loss: 1.3958, average\_loss: 1.7418  
 Epoch [39/50], Iter [305/418] Loss: 2.1885, average\_loss: 1.7496  
 Epoch [39/50], Iter [310/418] Loss: 1.7824, average\_loss: 1.7448  
 Epoch [39/50], Iter [315/418] Loss: 2.1840, average\_loss: 1.7464  
 Epoch [39/50], Iter [320/418] Loss: 2.6652, average\_loss: 1.7463  
 Epoch [39/50], Iter [325/418] Loss: 1.0084, average\_loss: 1.7420  
 Epoch [39/50], Iter [330/418] Loss: 2.5356, average\_loss: 1.7513  
 Epoch [39/50], Iter [335/418] Loss: 2.0134, average\_loss: 1.7492  
 Epoch [39/50], Iter [340/418] Loss: 1.5629, average\_loss: 1.7452  
 Epoch [39/50], Iter [345/418] Loss: 1.5104, average\_loss: 1.7461  
 Epoch [39/50], Iter [350/418] Loss: 1.7618, average\_loss: 1.7476  
 Epoch [39/50], Iter [355/418] Loss: 2.1696, average\_loss: 1.7524  
 Epoch [39/50], Iter [360/418] Loss: 2.3762, average\_loss: 1.7579  
 Epoch [39/50], Iter [365/418] Loss: 2.0709, average\_loss: 1.7600  
 Epoch [39/50], Iter [370/418] Loss: 1.8993, average\_loss: 1.7614  
 Epoch [39/50], Iter [375/418] Loss: 1.9138, average\_loss: 1.7623  
 Epoch [39/50], Iter [380/418] Loss: 1.7320, average\_loss: 1.7567  
 Epoch [39/50], Iter [385/418] Loss: 1.7862, average\_loss: 1.7570  
 Epoch [39/50], Iter [390/418] Loss: 1.8059, average\_loss: 1.7599  
 Epoch [39/50], Iter [395/418] Loss: 2.2369, average\_loss: 1.7576  
 Epoch [39/50], Iter [400/418] Loss: 1.8287, average\_loss: 1.7555  
 Epoch [39/50], Iter [405/418] Loss: 3.5678, average\_loss: 1.7584  
 Epoch [39/50], Iter [410/418] Loss: 1.3409, average\_loss: 1.7558  
 Epoch [39/50], Iter [415/418] Loss: 2.0712, average\_loss: 1.7572  
 Updating best test loss: 2.58471

Starting epoch 40 / 50

Learning Rate for this epoch: 0.0001

Epoch [40/50], Iter [5/418] Loss: 1.7644, average\_loss: 1.8483  
 Epoch [40/50], Iter [10/418] Loss: 1.2145, average\_loss: 1.6814  
 Epoch [40/50], Iter [15/418] Loss: 2.6206, average\_loss: 1.7390  
 Epoch [40/50], Iter [20/418] Loss: 1.0380, average\_loss: 1.7450  
 Epoch [40/50], Iter [25/418] Loss: 0.9227, average\_loss: 1.7756  
 Epoch [40/50], Iter [30/418] Loss: 2.1227, average\_loss: 1.9316  
 Epoch [40/50], Iter [35/418] Loss: 2.3639, average\_loss: 1.8927  
 Epoch [40/50], Iter [40/418] Loss: 2.1497, average\_loss: 1.8665  
 Epoch [40/50], Iter [45/418] Loss: 1.4718, average\_loss: 1.8593  
 Epoch [40/50], Iter [50/418] Loss: 2.6339, average\_loss: 1.8497  
 Epoch [40/50], Iter [55/418] Loss: 2.3497, average\_loss: 1.8598  
 Epoch [40/50], Iter [60/418] Loss: 1.3933, average\_loss: 1.8268  
 Epoch [40/50], Iter [65/418] Loss: 1.6094, average\_loss: 1.8038  
 Epoch [40/50], Iter [70/418] Loss: 2.1184, average\_loss: 1.7921  
 Epoch [40/50], Iter [75/418] Loss: 2.0998, average\_loss: 1.7955  
 Epoch [40/50], Iter [80/418] Loss: 1.4045, average\_loss: 1.7921



Epoch [40/50], Iter [85/418] Loss: 2.8781, average\_loss: 1.7960  
Epoch [40/50], Iter [90/418] Loss: 1.6823, average\_loss: 1.7813  
Epoch [40/50], Iter [95/418] Loss: 1.7022, average\_loss: 1.7885  
Epoch [40/50], Iter [100/418] Loss: 1.2786, average\_loss: 1.8005  
Epoch [40/50], Iter [105/418] Loss: 2.2089, average\_loss: 1.7929  
Epoch [40/50], Iter [110/418] Loss: 2.3145, average\_loss: 1.7949  
Epoch [40/50], Iter [115/418] Loss: 1.6413, average\_loss: 1.7838  
Epoch [40/50], Iter [120/418] Loss: 2.1749, average\_loss: 1.8005  
Epoch [40/50], Iter [125/418] Loss: 1.7096, average\_loss: 1.7982  
Epoch [40/50], Iter [130/418] Loss: 1.8299, average\_loss: 1.7910  
Epoch [40/50], Iter [135/418] Loss: 2.9952, average\_loss: 1.7961  
Epoch [40/50], Iter [140/418] Loss: 1.6077, average\_loss: 1.7808  
Epoch [40/50], Iter [145/418] Loss: 1.3751, average\_loss: 1.7687  
Epoch [40/50], Iter [150/418] Loss: 1.2722, average\_loss: 1.7621  
Epoch [40/50], Iter [155/418] Loss: 1.2074, average\_loss: 1.7640  
Epoch [40/50], Iter [160/418] Loss: 0.8564, average\_loss: 1.7548  
Epoch [40/50], Iter [165/418] Loss: 1.2043, average\_loss: 1.7630  
Epoch [40/50], Iter [170/418] Loss: 1.7383, average\_loss: 1.7553  
Epoch [40/50], Iter [175/418] Loss: 1.5354, average\_loss: 1.7565  
Epoch [40/50], Iter [180/418] Loss: 1.4171, average\_loss: 1.7530  
Epoch [40/50], Iter [185/418] Loss: 2.0259, average\_loss: 1.7460  
Epoch [40/50], Iter [190/418] Loss: 0.8342, average\_loss: 1.7422  
Epoch [40/50], Iter [195/418] Loss: 1.3480, average\_loss: 1.7404  
Epoch [40/50], Iter [200/418] Loss: 1.6028, average\_loss: 1.7359  
Epoch [40/50], Iter [205/418] Loss: 1.6679, average\_loss: 1.7359  
Epoch [40/50], Iter [210/418] Loss: 1.4545, average\_loss: 1.7287  
Epoch [40/50], Iter [215/418] Loss: 2.3790, average\_loss: 1.7269  
Epoch [40/50], Iter [220/418] Loss: 2.0035, average\_loss: 1.7267  
Epoch [40/50], Iter [225/418] Loss: 2.1317, average\_loss: 1.7287  
Epoch [40/50], Iter [230/418] Loss: 1.8062, average\_loss: 1.7264  
Epoch [40/50], Iter [235/418] Loss: 2.8044, average\_loss: 1.7327  
Epoch [40/50], Iter [240/418] Loss: 1.5000, average\_loss: 1.7368  
Epoch [40/50], Iter [245/418] Loss: 2.5043, average\_loss: 1.7396  
Epoch [40/50], Iter [250/418] Loss: 3.6679, average\_loss: 1.7461  
Epoch [40/50], Iter [255/418] Loss: 2.4447, average\_loss: 1.7478  
Epoch [40/50], Iter [260/418] Loss: 1.7740, average\_loss: 1.7497  
Epoch [40/50], Iter [265/418] Loss: 0.9796, average\_loss: 1.7465  
Epoch [40/50], Iter [270/418] Loss: 1.5989, average\_loss: 1.7463  
Epoch [40/50], Iter [275/418] Loss: 2.2466, average\_loss: 1.7476  
Epoch [40/50], Iter [280/418] Loss: 1.4205, average\_loss: 1.7436  
Epoch [40/50], Iter [285/418] Loss: 2.0483, average\_loss: 1.7410  
Epoch [40/50], Iter [290/418] Loss: 2.2342, average\_loss: 1.7431  
Epoch [40/50], Iter [295/418] Loss: 2.1553, average\_loss: 1.7443  
Epoch [40/50], Iter [300/418] Loss: 2.2527, average\_loss: 1.7461  
Epoch [40/50], Iter [305/418] Loss: 2.4422, average\_loss: 1.7478  
Epoch [40/50], Iter [310/418] Loss: 1.4904, average\_loss: 1.7489  
Epoch [40/50], Iter [315/418] Loss: 1.1839, average\_loss: 1.7438  
Epoch [40/50], Iter [320/418] Loss: 1.7947, average\_loss: 1.7471

Epoch [40/50], Iter [325/418] Loss: 1.3973, average\_loss: 1.7461  
 Epoch [40/50], Iter [330/418] Loss: 2.6379, average\_loss: 1.7534  
 Epoch [40/50], Iter [335/418] Loss: 1.4172, average\_loss: 1.7590  
 Epoch [40/50], Iter [340/418] Loss: 1.0289, average\_loss: 1.7535  
 Epoch [40/50], Iter [345/418] Loss: 2.5713, average\_loss: 1.7511  
 Epoch [40/50], Iter [350/418] Loss: 2.0033, average\_loss: 1.7491  
 Epoch [40/50], Iter [355/418] Loss: 1.4486, average\_loss: 1.7488  
 Epoch [40/50], Iter [360/418] Loss: 1.5669, average\_loss: 1.7459  
 Epoch [40/50], Iter [365/418] Loss: 2.3125, average\_loss: 1.7471  
 Epoch [40/50], Iter [370/418] Loss: 1.7210, average\_loss: 1.7491  
 Epoch [40/50], Iter [375/418] Loss: 1.8851, average\_loss: 1.7487  
 Epoch [40/50], Iter [380/418] Loss: 1.7677, average\_loss: 1.7455  
 Epoch [40/50], Iter [385/418] Loss: 1.3437, average\_loss: 1.7463  
 Epoch [40/50], Iter [390/418] Loss: 1.4338, average\_loss: 1.7451  
 Epoch [40/50], Iter [395/418] Loss: 1.3613, average\_loss: 1.7441  
 Epoch [40/50], Iter [400/418] Loss: 1.6418, average\_loss: 1.7443  
 Epoch [40/50], Iter [405/418] Loss: 1.5918, average\_loss: 1.7426  
 Epoch [40/50], Iter [410/418] Loss: 1.6483, average\_loss: 1.7389  
 Epoch [40/50], Iter [415/418] Loss: 0.5881, average\_loss: 1.7355  
 Updating best test loss: 2.57834

Starting epoch 41 / 50

Learning Rate for this epoch: 1e-05

Epoch [41/50], Iter [5/418] Loss: 1.5591, average\_loss: 1.6411  
 Epoch [41/50], Iter [10/418] Loss: 2.0127, average\_loss: 1.9480  
 Epoch [41/50], Iter [15/418] Loss: 1.8534, average\_loss: 1.9796  
 Epoch [41/50], Iter [20/418] Loss: 1.4625, average\_loss: 1.9458  
 Epoch [41/50], Iter [25/418] Loss: 2.4177, average\_loss: 1.9297  
 Epoch [41/50], Iter [30/418] Loss: 1.5174, average\_loss: 1.9068  
 Epoch [41/50], Iter [35/418] Loss: 1.8471, average\_loss: 1.8805  
 Epoch [41/50], Iter [40/418] Loss: 1.8178, average\_loss: 1.8883  
 Epoch [41/50], Iter [45/418] Loss: 1.5193, average\_loss: 1.8318  
 Epoch [41/50], Iter [50/418] Loss: 1.1487, average\_loss: 1.7955  
 Epoch [41/50], Iter [55/418] Loss: 2.0772, average\_loss: 1.7766  
 Epoch [41/50], Iter [60/418] Loss: 1.5262, average\_loss: 1.7572  
 Epoch [41/50], Iter [65/418] Loss: 1.1214, average\_loss: 1.7729  
 Epoch [41/50], Iter [70/418] Loss: 2.1863, average\_loss: 1.7751  
 Epoch [41/50], Iter [75/418] Loss: 2.0223, average\_loss: 1.7520  
 Epoch [41/50], Iter [80/418] Loss: 2.1981, average\_loss: 1.7582  
 Epoch [41/50], Iter [85/418] Loss: 2.1585, average\_loss: 1.7723  
 Epoch [41/50], Iter [90/418] Loss: 1.6322, average\_loss: 1.7831  
 Epoch [41/50], Iter [95/418] Loss: 1.3815, average\_loss: 1.7718  
 Epoch [41/50], Iter [100/418] Loss: 1.1192, average\_loss: 1.7619  
 Epoch [41/50], Iter [105/418] Loss: 1.5551, average\_loss: 1.7483  
 Epoch [41/50], Iter [110/418] Loss: 1.3819, average\_loss: 1.7613  
 Epoch [41/50], Iter [115/418] Loss: 1.7452, average\_loss: 1.7559  
 Epoch [41/50], Iter [120/418] Loss: 2.2268, average\_loss: 1.7579

Epoch [41/50], Iter [125/418] Loss: 2.4855, average\_loss: 1.7605  
 Epoch [41/50], Iter [130/418] Loss: 1.4250, average\_loss: 1.7560  
 Epoch [41/50], Iter [135/418] Loss: 1.7472, average\_loss: 1.7704  
 Epoch [41/50], Iter [140/418] Loss: 1.3363, average\_loss: 1.7733  
 Epoch [41/50], Iter [145/418] Loss: 0.8464, average\_loss: 1.7666  
 Epoch [41/50], Iter [150/418] Loss: 2.5780, average\_loss: 1.7768  
 Epoch [41/50], Iter [155/418] Loss: 1.2122, average\_loss: 1.7786  
 Epoch [41/50], Iter [160/418] Loss: 0.9874, average\_loss: 1.7692  
 Epoch [41/50], Iter [165/418] Loss: 1.8183, average\_loss: 1.7683  
 Epoch [41/50], Iter [170/418] Loss: 1.4346, average\_loss: 1.7638  
 Epoch [41/50], Iter [175/418] Loss: 2.3293, average\_loss: 1.7627  
 Epoch [41/50], Iter [180/418] Loss: 2.1090, average\_loss: 1.7658  
 Epoch [41/50], Iter [185/418] Loss: 2.3513, average\_loss: 1.7569  
 Epoch [41/50], Iter [190/418] Loss: 2.0065, average\_loss: 1.7572  
 Epoch [41/50], Iter [195/418] Loss: 1.6891, average\_loss: 1.7579  
 Epoch [41/50], Iter [200/418] Loss: 2.0939, average\_loss: 1.7527  
 Epoch [41/50], Iter [205/418] Loss: 1.3534, average\_loss: 1.7527  
 Epoch [41/50], Iter [210/418] Loss: 1.4433, average\_loss: 1.7643  
 Epoch [41/50], Iter [215/418] Loss: 2.5142, average\_loss: 1.7695  
 Epoch [41/50], Iter [220/418] Loss: 0.9621, average\_loss: 1.7634  
 Epoch [41/50], Iter [225/418] Loss: 1.6337, average\_loss: 1.7611  
 Epoch [41/50], Iter [230/418] Loss: 1.9870, average\_loss: 1.7591  
 Epoch [41/50], Iter [235/418] Loss: 1.1421, average\_loss: 1.7639  
 Epoch [41/50], Iter [240/418] Loss: 1.1386, average\_loss: 1.7657  
 Epoch [41/50], Iter [245/418] Loss: 1.5959, average\_loss: 1.7687  
 Epoch [41/50], Iter [250/418] Loss: 1.7843, average\_loss: 1.7744  
 Epoch [41/50], Iter [255/418] Loss: 2.4399, average\_loss: 1.7783  
 Epoch [41/50], Iter [260/418] Loss: 1.9057, average\_loss: 1.7762  
 Epoch [41/50], Iter [265/418] Loss: 1.5264, average\_loss: 1.7792  
 Epoch [41/50], Iter [270/418] Loss: 1.4687, average\_loss: 1.7772  
 Epoch [41/50], Iter [275/418] Loss: 0.9070, average\_loss: 1.7724  
 Epoch [41/50], Iter [280/418] Loss: 1.6011, average\_loss: 1.7736  
 Epoch [41/50], Iter [285/418] Loss: 1.2660, average\_loss: 1.7682  
 Epoch [41/50], Iter [290/418] Loss: 1.7330, average\_loss: 1.7676  
 Epoch [41/50], Iter [295/418] Loss: 2.4007, average\_loss: 1.7660  
 Epoch [41/50], Iter [300/418] Loss: 1.3874, average\_loss: 1.7585  
 Epoch [41/50], Iter [305/418] Loss: 1.2664, average\_loss: 1.7554  
 Epoch [41/50], Iter [310/418] Loss: 1.6262, average\_loss: 1.7657  
 Epoch [41/50], Iter [315/418] Loss: 1.7446, average\_loss: 1.7644  
 Epoch [41/50], Iter [320/418] Loss: 1.2565, average\_loss: 1.7675  
 Epoch [41/50], Iter [325/418] Loss: 1.7328, average\_loss: 1.7615  
 Epoch [41/50], Iter [330/418] Loss: 1.5946, average\_loss: 1.7623  
 Epoch [41/50], Iter [335/418] Loss: 1.3911, average\_loss: 1.7638  
 Epoch [41/50], Iter [340/418] Loss: 1.0131, average\_loss: 1.7587  
 Epoch [41/50], Iter [345/418] Loss: 1.8045, average\_loss: 1.7546  
 Epoch [41/50], Iter [350/418] Loss: 2.7190, average\_loss: 1.7580  
 Epoch [41/50], Iter [355/418] Loss: 1.3468, average\_loss: 1.7547  
 Epoch [41/50], Iter [360/418] Loss: 1.6098, average\_loss: 1.7506

```
Epoch [41/50], Iter [365/418] Loss: 1.7463, average_loss: 1.7520
Epoch [41/50], Iter [370/418] Loss: 1.1687, average_loss: 1.7502
Epoch [41/50], Iter [375/418] Loss: 1.9364, average_loss: 1.7465
Epoch [41/50], Iter [380/418] Loss: 1.0263, average_loss: 1.7430
Epoch [41/50], Iter [385/418] Loss: 2.3222, average_loss: 1.7428
Epoch [41/50], Iter [390/418] Loss: 1.5708, average_loss: 1.7441
Epoch [41/50], Iter [395/418] Loss: 1.2442, average_loss: 1.7429
Epoch [41/50], Iter [400/418] Loss: 1.3230, average_loss: 1.7394
Epoch [41/50], Iter [405/418] Loss: 1.5291, average_loss: 1.7420
Epoch [41/50], Iter [410/418] Loss: 1.6011, average_loss: 1.7394
Epoch [41/50], Iter [415/418] Loss: 1.8792, average_loss: 1.7414
```

---Evaluate model on test samples---

100%|

| 4950/4950 [02:30<00:00, 32.85it/s]

```
---class aeroplane ap 0.536389135579781---
---class bicycle ap 0.6248697152315321---
---class bird ap 0.5079506238285536---
---class boat ap 0.2774800877171798---
---class bottle ap 0.22034201228829858---
---class bus ap 0.6426463426933336---
---class car ap 0.687179565949683---
---class cat ap 0.7087691329817105---
---class chair ap 0.32420474759705514---
---class cow ap 0.5269105533445959---
---class diningtable ap 0.3705897388196978---
---class dog ap 0.6485816114041354---
---class horse ap 0.6483822500330603---
---class motorbike ap 0.6029979060755816---
---class person ap 0.557182107216482---
---class pottedplant ap 0.21329528583889404---
---class sheep ap 0.4617207706454124---
---class sofa ap 0.5058889057803728---
---class train ap 0.6546567343565473---
---class tvmonitor ap 0.49930296999882695---
---map 0.5109670098690365---
```

Starting epoch 42 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [42/50], Iter [5/418] Loss: 1.6087, average_loss: 1.8272
Epoch [42/50], Iter [10/418] Loss: 1.7073, average_loss: 1.7961
Epoch [42/50], Iter [15/418] Loss: 2.2018, average_loss: 1.8287
Epoch [42/50], Iter [20/418] Loss: 1.3091, average_loss: 1.7922
Epoch [42/50], Iter [25/418] Loss: 0.9463, average_loss: 1.7309
Epoch [42/50], Iter [30/418] Loss: 1.3460, average_loss: 1.6890
Epoch [42/50], Iter [35/418] Loss: 2.6583, average_loss: 1.7337
Epoch [42/50], Iter [40/418] Loss: 1.3427, average_loss: 1.7106
```

Epoch [42/50], Iter [45/418] Loss: 2.8645, average\_loss: 1.7605  
 Epoch [42/50], Iter [50/418] Loss: 1.9986, average\_loss: 1.7608  
 Epoch [42/50], Iter [55/418] Loss: 1.9001, average\_loss: 1.7393  
 Epoch [42/50], Iter [60/418] Loss: 1.4375, average\_loss: 1.7337  
 Epoch [42/50], Iter [65/418] Loss: 1.9444, average\_loss: 1.7064  
 Epoch [42/50], Iter [70/418] Loss: 1.8606, average\_loss: 1.7135  
 Epoch [42/50], Iter [75/418] Loss: 1.4714, average\_loss: 1.7052  
 Epoch [42/50], Iter [80/418] Loss: 1.6291, average\_loss: 1.7104  
 Epoch [42/50], Iter [85/418] Loss: 3.2918, average\_loss: 1.7176  
 Epoch [42/50], Iter [90/418] Loss: 1.5119, average\_loss: 1.7128  
 Epoch [42/50], Iter [95/418] Loss: 2.7775, average\_loss: 1.7330  
 Epoch [42/50], Iter [100/418] Loss: 1.7100, average\_loss: 1.7364  
 Epoch [42/50], Iter [105/418] Loss: 1.8340, average\_loss: 1.7484  
 Epoch [42/50], Iter [110/418] Loss: 1.4165, average\_loss: 1.7355  
 Epoch [42/50], Iter [115/418] Loss: 2.3409, average\_loss: 1.7441  
 Epoch [42/50], Iter [120/418] Loss: 1.6606, average\_loss: 1.7342  
 Epoch [42/50], Iter [125/418] Loss: 1.7471, average\_loss: 1.7403  
 Epoch [42/50], Iter [130/418] Loss: 1.4285, average\_loss: 1.7463  
 Epoch [42/50], Iter [135/418] Loss: 1.5768, average\_loss: 1.7489  
 Epoch [42/50], Iter [140/418] Loss: 2.4570, average\_loss: 1.7524  
 Epoch [42/50], Iter [145/418] Loss: 1.6393, average\_loss: 1.7590  
 Epoch [42/50], Iter [150/418] Loss: 1.7906, average\_loss: 1.7590  
 Epoch [42/50], Iter [155/418] Loss: 2.0038, average\_loss: 1.7586  
 Epoch [42/50], Iter [160/418] Loss: 1.5218, average\_loss: 1.7556  
 Epoch [42/50], Iter [165/418] Loss: 1.0147, average\_loss: 1.7505  
 Epoch [42/50], Iter [170/418] Loss: 1.5974, average\_loss: 1.7582  
 Epoch [42/50], Iter [175/418] Loss: 1.7302, average\_loss: 1.7526  
 Epoch [42/50], Iter [180/418] Loss: 2.2199, average\_loss: 1.7518  
 Epoch [42/50], Iter [185/418] Loss: 1.4712, average\_loss: 1.7506  
 Epoch [42/50], Iter [190/418] Loss: 0.9582, average\_loss: 1.7444  
 Epoch [42/50], Iter [195/418] Loss: 1.6457, average\_loss: 1.7444  
 Epoch [42/50], Iter [200/418] Loss: 1.6357, average\_loss: 1.7400  
 Epoch [42/50], Iter [205/418] Loss: 1.3989, average\_loss: 1.7375  
 Epoch [42/50], Iter [210/418] Loss: 2.1723, average\_loss: 1.7370  
 Epoch [42/50], Iter [215/418] Loss: 1.4035, average\_loss: 1.7402  
 Epoch [42/50], Iter [220/418] Loss: 1.2029, average\_loss: 1.7443  
 Epoch [42/50], Iter [225/418] Loss: 2.0195, average\_loss: 1.7484  
 Epoch [42/50], Iter [230/418] Loss: 1.4311, average\_loss: 1.7442  
 Epoch [42/50], Iter [235/418] Loss: 1.8622, average\_loss: 1.7422  
 Epoch [42/50], Iter [240/418] Loss: 1.3815, average\_loss: 1.7423  
 Epoch [42/50], Iter [245/418] Loss: 2.1313, average\_loss: 1.7406  
 Epoch [42/50], Iter [250/418] Loss: 1.9809, average\_loss: 1.7422  
 Epoch [42/50], Iter [255/418] Loss: 1.5114, average\_loss: 1.7418  
 Epoch [42/50], Iter [260/418] Loss: 1.6665, average\_loss: 1.7418  
 Epoch [42/50], Iter [265/418] Loss: 2.5160, average\_loss: 1.7398  
 Epoch [42/50], Iter [270/418] Loss: 1.5711, average\_loss: 1.7318  
 Epoch [42/50], Iter [275/418] Loss: 2.4913, average\_loss: 1.7294  
 Epoch [42/50], Iter [280/418] Loss: 1.3405, average\_loss: 1.7287

Epoch [42/50], Iter [285/418] Loss: 2.6147, average\_loss: 1.7303  
 Epoch [42/50], Iter [290/418] Loss: 1.9247, average\_loss: 1.7330  
 Epoch [42/50], Iter [295/418] Loss: 2.0551, average\_loss: 1.7377  
 Epoch [42/50], Iter [300/418] Loss: 1.5618, average\_loss: 1.7396  
 Epoch [42/50], Iter [305/418] Loss: 2.4570, average\_loss: 1.7401  
 Epoch [42/50], Iter [310/418] Loss: 1.9877, average\_loss: 1.7402  
 Epoch [42/50], Iter [315/418] Loss: 1.5710, average\_loss: 1.7455  
 Epoch [42/50], Iter [320/418] Loss: 1.3015, average\_loss: 1.7462  
 Epoch [42/50], Iter [325/418] Loss: 1.8850, average\_loss: 1.7430  
 Epoch [42/50], Iter [330/418] Loss: 1.4102, average\_loss: 1.7383  
 Epoch [42/50], Iter [335/418] Loss: 1.3240, average\_loss: 1.7388  
 Epoch [42/50], Iter [340/418] Loss: 1.1087, average\_loss: 1.7396  
 Epoch [42/50], Iter [345/418] Loss: 1.5175, average\_loss: 1.7397  
 Epoch [42/50], Iter [350/418] Loss: 1.6730, average\_loss: 1.7378  
 Epoch [42/50], Iter [355/418] Loss: 0.8426, average\_loss: 1.7345  
 Epoch [42/50], Iter [360/418] Loss: 1.4258, average\_loss: 1.7313  
 Epoch [42/50], Iter [365/418] Loss: 1.5108, average\_loss: 1.7328  
 Epoch [42/50], Iter [370/418] Loss: 1.8955, average\_loss: 1.7335  
 Epoch [42/50], Iter [375/418] Loss: 1.6020, average\_loss: 1.7340  
 Epoch [42/50], Iter [380/418] Loss: 1.1573, average\_loss: 1.7264  
 Epoch [42/50], Iter [385/418] Loss: 1.7257, average\_loss: 1.7317  
 Epoch [42/50], Iter [390/418] Loss: 1.4213, average\_loss: 1.7319  
 Epoch [42/50], Iter [395/418] Loss: 2.6530, average\_loss: 1.7320  
 Epoch [42/50], Iter [400/418] Loss: 1.6651, average\_loss: 1.7333  
 Epoch [42/50], Iter [405/418] Loss: 1.7203, average\_loss: 1.7365  
 Epoch [42/50], Iter [410/418] Loss: 1.6939, average\_loss: 1.7366  
 Epoch [42/50], Iter [415/418] Loss: 1.1486, average\_loss: 1.7349

Starting epoch 43 / 50

Learning Rate for this epoch: 1e-05

Epoch [43/50], Iter [5/418] Loss: 0.9727, average\_loss: 1.5225  
 Epoch [43/50], Iter [10/418] Loss: 2.3778, average\_loss: 1.5802  
 Epoch [43/50], Iter [15/418] Loss: 1.3697, average\_loss: 1.6758  
 Epoch [43/50], Iter [20/418] Loss: 1.9109, average\_loss: 1.6050  
 Epoch [43/50], Iter [25/418] Loss: 2.4033, average\_loss: 1.6306  
 Epoch [43/50], Iter [30/418] Loss: 2.5493, average\_loss: 1.7623  
 Epoch [43/50], Iter [35/418] Loss: 2.4539, average\_loss: 1.7570  
 Epoch [43/50], Iter [40/418] Loss: 2.3458, average\_loss: 1.7496  
 Epoch [43/50], Iter [45/418] Loss: 1.8836, average\_loss: 1.7616  
 Epoch [43/50], Iter [50/418] Loss: 2.5450, average\_loss: 1.7528  
 Epoch [43/50], Iter [55/418] Loss: 2.0106, average\_loss: 1.7529  
 Epoch [43/50], Iter [60/418] Loss: 2.0523, average\_loss: 1.7708  
 Epoch [43/50], Iter [65/418] Loss: 1.8578, average\_loss: 1.7639  
 Epoch [43/50], Iter [70/418] Loss: 1.8563, average\_loss: 1.7659  
 Epoch [43/50], Iter [75/418] Loss: 2.3791, average\_loss: 1.7654  
 Epoch [43/50], Iter [80/418] Loss: 1.8893, average\_loss: 1.7558  
 Epoch [43/50], Iter [85/418] Loss: 1.9739, average\_loss: 1.7680

Epoch [43/50], Iter [90/418] Loss: 1.2041, average\_loss: 1.7629  
 Epoch [43/50], Iter [95/418] Loss: 2.4733, average\_loss: 1.7707  
 Epoch [43/50], Iter [100/418] Loss: 1.1923, average\_loss: 1.7636  
 Epoch [43/50], Iter [105/418] Loss: 1.3966, average\_loss: 1.7655  
 Epoch [43/50], Iter [110/418] Loss: 2.0413, average\_loss: 1.7658  
 Epoch [43/50], Iter [115/418] Loss: 2.2814, average\_loss: 1.7574  
 Epoch [43/50], Iter [120/418] Loss: 1.6141, average\_loss: 1.7505  
 Epoch [43/50], Iter [125/418] Loss: 1.7899, average\_loss: 1.7501  
 Epoch [43/50], Iter [130/418] Loss: 1.9576, average\_loss: 1.7468  
 Epoch [43/50], Iter [135/418] Loss: 1.8404, average\_loss: 1.7442  
 Epoch [43/50], Iter [140/418] Loss: 1.4316, average\_loss: 1.7404  
 Epoch [43/50], Iter [145/418] Loss: 1.6414, average\_loss: 1.7452  
 Epoch [43/50], Iter [150/418] Loss: 3.5845, average\_loss: 1.7607  
 Epoch [43/50], Iter [155/418] Loss: 1.8317, average\_loss: 1.7560  
 Epoch [43/50], Iter [160/418] Loss: 3.1209, average\_loss: 1.7613  
 Epoch [43/50], Iter [165/418] Loss: 1.6962, average\_loss: 1.7583  
 Epoch [43/50], Iter [170/418] Loss: 1.8155, average\_loss: 1.7561  
 Epoch [43/50], Iter [175/418] Loss: 2.1987, average\_loss: 1.7600  
 Epoch [43/50], Iter [180/418] Loss: 1.7365, average\_loss: 1.7616  
 Epoch [43/50], Iter [185/418] Loss: 1.1768, average\_loss: 1.7573  
 Epoch [43/50], Iter [190/418] Loss: 0.8879, average\_loss: 1.7562  
 Epoch [43/50], Iter [195/418] Loss: 0.6572, average\_loss: 1.7411  
 Epoch [43/50], Iter [200/418] Loss: 2.2100, average\_loss: 1.7395  
 Epoch [43/50], Iter [205/418] Loss: 1.7456, average\_loss: 1.7322  
 Epoch [43/50], Iter [210/418] Loss: 0.8033, average\_loss: 1.7285  
 Epoch [43/50], Iter [215/418] Loss: 1.3586, average\_loss: 1.7254  
 Epoch [43/50], Iter [220/418] Loss: 2.2210, average\_loss: 1.7297  
 Epoch [43/50], Iter [225/418] Loss: 1.5703, average\_loss: 1.7285  
 Epoch [43/50], Iter [230/418] Loss: 1.9505, average\_loss: 1.7250  
 Epoch [43/50], Iter [235/418] Loss: 1.9074, average\_loss: 1.7222  
 Epoch [43/50], Iter [240/418] Loss: 1.4513, average\_loss: 1.7175  
 Epoch [43/50], Iter [245/418] Loss: 1.8101, average\_loss: 1.7139  
 Epoch [43/50], Iter [250/418] Loss: 1.3638, average\_loss: 1.7111  
 Epoch [43/50], Iter [255/418] Loss: 1.7216, average\_loss: 1.7102  
 Epoch [43/50], Iter [260/418] Loss: 1.3264, average\_loss: 1.7089  
 Epoch [43/50], Iter [265/418] Loss: 2.2853, average\_loss: 1.7143  
 Epoch [43/50], Iter [270/418] Loss: 1.3602, average\_loss: 1.7151  
 Epoch [43/50], Iter [275/418] Loss: 1.2272, average\_loss: 1.7092  
 Epoch [43/50], Iter [280/418] Loss: 1.9811, average\_loss: 1.7100  
 Epoch [43/50], Iter [285/418] Loss: 2.1135, average\_loss: 1.7119  
 Epoch [43/50], Iter [290/418] Loss: 2.2873, average\_loss: 1.7219  
 Epoch [43/50], Iter [295/418] Loss: 1.3072, average\_loss: 1.7220  
 Epoch [43/50], Iter [300/418] Loss: 2.2596, average\_loss: 1.7251  
 Epoch [43/50], Iter [305/418] Loss: 1.4916, average\_loss: 1.7302  
 Epoch [43/50], Iter [310/418] Loss: 1.7776, average\_loss: 1.7284  
 Epoch [43/50], Iter [315/418] Loss: 1.5103, average\_loss: 1.7240  
 Epoch [43/50], Iter [320/418] Loss: 1.7261, average\_loss: 1.7223  
 Epoch [43/50], Iter [325/418] Loss: 1.3564, average\_loss: 1.7222

Epoch [43/50], Iter [330/418] Loss: 2.2312, average\_loss: 1.7201  
Epoch [43/50], Iter [335/418] Loss: 1.3153, average\_loss: 1.7197  
Epoch [43/50], Iter [340/418] Loss: 1.8114, average\_loss: 1.7185  
Epoch [43/50], Iter [345/418] Loss: 1.9086, average\_loss: 1.7153  
Epoch [43/50], Iter [350/418] Loss: 1.6154, average\_loss: 1.7188  
Epoch [43/50], Iter [355/418] Loss: 2.4750, average\_loss: 1.7232  
Epoch [43/50], Iter [360/418] Loss: 1.1112, average\_loss: 1.7222  
Epoch [43/50], Iter [365/418] Loss: 1.2764, average\_loss: 1.7175  
Epoch [43/50], Iter [370/418] Loss: 1.8364, average\_loss: 1.7162  
Epoch [43/50], Iter [375/418] Loss: 3.2552, average\_loss: 1.7198  
Epoch [43/50], Iter [380/418] Loss: 0.9868, average\_loss: 1.7180  
Epoch [43/50], Iter [385/418] Loss: 1.5724, average\_loss: 1.7173  
Epoch [43/50], Iter [390/418] Loss: 1.4907, average\_loss: 1.7219  
Epoch [43/50], Iter [395/418] Loss: 3.1671, average\_loss: 1.7273  
Epoch [43/50], Iter [400/418] Loss: 1.8834, average\_loss: 1.7282  
Epoch [43/50], Iter [405/418] Loss: 2.1456, average\_loss: 1.7297  
Epoch [43/50], Iter [410/418] Loss: 0.9649, average\_loss: 1.7270  
Epoch [43/50], Iter [415/418] Loss: 1.9938, average\_loss: 1.7262

Starting epoch 44 / 50

Learning Rate for this epoch: 1e-05

Epoch [44/50], Iter [5/418] Loss: 0.8769, average\_loss: 1.5310  
Epoch [44/50], Iter [10/418] Loss: 1.9161, average\_loss: 1.4528  
Epoch [44/50], Iter [15/418] Loss: 1.2587, average\_loss: 1.4694  
Epoch [44/50], Iter [20/418] Loss: 2.2661, average\_loss: 1.5395  
Epoch [44/50], Iter [25/418] Loss: 1.5789, average\_loss: 1.5496  
Epoch [44/50], Iter [30/418] Loss: 1.5944, average\_loss: 1.5572  
Epoch [44/50], Iter [35/418] Loss: 1.2923, average\_loss: 1.6044  
Epoch [44/50], Iter [40/418] Loss: 1.4908, average\_loss: 1.5906  
Epoch [44/50], Iter [45/418] Loss: 1.5814, average\_loss: 1.6425  
Epoch [44/50], Iter [50/418] Loss: 2.3644, average\_loss: 1.6785  
Epoch [44/50], Iter [55/418] Loss: 0.9914, average\_loss: 1.6561  
Epoch [44/50], Iter [60/418] Loss: 1.7891, average\_loss: 1.6489  
Epoch [44/50], Iter [65/418] Loss: 1.6213, average\_loss: 1.6462  
Epoch [44/50], Iter [70/418] Loss: 1.8251, average\_loss: 1.6622  
Epoch [44/50], Iter [75/418] Loss: 1.9794, average\_loss: 1.6651  
Epoch [44/50], Iter [80/418] Loss: 1.5483, average\_loss: 1.6603  
Epoch [44/50], Iter [85/418] Loss: 2.2272, average\_loss: 1.6743  
Epoch [44/50], Iter [90/418] Loss: 1.5598, average\_loss: 1.6777  
Epoch [44/50], Iter [95/418] Loss: 1.3118, average\_loss: 1.6736  
Epoch [44/50], Iter [100/418] Loss: 2.0165, average\_loss: 1.6659  
Epoch [44/50], Iter [105/418] Loss: 1.3255, average\_loss: 1.6639  
Epoch [44/50], Iter [110/418] Loss: 2.0663, average\_loss: 1.6598  
Epoch [44/50], Iter [115/418] Loss: 1.5165, average\_loss: 1.6733  
Epoch [44/50], Iter [120/418] Loss: 2.0226, average\_loss: 1.6628  
Epoch [44/50], Iter [125/418] Loss: 1.1637, average\_loss: 1.6716  
Epoch [44/50], Iter [130/418] Loss: 1.3044, average\_loss: 1.6813



Epoch [44/50], Iter [135/418] Loss: 2.1843, average\_loss: 1.6923  
 Epoch [44/50], Iter [140/418] Loss: 1.5094, average\_loss: 1.6879  
 Epoch [44/50], Iter [145/418] Loss: 1.1624, average\_loss: 1.6902  
 Epoch [44/50], Iter [150/418] Loss: 2.1754, average\_loss: 1.6900  
 Epoch [44/50], Iter [155/418] Loss: 1.3289, average\_loss: 1.6915  
 Epoch [44/50], Iter [160/418] Loss: 2.0593, average\_loss: 1.6998  
 Epoch [44/50], Iter [165/418] Loss: 1.5694, average\_loss: 1.6921  
 Epoch [44/50], Iter [170/418] Loss: 0.9110, average\_loss: 1.6860  
 Epoch [44/50], Iter [175/418] Loss: 2.4422, average\_loss: 1.6933  
 Epoch [44/50], Iter [180/418] Loss: 0.9646, average\_loss: 1.6905  
 Epoch [44/50], Iter [185/418] Loss: 1.9751, average\_loss: 1.6974  
 Epoch [44/50], Iter [190/418] Loss: 1.2371, average\_loss: 1.6991  
 Epoch [44/50], Iter [195/418] Loss: 0.8040, average\_loss: 1.7062  
 Epoch [44/50], Iter [200/418] Loss: 2.1392, average\_loss: 1.7090  
 Epoch [44/50], Iter [205/418] Loss: 1.9392, average\_loss: 1.7083  
 Epoch [44/50], Iter [210/418] Loss: 1.8028, average\_loss: 1.7105  
 Epoch [44/50], Iter [215/418] Loss: 1.6534, average\_loss: 1.7200  
 Epoch [44/50], Iter [220/418] Loss: 2.0320, average\_loss: 1.7244  
 Epoch [44/50], Iter [225/418] Loss: 2.1297, average\_loss: 1.7277  
 Epoch [44/50], Iter [230/418] Loss: 1.8350, average\_loss: 1.7335  
 Epoch [44/50], Iter [235/418] Loss: 1.5742, average\_loss: 1.7297  
 Epoch [44/50], Iter [240/418] Loss: 1.7520, average\_loss: 1.7316  
 Epoch [44/50], Iter [245/418] Loss: 2.3393, average\_loss: 1.7321  
 Epoch [44/50], Iter [250/418] Loss: 2.2380, average\_loss: 1.7396  
 Epoch [44/50], Iter [255/418] Loss: 0.9825, average\_loss: 1.7419  
 Epoch [44/50], Iter [260/418] Loss: 1.4406, average\_loss: 1.7383  
 Epoch [44/50], Iter [265/418] Loss: 1.1461, average\_loss: 1.7329  
 Epoch [44/50], Iter [270/418] Loss: 1.6871, average\_loss: 1.7282  
 Epoch [44/50], Iter [275/418] Loss: 1.7942, average\_loss: 1.7315  
 Epoch [44/50], Iter [280/418] Loss: 1.6982, average\_loss: 1.7264  
 Epoch [44/50], Iter [285/418] Loss: 1.1605, average\_loss: 1.7227  
 Epoch [44/50], Iter [290/418] Loss: 1.1925, average\_loss: 1.7192  
 Epoch [44/50], Iter [295/418] Loss: 1.0966, average\_loss: 1.7212  
 Epoch [44/50], Iter [300/418] Loss: 1.6069, average\_loss: 1.7228  
 Epoch [44/50], Iter [305/418] Loss: 1.4635, average\_loss: 1.7168  
 Epoch [44/50], Iter [310/418] Loss: 1.9242, average\_loss: 1.7221  
 Epoch [44/50], Iter [315/418] Loss: 2.6045, average\_loss: 1.7204  
 Epoch [44/50], Iter [320/418] Loss: 1.8440, average\_loss: 1.7204  
 Epoch [44/50], Iter [325/418] Loss: 1.8246, average\_loss: 1.7230  
 Epoch [44/50], Iter [330/418] Loss: 1.2569, average\_loss: 1.7157  
 Epoch [44/50], Iter [335/418] Loss: 1.7343, average\_loss: 1.7175  
 Epoch [44/50], Iter [340/418] Loss: 1.2072, average\_loss: 1.7198  
 Epoch [44/50], Iter [345/418] Loss: 1.9671, average\_loss: 1.7176  
 Epoch [44/50], Iter [350/418] Loss: 1.8159, average\_loss: 1.7189  
 Epoch [44/50], Iter [355/418] Loss: 1.9687, average\_loss: 1.7200  
 Epoch [44/50], Iter [360/418] Loss: 2.6086, average\_loss: 1.7204  
 Epoch [44/50], Iter [365/418] Loss: 2.4906, average\_loss: 1.7223  
 Epoch [44/50], Iter [370/418] Loss: 2.4752, average\_loss: 1.7326

Epoch [44/50], Iter [375/418] Loss: 1.2104, average\_loss: 1.7305  
Epoch [44/50], Iter [380/418] Loss: 1.8054, average\_loss: 1.7316  
Epoch [44/50], Iter [385/418] Loss: 0.8215, average\_loss: 1.7288  
Epoch [44/50], Iter [390/418] Loss: 1.5925, average\_loss: 1.7321  
Epoch [44/50], Iter [395/418] Loss: 1.2949, average\_loss: 1.7303  
Epoch [44/50], Iter [400/418] Loss: 1.2270, average\_loss: 1.7259  
Epoch [44/50], Iter [405/418] Loss: 1.8260, average\_loss: 1.7233  
Epoch [44/50], Iter [410/418] Loss: 1.8110, average\_loss: 1.7219  
Epoch [44/50], Iter [415/418] Loss: 1.4665, average\_loss: 1.7226

Starting epoch 45 / 50

Learning Rate for this epoch: 1e-05

Epoch [45/50], Iter [5/418] Loss: 1.5671, average\_loss: 1.3871  
Epoch [45/50], Iter [10/418] Loss: 1.4505, average\_loss: 1.5004  
Epoch [45/50], Iter [15/418] Loss: 2.0936, average\_loss: 1.5743  
Epoch [45/50], Iter [20/418] Loss: 1.3360, average\_loss: 1.5433  
Epoch [45/50], Iter [25/418] Loss: 2.4937, average\_loss: 1.6131  
Epoch [45/50], Iter [30/418] Loss: 2.3985, average\_loss: 1.6047  
Epoch [45/50], Iter [35/418] Loss: 1.2922, average\_loss: 1.6271  
Epoch [45/50], Iter [40/418] Loss: 1.2417, average\_loss: 1.6284  
Epoch [45/50], Iter [45/418] Loss: 1.2608, average\_loss: 1.6576  
Epoch [45/50], Iter [50/418] Loss: 1.5877, average\_loss: 1.6489  
Epoch [45/50], Iter [55/418] Loss: 1.2235, average\_loss: 1.6599  
Epoch [45/50], Iter [60/418] Loss: 1.2359, average\_loss: 1.6500  
Epoch [45/50], Iter [65/418] Loss: 1.8836, average\_loss: 1.6441  
Epoch [45/50], Iter [70/418] Loss: 1.5783, average\_loss: 1.6689  
Epoch [45/50], Iter [75/418] Loss: 1.9057, average\_loss: 1.6629  
Epoch [45/50], Iter [80/418] Loss: 1.7906, average\_loss: 1.6657  
Epoch [45/50], Iter [85/418] Loss: 2.1950, average\_loss: 1.6734  
Epoch [45/50], Iter [90/418] Loss: 1.6163, average\_loss: 1.6762  
Epoch [45/50], Iter [95/418] Loss: 1.2320, average\_loss: 1.6717  
Epoch [45/50], Iter [100/418] Loss: 1.4547, average\_loss: 1.6642  
Epoch [45/50], Iter [105/418] Loss: 1.4752, average\_loss: 1.6598  
Epoch [45/50], Iter [110/418] Loss: 1.2128, average\_loss: 1.6531  
Epoch [45/50], Iter [115/418] Loss: 1.9625, average\_loss: 1.6587  
Epoch [45/50], Iter [120/418] Loss: 1.5066, average\_loss: 1.6568  
Epoch [45/50], Iter [125/418] Loss: 1.5984, average\_loss: 1.6489  
Epoch [45/50], Iter [130/418] Loss: 1.7252, average\_loss: 1.6395  
Epoch [45/50], Iter [135/418] Loss: 1.5011, average\_loss: 1.6339  
Epoch [45/50], Iter [140/418] Loss: 2.6730, average\_loss: 1.6381  
Epoch [45/50], Iter [145/418] Loss: 1.6823, average\_loss: 1.6580  
Epoch [45/50], Iter [150/418] Loss: 2.2875, average\_loss: 1.6714  
Epoch [45/50], Iter [155/418] Loss: 1.4292, average\_loss: 1.6684  
Epoch [45/50], Iter [160/418] Loss: 0.9845, average\_loss: 1.6619  
Epoch [45/50], Iter [165/418] Loss: 1.5269, average\_loss: 1.6610  
Epoch [45/50], Iter [170/418] Loss: 2.2601, average\_loss: 1.6637  
Epoch [45/50], Iter [175/418] Loss: 1.1370, average\_loss: 1.6557

Epoch [45/50], Iter [180/418] Loss: 1.9607, average\_loss: 1.6697  
 Epoch [45/50], Iter [185/418] Loss: 1.8157, average\_loss: 1.6757  
 Epoch [45/50], Iter [190/418] Loss: 1.6004, average\_loss: 1.6676  
 Epoch [45/50], Iter [195/418] Loss: 1.4118, average\_loss: 1.6846  
 Epoch [45/50], Iter [200/418] Loss: 1.9022, average\_loss: 1.6844  
 Epoch [45/50], Iter [205/418] Loss: 1.2402, average\_loss: 1.6819  
 Epoch [45/50], Iter [210/418] Loss: 1.5600, average\_loss: 1.6794  
 Epoch [45/50], Iter [215/418] Loss: 1.9012, average\_loss: 1.6825  
 Epoch [45/50], Iter [220/418] Loss: 0.9967, average\_loss: 1.6800  
 Epoch [45/50], Iter [225/418] Loss: 1.8415, average\_loss: 1.6791  
 Epoch [45/50], Iter [230/418] Loss: 0.7437, average\_loss: 1.6743  
 Epoch [45/50], Iter [235/418] Loss: 1.1053, average\_loss: 1.6798  
 Epoch [45/50], Iter [240/418] Loss: 0.7988, average\_loss: 1.6672  
 Epoch [45/50], Iter [245/418] Loss: 1.7569, average\_loss: 1.6658  
 Epoch [45/50], Iter [250/418] Loss: 0.9084, average\_loss: 1.6617  
 Epoch [45/50], Iter [255/418] Loss: 1.5702, average\_loss: 1.6608  
 Epoch [45/50], Iter [260/418] Loss: 1.1049, average\_loss: 1.6581  
 Epoch [45/50], Iter [265/418] Loss: 0.9804, average\_loss: 1.6527  
 Epoch [45/50], Iter [270/418] Loss: 1.1358, average\_loss: 1.6500  
 Epoch [45/50], Iter [275/418] Loss: 1.2716, average\_loss: 1.6510  
 Epoch [45/50], Iter [280/418] Loss: 1.7800, average\_loss: 1.6494  
 Epoch [45/50], Iter [285/418] Loss: 1.6147, average\_loss: 1.6472  
 Epoch [45/50], Iter [290/418] Loss: 3.0359, average\_loss: 1.6552  
 Epoch [45/50], Iter [295/418] Loss: 1.9710, average\_loss: 1.6578  
 Epoch [45/50], Iter [300/418] Loss: 0.8905, average\_loss: 1.6559  
 Epoch [45/50], Iter [305/418] Loss: 1.6695, average\_loss: 1.6536  
 Epoch [45/50], Iter [310/418] Loss: 2.0293, average\_loss: 1.6570  
 Epoch [45/50], Iter [315/418] Loss: 2.1204, average\_loss: 1.6647  
 Epoch [45/50], Iter [320/418] Loss: 2.3670, average\_loss: 1.6676  
 Epoch [45/50], Iter [325/418] Loss: 2.2695, average\_loss: 1.6635  
 Epoch [45/50], Iter [330/418] Loss: 1.7843, average\_loss: 1.6659  
 Epoch [45/50], Iter [335/418] Loss: 2.2707, average\_loss: 1.6721  
 Epoch [45/50], Iter [340/418] Loss: 1.8625, average\_loss: 1.6730  
 Epoch [45/50], Iter [345/418] Loss: 1.9207, average\_loss: 1.6753  
 Epoch [45/50], Iter [350/418] Loss: 2.5397, average\_loss: 1.6725  
 Epoch [45/50], Iter [355/418] Loss: 1.1574, average\_loss: 1.6736  
 Epoch [45/50], Iter [360/418] Loss: 1.8146, average\_loss: 1.6821  
 Epoch [45/50], Iter [365/418] Loss: 2.6241, average\_loss: 1.6803  
 Epoch [45/50], Iter [370/418] Loss: 2.5399, average\_loss: 1.6824  
 Epoch [45/50], Iter [375/418] Loss: 1.0309, average\_loss: 1.6855  
 Epoch [45/50], Iter [380/418] Loss: 2.3432, average\_loss: 1.6880  
 Epoch [45/50], Iter [385/418] Loss: 1.7542, average\_loss: 1.6860  
 Epoch [45/50], Iter [390/418] Loss: 1.8622, average\_loss: 1.6861  
 Epoch [45/50], Iter [395/418] Loss: 1.1187, average\_loss: 1.6858  
 Epoch [45/50], Iter [400/418] Loss: 0.9504, average\_loss: 1.6829  
 Epoch [45/50], Iter [405/418] Loss: 1.5832, average\_loss: 1.6827  
 Epoch [45/50], Iter [410/418] Loss: 1.7778, average\_loss: 1.6847  
 Epoch [45/50], Iter [415/418] Loss: 1.3443, average\_loss: 1.6862

Starting epoch 46 / 50

Learning Rate for this epoch: 1e-05

Epoch [46/50], Iter [5/418] Loss: 1.7651, average\_loss: 1.4781  
Epoch [46/50], Iter [10/418] Loss: 2.2195, average\_loss: 1.7089  
Epoch [46/50], Iter [15/418] Loss: 0.9727, average\_loss: 1.6070  
Epoch [46/50], Iter [20/418] Loss: 1.0888, average\_loss: 1.6407  
Epoch [46/50], Iter [25/418] Loss: 1.3484, average\_loss: 1.6336  
Epoch [46/50], Iter [30/418] Loss: 3.1033, average\_loss: 1.7260  
Epoch [46/50], Iter [35/418] Loss: 1.2258, average\_loss: 1.7423  
Epoch [46/50], Iter [40/418] Loss: 1.3637, average\_loss: 1.7030  
Epoch [46/50], Iter [45/418] Loss: 2.8918, average\_loss: 1.7282  
Epoch [46/50], Iter [50/418] Loss: 2.5083, average\_loss: 1.7377  
Epoch [46/50], Iter [55/418] Loss: 3.0216, average\_loss: 1.7590  
Epoch [46/50], Iter [60/418] Loss: 2.0843, average\_loss: 1.7635  
Epoch [46/50], Iter [65/418] Loss: 1.9390, average\_loss: 1.7455  
Epoch [46/50], Iter [70/418] Loss: 1.4675, average\_loss: 1.7468  
Epoch [46/50], Iter [75/418] Loss: 2.4162, average\_loss: 1.7327  
Epoch [46/50], Iter [80/418] Loss: 1.3655, average\_loss: 1.7189  
Epoch [46/50], Iter [85/418] Loss: 1.3819, average\_loss: 1.7169  
Epoch [46/50], Iter [90/418] Loss: 1.6119, average\_loss: 1.7143  
Epoch [46/50], Iter [95/418] Loss: 2.2304, average\_loss: 1.7217  
Epoch [46/50], Iter [100/418] Loss: 1.1804, average\_loss: 1.7210  
Epoch [46/50], Iter [105/418] Loss: 1.2307, average\_loss: 1.7226  
Epoch [46/50], Iter [110/418] Loss: 1.9053, average\_loss: 1.7308  
Epoch [46/50], Iter [115/418] Loss: 1.6657, average\_loss: 1.7383  
Epoch [46/50], Iter [120/418] Loss: 1.4380, average\_loss: 1.7347  
Epoch [46/50], Iter [125/418] Loss: 1.6218, average\_loss: 1.7501  
Epoch [46/50], Iter [130/418] Loss: 2.1677, average\_loss: 1.7693  
Epoch [46/50], Iter [135/418] Loss: 1.4904, average\_loss: 1.7601  
Epoch [46/50], Iter [140/418] Loss: 1.9277, average\_loss: 1.7497  
Epoch [46/50], Iter [145/418] Loss: 1.2244, average\_loss: 1.7634  
Epoch [46/50], Iter [150/418] Loss: 2.2424, average\_loss: 1.7605  
Epoch [46/50], Iter [155/418] Loss: 1.6116, average\_loss: 1.7568  
Epoch [46/50], Iter [160/418] Loss: 1.8728, average\_loss: 1.7640  
Epoch [46/50], Iter [165/418] Loss: 1.6674, average\_loss: 1.7570  
Epoch [46/50], Iter [170/418] Loss: 2.1734, average\_loss: 1.7561  
Epoch [46/50], Iter [175/418] Loss: 1.2217, average\_loss: 1.7551  
Epoch [46/50], Iter [180/418] Loss: 0.9122, average\_loss: 1.7474  
Epoch [46/50], Iter [185/418] Loss: 1.8665, average\_loss: 1.7485  
Epoch [46/50], Iter [190/418] Loss: 1.9394, average\_loss: 1.7432  
Epoch [46/50], Iter [195/418] Loss: 1.7190, average\_loss: 1.7431  
Epoch [46/50], Iter [200/418] Loss: 1.5432, average\_loss: 1.7525  
Epoch [46/50], Iter [205/418] Loss: 2.8527, average\_loss: 1.7512  
Epoch [46/50], Iter [210/418] Loss: 1.8169, average\_loss: 1.7535  
Epoch [46/50], Iter [215/418] Loss: 1.8435, average\_loss: 1.7550  
Epoch [46/50], Iter [220/418] Loss: 1.8569, average\_loss: 1.7566

```

Epoch [46/50], Iter [225/418] Loss: 1.5798, average_loss: 1.7544
Epoch [46/50], Iter [230/418] Loss: 0.9884, average_loss: 1.7452
Epoch [46/50], Iter [235/418] Loss: 1.2120, average_loss: 1.7417
Epoch [46/50], Iter [240/418] Loss: 1.4965, average_loss: 1.7414
Epoch [46/50], Iter [245/418] Loss: 2.3849, average_loss: 1.7435
Epoch [46/50], Iter [250/418] Loss: 2.1993, average_loss: 1.7428
Epoch [46/50], Iter [255/418] Loss: 1.4773, average_loss: 1.7399
Epoch [46/50], Iter [260/418] Loss: 2.1284, average_loss: 1.7332
Epoch [46/50], Iter [265/418] Loss: 1.7337, average_loss: 1.7269
Epoch [46/50], Iter [270/418] Loss: 2.0908, average_loss: 1.7312
Epoch [46/50], Iter [275/418] Loss: 1.3603, average_loss: 1.7227
Epoch [46/50], Iter [280/418] Loss: 1.7696, average_loss: 1.7267
Epoch [46/50], Iter [285/418] Loss: 2.1079, average_loss: 1.7291
Epoch [46/50], Iter [290/418] Loss: 1.0374, average_loss: 1.7276
Epoch [46/50], Iter [295/418] Loss: 1.4818, average_loss: 1.7251
Epoch [46/50], Iter [300/418] Loss: 1.2553, average_loss: 1.7228
Epoch [46/50], Iter [305/418] Loss: 3.6814, average_loss: 1.7257
Epoch [46/50], Iter [310/418] Loss: 2.2905, average_loss: 1.7267
Epoch [46/50], Iter [315/418] Loss: 1.8495, average_loss: 1.7239
Epoch [46/50], Iter [320/418] Loss: 1.4886, average_loss: 1.7230
Epoch [46/50], Iter [325/418] Loss: 1.3760, average_loss: 1.7203
Epoch [46/50], Iter [330/418] Loss: 1.8950, average_loss: 1.7197
Epoch [46/50], Iter [335/418] Loss: 1.5476, average_loss: 1.7244
Epoch [46/50], Iter [340/418] Loss: 1.0536, average_loss: 1.7252
Epoch [46/50], Iter [345/418] Loss: 2.6159, average_loss: 1.7242
Epoch [46/50], Iter [350/418] Loss: 1.7096, average_loss: 1.7234
Epoch [46/50], Iter [355/418] Loss: 2.0382, average_loss: 1.7239
Epoch [46/50], Iter [360/418] Loss: 2.1914, average_loss: 1.7237
Epoch [46/50], Iter [365/418] Loss: 1.6497, average_loss: 1.7208
Epoch [46/50], Iter [370/418] Loss: 3.8207, average_loss: 1.7281
Epoch [46/50], Iter [375/418] Loss: 1.4842, average_loss: 1.7308
Epoch [46/50], Iter [380/418] Loss: 1.9581, average_loss: 1.7315
Epoch [46/50], Iter [385/418] Loss: 1.4090, average_loss: 1.7356
Epoch [46/50], Iter [390/418] Loss: 0.9267, average_loss: 1.7309
Epoch [46/50], Iter [395/418] Loss: 1.4609, average_loss: 1.7301
Epoch [46/50], Iter [400/418] Loss: 2.3423, average_loss: 1.7317
Epoch [46/50], Iter [405/418] Loss: 1.4084, average_loss: 1.7317
Epoch [46/50], Iter [410/418] Loss: 1.7044, average_loss: 1.7320
Epoch [46/50], Iter [415/418] Loss: 1.8692, average_loss: 1.7373

```

---Evaluate model on test samples---

100%|

| 4950/4950 [02:31<00:00, 32.67it/s]

---class aeroplane ap 0.5568121902511313---

---class bicycle ap 0.6263026977988786---

---class bird ap 0.5170971204547464---

---class boat ap 0.2920033716832511---

---class bottle ap 0.24616003295621794---

```

---class bus ap 0.6595053067288326---
---class car ap 0.7030972053072558---
---class cat ap 0.7083429000088228---
---class chair ap 0.3004780726682931---
---class cow ap 0.5644882379823999---
---class diningtable ap 0.3884388134381002---
---class dog ap 0.6294547014830273---
---class horse ap 0.6782739121546885---
---class motorbike ap 0.6304760277714454---
---class person ap 0.571868283374416---
---class pottedplant ap 0.21395678705290905---
---class sheep ap 0.48437587779335917---
---class sofa ap 0.5145873515651925---
---class train ap 0.6662720257036299---
---class tvmonitor ap 0.5075489663090084---
---map 0.5229769941242803---

```

Starting epoch 47 / 50

Learning Rate for this epoch: 1e-05

```

Epoch [47/50], Iter [5/418] Loss: 1.2987, average_loss: 1.8673
Epoch [47/50], Iter [10/418] Loss: 1.5955, average_loss: 1.7622
Epoch [47/50], Iter [15/418] Loss: 1.2586, average_loss: 1.6786
Epoch [47/50], Iter [20/418] Loss: 2.3040, average_loss: 1.6515
Epoch [47/50], Iter [25/418] Loss: 2.3531, average_loss: 1.6351
Epoch [47/50], Iter [30/418] Loss: 1.1475, average_loss: 1.6795
Epoch [47/50], Iter [35/418] Loss: 1.8768, average_loss: 1.6952
Epoch [47/50], Iter [40/418] Loss: 1.3466, average_loss: 1.7012
Epoch [47/50], Iter [45/418] Loss: 1.8529, average_loss: 1.6931
Epoch [47/50], Iter [50/418] Loss: 1.6929, average_loss: 1.7234
Epoch [47/50], Iter [55/418] Loss: 2.1153, average_loss: 1.7504
Epoch [47/50], Iter [60/418] Loss: 2.2697, average_loss: 1.7436
Epoch [47/50], Iter [65/418] Loss: 1.6451, average_loss: 1.7098
Epoch [47/50], Iter [70/418] Loss: 1.1916, average_loss: 1.7064
Epoch [47/50], Iter [75/418] Loss: 2.1488, average_loss: 1.7101
Epoch [47/50], Iter [80/418] Loss: 1.6168, average_loss: 1.7289
Epoch [47/50], Iter [85/418] Loss: 1.9072, average_loss: 1.7308
Epoch [47/50], Iter [90/418] Loss: 1.4611, average_loss: 1.7442
Epoch [47/50], Iter [95/418] Loss: 2.0102, average_loss: 1.7355
Epoch [47/50], Iter [100/418] Loss: 1.4821, average_loss: 1.7286
Epoch [47/50], Iter [105/418] Loss: 1.4953, average_loss: 1.7364
Epoch [47/50], Iter [110/418] Loss: 2.7924, average_loss: 1.7360
Epoch [47/50], Iter [115/418] Loss: 1.6792, average_loss: 1.7369
Epoch [47/50], Iter [120/418] Loss: 1.8718, average_loss: 1.7313
Epoch [47/50], Iter [125/418] Loss: 2.7667, average_loss: 1.7513
Epoch [47/50], Iter [130/418] Loss: 1.5881, average_loss: 1.7583
Epoch [47/50], Iter [135/418] Loss: 1.8022, average_loss: 1.7576
Epoch [47/50], Iter [140/418] Loss: 2.6181, average_loss: 1.7616

```

Epoch [47/50], Iter [145/418] Loss: 1.0685, average\_loss: 1.7643  
 Epoch [47/50], Iter [150/418] Loss: 1.4228, average\_loss: 1.7569  
 Epoch [47/50], Iter [155/418] Loss: 1.6855, average\_loss: 1.7539  
 Epoch [47/50], Iter [160/418] Loss: 1.5926, average\_loss: 1.7501  
 Epoch [47/50], Iter [165/418] Loss: 1.9180, average\_loss: 1.7479  
 Epoch [47/50], Iter [170/418] Loss: 1.4244, average\_loss: 1.7431  
 Epoch [47/50], Iter [175/418] Loss: 1.5877, average\_loss: 1.7390  
 Epoch [47/50], Iter [180/418] Loss: 1.9892, average\_loss: 1.7447  
 Epoch [47/50], Iter [185/418] Loss: 1.0467, average\_loss: 1.7355  
 Epoch [47/50], Iter [190/418] Loss: 1.6762, average\_loss: 1.7357  
 Epoch [47/50], Iter [195/418] Loss: 1.3304, average\_loss: 1.7301  
 Epoch [47/50], Iter [200/418] Loss: 2.3652, average\_loss: 1.7272  
 Epoch [47/50], Iter [205/418] Loss: 1.4081, average\_loss: 1.7274  
 Epoch [47/50], Iter [210/418] Loss: 1.1177, average\_loss: 1.7236  
 Epoch [47/50], Iter [215/418] Loss: 1.2669, average\_loss: 1.7173  
 Epoch [47/50], Iter [220/418] Loss: 1.3033, average\_loss: 1.7198  
 Epoch [47/50], Iter [225/418] Loss: 0.9639, average\_loss: 1.7138  
 Epoch [47/50], Iter [230/418] Loss: 1.4054, average\_loss: 1.7054  
 Epoch [47/50], Iter [235/418] Loss: 1.3079, average\_loss: 1.7016  
 Epoch [47/50], Iter [240/418] Loss: 1.8586, average\_loss: 1.7072  
 Epoch [47/50], Iter [245/418] Loss: 3.6089, average\_loss: 1.7115  
 Epoch [47/50], Iter [250/418] Loss: 0.8203, average\_loss: 1.7070  
 Epoch [47/50], Iter [255/418] Loss: 2.9594, average\_loss: 1.7159  
 Epoch [47/50], Iter [260/418] Loss: 1.3477, average\_loss: 1.7133  
 Epoch [47/50], Iter [265/418] Loss: 1.0165, average\_loss: 1.7030  
 Epoch [47/50], Iter [270/418] Loss: 1.8737, average\_loss: 1.7066  
 Epoch [47/50], Iter [275/418] Loss: 1.1728, average\_loss: 1.7032  
 Epoch [47/50], Iter [280/418] Loss: 1.9314, average\_loss: 1.7010  
 Epoch [47/50], Iter [285/418] Loss: 1.7562, average\_loss: 1.7041  
 Epoch [47/50], Iter [290/418] Loss: 2.0359, average\_loss: 1.7015  
 Epoch [47/50], Iter [295/418] Loss: 2.0025, average\_loss: 1.7026  
 Epoch [47/50], Iter [300/418] Loss: 2.9112, average\_loss: 1.7104  
 Epoch [47/50], Iter [305/418] Loss: 1.5535, average\_loss: 1.7103  
 Epoch [47/50], Iter [310/418] Loss: 1.7570, average\_loss: 1.7058  
 Epoch [47/50], Iter [315/418] Loss: 2.3541, average\_loss: 1.7095  
 Epoch [47/50], Iter [320/418] Loss: 2.0198, average\_loss: 1.7147  
 Epoch [47/50], Iter [325/418] Loss: 1.0868, average\_loss: 1.7124  
 Epoch [47/50], Iter [330/418] Loss: 1.8041, average\_loss: 1.7140  
 Epoch [47/50], Iter [335/418] Loss: 3.0849, average\_loss: 1.7191  
 Epoch [47/50], Iter [340/418] Loss: 2.2622, average\_loss: 1.7274  
 Epoch [47/50], Iter [345/418] Loss: 2.1398, average\_loss: 1.7296  
 Epoch [47/50], Iter [350/418] Loss: 1.6176, average\_loss: 1.7228  
 Epoch [47/50], Iter [355/418] Loss: 1.4826, average\_loss: 1.7199  
 Epoch [47/50], Iter [360/418] Loss: 2.2445, average\_loss: 1.7232  
 Epoch [47/50], Iter [365/418] Loss: 1.1408, average\_loss: 1.7231  
 Epoch [47/50], Iter [370/418] Loss: 2.7682, average\_loss: 1.7259  
 Epoch [47/50], Iter [375/418] Loss: 1.9648, average\_loss: 1.7268  
 Epoch [47/50], Iter [380/418] Loss: 1.4847, average\_loss: 1.7245

Epoch [47/50], Iter [385/418] Loss: 0.9889, average\_loss: 1.7198  
Epoch [47/50], Iter [390/418] Loss: 2.7149, average\_loss: 1.7196  
Epoch [47/50], Iter [395/418] Loss: 1.8298, average\_loss: 1.7203  
Epoch [47/50], Iter [400/418] Loss: 2.4183, average\_loss: 1.7232  
Epoch [47/50], Iter [405/418] Loss: 2.3363, average\_loss: 1.7216  
Epoch [47/50], Iter [410/418] Loss: 2.3886, average\_loss: 1.7232  
Epoch [47/50], Iter [415/418] Loss: 1.4283, average\_loss: 1.7209

Starting epoch 48 / 50

Learning Rate for this epoch: 1e-05

Epoch [48/50], Iter [5/418] Loss: 2.4424, average\_loss: 2.0868  
Epoch [48/50], Iter [10/418] Loss: 2.2001, average\_loss: 1.8924  
Epoch [48/50], Iter [15/418] Loss: 1.5755, average\_loss: 1.7424  
Epoch [48/50], Iter [20/418] Loss: 0.8539, average\_loss: 1.7435  
Epoch [48/50], Iter [25/418] Loss: 2.2243, average\_loss: 1.7317  
Epoch [48/50], Iter [30/418] Loss: 1.6755, average\_loss: 1.7846  
Epoch [48/50], Iter [35/418] Loss: 1.7147, average\_loss: 1.8273  
Epoch [48/50], Iter [40/418] Loss: 1.5174, average\_loss: 1.8243  
Epoch [48/50], Iter [45/418] Loss: 2.0598, average\_loss: 1.8007  
Epoch [48/50], Iter [50/418] Loss: 0.8819, average\_loss: 1.7471  
Epoch [48/50], Iter [55/418] Loss: 1.9873, average\_loss: 1.7651  
Epoch [48/50], Iter [60/418] Loss: 1.8041, average\_loss: 1.7465  
Epoch [48/50], Iter [65/418] Loss: 1.5189, average\_loss: 1.7174  
Epoch [48/50], Iter [70/418] Loss: 1.4731, average\_loss: 1.7062  
Epoch [48/50], Iter [75/418] Loss: 1.2437, average\_loss: 1.6944  
Epoch [48/50], Iter [80/418] Loss: 1.7235, average\_loss: 1.6832  
Epoch [48/50], Iter [85/418] Loss: 4.8134, average\_loss: 1.7319  
Epoch [48/50], Iter [90/418] Loss: 1.6805, average\_loss: 1.7476  
Epoch [48/50], Iter [95/418] Loss: 1.2767, average\_loss: 1.7237  
Epoch [48/50], Iter [100/418] Loss: 2.6505, average\_loss: 1.7414  
Epoch [48/50], Iter [105/418] Loss: 3.3686, average\_loss: 1.7613  
Epoch [48/50], Iter [110/418] Loss: 1.0663, average\_loss: 1.7385  
Epoch [48/50], Iter [115/418] Loss: 2.3652, average\_loss: 1.7454  
Epoch [48/50], Iter [120/418] Loss: 1.7236, average\_loss: 1.7535  
Epoch [48/50], Iter [125/418] Loss: 1.6605, average\_loss: 1.7549  
Epoch [48/50], Iter [130/418] Loss: 1.4382, average\_loss: 1.7500  
Epoch [48/50], Iter [135/418] Loss: 1.4148, average\_loss: 1.7545  
Epoch [48/50], Iter [140/418] Loss: 2.1440, average\_loss: 1.7585  
Epoch [48/50], Iter [145/418] Loss: 1.3610, average\_loss: 1.7444  
Epoch [48/50], Iter [150/418] Loss: 2.2331, average\_loss: 1.7433  
Epoch [48/50], Iter [155/418] Loss: 1.3159, average\_loss: 1.7360  
Epoch [48/50], Iter [160/418] Loss: 1.2873, average\_loss: 1.7392  
Epoch [48/50], Iter [165/418] Loss: 1.5936, average\_loss: 1.7314  
Epoch [48/50], Iter [170/418] Loss: 1.7816, average\_loss: 1.7296  
Epoch [48/50], Iter [175/418] Loss: 1.6482, average\_loss: 1.7310  
Epoch [48/50], Iter [180/418] Loss: 1.3324, average\_loss: 1.7156  
Epoch [48/50], Iter [185/418] Loss: 1.7005, average\_loss: 1.7150



Epoch [48/50], Iter [190/418] Loss: 1.4812, average\_loss: 1.7073  
 Epoch [48/50], Iter [195/418] Loss: 1.1346, average\_loss: 1.7079  
 Epoch [48/50], Iter [200/418] Loss: 1.4060, average\_loss: 1.6998  
 Epoch [48/50], Iter [205/418] Loss: 1.3151, average\_loss: 1.6935  
 Epoch [48/50], Iter [210/418] Loss: 3.2884, average\_loss: 1.6918  
 Epoch [48/50], Iter [215/418] Loss: 1.8325, average\_loss: 1.6880  
 Epoch [48/50], Iter [220/418] Loss: 1.1764, average\_loss: 1.6875  
 Epoch [48/50], Iter [225/418] Loss: 1.4859, average\_loss: 1.6894  
 Epoch [48/50], Iter [230/418] Loss: 1.5631, average\_loss: 1.6849  
 Epoch [48/50], Iter [235/418] Loss: 1.8428, average\_loss: 1.6792  
 Epoch [48/50], Iter [240/418] Loss: 2.3114, average\_loss: 1.6902  
 Epoch [48/50], Iter [245/418] Loss: 1.2486, average\_loss: 1.7026  
 Epoch [48/50], Iter [250/418] Loss: 2.6309, average\_loss: 1.7082  
 Epoch [48/50], Iter [255/418] Loss: 2.0702, average\_loss: 1.7128  
 Epoch [48/50], Iter [260/418] Loss: 1.6734, average\_loss: 1.7120  
 Epoch [48/50], Iter [265/418] Loss: 1.3038, average\_loss: 1.7069  
 Epoch [48/50], Iter [270/418] Loss: 2.3155, average\_loss: 1.7121  
 Epoch [48/50], Iter [275/418] Loss: 1.7745, average\_loss: 1.7095  
 Epoch [48/50], Iter [280/418] Loss: 1.3670, average\_loss: 1.7107  
 Epoch [48/50], Iter [285/418] Loss: 1.6118, average\_loss: 1.7071  
 Epoch [48/50], Iter [290/418] Loss: 1.6230, average\_loss: 1.7056  
 Epoch [48/50], Iter [295/418] Loss: 1.6122, average\_loss: 1.7073  
 Epoch [48/50], Iter [300/418] Loss: 1.3861, average\_loss: 1.7052  
 Epoch [48/50], Iter [305/418] Loss: 1.3926, average\_loss: 1.7008  
 Epoch [48/50], Iter [310/418] Loss: 1.1521, average\_loss: 1.6986  
 Epoch [48/50], Iter [315/418] Loss: 1.4402, average\_loss: 1.6949  
 Epoch [48/50], Iter [320/418] Loss: 2.4514, average\_loss: 1.6918  
 Epoch [48/50], Iter [325/418] Loss: 2.1599, average\_loss: 1.6908  
 Epoch [48/50], Iter [330/418] Loss: 2.3861, average\_loss: 1.6959  
 Epoch [48/50], Iter [335/418] Loss: 2.1067, average\_loss: 1.6974  
 Epoch [48/50], Iter [340/418] Loss: 1.1053, average\_loss: 1.6906  
 Epoch [48/50], Iter [345/418] Loss: 1.2859, average\_loss: 1.6892  
 Epoch [48/50], Iter [350/418] Loss: 2.2827, average\_loss: 1.6907  
 Epoch [48/50], Iter [355/418] Loss: 1.8100, average\_loss: 1.6875  
 Epoch [48/50], Iter [360/418] Loss: 2.6595, average\_loss: 1.6898  
 Epoch [48/50], Iter [365/418] Loss: 1.5359, average\_loss: 1.6891  
 Epoch [48/50], Iter [370/418] Loss: 1.4148, average\_loss: 1.6841  
 Epoch [48/50], Iter [375/418] Loss: 1.4172, average\_loss: 1.6851  
 Epoch [48/50], Iter [380/418] Loss: 1.4919, average\_loss: 1.6805  
 Epoch [48/50], Iter [385/418] Loss: 1.5783, average\_loss: 1.6766  
 Epoch [48/50], Iter [390/418] Loss: 1.4036, average\_loss: 1.6818  
 Epoch [48/50], Iter [395/418] Loss: 1.3901, average\_loss: 1.6825  
 Epoch [48/50], Iter [400/418] Loss: 1.2884, average\_loss: 1.6808  
 Epoch [48/50], Iter [405/418] Loss: 1.3490, average\_loss: 1.6822  
 Epoch [48/50], Iter [410/418] Loss: 1.0032, average\_loss: 1.6874  
 Epoch [48/50], Iter [415/418] Loss: 1.4051, average\_loss: 1.6857

Starting epoch 49 / 50

Learning Rate for this epoch: 1e-05

Epoch [49/50],	Iter [5/418]	Loss: 2.1261,	average_loss: 1.5709
Epoch [49/50],	Iter [10/418]	Loss: 1.4635,	average_loss: 1.7686
Epoch [49/50],	Iter [15/418]	Loss: 1.3738,	average_loss: 1.9095
Epoch [49/50],	Iter [20/418]	Loss: 1.2931,	average_loss: 1.9384
Epoch [49/50],	Iter [25/418]	Loss: 1.6723,	average_loss: 1.8382
Epoch [49/50],	Iter [30/418]	Loss: 1.5570,	average_loss: 1.8138
Epoch [49/50],	Iter [35/418]	Loss: 2.1855,	average_loss: 1.9156
Epoch [49/50],	Iter [40/418]	Loss: 1.7082,	average_loss: 1.8845
Epoch [49/50],	Iter [45/418]	Loss: 1.7000,	average_loss: 1.8719
Epoch [49/50],	Iter [50/418]	Loss: 1.7013,	average_loss: 1.8682
Epoch [49/50],	Iter [55/418]	Loss: 2.0858,	average_loss: 1.8623
Epoch [49/50],	Iter [60/418]	Loss: 1.2406,	average_loss: 1.8583
Epoch [49/50],	Iter [65/418]	Loss: 1.2700,	average_loss: 1.8326
Epoch [49/50],	Iter [70/418]	Loss: 1.0967,	average_loss: 1.8243
Epoch [49/50],	Iter [75/418]	Loss: 1.3994,	average_loss: 1.8257
Epoch [49/50],	Iter [80/418]	Loss: 3.7315,	average_loss: 1.8316
Epoch [49/50],	Iter [85/418]	Loss: 1.7940,	average_loss: 1.8191
Epoch [49/50],	Iter [90/418]	Loss: 1.2621,	average_loss: 1.8160
Epoch [49/50],	Iter [95/418]	Loss: 1.3516,	average_loss: 1.7950
Epoch [49/50],	Iter [100/418]	Loss: 2.0804,	average_loss: 1.7884
Epoch [49/50],	Iter [105/418]	Loss: 1.0099,	average_loss: 1.7798
Epoch [49/50],	Iter [110/418]	Loss: 1.1441,	average_loss: 1.7744
Epoch [49/50],	Iter [115/418]	Loss: 1.2761,	average_loss: 1.7593
Epoch [49/50],	Iter [120/418]	Loss: 1.8330,	average_loss: 1.7628
Epoch [49/50],	Iter [125/418]	Loss: 1.2398,	average_loss: 1.7584
Epoch [49/50],	Iter [130/418]	Loss: 1.3294,	average_loss: 1.7404
Epoch [49/50],	Iter [135/418]	Loss: 2.7775,	average_loss: 1.7412
Epoch [49/50],	Iter [140/418]	Loss: 1.6758,	average_loss: 1.7392
Epoch [49/50],	Iter [145/418]	Loss: 1.7923,	average_loss: 1.7396
Epoch [49/50],	Iter [150/418]	Loss: 1.9715,	average_loss: 1.7295
Epoch [49/50],	Iter [155/418]	Loss: 1.4277,	average_loss: 1.7255
Epoch [49/50],	Iter [160/418]	Loss: 1.4240,	average_loss: 1.7171
Epoch [49/50],	Iter [165/418]	Loss: 2.2971,	average_loss: 1.7189
Epoch [49/50],	Iter [170/418]	Loss: 2.4883,	average_loss: 1.7106
Epoch [49/50],	Iter [175/418]	Loss: 1.2352,	average_loss: 1.7196
Epoch [49/50],	Iter [180/418]	Loss: 0.9717,	average_loss: 1.7235
Epoch [49/50],	Iter [185/418]	Loss: 2.6886,	average_loss: 1.7252
Epoch [49/50],	Iter [190/418]	Loss: 1.4384,	average_loss: 1.7363
Epoch [49/50],	Iter [195/418]	Loss: 2.8327,	average_loss: 1.7474
Epoch [49/50],	Iter [200/418]	Loss: 1.5791,	average_loss: 1.7454
Epoch [49/50],	Iter [205/418]	Loss: 2.0287,	average_loss: 1.7588
Epoch [49/50],	Iter [210/418]	Loss: 1.5959,	average_loss: 1.7581
Epoch [49/50],	Iter [215/418]	Loss: 2.2633,	average_loss: 1.7502
Epoch [49/50],	Iter [220/418]	Loss: 1.4951,	average_loss: 1.7506
Epoch [49/50],	Iter [225/418]	Loss: 2.4099,	average_loss: 1.7511
Epoch [49/50],	Iter [230/418]	Loss: 1.1975,	average_loss: 1.7476

Epoch [49/50], Iter [235/418] Loss: 1.1161, average\_loss: 1.7384  
Epoch [49/50], Iter [240/418] Loss: 1.6816, average\_loss: 1.7329  
Epoch [49/50], Iter [245/418] Loss: 1.6341, average\_loss: 1.7323  
Epoch [49/50], Iter [250/418] Loss: 1.5134, average\_loss: 1.7305  
Epoch [49/50], Iter [255/418] Loss: 2.1755, average\_loss: 1.7362  
Epoch [49/50], Iter [260/418] Loss: 2.9580, average\_loss: 1.7403  
Epoch [49/50], Iter [265/418] Loss: 1.8336, average\_loss: 1.7358  
Epoch [49/50], Iter [270/418] Loss: 1.3484, average\_loss: 1.7268  
Epoch [49/50], Iter [275/418] Loss: 1.6804, average\_loss: 1.7246  
Epoch [49/50], Iter [280/418] Loss: 2.1330, average\_loss: 1.7268  
Epoch [49/50], Iter [285/418] Loss: 1.9539, average\_loss: 1.7278  
Epoch [49/50], Iter [290/418] Loss: 2.6166, average\_loss: 1.7315  
Epoch [49/50], Iter [295/418] Loss: 2.3206, average\_loss: 1.7353  
Epoch [49/50], Iter [300/418] Loss: 2.2902, average\_loss: 1.7349  
Epoch [49/50], Iter [305/418] Loss: 1.8309, average\_loss: 1.7334  
Epoch [49/50], Iter [310/418] Loss: 1.1006, average\_loss: 1.7292  
Epoch [49/50], Iter [315/418] Loss: 1.7429, average\_loss: 1.7269  
Epoch [49/50], Iter [320/418] Loss: 1.3732, average\_loss: 1.7208  
Epoch [49/50], Iter [325/418] Loss: 1.4571, average\_loss: 1.7221  
Epoch [49/50], Iter [330/418] Loss: 1.8762, average\_loss: 1.7305  
Epoch [49/50], Iter [335/418] Loss: 1.4629, average\_loss: 1.7303  
Epoch [49/50], Iter [340/418] Loss: 2.6764, average\_loss: 1.7311  
Epoch [49/50], Iter [345/418] Loss: 1.9288, average\_loss: 1.7327  
Epoch [49/50], Iter [350/418] Loss: 1.6007, average\_loss: 1.7286  
Epoch [49/50], Iter [355/418] Loss: 1.3301, average\_loss: 1.7290  
Epoch [49/50], Iter [360/418] Loss: 1.5789, average\_loss: 1.7265  
Epoch [49/50], Iter [365/418] Loss: 1.5703, average\_loss: 1.7272  
Epoch [49/50], Iter [370/418] Loss: 1.2131, average\_loss: 1.7291  
Epoch [49/50], Iter [375/418] Loss: 1.0849, average\_loss: 1.7297  
Epoch [49/50], Iter [380/418] Loss: 1.9169, average\_loss: 1.7313  
Epoch [49/50], Iter [385/418] Loss: 2.2620, average\_loss: 1.7305  
Epoch [49/50], Iter [390/418] Loss: 2.0015, average\_loss: 1.7386  
Epoch [49/50], Iter [395/418] Loss: 0.9164, average\_loss: 1.7371  
Epoch [49/50], Iter [400/418] Loss: 1.4788, average\_loss: 1.7378  
Epoch [49/50], Iter [405/418] Loss: 1.0403, average\_loss: 1.7316  
Epoch [49/50], Iter [410/418] Loss: 1.4900, average\_loss: 1.7313  
Epoch [49/50], Iter [415/418] Loss: 2.6713, average\_loss: 1.7333

Starting epoch 50 / 50

Learning Rate for this epoch: 1e-05

Epoch [50/50], Iter [5/418] Loss: 1.9201, average\_loss: 1.6976  
Epoch [50/50], Iter [10/418] Loss: 2.3419, average\_loss: 1.7439  
Epoch [50/50], Iter [15/418] Loss: 2.0072, average\_loss: 1.7231  
Epoch [50/50], Iter [20/418] Loss: 1.3488, average\_loss: 1.6679  
Epoch [50/50], Iter [25/418] Loss: 1.5882, average\_loss: 1.6825  
Epoch [50/50], Iter [30/418] Loss: 1.6077, average\_loss: 1.6724  
Epoch [50/50], Iter [35/418] Loss: 2.1330, average\_loss: 1.6726

Epoch [50/50], Iter [40/418] Loss: 1.0268, average\_loss: 1.6744  
 Epoch [50/50], Iter [45/418] Loss: 1.7403, average\_loss: 1.7058  
 Epoch [50/50], Iter [50/418] Loss: 1.2489, average\_loss: 1.6790  
 Epoch [50/50], Iter [55/418] Loss: 2.1598, average\_loss: 1.7023  
 Epoch [50/50], Iter [60/418] Loss: 2.0704, average\_loss: 1.7125  
 Epoch [50/50], Iter [65/418] Loss: 1.8956, average\_loss: 1.7103  
 Epoch [50/50], Iter [70/418] Loss: 1.8428, average\_loss: 1.7176  
 Epoch [50/50], Iter [75/418] Loss: 1.1740, average\_loss: 1.6961  
 Epoch [50/50], Iter [80/418] Loss: 0.8026, average\_loss: 1.6858  
 Epoch [50/50], Iter [85/418] Loss: 1.4556, average\_loss: 1.7101  
 Epoch [50/50], Iter [90/418] Loss: 2.3676, average\_loss: 1.7205  
 Epoch [50/50], Iter [95/418] Loss: 1.4178, average\_loss: 1.7092  
 Epoch [50/50], Iter [100/418] Loss: 1.3841, average\_loss: 1.6986  
 Epoch [50/50], Iter [105/418] Loss: 2.3767, average\_loss: 1.7032  
 Epoch [50/50], Iter [110/418] Loss: 2.4464, average\_loss: 1.7046  
 Epoch [50/50], Iter [115/418] Loss: 0.7114, average\_loss: 1.7084  
 Epoch [50/50], Iter [120/418] Loss: 1.8930, average\_loss: 1.7183  
 Epoch [50/50], Iter [125/418] Loss: 1.3707, average\_loss: 1.7081  
 Epoch [50/50], Iter [130/418] Loss: 1.0595, average\_loss: 1.6994  
 Epoch [50/50], Iter [135/418] Loss: 1.3826, average\_loss: 1.6951  
 Epoch [50/50], Iter [140/418] Loss: 1.3723, average\_loss: 1.7001  
 Epoch [50/50], Iter [145/418] Loss: 1.2830, average\_loss: 1.6902  
 Epoch [50/50], Iter [150/418] Loss: 0.9231, average\_loss: 1.7011  
 Epoch [50/50], Iter [155/418] Loss: 1.4799, average\_loss: 1.6860  
 Epoch [50/50], Iter [160/418] Loss: 1.5095, average\_loss: 1.6833  
 Epoch [50/50], Iter [165/418] Loss: 1.9671, average\_loss: 1.6884  
 Epoch [50/50], Iter [170/418] Loss: 1.8097, average\_loss: 1.6893  
 Epoch [50/50], Iter [175/418] Loss: 2.1463, average\_loss: 1.6874  
 Epoch [50/50], Iter [180/418] Loss: 0.7499, average\_loss: 1.6804  
 Epoch [50/50], Iter [185/418] Loss: 1.5773, average\_loss: 1.6770  
 Epoch [50/50], Iter [190/418] Loss: 1.2618, average\_loss: 1.6738  
 Epoch [50/50], Iter [195/418] Loss: 3.0676, average\_loss: 1.6781  
 Epoch [50/50], Iter [200/418] Loss: 1.2631, average\_loss: 1.6756  
 Epoch [50/50], Iter [205/418] Loss: 1.9427, average\_loss: 1.6901  
 Epoch [50/50], Iter [210/418] Loss: 2.0781, average\_loss: 1.6966  
 Epoch [50/50], Iter [215/418] Loss: 1.0379, average\_loss: 1.6919  
 Epoch [50/50], Iter [220/418] Loss: 1.9411, average\_loss: 1.6961  
 Epoch [50/50], Iter [225/418] Loss: 1.7036, average\_loss: 1.6920  
 Epoch [50/50], Iter [230/418] Loss: 2.4357, average\_loss: 1.6996  
 Epoch [50/50], Iter [235/418] Loss: 1.3497, average\_loss: 1.6978  
 Epoch [50/50], Iter [240/418] Loss: 1.5485, average\_loss: 1.6982  
 Epoch [50/50], Iter [245/418] Loss: 1.2414, average\_loss: 1.6990  
 Epoch [50/50], Iter [250/418] Loss: 1.0804, average\_loss: 1.6924  
 Epoch [50/50], Iter [255/418] Loss: 1.3470, average\_loss: 1.6922  
 Epoch [50/50], Iter [260/418] Loss: 2.2296, average\_loss: 1.6966  
 Epoch [50/50], Iter [265/418] Loss: 2.8218, average\_loss: 1.6980  
 Epoch [50/50], Iter [270/418] Loss: 2.1839, average\_loss: 1.6991  
 Epoch [50/50], Iter [275/418] Loss: 1.9042, average\_loss: 1.7050

```
Epoch [50/50], Iter [280/418] Loss: 1.5294, average_loss: 1.7046
Epoch [50/50], Iter [285/418] Loss: 1.4887, average_loss: 1.7114
Epoch [50/50], Iter [290/418] Loss: 1.2029, average_loss: 1.7089
Epoch [50/50], Iter [295/418] Loss: 1.5454, average_loss: 1.7089
Epoch [50/50], Iter [300/418] Loss: 2.0601, average_loss: 1.7159
Epoch [50/50], Iter [305/418] Loss: 1.6441, average_loss: 1.7127
Epoch [50/50], Iter [310/418] Loss: 1.8170, average_loss: 1.7203
Epoch [50/50], Iter [315/418] Loss: 1.6138, average_loss: 1.7155
Epoch [50/50], Iter [320/418] Loss: 1.4985, average_loss: 1.7150
Epoch [50/50], Iter [325/418] Loss: 1.5726, average_loss: 1.7125
Epoch [50/50], Iter [330/418] Loss: 2.4887, average_loss: 1.7094
Epoch [50/50], Iter [335/418] Loss: 1.4458, average_loss: 1.7067
Epoch [50/50], Iter [340/418] Loss: 1.2482, average_loss: 1.7042
Epoch [50/50], Iter [345/418] Loss: 2.1577, average_loss: 1.7070
Epoch [50/50], Iter [350/418] Loss: 1.2429, average_loss: 1.7036
Epoch [50/50], Iter [355/418] Loss: 1.6003, average_loss: 1.7033
Epoch [50/50], Iter [360/418] Loss: 1.9431, average_loss: 1.7053
Epoch [50/50], Iter [365/418] Loss: 2.1102, average_loss: 1.7111
Epoch [50/50], Iter [370/418] Loss: 1.6580, average_loss: 1.7158
Epoch [50/50], Iter [375/418] Loss: 1.2279, average_loss: 1.7098
Epoch [50/50], Iter [380/418] Loss: 1.0957, average_loss: 1.7100
Epoch [50/50], Iter [385/418] Loss: 1.2706, average_loss: 1.7080
Epoch [50/50], Iter [390/418] Loss: 1.9639, average_loss: 1.7104
Epoch [50/50], Iter [395/418] Loss: 1.6078, average_loss: 1.7055
Epoch [50/50], Iter [400/418] Loss: 1.3494, average_loss: 1.7020
Epoch [50/50], Iter [405/418] Loss: 2.4662, average_loss: 1.7032
Epoch [50/50], Iter [410/418] Loss: 1.5323, average_loss: 1.7099
Epoch [50/50], Iter [415/418] Loss: 3.1400, average_loss: 1.7133
```

## 2 View example predictions

```
[10]: net.eval()

# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

print('predicting...')
result = predict_image(net, image_name, root_img_directory=file_root_test)
for left_up, right_bottom, class_name, _, prob in result:
    color = COLORS[VOC_CLASSES.index(class_name)]
    cv2.rectangle(image, left_up, right_bottom, color, 2)
    label = class_name + str(round(prob, 2))
    text_size, baseline = cv2.getTextSize(label, cv2.FONT_HERSHEY_SIMPLEX, 0.4, 1)
    ↪1)
```

```

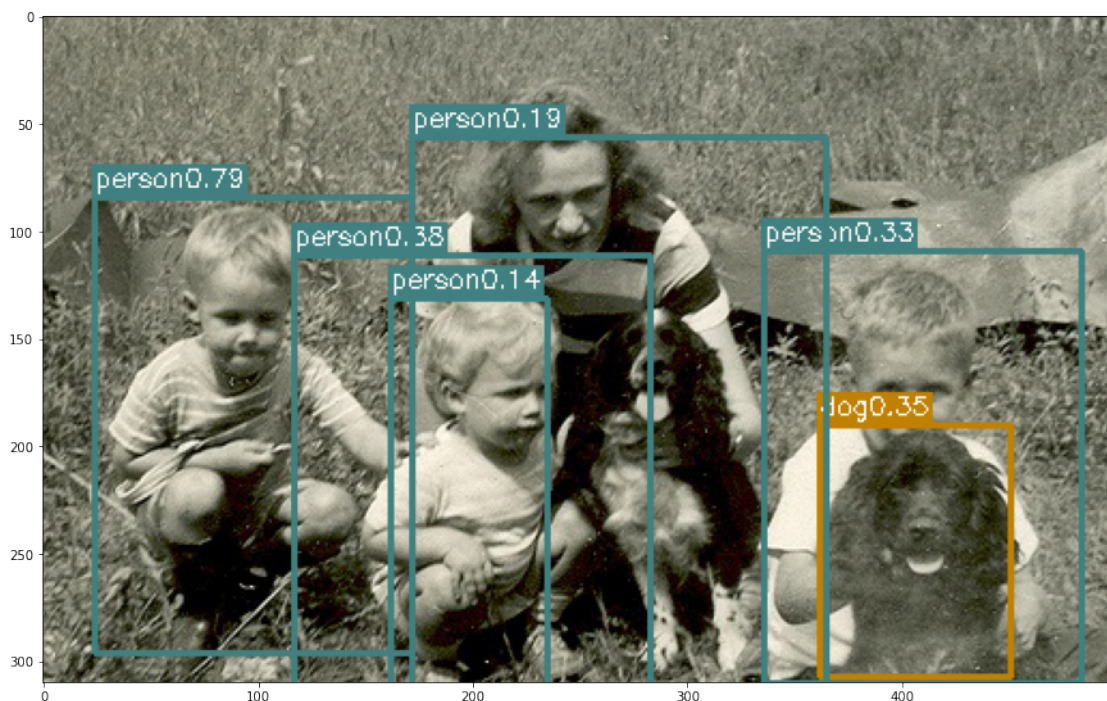
    p1 = (left_up[0], left_up[1] - text_size[1])
    cv2.rectangle(image, (p1[0] - 2 // 2, p1[1] - 2 - baseline), (p1[0] +
→text_size[0], p1[1] + text_size[1]),
                color, -1)
    cv2.putText(image, label, (p1[0], p1[1] + baseline), cv2.
→FONT_HERSHEY_SIMPLEX, 0.4, (255, 255, 255), 1, 8)

plt.figure(figsize = (15,15))
plt.imshow(image)

```

predicting...

[10]: <matplotlib.image.AxesImage at 0x1de83d087f0>



## 2.1 Evaluate on Test

To evaluate detection results we use mAP (mean of average precision over each class)

[11]: `test_aps = evaluate(net, test_dataset_file=annotation_file_test)`

---Evaluate model on test samples---

100%|  
| 4950/4950 [02:35<00:00, 31.79it/s]

```

---class aeroplane ap 0.5352900174432971---
---class bicycle ap 0.6335592431913784---
---class bird ap 0.5387581045697483---
---class boat ap 0.28355504220151023---
---class bottle ap 0.2573056847736956---
---class bus ap 0.6473556218325267---
---class car ap 0.6943787064663378---
---class cat ap 0.7074850870452833---
---class chair ap 0.30946662921244966---
---class cow ap 0.576450544116456---
---class diningtable ap 0.3864492836625337---
---class dog ap 0.6297963062938732---
---class horse ap 0.6874183375183067---
---class motorbike ap 0.6327913194185217---
---class person ap 0.5739780839761537---
---class pottedplant ap 0.21212462631944667---
---class sheep ap 0.46526125677409236---
---class sofa ap 0.5048238949770311---
---class train ap 0.6511643270700493---
---class tvmonitor ap 0.4943055934864307---
---map 0.521085885517456---

```

```
[13]: output_submission_csv('my_solution.csv', test_aps)
```

```
[15]: net.eval()
```

```

# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

print('predicting...')
result = predict_image(net, image_name, root_img_directory=file_root_test)
for left_up, right_bottom, class_name, _, prob in result:
    color = COLORS[VOC_CLASSES.index(class_name)]
    cv2.rectangle(image, left_up, right_bottom, color, 2)
    label = class_name + str(round(prob, 2))
    text_size, baseline = cv2.getTextSize(label, cv2.FONT_HERSHEY_SIMPLEX, 0.4, 1)
    p1 = (left_up[0], left_up[1] - text_size[1])
    cv2.rectangle(image, (p1[0] - 2 // 2, p1[1] - 2 - baseline), (p1[0] +
    text_size[0], p1[1] + text_size[1]),
    color, -1)
    cv2.putText(image, label, (p1[0], p1[1] + baseline), cv2.
    FONT_HERSHEY_SIMPLEX, 0.4, (255, 255, 255), 1, 8)

plt.figure(figsize = (15,15))

```

```
plt.imshow(image)
```

predicting...

[15]: <matplotlib.image.AxesImage at 0x1de6a724580>

