

```
In [3]: from google.colab import drive
drive.mount('/content/gdrive', force_remount=True)
import os
os.chdir("/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2")
!pip3 install torch torchvision
import torch
a = torch.Tensor([1]).cuda()
print(a)
```

```
Mounted at /content/gdrive
Requirement already satisfied: torch in /usr/local/lib/python3.6/dist-packages (1.6.0+cu101)
Requirement already satisfied: torchvision in /usr/local/lib/python3.6/dist-packages (0.7.0+cu101)
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from torchvision) (1.18.5)
Requirement already satisfied: future in /usr/local/lib/python3.6/dist-packages (from torchvision) (0.16.0)
Requirement already satisfied: pillow>=4.1.1 in /usr/local/lib/python3.6/dist-packages (from torchvision) (7.0.0)
tensor([1.], device='cuda:0')
```

```
In [4]: import os
import random

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
```

Initialization

```
In [5]: device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
```

```
In [6]: # 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` (<https://pytorch.org/docs/stable/torchvision/models.html>) 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.

```
In [7]: 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

Downloading: "https://download.pytorch.org/models/resnet50-19c8e357.pth" to /root/.cache/torch/hub/checkpoints/resnet50-19c8e357.pth

```
In [8]: learning_rate = 0.001
num_epochs = 50
batch_size = 24

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

```
In [9]: criterion = YoloLoss(S, B, lambda_coord, lambda_noobj)
optimizer = torch.optim.SGD(net.parameters(), lr=learning_rate, momentum=0.9, weight_decay=5e-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 expected output. We also use a decoder which allows us to convert the opposite direction into image coordinate bounding boxes.

```
In [10]: 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
```

```
In [11]: 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
```

Train detector

```

In [12]: 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_lo
ss / (i+1)))

    # 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

```
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/yolo_loss.py:127: UserWarning:
indexing with dtype torch.uint8 is now deprecated, please use a dtype torch.bool instead. (Triggered internally at /pytorch/aten/src/ATen/native/IndexingUtils.h:20.)
  no_object_prediction_c = no_object_prediction[no_object_prediction_mask]
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/yolo_loss.py:128: UserWarning:
indexing with dtype torch.uint8 is now deprecated, please use a dtype torch.bool instead. (Triggered internally at /pytorch/aten/src/ATen/native/IndexingUtils.h:20.)
  no_object_target_c = no_object_target[no_object_prediction_mask]
/usr/local/lib/python3.6/dist-packages/torch/nn/_reduction.py:44: UserWarning: size_average and reduce args will be deprecated, please use reduction='sum' instead.
  warnings.warn(warning.format(ret))
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/yolo_loss.py:249: UserWarning:
indexing with dtype torch.uint8 is now deprecated, please use a dtype torch.bool instead. (Triggered internally at /pytorch/aten/src/ATen/native/IndexingUtils.h:20.)
  box_pred_response = bounding_box_pred[contains_object_response_mask].view(-1,5)
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/yolo_loss.py:250: UserWarning:
indexing with dtype torch.uint8 is now deprecated, please use a dtype torch.bool instead. (Triggered internally at /pytorch/aten/src/ATen/native/IndexingUtils.h:20.)
  box_target_response_iou = box_target_iou[contains_object_response_mask].view(-1,5)
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/yolo_loss.py:251: UserWarning:
indexing with dtype torch.uint8 is now deprecated, please use a dtype torch.bool instead. (Triggered internally at /pytorch/aten/src/ATen/native/IndexingUtils.h:20.)
  box_target_response = bounding_box_target[contains_object_response_mask].view(-1,5)
```

Epoch [1/50], Iter [5/209] Loss: 61.3833, average_loss: 71.0393
Epoch [1/50], Iter [10/209] Loss: 43.5493, average_loss: 63.6080
Epoch [1/50], Iter [15/209] Loss: 24.3080, average_loss: 53.1566
Epoch [1/50], Iter [20/209] Loss: 14.4515, average_loss: 45.0237
Epoch [1/50], Iter [25/209] Loss: 9.7839, average_loss: 38.8002
Epoch [1/50], Iter [30/209] Loss: 10.3076, average_loss: 34.1920
Epoch [1/50], Iter [35/209] Loss: 13.4134, average_loss: 30.9554
Epoch [1/50], Iter [40/209] Loss: 7.9688, average_loss: 28.2819
Epoch [1/50], Iter [45/209] Loss: 7.9212, average_loss: 26.0658
Epoch [1/50], Iter [50/209] Loss: 8.1076, average_loss: 24.2831
Epoch [1/50], Iter [55/209] Loss: 7.0433, average_loss: 22.7430
Epoch [1/50], Iter [60/209] Loss: 8.3195, average_loss: 21.4984
Epoch [1/50], Iter [65/209] Loss: 6.3860, average_loss: 20.4528
Epoch [1/50], Iter [70/209] Loss: 8.3182, average_loss: 19.5802
Epoch [1/50], Iter [75/209] Loss: 7.0520, average_loss: 18.7179
Epoch [1/50], Iter [80/209] Loss: 9.3236, average_loss: 18.0749
Epoch [1/50], Iter [85/209] Loss: 9.5357, average_loss: 17.4655
Epoch [1/50], Iter [90/209] Loss: 7.0903, average_loss: 16.9009
Epoch [1/50], Iter [95/209] Loss: 5.9050, average_loss: 16.3802
Epoch [1/50], Iter [100/209] Loss: 5.4181, average_loss: 15.9185
Epoch [1/50], Iter [105/209] Loss: 5.6463, average_loss: 15.4895
Epoch [1/50], Iter [110/209] Loss: 6.0936, average_loss: 15.0600
Epoch [1/50], Iter [115/209] Loss: 6.5684, average_loss: 14.6803
Epoch [1/50], Iter [120/209] Loss: 4.7196, average_loss: 14.3111
Epoch [1/50], Iter [125/209] Loss: 8.2107, average_loss: 14.0386
Epoch [1/50], Iter [130/209] Loss: 5.3420, average_loss: 13.7812
Epoch [1/50], Iter [135/209] Loss: 5.5750, average_loss: 13.4725
Epoch [1/50], Iter [140/209] Loss: 4.6416, average_loss: 13.2064
Epoch [1/50], Iter [145/209] Loss: 7.4164, average_loss: 12.9641
Epoch [1/50], Iter [150/209] Loss: 7.5379, average_loss: 12.7334
Epoch [1/50], Iter [155/209] Loss: 5.0346, average_loss: 12.4901
Epoch [1/50], Iter [160/209] Loss: 6.3297, average_loss: 12.2808
Epoch [1/50], Iter [165/209] Loss: 6.7887, average_loss: 12.0917
Epoch [1/50], Iter [170/209] Loss: 5.5304, average_loss: 11.9012
Epoch [1/50], Iter [175/209] Loss: 4.3530, average_loss: 11.7106
Epoch [1/50], Iter [180/209] Loss: 6.4540, average_loss: 11.5480
Epoch [1/50], Iter [185/209] Loss: 5.8050, average_loss: 11.3841
Epoch [1/50], Iter [190/209] Loss: 4.8813, average_loss: 11.2319
Epoch [1/50], Iter [195/209] Loss: 6.2947, average_loss: 11.0911
Epoch [1/50], Iter [200/209] Loss: 4.9609, average_loss: 10.9639
Epoch [1/50], Iter [205/209] Loss: 6.6033, average_loss: 10.8203
Updating best test loss: 5.56547

Starting epoch 2 / 50

Learning Rate for this epoch: 0.001

Epoch [2/50], Iter [5/209] Loss: 4.7458, average_loss: 5.0936
Epoch [2/50], Iter [10/209] Loss: 5.0857, average_loss: 4.9604
Epoch [2/50], Iter [15/209] Loss: 3.9562, average_loss: 5.0908
Epoch [2/50], Iter [20/209] Loss: 6.1923, average_loss: 5.1419
Epoch [2/50], Iter [25/209] Loss: 5.6116, average_loss: 5.2007
Epoch [2/50], Iter [30/209] Loss: 4.7867, average_loss: 5.3400
Epoch [2/50], Iter [35/209] Loss: 5.6774, average_loss: 5.3212
Epoch [2/50], Iter [40/209] Loss: 5.5708, average_loss: 5.3316
Epoch [2/50], Iter [45/209] Loss: 5.3957, average_loss: 5.3494
Epoch [2/50], Iter [50/209] Loss: 6.3796, average_loss: 5.3307
Epoch [2/50], Iter [55/209] Loss: 5.7313, average_loss: 5.3070
Epoch [2/50], Iter [60/209] Loss: 5.7799, average_loss: 5.3156
Epoch [2/50], Iter [65/209] Loss: 3.7776, average_loss: 5.2944
Epoch [2/50], Iter [70/209] Loss: 5.1827, average_loss: 5.3115
Epoch [2/50], Iter [75/209] Loss: 5.7183, average_loss: 5.3373
Epoch [2/50], Iter [80/209] Loss: 5.0238, average_loss: 5.3086

```
Epoch [2/50], Iter [85/209] Loss: 4.7879, average_loss: 5.2897
Epoch [2/50], Iter [90/209] Loss: 4.9005, average_loss: 5.2570
Epoch [2/50], Iter [95/209] Loss: 4.7639, average_loss: 5.2554
Epoch [2/50], Iter [100/209] Loss: 5.1184, average_loss: 5.2421
Epoch [2/50], Iter [105/209] Loss: 4.2226, average_loss: 5.1996
Epoch [2/50], Iter [110/209] Loss: 3.7065, average_loss: 5.1654
Epoch [2/50], Iter [115/209] Loss: 4.6170, average_loss: 5.1699
Epoch [2/50], Iter [120/209] Loss: 5.0205, average_loss: 5.1328
Epoch [2/50], Iter [125/209] Loss: 5.3084, average_loss: 5.1177
Epoch [2/50], Iter [130/209] Loss: 5.6785, average_loss: 5.1222
Epoch [2/50], Iter [135/209] Loss: 5.0385, average_loss: 5.1265
Epoch [2/50], Iter [140/209] Loss: 4.7619, average_loss: 5.1093
Epoch [2/50], Iter [145/209] Loss: 3.4883, average_loss: 5.0901
Epoch [2/50], Iter [150/209] Loss: 4.6222, average_loss: 5.0828
Epoch [2/50], Iter [155/209] Loss: 4.0188, average_loss: 5.0608
Epoch [2/50], Iter [160/209] Loss: 5.4757, average_loss: 5.0542
Epoch [2/50], Iter [165/209] Loss: 4.3615, average_loss: 5.0483
Epoch [2/50], Iter [170/209] Loss: 4.9793, average_loss: 5.0454
Epoch [2/50], Iter [175/209] Loss: 4.8161, average_loss: 5.0387
Epoch [2/50], Iter [180/209] Loss: 4.8702, average_loss: 5.0255
Epoch [2/50], Iter [185/209] Loss: 3.5994, average_loss: 5.0132
Epoch [2/50], Iter [190/209] Loss: 3.7481, average_loss: 5.0003
Epoch [2/50], Iter [195/209] Loss: 3.4758, average_loss: 4.9816
Epoch [2/50], Iter [200/209] Loss: 4.1870, average_loss: 4.9851
Epoch [2/50], Iter [205/209] Loss: 6.5510, average_loss: 4.9790
Updating best test loss: 4.71622
```

Starting epoch 3 / 50

Learning Rate for this epoch: 0.001

```
Epoch [3/50], Iter [5/209] Loss: 4.2510, average_loss: 4.2828
Epoch [3/50], Iter [10/209] Loss: 3.7881, average_loss: 4.7503
Epoch [3/50], Iter [15/209] Loss: 4.0520, average_loss: 4.6411
Epoch [3/50], Iter [20/209] Loss: 4.1567, average_loss: 4.4900
Epoch [3/50], Iter [25/209] Loss: 4.7384, average_loss: 4.6360
Epoch [3/50], Iter [30/209] Loss: 7.3717, average_loss: 4.7190
Epoch [3/50], Iter [35/209] Loss: 5.4184, average_loss: 4.7530
Epoch [3/50], Iter [40/209] Loss: 4.1932, average_loss: 4.7473
Epoch [3/50], Iter [45/209] Loss: 3.3596, average_loss: 4.6500
Epoch [3/50], Iter [50/209] Loss: 5.2978, average_loss: 4.6463
Epoch [3/50], Iter [55/209] Loss: 4.6098, average_loss: 4.5874
Epoch [3/50], Iter [60/209] Loss: 3.2867, average_loss: 4.5320
Epoch [3/50], Iter [65/209] Loss: 4.8738, average_loss: 4.5319
Epoch [3/50], Iter [70/209] Loss: 5.9411, average_loss: 4.5502
Epoch [3/50], Iter [75/209] Loss: 5.5786, average_loss: 4.5925
Epoch [3/50], Iter [80/209] Loss: 4.0639, average_loss: 4.5488
Epoch [3/50], Iter [85/209] Loss: 5.2923, average_loss: 4.5451
Epoch [3/50], Iter [90/209] Loss: 4.5910, average_loss: 4.5372
Epoch [3/50], Iter [95/209] Loss: 3.7759, average_loss: 4.5356
Epoch [3/50], Iter [100/209] Loss: 5.7173, average_loss: 4.5316
Epoch [3/50], Iter [105/209] Loss: 4.9295, average_loss: 4.5035
Epoch [3/50], Iter [110/209] Loss: 3.5564, average_loss: 4.4809
Epoch [3/50], Iter [115/209] Loss: 4.8215, average_loss: 4.4918
Epoch [3/50], Iter [120/209] Loss: 4.2682, average_loss: 4.5039
Epoch [3/50], Iter [125/209] Loss: 4.2608, average_loss: 4.4874
Epoch [3/50], Iter [130/209] Loss: 4.2390, average_loss: 4.4826
Epoch [3/50], Iter [135/209] Loss: 3.5837, average_loss: 4.4566
Epoch [3/50], Iter [140/209] Loss: 4.1869, average_loss: 4.4554
Epoch [3/50], Iter [145/209] Loss: 3.1940, average_loss: 4.4279
Epoch [3/50], Iter [150/209] Loss: 3.6210, average_loss: 4.4364
Epoch [3/50], Iter [155/209] Loss: 4.3236, average_loss: 4.4340
Epoch [3/50], Iter [160/209] Loss: 4.2704, average_loss: 4.4241
Epoch [3/50], Iter [165/209] Loss: 4.4166, average_loss: 4.4346
```

```
Epoch [3/50], Iter [170/209] Loss: 4.0302, average_loss: 4.4295
Epoch [3/50], Iter [175/209] Loss: 4.1991, average_loss: 4.4292
Epoch [3/50], Iter [180/209] Loss: 3.9791, average_loss: 4.4079
Epoch [3/50], Iter [185/209] Loss: 2.8662, average_loss: 4.4007
Epoch [3/50], Iter [190/209] Loss: 3.3967, average_loss: 4.4032
Epoch [3/50], Iter [195/209] Loss: 3.9747, average_loss: 4.3924
Epoch [3/50], Iter [200/209] Loss: 5.1175, average_loss: 4.3896
Epoch [3/50], Iter [205/209] Loss: 4.3706, average_loss: 4.3847
Updating best test loss: 4.33614
```

Starting epoch 4 / 50

Learning Rate for this epoch: 0.001

```
Epoch [4/50], Iter [5/209] Loss: 4.8308, average_loss: 4.4536
Epoch [4/50], Iter [10/209] Loss: 4.3085, average_loss: 4.1309
Epoch [4/50], Iter [15/209] Loss: 4.3552, average_loss: 4.1010
Epoch [4/50], Iter [20/209] Loss: 5.0847, average_loss: 4.3234
Epoch [4/50], Iter [25/209] Loss: 4.6215, average_loss: 4.3311
Epoch [4/50], Iter [30/209] Loss: 5.5088, average_loss: 4.2637
Epoch [4/50], Iter [35/209] Loss: 3.8998, average_loss: 4.2407
Epoch [4/50], Iter [40/209] Loss: 2.8818, average_loss: 4.1639
Epoch [4/50], Iter [45/209] Loss: 4.4147, average_loss: 4.1237
Epoch [4/50], Iter [50/209] Loss: 4.4875, average_loss: 4.1119
Epoch [4/50], Iter [55/209] Loss: 4.7082, average_loss: 4.1591
Epoch [4/50], Iter [60/209] Loss: 4.0063, average_loss: 4.1245
Epoch [4/50], Iter [65/209] Loss: 5.8385, average_loss: 4.1451
Epoch [4/50], Iter [70/209] Loss: 3.9083, average_loss: 4.1441
Epoch [4/50], Iter [75/209] Loss: 3.4412, average_loss: 4.1161
Epoch [4/50], Iter [80/209] Loss: 3.1741, average_loss: 4.0796
Epoch [4/50], Iter [85/209] Loss: 3.9580, average_loss: 4.0619
Epoch [4/50], Iter [90/209] Loss: 3.3619, average_loss: 4.0373
Epoch [4/50], Iter [95/209] Loss: 3.1913, average_loss: 4.0098
Epoch [4/50], Iter [100/209] Loss: 5.0439, average_loss: 4.0040
Epoch [4/50], Iter [105/209] Loss: 4.0273, average_loss: 4.0360
Epoch [4/50], Iter [110/209] Loss: 3.9892, average_loss: 4.0321
Epoch [4/50], Iter [115/209] Loss: 3.1455, average_loss: 4.0365
Epoch [4/50], Iter [120/209] Loss: 4.1416, average_loss: 4.0504
Epoch [4/50], Iter [125/209] Loss: 3.3566, average_loss: 4.0431
Epoch [4/50], Iter [130/209] Loss: 3.3309, average_loss: 4.0362
Epoch [4/50], Iter [135/209] Loss: 4.2380, average_loss: 4.0382
Epoch [4/50], Iter [140/209] Loss: 4.0570, average_loss: 4.0270
Epoch [4/50], Iter [145/209] Loss: 4.7024, average_loss: 4.0353
Epoch [4/50], Iter [150/209] Loss: 3.9324, average_loss: 4.0321
Epoch [4/50], Iter [155/209] Loss: 4.1950, average_loss: 4.0230
Epoch [4/50], Iter [160/209] Loss: 4.2990, average_loss: 4.0068
Epoch [4/50], Iter [165/209] Loss: 4.5277, average_loss: 4.0150
Epoch [4/50], Iter [170/209] Loss: 3.4699, average_loss: 4.0223
Epoch [4/50], Iter [175/209] Loss: 3.3837, average_loss: 4.0367
Epoch [4/50], Iter [180/209] Loss: 4.9498, average_loss: 4.0379
Epoch [4/50], Iter [185/209] Loss: 4.3561, average_loss: 4.0281
Epoch [4/50], Iter [190/209] Loss: 5.3701, average_loss: 4.0326
Epoch [4/50], Iter [195/209] Loss: 4.6434, average_loss: 4.0394
Epoch [4/50], Iter [200/209] Loss: 3.1058, average_loss: 4.0439
Epoch [4/50], Iter [205/209] Loss: 3.1722, average_loss: 4.0364
Updating best test loss: 4.02932
```

Starting epoch 5 / 50

Learning Rate for this epoch: 0.001

```
Epoch [5/50], Iter [5/209] Loss: 4.2286, average_loss: 3.5975
Epoch [5/50], Iter [10/209] Loss: 3.7163, average_loss: 3.6587
Epoch [5/50], Iter [15/209] Loss: 3.7840, average_loss: 3.6173
Epoch [5/50], Iter [20/209] Loss: 3.7693, average_loss: 3.8442
```



```
Epoch [5/50], Iter [25/209] Loss: 4.0434, average_loss: 3.8264
Epoch [5/50], Iter [30/209] Loss: 3.5959, average_loss: 3.7594
Epoch [5/50], Iter [35/209] Loss: 3.7514, average_loss: 3.8200
Epoch [5/50], Iter [40/209] Loss: 4.0966, average_loss: 3.8363
Epoch [5/50], Iter [45/209] Loss: 5.2197, average_loss: 3.8472
Epoch [5/50], Iter [50/209] Loss: 2.6431, average_loss: 3.7711
Epoch [5/50], Iter [55/209] Loss: 2.9846, average_loss: 3.7865
Epoch [5/50], Iter [60/209] Loss: 4.1678, average_loss: 3.7720
Epoch [5/50], Iter [65/209] Loss: 4.1300, average_loss: 3.7566
Epoch [5/50], Iter [70/209] Loss: 4.5069, average_loss: 3.7601
Epoch [5/50], Iter [75/209] Loss: 3.6258, average_loss: 3.7811
Epoch [5/50], Iter [80/209] Loss: 3.8153, average_loss: 3.7900
Epoch [5/50], Iter [85/209] Loss: 3.6753, average_loss: 3.7703
Epoch [5/50], Iter [90/209] Loss: 3.8519, average_loss: 3.7876
Epoch [5/50], Iter [95/209] Loss: 3.6094, average_loss: 3.7814
Epoch [5/50], Iter [100/209] Loss: 4.3850, average_loss: 3.7909
Epoch [5/50], Iter [105/209] Loss: 4.4008, average_loss: 3.8053
Epoch [5/50], Iter [110/209] Loss: 2.7901, average_loss: 3.7865
Epoch [5/50], Iter [115/209] Loss: 3.8882, average_loss: 3.7943
Epoch [5/50], Iter [120/209] Loss: 3.8666, average_loss: 3.7831
Epoch [5/50], Iter [125/209] Loss: 3.2460, average_loss: 3.7754
Epoch [5/50], Iter [130/209] Loss: 4.3867, average_loss: 3.7773
Epoch [5/50], Iter [135/209] Loss: 3.9485, average_loss: 3.7834
Epoch [5/50], Iter [140/209] Loss: 3.1708, average_loss: 3.7860
Epoch [5/50], Iter [145/209] Loss: 4.3565, average_loss: 3.7815
Epoch [5/50], Iter [150/209] Loss: 3.9212, average_loss: 3.7841
Epoch [5/50], Iter [155/209] Loss: 3.4939, average_loss: 3.7824
Epoch [5/50], Iter [160/209] Loss: 3.4429, average_loss: 3.7807
Epoch [5/50], Iter [165/209] Loss: 4.0883, average_loss: 3.7686
Epoch [5/50], Iter [170/209] Loss: 3.1729, average_loss: 3.7690
Epoch [5/50], Iter [175/209] Loss: 3.3363, average_loss: 3.7623
Epoch [5/50], Iter [180/209] Loss: 3.5355, average_loss: 3.7460
Epoch [5/50], Iter [185/209] Loss: 2.7672, average_loss: 3.7354
Epoch [5/50], Iter [190/209] Loss: 4.1735, average_loss: 3.7357
Epoch [5/50], Iter [195/209] Loss: 4.2278, average_loss: 3.7416
Epoch [5/50], Iter [200/209] Loss: 4.2621, average_loss: 3.7511
Epoch [5/50], Iter [205/209] Loss: 3.2930, average_loss: 3.7520
Updating best test loss: 3.79355
```

Starting epoch 6 / 50

Learning Rate for this epoch: 0.001

```
Epoch [6/50], Iter [5/209] Loss: 3.2417, average_loss: 3.9429
Epoch [6/50], Iter [10/209] Loss: 3.2395, average_loss: 3.6027
Epoch [6/50], Iter [15/209] Loss: 3.2355, average_loss: 3.6065
Epoch [6/50], Iter [20/209] Loss: 3.7704, average_loss: 3.5716
Epoch [6/50], Iter [25/209] Loss: 3.3122, average_loss: 3.5137
Epoch [6/50], Iter [30/209] Loss: 3.0889, average_loss: 3.5016
Epoch [6/50], Iter [35/209] Loss: 3.8345, average_loss: 3.5328
Epoch [6/50], Iter [40/209] Loss: 3.4180, average_loss: 3.5007
Epoch [6/50], Iter [45/209] Loss: 3.3743, average_loss: 3.5351
Epoch [6/50], Iter [50/209] Loss: 4.0248, average_loss: 3.5472
Epoch [6/50], Iter [55/209] Loss: 2.7181, average_loss: 3.5660
Epoch [6/50], Iter [60/209] Loss: 3.0990, average_loss: 3.5458
Epoch [6/50], Iter [65/209] Loss: 4.1809, average_loss: 3.5326
Epoch [6/50], Iter [70/209] Loss: 3.3028, average_loss: 3.5314
Epoch [6/50], Iter [75/209] Loss: 2.2546, average_loss: 3.5057
Epoch [6/50], Iter [80/209] Loss: 2.7868, average_loss: 3.4737
Epoch [6/50], Iter [85/209] Loss: 3.6839, average_loss: 3.4656
Epoch [6/50], Iter [90/209] Loss: 4.3112, average_loss: 3.4770
Epoch [6/50], Iter [95/209] Loss: 3.8921, average_loss: 3.4755
Epoch [6/50], Iter [100/209] Loss: 4.5224, average_loss: 3.4862
Epoch [6/50], Iter [105/209] Loss: 3.4146, average_loss: 3.4531
```

```
Epoch [6/50], Iter [110/209] Loss: 3.3720, average_loss: 3.4578
Epoch [6/50], Iter [115/209] Loss: 5.6564, average_loss: 3.4943
Epoch [6/50], Iter [120/209] Loss: 2.7703, average_loss: 3.4667
Epoch [6/50], Iter [125/209] Loss: 4.2197, average_loss: 3.4760
Epoch [6/50], Iter [130/209] Loss: 3.0285, average_loss: 3.4980
Epoch [6/50], Iter [135/209] Loss: 3.4806, average_loss: 3.5178
Epoch [6/50], Iter [140/209] Loss: 4.3321, average_loss: 3.5185
Epoch [6/50], Iter [145/209] Loss: 3.1905, average_loss: 3.5118
Epoch [6/50], Iter [150/209] Loss: 3.2057, average_loss: 3.5087
Epoch [6/50], Iter [155/209] Loss: 3.7412, average_loss: 3.5052
Epoch [6/50], Iter [160/209] Loss: 3.8649, average_loss: 3.4965
Epoch [6/50], Iter [165/209] Loss: 4.2436, average_loss: 3.4939
Epoch [6/50], Iter [170/209] Loss: 3.7672, average_loss: 3.5066
Epoch [6/50], Iter [175/209] Loss: 3.5593, average_loss: 3.5158
Epoch [6/50], Iter [180/209] Loss: 4.2045, average_loss: 3.5306
Epoch [6/50], Iter [185/209] Loss: 2.7621, average_loss: 3.5325
Epoch [6/50], Iter [190/209] Loss: 3.3871, average_loss: 3.5309
Epoch [6/50], Iter [195/209] Loss: 3.8905, average_loss: 3.5337
Epoch [6/50], Iter [200/209] Loss: 3.7412, average_loss: 3.5306
Epoch [6/50], Iter [205/209] Loss: 4.1136, average_loss: 3.5273
Updating best test loss: 3.62761
```

Starting epoch 7 / 50

Learning Rate for this epoch: 0.001

```
Epoch [7/50], Iter [5/209] Loss: 3.7764, average_loss: 3.4866
Epoch [7/50], Iter [10/209] Loss: 4.2146, average_loss: 3.4320
Epoch [7/50], Iter [15/209] Loss: 2.3044, average_loss: 3.3283
Epoch [7/50], Iter [20/209] Loss: 2.8724, average_loss: 3.3773
Epoch [7/50], Iter [25/209] Loss: 3.2586, average_loss: 3.3676
Epoch [7/50], Iter [30/209] Loss: 2.8003, average_loss: 3.3823
Epoch [7/50], Iter [35/209] Loss: 3.4825, average_loss: 3.3302
Epoch [7/50], Iter [40/209] Loss: 3.2734, average_loss: 3.2978
Epoch [7/50], Iter [45/209] Loss: 3.1847, average_loss: 3.3437
Epoch [7/50], Iter [50/209] Loss: 3.5425, average_loss: 3.3308
Epoch [7/50], Iter [55/209] Loss: 2.9310, average_loss: 3.3498
Epoch [7/50], Iter [60/209] Loss: 4.1835, average_loss: 3.3603
Epoch [7/50], Iter [65/209] Loss: 3.4492, average_loss: 3.3560
Epoch [7/50], Iter [70/209] Loss: 3.5269, average_loss: 3.3620
Epoch [7/50], Iter [75/209] Loss: 2.5875, average_loss: 3.3388
Epoch [7/50], Iter [80/209] Loss: 5.1009, average_loss: 3.3513
Epoch [7/50], Iter [85/209] Loss: 4.0816, average_loss: 3.3902
Epoch [7/50], Iter [90/209] Loss: 3.8160, average_loss: 3.4206
Epoch [7/50], Iter [95/209] Loss: 2.9784, average_loss: 3.4251
Epoch [7/50], Iter [100/209] Loss: 2.8163, average_loss: 3.4027
Epoch [7/50], Iter [105/209] Loss: 3.4530, average_loss: 3.4070
Epoch [7/50], Iter [110/209] Loss: 3.3513, average_loss: 3.3993
Epoch [7/50], Iter [115/209] Loss: 3.4215, average_loss: 3.3968
Epoch [7/50], Iter [120/209] Loss: 3.5582, average_loss: 3.3999
Epoch [7/50], Iter [125/209] Loss: 2.7668, average_loss: 3.3836
Epoch [7/50], Iter [130/209] Loss: 5.0026, average_loss: 3.3752
Epoch [7/50], Iter [135/209] Loss: 2.3379, average_loss: 3.3647
Epoch [7/50], Iter [140/209] Loss: 3.0208, average_loss: 3.3529
Epoch [7/50], Iter [145/209] Loss: 3.0849, average_loss: 3.3592
Epoch [7/50], Iter [150/209] Loss: 2.6885, average_loss: 3.3639
Epoch [7/50], Iter [155/209] Loss: 2.1763, average_loss: 3.3457
Epoch [7/50], Iter [160/209] Loss: 2.3535, average_loss: 3.3512
Epoch [7/50], Iter [165/209] Loss: 4.0253, average_loss: 3.3529
Epoch [7/50], Iter [170/209] Loss: 3.1098, average_loss: 3.3598
Epoch [7/50], Iter [175/209] Loss: 3.0293, average_loss: 3.3632
Epoch [7/50], Iter [180/209] Loss: 3.3399, average_loss: 3.3627
Epoch [7/50], Iter [185/209] Loss: 3.0410, average_loss: 3.3558
Epoch [7/50], Iter [190/209] Loss: 3.4559, average_loss: 3.3503
```

Epoch [7/50], Iter [195/209] Loss: 3.7710, average_loss: 3.3460
Epoch [7/50], Iter [200/209] Loss: 2.8825, average_loss: 3.3432
Epoch [7/50], Iter [205/209] Loss: 2.9146, average_loss: 3.3443
Updating best test loss: 3.48491

Starting epoch 8 / 50

Learning Rate for this epoch: 0.001

Epoch [8/50], Iter [5/209] Loss: 2.7247, average_loss: 3.0141
Epoch [8/50], Iter [10/209] Loss: 3.2746, average_loss: 3.2993
Epoch [8/50], Iter [15/209] Loss: 3.9682, average_loss: 3.2619
Epoch [8/50], Iter [20/209] Loss: 2.6549, average_loss: 3.2285
Epoch [8/50], Iter [25/209] Loss: 2.9140, average_loss: 3.2289
Epoch [8/50], Iter [30/209] Loss: 3.1713, average_loss: 3.2563
Epoch [8/50], Iter [35/209] Loss: 3.1887, average_loss: 3.2462
Epoch [8/50], Iter [40/209] Loss: 2.5979, average_loss: 3.2222
Epoch [8/50], Iter [45/209] Loss: 3.6351, average_loss: 3.2591
Epoch [8/50], Iter [50/209] Loss: 3.5404, average_loss: 3.2660
Epoch [8/50], Iter [55/209] Loss: 3.8029, average_loss: 3.2669
Epoch [8/50], Iter [60/209] Loss: 4.5177, average_loss: 3.2712
Epoch [8/50], Iter [65/209] Loss: 3.2520, average_loss: 3.2795
Epoch [8/50], Iter [70/209] Loss: 3.0063, average_loss: 3.2698
Epoch [8/50], Iter [75/209] Loss: 2.9354, average_loss: 3.2507
Epoch [8/50], Iter [80/209] Loss: 2.7077, average_loss: 3.2320
Epoch [8/50], Iter [85/209] Loss: 4.0759, average_loss: 3.2262
Epoch [8/50], Iter [90/209] Loss: 2.5592, average_loss: 3.2360
Epoch [8/50], Iter [95/209] Loss: 3.6591, average_loss: 3.2355
Epoch [8/50], Iter [100/209] Loss: 2.6741, average_loss: 3.2040
Epoch [8/50], Iter [105/209] Loss: 3.0515, average_loss: 3.1929
Epoch [8/50], Iter [110/209] Loss: 1.9546, average_loss: 3.1791
Epoch [8/50], Iter [115/209] Loss: 3.3370, average_loss: 3.1816
Epoch [8/50], Iter [120/209] Loss: 2.8233, average_loss: 3.1718
Epoch [8/50], Iter [125/209] Loss: 3.2038, average_loss: 3.1940
Epoch [8/50], Iter [130/209] Loss: 2.7577, average_loss: 3.1933
Epoch [8/50], Iter [135/209] Loss: 2.9078, average_loss: 3.1959
Epoch [8/50], Iter [140/209] Loss: 3.8532, average_loss: 3.1900
Epoch [8/50], Iter [145/209] Loss: 3.3613, average_loss: 3.1974
Epoch [8/50], Iter [150/209] Loss: 2.7442, average_loss: 3.2031
Epoch [8/50], Iter [155/209] Loss: 2.7339, average_loss: 3.1975
Epoch [8/50], Iter [160/209] Loss: 2.7636, average_loss: 3.2102
Epoch [8/50], Iter [165/209] Loss: 2.8854, average_loss: 3.2148
Epoch [8/50], Iter [170/209] Loss: 3.5979, average_loss: 3.2159
Epoch [8/50], Iter [175/209] Loss: 2.5590, average_loss: 3.2310
Epoch [8/50], Iter [180/209] Loss: 2.6433, average_loss: 3.2241
Epoch [8/50], Iter [185/209] Loss: 3.5211, average_loss: 3.2241
Epoch [8/50], Iter [190/209] Loss: 2.6741, average_loss: 3.2299
Epoch [8/50], Iter [195/209] Loss: 2.4623, average_loss: 3.2256
Epoch [8/50], Iter [200/209] Loss: 3.5844, average_loss: 3.2252
Epoch [8/50], Iter [205/209] Loss: 3.6512, average_loss: 3.2303
Updating best test loss: 3.39725

Starting epoch 9 / 50

Learning Rate for this epoch: 0.001

Epoch [9/50], Iter [5/209] Loss: 3.6419, average_loss: 3.1506
Epoch [9/50], Iter [10/209] Loss: 2.4120, average_loss: 2.9987
Epoch [9/50], Iter [15/209] Loss: 2.9670, average_loss: 3.0037
Epoch [9/50], Iter [20/209] Loss: 1.9429, average_loss: 2.9613
Epoch [9/50], Iter [25/209] Loss: 2.1695, average_loss: 2.9280
Epoch [9/50], Iter [30/209] Loss: 3.0258, average_loss: 2.9520
Epoch [9/50], Iter [35/209] Loss: 3.1514, average_loss: 3.0130
Epoch [9/50], Iter [40/209] Loss: 3.2988, average_loss: 3.0307
Epoch [9/50], Iter [45/209] Loss: 2.5062, average_loss: 3.0163

```
Epoch [9/50], Iter [50/209] Loss: 2.8370, average_loss: 3.0007
Epoch [9/50], Iter [55/209] Loss: 3.3826, average_loss: 2.9998
Epoch [9/50], Iter [60/209] Loss: 3.2194, average_loss: 3.0018
Epoch [9/50], Iter [65/209] Loss: 2.8806, average_loss: 3.0039
Epoch [9/50], Iter [70/209] Loss: 3.9045, average_loss: 3.0345
Epoch [9/50], Iter [75/209] Loss: 3.5233, average_loss: 3.0256
Epoch [9/50], Iter [80/209] Loss: 2.6539, average_loss: 3.0273
Epoch [9/50], Iter [85/209] Loss: 4.7734, average_loss: 3.0699
Epoch [9/50], Iter [90/209] Loss: 2.9903, average_loss: 3.0831
Epoch [9/50], Iter [95/209] Loss: 3.4349, average_loss: 3.0920
Epoch [9/50], Iter [100/209] Loss: 4.2557, average_loss: 3.1099
Epoch [9/50], Iter [105/209] Loss: 3.9541, average_loss: 3.1186
Epoch [9/50], Iter [110/209] Loss: 3.1880, average_loss: 3.1194
Epoch [9/50], Iter [115/209] Loss: 2.8120, average_loss: 3.1159
Epoch [9/50], Iter [120/209] Loss: 2.6236, average_loss: 3.1080
Epoch [9/50], Iter [125/209] Loss: 3.1871, average_loss: 3.1191
Epoch [9/50], Iter [130/209] Loss: 2.5909, average_loss: 3.1112
Epoch [9/50], Iter [135/209] Loss: 2.7579, average_loss: 3.1027
Epoch [9/50], Iter [140/209] Loss: 3.1690, average_loss: 3.0973
Epoch [9/50], Iter [145/209] Loss: 3.3594, average_loss: 3.0911
Epoch [9/50], Iter [150/209] Loss: 2.5756, average_loss: 3.0886
Epoch [9/50], Iter [155/209] Loss: 2.8269, average_loss: 3.0872
Epoch [9/50], Iter [160/209] Loss: 3.5108, average_loss: 3.0844
Epoch [9/50], Iter [165/209] Loss: 2.5795, average_loss: 3.0767
Epoch [9/50], Iter [170/209] Loss: 3.5090, average_loss: 3.0775
Epoch [9/50], Iter [175/209] Loss: 3.3015, average_loss: 3.0865
Epoch [9/50], Iter [180/209] Loss: 3.4996, average_loss: 3.0916
Epoch [9/50], Iter [185/209] Loss: 3.1742, average_loss: 3.0977
Epoch [9/50], Iter [190/209] Loss: 2.3591, average_loss: 3.0962
Epoch [9/50], Iter [195/209] Loss: 2.3748, average_loss: 3.0933
Epoch [9/50], Iter [200/209] Loss: 2.6671, average_loss: 3.0892
Epoch [9/50], Iter [205/209] Loss: 3.1259, average_loss: 3.0847
Updating best test loss: 3.31746
```

Starting epoch 10 / 50

Learning Rate for this epoch: 0.001

```
Epoch [10/50], Iter [5/209] Loss: 2.1066, average_loss: 2.5789
Epoch [10/50], Iter [10/209] Loss: 3.4061, average_loss: 2.7167
Epoch [10/50], Iter [15/209] Loss: 2.9329, average_loss: 2.7640
Epoch [10/50], Iter [20/209] Loss: 2.6336, average_loss: 2.8176
Epoch [10/50], Iter [25/209] Loss: 2.8605, average_loss: 2.9380
Epoch [10/50], Iter [30/209] Loss: 2.6780, average_loss: 2.8806
Epoch [10/50], Iter [35/209] Loss: 2.7517, average_loss: 2.8872
Epoch [10/50], Iter [40/209] Loss: 3.6110, average_loss: 2.9347
Epoch [10/50], Iter [45/209] Loss: 2.3122, average_loss: 2.9116
Epoch [10/50], Iter [50/209] Loss: 2.1878, average_loss: 2.8850
Epoch [10/50], Iter [55/209] Loss: 2.8935, average_loss: 2.8667
Epoch [10/50], Iter [60/209] Loss: 3.5919, average_loss: 2.8944
Epoch [10/50], Iter [65/209] Loss: 2.7702, average_loss: 2.8985
Epoch [10/50], Iter [70/209] Loss: 3.5401, average_loss: 2.9200
Epoch [10/50], Iter [75/209] Loss: 3.0922, average_loss: 2.9124
Epoch [10/50], Iter [80/209] Loss: 2.3474, average_loss: 2.9101
Epoch [10/50], Iter [85/209] Loss: 2.4049, average_loss: 2.9208
Epoch [10/50], Iter [90/209] Loss: 2.6017, average_loss: 2.9255
Epoch [10/50], Iter [95/209] Loss: 2.2825, average_loss: 2.9290
Epoch [10/50], Iter [100/209] Loss: 3.3437, average_loss: 2.9558
Epoch [10/50], Iter [105/209] Loss: 3.0023, average_loss: 2.9563
Epoch [10/50], Iter [110/209] Loss: 2.8535, average_loss: 2.9547
Epoch [10/50], Iter [115/209] Loss: 2.9278, average_loss: 2.9398
Epoch [10/50], Iter [120/209] Loss: 3.4405, average_loss: 2.9335
Epoch [10/50], Iter [125/209] Loss: 2.3654, average_loss: 2.9340
Epoch [10/50], Iter [130/209] Loss: 2.8605, average_loss: 2.9311
```

Epoch [10/50], Iter [135/209] Loss: 3.0880, average_loss: 2.9336
Epoch [10/50], Iter [140/209] Loss: 3.8389, average_loss: 2.9336
Epoch [10/50], Iter [145/209] Loss: 3.2335, average_loss: 2.9381
Epoch [10/50], Iter [150/209] Loss: 3.2215, average_loss: 2.9427
Epoch [10/50], Iter [155/209] Loss: 2.6750, average_loss: 2.9445
Epoch [10/50], Iter [160/209] Loss: 2.8121, average_loss: 2.9412
Epoch [10/50], Iter [165/209] Loss: 3.7971, average_loss: 2.9463
Epoch [10/50], Iter [170/209] Loss: 3.1208, average_loss: 2.9468
Epoch [10/50], Iter [175/209] Loss: 4.2280, average_loss: 2.9510
Epoch [10/50], Iter [180/209] Loss: 2.4037, average_loss: 2.9500
Epoch [10/50], Iter [185/209] Loss: 2.2110, average_loss: 2.9453
Epoch [10/50], Iter [190/209] Loss: 1.8004, average_loss: 2.9448
Epoch [10/50], Iter [195/209] Loss: 2.8186, average_loss: 2.9425
Epoch [10/50], Iter [200/209] Loss: 2.6130, average_loss: 2.9459
Epoch [10/50], Iter [205/209] Loss: 4.8008, average_loss: 2.9546
Updating best test loss: 3.24361

Starting epoch 11 / 50

Learning Rate for this epoch: 0.001

Epoch [11/50], Iter [5/209] Loss: 2.7623, average_loss: 2.9701
Epoch [11/50], Iter [10/209] Loss: 2.9115, average_loss: 2.9626
Epoch [11/50], Iter [15/209] Loss: 2.4943, average_loss: 2.9419
Epoch [11/50], Iter [20/209] Loss: 2.5866, average_loss: 2.8430
Epoch [11/50], Iter [25/209] Loss: 3.4392, average_loss: 2.9078
Epoch [11/50], Iter [30/209] Loss: 2.7034, average_loss: 2.9191
Epoch [11/50], Iter [35/209] Loss: 3.5571, average_loss: 2.9049
Epoch [11/50], Iter [40/209] Loss: 2.6341, average_loss: 2.8887
Epoch [11/50], Iter [45/209] Loss: 2.8429, average_loss: 2.9556
Epoch [11/50], Iter [50/209] Loss: 3.5225, average_loss: 2.9538
Epoch [11/50], Iter [55/209] Loss: 2.6103, average_loss: 2.9200
Epoch [11/50], Iter [60/209] Loss: 3.6360, average_loss: 2.9345
Epoch [11/50], Iter [65/209] Loss: 2.0395, average_loss: 2.9002
Epoch [11/50], Iter [70/209] Loss: 2.4973, average_loss: 2.9086
Epoch [11/50], Iter [75/209] Loss: 2.0238, average_loss: 2.8908
Epoch [11/50], Iter [80/209] Loss: 2.3763, average_loss: 2.8888
Epoch [11/50], Iter [85/209] Loss: 2.3564, average_loss: 2.8607
Epoch [11/50], Iter [90/209] Loss: 3.1341, average_loss: 2.8582
Epoch [11/50], Iter [95/209] Loss: 3.8338, average_loss: 2.8572
Epoch [11/50], Iter [100/209] Loss: 2.7583, average_loss: 2.8493
Epoch [11/50], Iter [105/209] Loss: 2.1188, average_loss: 2.8475
Epoch [11/50], Iter [110/209] Loss: 2.6237, average_loss: 2.8564
Epoch [11/50], Iter [115/209] Loss: 3.7817, average_loss: 2.8783
Epoch [11/50], Iter [120/209] Loss: 3.6806, average_loss: 2.9001
Epoch [11/50], Iter [125/209] Loss: 2.7813, average_loss: 2.8947
Epoch [11/50], Iter [130/209] Loss: 1.9869, average_loss: 2.8865
Epoch [11/50], Iter [135/209] Loss: 3.0376, average_loss: 2.8821
Epoch [11/50], Iter [140/209] Loss: 3.1397, average_loss: 2.8736
Epoch [11/50], Iter [145/209] Loss: 1.8535, average_loss: 2.8686
Epoch [11/50], Iter [150/209] Loss: 2.4507, average_loss: 2.8739
Epoch [11/50], Iter [155/209] Loss: 4.0439, average_loss: 2.8805
Epoch [11/50], Iter [160/209] Loss: 2.2760, average_loss: 2.8719
Epoch [11/50], Iter [165/209] Loss: 2.9665, average_loss: 2.8700
Epoch [11/50], Iter [170/209] Loss: 2.5429, average_loss: 2.8683
Epoch [11/50], Iter [175/209] Loss: 2.1436, average_loss: 2.8745
Epoch [11/50], Iter [180/209] Loss: 2.3896, average_loss: 2.8753
Epoch [11/50], Iter [185/209] Loss: 2.1917, average_loss: 2.8693
Epoch [11/50], Iter [190/209] Loss: 4.0165, average_loss: 2.8710
Epoch [11/50], Iter [195/209] Loss: 3.6587, average_loss: 2.8741
Epoch [11/50], Iter [200/209] Loss: 2.6025, average_loss: 2.8871
Epoch [11/50], Iter [205/209] Loss: 2.9681, average_loss: 2.8884
Updating best test loss: 3.21302

Starting epoch 12 / 50

Learning Rate for this epoch: 0.001

Epoch [12/50], Iter [5/209] Loss: 3.2522, average_loss: 2.6714
Epoch [12/50], Iter [10/209] Loss: 3.5193, average_loss: 2.6474
Epoch [12/50], Iter [15/209] Loss: 3.6592, average_loss: 2.7422
Epoch [12/50], Iter [20/209] Loss: 2.8004, average_loss: 2.8000
Epoch [12/50], Iter [25/209] Loss: 2.7248, average_loss: 2.8085
Epoch [12/50], Iter [30/209] Loss: 3.2256, average_loss: 2.8135
Epoch [12/50], Iter [35/209] Loss: 2.7463, average_loss: 2.8019
Epoch [12/50], Iter [40/209] Loss: 2.6780, average_loss: 2.8289
Epoch [12/50], Iter [45/209] Loss: 2.9085, average_loss: 2.8583
Epoch [12/50], Iter [50/209] Loss: 3.2003, average_loss: 2.8897
Epoch [12/50], Iter [55/209] Loss: 3.7776, average_loss: 2.8719
Epoch [12/50], Iter [60/209] Loss: 3.3831, average_loss: 2.8871
Epoch [12/50], Iter [65/209] Loss: 3.4262, average_loss: 2.8924
Epoch [12/50], Iter [70/209] Loss: 2.7600, average_loss: 2.8754
Epoch [12/50], Iter [75/209] Loss: 3.9723, average_loss: 2.9062
Epoch [12/50], Iter [80/209] Loss: 3.8376, average_loss: 2.9355
Epoch [12/50], Iter [85/209] Loss: 2.7214, average_loss: 2.9346
Epoch [12/50], Iter [90/209] Loss: 4.2404, average_loss: 2.9401
Epoch [12/50], Iter [95/209] Loss: 2.3453, average_loss: 2.9195
Epoch [12/50], Iter [100/209] Loss: 2.5682, average_loss: 2.9156
Epoch [12/50], Iter [105/209] Loss: 3.8433, average_loss: 2.9084
Epoch [12/50], Iter [110/209] Loss: 2.2589, average_loss: 2.8889
Epoch [12/50], Iter [115/209] Loss: 3.2513, average_loss: 2.8913
Epoch [12/50], Iter [120/209] Loss: 2.5879, average_loss: 2.9003
Epoch [12/50], Iter [125/209] Loss: 2.8008, average_loss: 2.8975
Epoch [12/50], Iter [130/209] Loss: 2.9945, average_loss: 2.8898
Epoch [12/50], Iter [135/209] Loss: 2.7346, average_loss: 2.8917
Epoch [12/50], Iter [140/209] Loss: 2.4364, average_loss: 2.8886
Epoch [12/50], Iter [145/209] Loss: 3.7577, average_loss: 2.8823
Epoch [12/50], Iter [150/209] Loss: 2.8199, average_loss: 2.8820
Epoch [12/50], Iter [155/209] Loss: 3.0532, average_loss: 2.8787
Epoch [12/50], Iter [160/209] Loss: 2.3465, average_loss: 2.8735
Epoch [12/50], Iter [165/209] Loss: 2.7517, average_loss: 2.8550
Epoch [12/50], Iter [170/209] Loss: 2.6579, average_loss: 2.8594
Epoch [12/50], Iter [175/209] Loss: 3.2328, average_loss: 2.8581
Epoch [12/50], Iter [180/209] Loss: 4.6947, average_loss: 2.8646
Epoch [12/50], Iter [185/209] Loss: 2.7174, average_loss: 2.8611
Epoch [12/50], Iter [190/209] Loss: 2.8563, average_loss: 2.8557
Epoch [12/50], Iter [195/209] Loss: 1.9867, average_loss: 2.8463
Epoch [12/50], Iter [200/209] Loss: 3.3155, average_loss: 2.8497
Epoch [12/50], Iter [205/209] Loss: 2.1859, average_loss: 2.8491
Updating best test loss: 3.14045

Starting epoch 13 / 50

Learning Rate for this epoch: 0.001

Epoch [13/50], Iter [5/209] Loss: 2.3336, average_loss: 2.6528
Epoch [13/50], Iter [10/209] Loss: 2.2153, average_loss: 2.5490
Epoch [13/50], Iter [15/209] Loss: 3.0941, average_loss: 2.6237
Epoch [13/50], Iter [20/209] Loss: 2.8285, average_loss: 2.5642
Epoch [13/50], Iter [25/209] Loss: 2.1414, average_loss: 2.6610
Epoch [13/50], Iter [30/209] Loss: 2.4244, average_loss: 2.6913
Epoch [13/50], Iter [35/209] Loss: 2.4572, average_loss: 2.7379
Epoch [13/50], Iter [40/209] Loss: 2.6618, average_loss: 2.7015
Epoch [13/50], Iter [45/209] Loss: 3.8437, average_loss: 2.7145
Epoch [13/50], Iter [50/209] Loss: 2.8830, average_loss: 2.7154
Epoch [13/50], Iter [55/209] Loss: 3.3147, average_loss: 2.7580
Epoch [13/50], Iter [60/209] Loss: 2.6879, average_loss: 2.7204
Epoch [13/50], Iter [65/209] Loss: 2.6973, average_loss: 2.7285
Epoch [13/50], Iter [70/209] Loss: 2.7094, average_loss: 2.7358

```
Epoch [13/50], Iter [75/209] Loss: 2.1848, average_loss: 2.7456
Epoch [13/50], Iter [80/209] Loss: 2.1822, average_loss: 2.7373
Epoch [13/50], Iter [85/209] Loss: 3.0415, average_loss: 2.7373
Epoch [13/50], Iter [90/209] Loss: 2.4664, average_loss: 2.7462
Epoch [13/50], Iter [95/209] Loss: 3.3276, average_loss: 2.7498
Epoch [13/50], Iter [100/209] Loss: 2.1326, average_loss: 2.7262
Epoch [13/50], Iter [105/209] Loss: 2.7395, average_loss: 2.7256
Epoch [13/50], Iter [110/209] Loss: 2.5962, average_loss: 2.7283
Epoch [13/50], Iter [115/209] Loss: 2.4597, average_loss: 2.7246
Epoch [13/50], Iter [120/209] Loss: 2.7665, average_loss: 2.7278
Epoch [13/50], Iter [125/209] Loss: 2.8868, average_loss: 2.7082
Epoch [13/50], Iter [130/209] Loss: 2.9673, average_loss: 2.7292
Epoch [13/50], Iter [135/209] Loss: 2.9972, average_loss: 2.7321
Epoch [13/50], Iter [140/209] Loss: 2.9849, average_loss: 2.7272
Epoch [13/50], Iter [145/209] Loss: 2.6909, average_loss: 2.7315
Epoch [13/50], Iter [150/209] Loss: 2.4444, average_loss: 2.7209
Epoch [13/50], Iter [155/209] Loss: 2.2744, average_loss: 2.7214
Epoch [13/50], Iter [160/209] Loss: 2.6701, average_loss: 2.7219
Epoch [13/50], Iter [165/209] Loss: 2.6996, average_loss: 2.7177
Epoch [13/50], Iter [170/209] Loss: 2.4548, average_loss: 2.7233
Epoch [13/50], Iter [175/209] Loss: 2.4442, average_loss: 2.7237
Epoch [13/50], Iter [180/209] Loss: 2.8621, average_loss: 2.7250
Epoch [13/50], Iter [185/209] Loss: 3.0606, average_loss: 2.7395
Epoch [13/50], Iter [190/209] Loss: 2.4884, average_loss: 2.7400
Epoch [13/50], Iter [195/209] Loss: 2.3671, average_loss: 2.7344
Epoch [13/50], Iter [200/209] Loss: 2.5651, average_loss: 2.7390
Epoch [13/50], Iter [205/209] Loss: 2.6711, average_loss: 2.7349
```

Starting epoch 14 / 50

Learning Rate for this epoch: 0.001

```
Epoch [14/50], Iter [5/209] Loss: 2.8352, average_loss: 2.6350
Epoch [14/50], Iter [10/209] Loss: 1.8212, average_loss: 2.7733
Epoch [14/50], Iter [15/209] Loss: 2.9329, average_loss: 2.7749
Epoch [14/50], Iter [20/209] Loss: 2.8306, average_loss: 2.7600
Epoch [14/50], Iter [25/209] Loss: 2.8935, average_loss: 2.7627
Epoch [14/50], Iter [30/209] Loss: 2.4122, average_loss: 2.7805
Epoch [14/50], Iter [35/209] Loss: 2.1662, average_loss: 2.7541
Epoch [14/50], Iter [40/209] Loss: 2.4322, average_loss: 2.7420
Epoch [14/50], Iter [45/209] Loss: 2.7733, average_loss: 2.7411
Epoch [14/50], Iter [50/209] Loss: 3.0223, average_loss: 2.7621
Epoch [14/50], Iter [55/209] Loss: 2.7598, average_loss: 2.7474
Epoch [14/50], Iter [60/209] Loss: 2.2055, average_loss: 2.7630
Epoch [14/50], Iter [65/209] Loss: 3.0157, average_loss: 2.7688
Epoch [14/50], Iter [70/209] Loss: 1.8937, average_loss: 2.7296
Epoch [14/50], Iter [75/209] Loss: 2.4929, average_loss: 2.7397
Epoch [14/50], Iter [80/209] Loss: 3.0460, average_loss: 2.7174
Epoch [14/50], Iter [85/209] Loss: 2.3011, average_loss: 2.7059
Epoch [14/50], Iter [90/209] Loss: 1.6893, average_loss: 2.7045
Epoch [14/50], Iter [95/209] Loss: 2.1597, average_loss: 2.6928
Epoch [14/50], Iter [100/209] Loss: 3.1667, average_loss: 2.6854
Epoch [14/50], Iter [105/209] Loss: 2.3382, average_loss: 2.6806
Epoch [14/50], Iter [110/209] Loss: 2.2616, average_loss: 2.6768
Epoch [14/50], Iter [115/209] Loss: 3.2818, average_loss: 2.6772
Epoch [14/50], Iter [120/209] Loss: 2.4439, average_loss: 2.6842
Epoch [14/50], Iter [125/209] Loss: 2.2272, average_loss: 2.6952
Epoch [14/50], Iter [130/209] Loss: 2.4701, average_loss: 2.6967
Epoch [14/50], Iter [135/209] Loss: 3.2274, average_loss: 2.6961
Epoch [14/50], Iter [140/209] Loss: 3.0204, average_loss: 2.6990
Epoch [14/50], Iter [145/209] Loss: 2.8360, average_loss: 2.6999
Epoch [14/50], Iter [150/209] Loss: 2.7585, average_loss: 2.6978
Epoch [14/50], Iter [155/209] Loss: 2.6395, average_loss: 2.6973
Epoch [14/50], Iter [160/209] Loss: 2.4745, average_loss: 2.6961
```

Epoch [14/50], Iter [165/209] Loss: 3.9632, average_loss: 2.7036
Epoch [14/50], Iter [170/209] Loss: 3.8255, average_loss: 2.7124
Epoch [14/50], Iter [175/209] Loss: 2.2019, average_loss: 2.7139
Epoch [14/50], Iter [180/209] Loss: 2.6898, average_loss: 2.7091
Epoch [14/50], Iter [185/209] Loss: 2.2905, average_loss: 2.7133
Epoch [14/50], Iter [190/209] Loss: 2.4864, average_loss: 2.7042
Epoch [14/50], Iter [195/209] Loss: 2.7831, average_loss: 2.7023
Epoch [14/50], Iter [200/209] Loss: 2.3154, average_loss: 2.6934
Epoch [14/50], Iter [205/209] Loss: 2.6020, average_loss: 2.6919
Updating best test loss: 3.07413

Starting epoch 15 / 50

Learning Rate for this epoch: 0.001

Epoch [15/50], Iter [5/209] Loss: 3.8652, average_loss: 2.8639
Epoch [15/50], Iter [10/209] Loss: 2.1379, average_loss: 2.7146
Epoch [15/50], Iter [15/209] Loss: 2.5134, average_loss: 2.6141
Epoch [15/50], Iter [20/209] Loss: 1.9502, average_loss: 2.5424
Epoch [15/50], Iter [25/209] Loss: 3.0277, average_loss: 2.5636
Epoch [15/50], Iter [30/209] Loss: 2.3227, average_loss: 2.5707
Epoch [15/50], Iter [35/209] Loss: 2.6571, average_loss: 2.5569
Epoch [15/50], Iter [40/209] Loss: 2.4565, average_loss: 2.5726
Epoch [15/50], Iter [45/209] Loss: 2.9777, average_loss: 2.5832
Epoch [15/50], Iter [50/209] Loss: 2.6780, average_loss: 2.5825
Epoch [15/50], Iter [55/209] Loss: 1.8433, average_loss: 2.5889
Epoch [15/50], Iter [60/209] Loss: 3.1452, average_loss: 2.6045
Epoch [15/50], Iter [65/209] Loss: 2.3859, average_loss: 2.6009
Epoch [15/50], Iter [70/209] Loss: 2.2629, average_loss: 2.5839
Epoch [15/50], Iter [75/209] Loss: 2.5558, average_loss: 2.5861
Epoch [15/50], Iter [80/209] Loss: 3.3056, average_loss: 2.5906
Epoch [15/50], Iter [85/209] Loss: 2.7856, average_loss: 2.6105
Epoch [15/50], Iter [90/209] Loss: 1.9021, average_loss: 2.6259
Epoch [15/50], Iter [95/209] Loss: 3.9521, average_loss: 2.6238
Epoch [15/50], Iter [100/209] Loss: 2.1278, average_loss: 2.6185
Epoch [15/50], Iter [105/209] Loss: 2.0039, average_loss: 2.6086
Epoch [15/50], Iter [110/209] Loss: 2.3665, average_loss: 2.6047
Epoch [15/50], Iter [115/209] Loss: 2.5490, average_loss: 2.6026
Epoch [15/50], Iter [120/209] Loss: 2.7039, average_loss: 2.6029
Epoch [15/50], Iter [125/209] Loss: 2.7958, average_loss: 2.6069
Epoch [15/50], Iter [130/209] Loss: 2.6141, average_loss: 2.6103
Epoch [15/50], Iter [135/209] Loss: 2.5474, average_loss: 2.6052
Epoch [15/50], Iter [140/209] Loss: 2.6331, average_loss: 2.6053
Epoch [15/50], Iter [145/209] Loss: 2.8050, average_loss: 2.6033
Epoch [15/50], Iter [150/209] Loss: 3.1751, average_loss: 2.6089
Epoch [15/50], Iter [155/209] Loss: 2.1801, average_loss: 2.6060
Epoch [15/50], Iter [160/209] Loss: 2.8383, average_loss: 2.6066
Epoch [15/50], Iter [165/209] Loss: 2.8507, average_loss: 2.6072
Epoch [15/50], Iter [170/209] Loss: 2.2505, average_loss: 2.6065
Epoch [15/50], Iter [175/209] Loss: 2.3865, average_loss: 2.6121
Epoch [15/50], Iter [180/209] Loss: 3.1578, average_loss: 2.6053
Epoch [15/50], Iter [185/209] Loss: 4.0583, average_loss: 2.6207
Epoch [15/50], Iter [190/209] Loss: 2.4820, average_loss: 2.6245
Epoch [15/50], Iter [195/209] Loss: 2.4932, average_loss: 2.6336
Epoch [15/50], Iter [200/209] Loss: 4.0582, average_loss: 2.6309
Epoch [15/50], Iter [205/209] Loss: 2.5041, average_loss: 2.6285
Updating best test loss: 3.03580

Starting epoch 16 / 50

Learning Rate for this epoch: 0.001

Epoch [16/50], Iter [5/209] Loss: 1.7673, average_loss: 2.6463
Epoch [16/50], Iter [10/209] Loss: 3.2084, average_loss: 2.7096
Epoch [16/50], Iter [15/209] Loss: 3.1701, average_loss: 2.6818

Epoch [16/50], Iter [20/209] Loss: 2.5792, average_loss: 2.6396
Epoch [16/50], Iter [25/209] Loss: 1.6545, average_loss: 2.5800
Epoch [16/50], Iter [30/209] Loss: 3.0499, average_loss: 2.6046
Epoch [16/50], Iter [35/209] Loss: 2.7537, average_loss: 2.6266
Epoch [16/50], Iter [40/209] Loss: 2.7281, average_loss: 2.6298
Epoch [16/50], Iter [45/209] Loss: 1.9151, average_loss: 2.6067
Epoch [16/50], Iter [50/209] Loss: 1.8320, average_loss: 2.6125
Epoch [16/50], Iter [55/209] Loss: 2.4375, average_loss: 2.6165
Epoch [16/50], Iter [60/209] Loss: 2.4503, average_loss: 2.6239
Epoch [16/50], Iter [65/209] Loss: 2.1638, average_loss: 2.6147
Epoch [16/50], Iter [70/209] Loss: 3.1187, average_loss: 2.6111
Epoch [16/50], Iter [75/209] Loss: 1.9006, average_loss: 2.5967
Epoch [16/50], Iter [80/209] Loss: 3.0649, average_loss: 2.5877
Epoch [16/50], Iter [85/209] Loss: 3.1345, average_loss: 2.5992
Epoch [16/50], Iter [90/209] Loss: 3.1622, average_loss: 2.6094
Epoch [16/50], Iter [95/209] Loss: 3.0038, average_loss: 2.6292
Epoch [16/50], Iter [100/209] Loss: 3.2521, average_loss: 2.6214
Epoch [16/50], Iter [105/209] Loss: 2.2355, average_loss: 2.6181
Epoch [16/50], Iter [110/209] Loss: 2.2051, average_loss: 2.6042
Epoch [16/50], Iter [115/209] Loss: 2.1461, average_loss: 2.5967
Epoch [16/50], Iter [120/209] Loss: 1.8123, average_loss: 2.5831
Epoch [16/50], Iter [125/209] Loss: 2.3568, average_loss: 2.5592
Epoch [16/50], Iter [130/209] Loss: 2.3119, average_loss: 2.5748
Epoch [16/50], Iter [135/209] Loss: 2.7436, average_loss: 2.5695
Epoch [16/50], Iter [140/209] Loss: 2.6093, average_loss: 2.5659
Epoch [16/50], Iter [145/209] Loss: 2.6837, average_loss: 2.5635
Epoch [16/50], Iter [150/209] Loss: 2.4211, average_loss: 2.5542
Epoch [16/50], Iter [155/209] Loss: 2.4263, average_loss: 2.5585
Epoch [16/50], Iter [160/209] Loss: 1.9827, average_loss: 2.5511
Epoch [16/50], Iter [165/209] Loss: 1.9297, average_loss: 2.5319
Epoch [16/50], Iter [170/209] Loss: 2.0696, average_loss: 2.5273
Epoch [16/50], Iter [175/209] Loss: 2.1621, average_loss: 2.5243
Epoch [16/50], Iter [180/209] Loss: 3.1481, average_loss: 2.5326
Epoch [16/50], Iter [185/209] Loss: 2.9553, average_loss: 2.5340
Epoch [16/50], Iter [190/209] Loss: 2.3430, average_loss: 2.5324
Epoch [16/50], Iter [195/209] Loss: 2.9614, average_loss: 2.5298
Epoch [16/50], Iter [200/209] Loss: 1.6161, average_loss: 2.5273
Epoch [16/50], Iter [205/209] Loss: 3.0791, average_loss: 2.5328
Updating best test loss: 2.97056

Starting epoch 17 / 50

Learning Rate for this epoch: 0.001

Epoch [17/50], Iter [5/209] Loss: 2.6941, average_loss: 2.8137
Epoch [17/50], Iter [10/209] Loss: 2.3423, average_loss: 2.7687
Epoch [17/50], Iter [15/209] Loss: 2.6699, average_loss: 2.6892
Epoch [17/50], Iter [20/209] Loss: 3.6512, average_loss: 2.8325
Epoch [17/50], Iter [25/209] Loss: 2.5726, average_loss: 2.6959
Epoch [17/50], Iter [30/209] Loss: 3.0497, average_loss: 2.6678
Epoch [17/50], Iter [35/209] Loss: 2.4792, average_loss: 2.5842
Epoch [17/50], Iter [40/209] Loss: 2.4969, average_loss: 2.5729
Epoch [17/50], Iter [45/209] Loss: 2.6403, average_loss: 2.5743
Epoch [17/50], Iter [50/209] Loss: 2.5381, average_loss: 2.5725
Epoch [17/50], Iter [55/209] Loss: 2.6210, average_loss: 2.5570
Epoch [17/50], Iter [60/209] Loss: 2.8644, average_loss: 2.5451
Epoch [17/50], Iter [65/209] Loss: 2.6298, average_loss: 2.5287
Epoch [17/50], Iter [70/209] Loss: 2.5544, average_loss: 2.5116
Epoch [17/50], Iter [75/209] Loss: 2.3942, average_loss: 2.5101
Epoch [17/50], Iter [80/209] Loss: 3.6274, average_loss: 2.5480
Epoch [17/50], Iter [85/209] Loss: 2.6030, average_loss: 2.5392
Epoch [17/50], Iter [90/209] Loss: 2.5130, average_loss: 2.5469
Epoch [17/50], Iter [95/209] Loss: 2.7933, average_loss: 2.5414
Epoch [17/50], Iter [100/209] Loss: 2.2638, average_loss: 2.5332

```
Epoch [17/50], Iter [105/209] Loss: 2.4237, average_loss: 2.5346
Epoch [17/50], Iter [110/209] Loss: 2.2034, average_loss: 2.5206
Epoch [17/50], Iter [115/209] Loss: 1.9170, average_loss: 2.5080
Epoch [17/50], Iter [120/209] Loss: 2.7061, average_loss: 2.5170
Epoch [17/50], Iter [125/209] Loss: 1.8595, average_loss: 2.5235
Epoch [17/50], Iter [130/209] Loss: 1.9600, average_loss: 2.5160
Epoch [17/50], Iter [135/209] Loss: 2.7428, average_loss: 2.5119
Epoch [17/50], Iter [140/209] Loss: 2.2011, average_loss: 2.5158
Epoch [17/50], Iter [145/209] Loss: 2.9484, average_loss: 2.5157
Epoch [17/50], Iter [150/209] Loss: 2.9688, average_loss: 2.5220
Epoch [17/50], Iter [155/209] Loss: 2.4580, average_loss: 2.5184
Epoch [17/50], Iter [160/209] Loss: 1.7571, average_loss: 2.5200
Epoch [17/50], Iter [165/209] Loss: 2.2461, average_loss: 2.5153
Epoch [17/50], Iter [170/209] Loss: 2.6857, average_loss: 2.5095
Epoch [17/50], Iter [175/209] Loss: 2.2298, average_loss: 2.5025
Epoch [17/50], Iter [180/209] Loss: 2.1825, average_loss: 2.5038
Epoch [17/50], Iter [185/209] Loss: 2.5322, average_loss: 2.5105
Epoch [17/50], Iter [190/209] Loss: 2.5967, average_loss: 2.5078
Epoch [17/50], Iter [195/209] Loss: 2.7126, average_loss: 2.5186
Epoch [17/50], Iter [200/209] Loss: 2.0984, average_loss: 2.5176
Epoch [17/50], Iter [205/209] Loss: 3.0016, average_loss: 2.5221
Updating best test loss: 2.96920
```

Starting epoch 18 / 50

Learning Rate for this epoch: 0.001

```
Epoch [18/50], Iter [5/209] Loss: 2.3011, average_loss: 2.4732
Epoch [18/50], Iter [10/209] Loss: 1.6417, average_loss: 2.3312
Epoch [18/50], Iter [15/209] Loss: 2.4988, average_loss: 2.4180
Epoch [18/50], Iter [20/209] Loss: 2.4235, average_loss: 2.4083
Epoch [18/50], Iter [25/209] Loss: 1.8982, average_loss: 2.3588
Epoch [18/50], Iter [30/209] Loss: 2.2921, average_loss: 2.4000
Epoch [18/50], Iter [35/209] Loss: 2.8340, average_loss: 2.3842
Epoch [18/50], Iter [40/209] Loss: 2.5078, average_loss: 2.4243
Epoch [18/50], Iter [45/209] Loss: 2.1623, average_loss: 2.4092
Epoch [18/50], Iter [50/209] Loss: 2.0484, average_loss: 2.4202
Epoch [18/50], Iter [55/209] Loss: 2.0084, average_loss: 2.3886
Epoch [18/50], Iter [60/209] Loss: 1.8833, average_loss: 2.3824
Epoch [18/50], Iter [65/209] Loss: 2.5008, average_loss: 2.3966
Epoch [18/50], Iter [70/209] Loss: 3.3332, average_loss: 2.4187
Epoch [18/50], Iter [75/209] Loss: 2.3713, average_loss: 2.4263
Epoch [18/50], Iter [80/209] Loss: 2.4121, average_loss: 2.4236
Epoch [18/50], Iter [85/209] Loss: 1.7805, average_loss: 2.4231
Epoch [18/50], Iter [90/209] Loss: 2.8819, average_loss: 2.4283
Epoch [18/50], Iter [95/209] Loss: 2.9433, average_loss: 2.4295
Epoch [18/50], Iter [100/209] Loss: 3.1522, average_loss: 2.4272
Epoch [18/50], Iter [105/209] Loss: 2.5466, average_loss: 2.4258
Epoch [18/50], Iter [110/209] Loss: 2.3476, average_loss: 2.4197
Epoch [18/50], Iter [115/209] Loss: 2.7363, average_loss: 2.4264
Epoch [18/50], Iter [120/209] Loss: 4.1291, average_loss: 2.4420
Epoch [18/50], Iter [125/209] Loss: 2.2880, average_loss: 2.4397
Epoch [18/50], Iter [130/209] Loss: 1.7559, average_loss: 2.4256
Epoch [18/50], Iter [135/209] Loss: 2.2196, average_loss: 2.4238
Epoch [18/50], Iter [140/209] Loss: 2.7389, average_loss: 2.4346
Epoch [18/50], Iter [145/209] Loss: 1.9515, average_loss: 2.4241
Epoch [18/50], Iter [150/209] Loss: 2.4741, average_loss: 2.4137
Epoch [18/50], Iter [155/209] Loss: 2.6745, average_loss: 2.4179
Epoch [18/50], Iter [160/209] Loss: 2.6026, average_loss: 2.4160
Epoch [18/50], Iter [165/209] Loss: 2.2859, average_loss: 2.4151
Epoch [18/50], Iter [170/209] Loss: 3.1547, average_loss: 2.4220
Epoch [18/50], Iter [175/209] Loss: 2.6249, average_loss: 2.4298
Epoch [18/50], Iter [180/209] Loss: 2.0134, average_loss: 2.4351
Epoch [18/50], Iter [185/209] Loss: 2.6397, average_loss: 2.4371
```

Epoch [18/50], Iter [190/209] Loss: 1.9558, average_loss: 2.4411
Epoch [18/50], Iter [195/209] Loss: 2.3560, average_loss: 2.4384
Epoch [18/50], Iter [200/209] Loss: 2.5636, average_loss: 2.4404
Epoch [18/50], Iter [205/209] Loss: 1.9945, average_loss: 2.4456

Starting epoch 19 / 50

Learning Rate for this epoch: 0.001

Epoch [19/50], Iter [5/209] Loss: 2.3215, average_loss: 2.3335
Epoch [19/50], Iter [10/209] Loss: 1.8874, average_loss: 2.2488
Epoch [19/50], Iter [15/209] Loss: 1.4179, average_loss: 2.1559
Epoch [19/50], Iter [20/209] Loss: 1.9385, average_loss: 2.1906
Epoch [19/50], Iter [25/209] Loss: 2.0997, average_loss: 2.2477
Epoch [19/50], Iter [30/209] Loss: 2.4769, average_loss: 2.3523
Epoch [19/50], Iter [35/209] Loss: 2.2932, average_loss: 2.3679
Epoch [19/50], Iter [40/209] Loss: 2.4529, average_loss: 2.3758
Epoch [19/50], Iter [45/209] Loss: 2.3845, average_loss: 2.3762
Epoch [19/50], Iter [50/209] Loss: 2.3440, average_loss: 2.3406
Epoch [19/50], Iter [55/209] Loss: 2.7834, average_loss: 2.3481
Epoch [19/50], Iter [60/209] Loss: 2.3516, average_loss: 2.3222
Epoch [19/50], Iter [65/209] Loss: 2.0676, average_loss: 2.3516
Epoch [19/50], Iter [70/209] Loss: 2.4980, average_loss: 2.4111
Epoch [19/50], Iter [75/209] Loss: 2.3976, average_loss: 2.4131
Epoch [19/50], Iter [80/209] Loss: 1.9826, average_loss: 2.4125
Epoch [19/50], Iter [85/209] Loss: 2.4763, average_loss: 2.3951
Epoch [19/50], Iter [90/209] Loss: 1.7501, average_loss: 2.3871
Epoch [19/50], Iter [95/209] Loss: 2.3356, average_loss: 2.3828
Epoch [19/50], Iter [100/209] Loss: 1.8532, average_loss: 2.3758
Epoch [19/50], Iter [105/209] Loss: 2.2967, average_loss: 2.3819
Epoch [19/50], Iter [110/209] Loss: 2.0902, average_loss: 2.3934
Epoch [19/50], Iter [115/209] Loss: 2.2793, average_loss: 2.3847
Epoch [19/50], Iter [120/209] Loss: 2.6022, average_loss: 2.3885
Epoch [19/50], Iter [125/209] Loss: 1.9592, average_loss: 2.3859
Epoch [19/50], Iter [130/209] Loss: 2.1256, average_loss: 2.3744
Epoch [19/50], Iter [135/209] Loss: 2.3244, average_loss: 2.3786
Epoch [19/50], Iter [140/209] Loss: 1.6986, average_loss: 2.3834
Epoch [19/50], Iter [145/209] Loss: 2.5068, average_loss: 2.3888
Epoch [19/50], Iter [150/209] Loss: 2.6509, average_loss: 2.3854
Epoch [19/50], Iter [155/209] Loss: 2.6380, average_loss: 2.3929
Epoch [19/50], Iter [160/209] Loss: 2.3003, average_loss: 2.3980
Epoch [19/50], Iter [165/209] Loss: 2.4372, average_loss: 2.4019
Epoch [19/50], Iter [170/209] Loss: 2.9184, average_loss: 2.4065
Epoch [19/50], Iter [175/209] Loss: 2.4152, average_loss: 2.4048
Epoch [19/50], Iter [180/209] Loss: 1.9911, average_loss: 2.4077
Epoch [19/50], Iter [185/209] Loss: 3.6134, average_loss: 2.4162
Epoch [19/50], Iter [190/209] Loss: 2.2194, average_loss: 2.4123
Epoch [19/50], Iter [195/209] Loss: 2.4367, average_loss: 2.4090
Epoch [19/50], Iter [200/209] Loss: 2.9006, average_loss: 2.4131
Epoch [19/50], Iter [205/209] Loss: 3.2086, average_loss: 2.4184
Updating best test loss: 2.91726

Starting epoch 20 / 50

Learning Rate for this epoch: 0.001

Epoch [20/50], Iter [5/209] Loss: 2.7928, average_loss: 2.4655
Epoch [20/50], Iter [10/209] Loss: 2.5459, average_loss: 2.5333
Epoch [20/50], Iter [15/209] Loss: 1.7550, average_loss: 2.4375
Epoch [20/50], Iter [20/209] Loss: 3.4960, average_loss: 2.5429
Epoch [20/50], Iter [25/209] Loss: 2.5028, average_loss: 2.4926
Epoch [20/50], Iter [30/209] Loss: 1.8160, average_loss: 2.4420
Epoch [20/50], Iter [35/209] Loss: 2.0729, average_loss: 2.4735
Epoch [20/50], Iter [40/209] Loss: 2.2472, average_loss: 2.4983
Epoch [20/50], Iter [45/209] Loss: 3.6174, average_loss: 2.5014

Epoch [20/50], Iter [50/209] Loss: 2.6653, average_loss: 2.4816
Epoch [20/50], Iter [55/209] Loss: 2.0575, average_loss: 2.4416
Epoch [20/50], Iter [60/209] Loss: 1.9283, average_loss: 2.4172
Epoch [20/50], Iter [65/209] Loss: 2.0697, average_loss: 2.4037
Epoch [20/50], Iter [70/209] Loss: 2.6582, average_loss: 2.4303
Epoch [20/50], Iter [75/209] Loss: 2.9599, average_loss: 2.4258
Epoch [20/50], Iter [80/209] Loss: 2.2112, average_loss: 2.4181
Epoch [20/50], Iter [85/209] Loss: 2.2284, average_loss: 2.4306
Epoch [20/50], Iter [90/209] Loss: 2.0700, average_loss: 2.4275
Epoch [20/50], Iter [95/209] Loss: 2.4267, average_loss: 2.4183
Epoch [20/50], Iter [100/209] Loss: 1.5257, average_loss: 2.4239
Epoch [20/50], Iter [105/209] Loss: 2.2923, average_loss: 2.4246
Epoch [20/50], Iter [110/209] Loss: 1.9983, average_loss: 2.4045
Epoch [20/50], Iter [115/209] Loss: 1.9194, average_loss: 2.4020
Epoch [20/50], Iter [120/209] Loss: 1.9571, average_loss: 2.4022
Epoch [20/50], Iter [125/209] Loss: 3.7709, average_loss: 2.4069
Epoch [20/50], Iter [130/209] Loss: 2.9387, average_loss: 2.4092
Epoch [20/50], Iter [135/209] Loss: 2.8745, average_loss: 2.3972
Epoch [20/50], Iter [140/209] Loss: 3.0471, average_loss: 2.4041
Epoch [20/50], Iter [145/209] Loss: 2.3073, average_loss: 2.3993
Epoch [20/50], Iter [150/209] Loss: 3.2110, average_loss: 2.4077
Epoch [20/50], Iter [155/209] Loss: 1.7685, average_loss: 2.4029
Epoch [20/50], Iter [160/209] Loss: 1.8498, average_loss: 2.3996
Epoch [20/50], Iter [165/209] Loss: 2.8134, average_loss: 2.4082
Epoch [20/50], Iter [170/209] Loss: 2.7080, average_loss: 2.4140
Epoch [20/50], Iter [175/209] Loss: 1.8989, average_loss: 2.4052
Epoch [20/50], Iter [180/209] Loss: 1.7624, average_loss: 2.3944
Epoch [20/50], Iter [185/209] Loss: 2.1570, average_loss: 2.3868
Epoch [20/50], Iter [190/209] Loss: 2.2463, average_loss: 2.3776
Epoch [20/50], Iter [195/209] Loss: 2.2099, average_loss: 2.3817
Epoch [20/50], Iter [200/209] Loss: 2.2474, average_loss: 2.3853
Epoch [20/50], Iter [205/209] Loss: 2.3890, average_loss: 2.3888
Updating best test loss: 2.89333

Starting epoch 21 / 50

Learning Rate for this epoch: 0.001

Epoch [21/50], Iter [5/209] Loss: 1.7541, average_loss: 2.2038
Epoch [21/50], Iter [10/209] Loss: 2.0493, average_loss: 2.2501
Epoch [21/50], Iter [15/209] Loss: 2.1278, average_loss: 2.2192
Epoch [21/50], Iter [20/209] Loss: 2.2135, average_loss: 2.1952
Epoch [21/50], Iter [25/209] Loss: 2.0007, average_loss: 2.2095
Epoch [21/50], Iter [30/209] Loss: 2.8035, average_loss: 2.2402
Epoch [21/50], Iter [35/209] Loss: 1.9528, average_loss: 2.2489
Epoch [21/50], Iter [40/209] Loss: 2.7919, average_loss: 2.2658
Epoch [21/50], Iter [45/209] Loss: 2.2746, average_loss: 2.3152
Epoch [21/50], Iter [50/209] Loss: 2.4784, average_loss: 2.2855
Epoch [21/50], Iter [55/209] Loss: 1.8731, average_loss: 2.2952
Epoch [21/50], Iter [60/209] Loss: 2.1223, average_loss: 2.3046
Epoch [21/50], Iter [65/209] Loss: 2.4620, average_loss: 2.2941
Epoch [21/50], Iter [70/209] Loss: 2.4610, average_loss: 2.3012
Epoch [21/50], Iter [75/209] Loss: 2.5180, average_loss: 2.3289
Epoch [21/50], Iter [80/209] Loss: 2.3470, average_loss: 2.3242
Epoch [21/50], Iter [85/209] Loss: 2.4830, average_loss: 2.3209
Epoch [21/50], Iter [90/209] Loss: 2.2773, average_loss: 2.3318
Epoch [21/50], Iter [95/209] Loss: 2.3282, average_loss: 2.3258
Epoch [21/50], Iter [100/209] Loss: 2.6523, average_loss: 2.3291
Epoch [21/50], Iter [105/209] Loss: 2.5524, average_loss: 2.3443
Epoch [21/50], Iter [110/209] Loss: 2.6391, average_loss: 2.3520
Epoch [21/50], Iter [115/209] Loss: 1.9175, average_loss: 2.3345
Epoch [21/50], Iter [120/209] Loss: 2.6384, average_loss: 2.3435
Epoch [21/50], Iter [125/209] Loss: 2.5496, average_loss: 2.3329
Epoch [21/50], Iter [130/209] Loss: 2.1831, average_loss: 2.3346

Epoch [21/50], Iter [135/209] Loss: 2.4880, average_loss: 2.3338
Epoch [21/50], Iter [140/209] Loss: 2.1034, average_loss: 2.3300
Epoch [21/50], Iter [145/209] Loss: 2.0745, average_loss: 2.3260
Epoch [21/50], Iter [150/209] Loss: 2.2857, average_loss: 2.3313
Epoch [21/50], Iter [155/209] Loss: 2.3909, average_loss: 2.3246
Epoch [21/50], Iter [160/209] Loss: 4.1667, average_loss: 2.3437
Epoch [21/50], Iter [165/209] Loss: 2.2719, average_loss: 2.3443
Epoch [21/50], Iter [170/209] Loss: 2.3694, average_loss: 2.3364
Epoch [21/50], Iter [175/209] Loss: 2.5964, average_loss: 2.3394
Epoch [21/50], Iter [180/209] Loss: 1.7517, average_loss: 2.3299
Epoch [21/50], Iter [185/209] Loss: 1.8482, average_loss: 2.3289
Epoch [21/50], Iter [190/209] Loss: 2.2820, average_loss: 2.3270
Epoch [21/50], Iter [195/209] Loss: 2.4661, average_loss: 2.3271
Epoch [21/50], Iter [200/209] Loss: 2.0954, average_loss: 2.3305
Epoch [21/50], Iter [205/209] Loss: 1.6464, average_loss: 2.3270
Updating best test loss: 2.87365

Starting epoch 22 / 50

Learning Rate for this epoch: 0.001

Epoch [22/50], Iter [5/209] Loss: 2.2816, average_loss: 2.5517
Epoch [22/50], Iter [10/209] Loss: 2.4781, average_loss: 2.5188
Epoch [22/50], Iter [15/209] Loss: 2.7441, average_loss: 2.3902
Epoch [22/50], Iter [20/209] Loss: 1.9210, average_loss: 2.3789
Epoch [22/50], Iter [25/209] Loss: 1.7552, average_loss: 2.3616
Epoch [22/50], Iter [30/209] Loss: 3.2203, average_loss: 2.3835
Epoch [22/50], Iter [35/209] Loss: 2.2072, average_loss: 2.3247
Epoch [22/50], Iter [40/209] Loss: 2.1956, average_loss: 2.3280
Epoch [22/50], Iter [45/209] Loss: 2.2546, average_loss: 2.3204
Epoch [22/50], Iter [50/209] Loss: 2.4157, average_loss: 2.3380
Epoch [22/50], Iter [55/209] Loss: 2.3440, average_loss: 2.3401
Epoch [22/50], Iter [60/209] Loss: 1.2764, average_loss: 2.3131
Epoch [22/50], Iter [65/209] Loss: 2.4226, average_loss: 2.3162
Epoch [22/50], Iter [70/209] Loss: 2.5127, average_loss: 2.3295
Epoch [22/50], Iter [75/209] Loss: 2.4791, average_loss: 2.3283
Epoch [22/50], Iter [80/209] Loss: 1.8555, average_loss: 2.3132
Epoch [22/50], Iter [85/209] Loss: 2.3109, average_loss: 2.2994
Epoch [22/50], Iter [90/209] Loss: 1.3702, average_loss: 2.2927
Epoch [22/50], Iter [95/209] Loss: 2.4211, average_loss: 2.2930
Epoch [22/50], Iter [100/209] Loss: 1.6958, average_loss: 2.2948
Epoch [22/50], Iter [105/209] Loss: 2.7325, average_loss: 2.2916
Epoch [22/50], Iter [110/209] Loss: 1.8109, average_loss: 2.2857
Epoch [22/50], Iter [115/209] Loss: 1.5814, average_loss: 2.2751
Epoch [22/50], Iter [120/209] Loss: 1.9667, average_loss: 2.2672
Epoch [22/50], Iter [125/209] Loss: 1.5891, average_loss: 2.2625
Epoch [22/50], Iter [130/209] Loss: 1.9566, average_loss: 2.2558
Epoch [22/50], Iter [135/209] Loss: 1.7239, average_loss: 2.2537
Epoch [22/50], Iter [140/209] Loss: 2.4317, average_loss: 2.2450
Epoch [22/50], Iter [145/209] Loss: 1.8241, average_loss: 2.2499
Epoch [22/50], Iter [150/209] Loss: 3.3354, average_loss: 2.2592
Epoch [22/50], Iter [155/209] Loss: 2.6103, average_loss: 2.2595
Epoch [22/50], Iter [160/209] Loss: 2.3695, average_loss: 2.2699
Epoch [22/50], Iter [165/209] Loss: 2.9241, average_loss: 2.2679
Epoch [22/50], Iter [170/209] Loss: 1.9855, average_loss: 2.2579
Epoch [22/50], Iter [175/209] Loss: 2.1110, average_loss: 2.2552
Epoch [22/50], Iter [180/209] Loss: 2.0000, average_loss: 2.2554
Epoch [22/50], Iter [185/209] Loss: 3.2507, average_loss: 2.2690
Epoch [22/50], Iter [190/209] Loss: 2.4493, average_loss: 2.2787
Epoch [22/50], Iter [195/209] Loss: 2.3603, average_loss: 2.2856
Epoch [22/50], Iter [200/209] Loss: 2.0435, average_loss: 2.2838
Epoch [22/50], Iter [205/209] Loss: 2.4128, average_loss: 2.2845
Updating best test loss: 2.87009

Starting epoch 23 / 50

Learning Rate for this epoch: 0.001

Epoch [23/50], Iter [5/209] Loss: 1.7835, average_loss: 2.1955
Epoch [23/50], Iter [10/209] Loss: 1.7727, average_loss: 1.9561
Epoch [23/50], Iter [15/209] Loss: 1.3579, average_loss: 1.9590
Epoch [23/50], Iter [20/209] Loss: 2.2233, average_loss: 2.0184
Epoch [23/50], Iter [25/209] Loss: 2.1043, average_loss: 2.0614
Epoch [23/50], Iter [30/209] Loss: 1.9938, average_loss: 2.0809
Epoch [23/50], Iter [35/209] Loss: 1.9609, average_loss: 2.0316
Epoch [23/50], Iter [40/209] Loss: 1.9878, average_loss: 2.0349
Epoch [23/50], Iter [45/209] Loss: 2.1057, average_loss: 2.0370
Epoch [23/50], Iter [50/209] Loss: 1.9746, average_loss: 2.0525
Epoch [23/50], Iter [55/209] Loss: 3.4226, average_loss: 2.0653
Epoch [23/50], Iter [60/209] Loss: 2.1594, average_loss: 2.0902
Epoch [23/50], Iter [65/209] Loss: 2.2942, average_loss: 2.1357
Epoch [23/50], Iter [70/209] Loss: 2.6428, average_loss: 2.1392
Epoch [23/50], Iter [75/209] Loss: 2.2718, average_loss: 2.1458
Epoch [23/50], Iter [80/209] Loss: 1.8418, average_loss: 2.1549
Epoch [23/50], Iter [85/209] Loss: 1.6164, average_loss: 2.1494
Epoch [23/50], Iter [90/209] Loss: 2.3204, average_loss: 2.1521
Epoch [23/50], Iter [95/209] Loss: 2.2859, average_loss: 2.1690
Epoch [23/50], Iter [100/209] Loss: 2.4920, average_loss: 2.1662
Epoch [23/50], Iter [105/209] Loss: 1.9333, average_loss: 2.1747
Epoch [23/50], Iter [110/209] Loss: 1.7538, average_loss: 2.1806
Epoch [23/50], Iter [115/209] Loss: 2.4182, average_loss: 2.1969
Epoch [23/50], Iter [120/209] Loss: 3.1123, average_loss: 2.1959
Epoch [23/50], Iter [125/209] Loss: 2.1794, average_loss: 2.1861
Epoch [23/50], Iter [130/209] Loss: 3.9333, average_loss: 2.1923
Epoch [23/50], Iter [135/209] Loss: 2.7003, average_loss: 2.1984
Epoch [23/50], Iter [140/209] Loss: 2.3654, average_loss: 2.2135
Epoch [23/50], Iter [145/209] Loss: 1.5219, average_loss: 2.2119
Epoch [23/50], Iter [150/209] Loss: 3.1652, average_loss: 2.2201
Epoch [23/50], Iter [155/209] Loss: 3.0749, average_loss: 2.2249
Epoch [23/50], Iter [160/209] Loss: 1.9205, average_loss: 2.2289
Epoch [23/50], Iter [165/209] Loss: 2.8663, average_loss: 2.2333
Epoch [23/50], Iter [170/209] Loss: 2.1905, average_loss: 2.2339
Epoch [23/50], Iter [175/209] Loss: 1.9180, average_loss: 2.2342
Epoch [23/50], Iter [180/209] Loss: 2.6653, average_loss: 2.2319
Epoch [23/50], Iter [185/209] Loss: 2.4261, average_loss: 2.2323
Epoch [23/50], Iter [190/209] Loss: 2.6405, average_loss: 2.2383
Epoch [23/50], Iter [195/209] Loss: 2.3667, average_loss: 2.2402
Epoch [23/50], Iter [200/209] Loss: 2.2309, average_loss: 2.2442
Epoch [23/50], Iter [205/209] Loss: 2.0193, average_loss: 2.2429
Updating best test loss: 2.83030

Starting epoch 24 / 50

Learning Rate for this epoch: 0.001

Epoch [24/50], Iter [5/209] Loss: 2.2341, average_loss: 2.4022
Epoch [24/50], Iter [10/209] Loss: 2.5611, average_loss: 2.3254
Epoch [24/50], Iter [15/209] Loss: 2.4031, average_loss: 2.2576
Epoch [24/50], Iter [20/209] Loss: 2.6004, average_loss: 2.2423
Epoch [24/50], Iter [25/209] Loss: 2.6035, average_loss: 2.2319
Epoch [24/50], Iter [30/209] Loss: 2.9880, average_loss: 2.2568
Epoch [24/50], Iter [35/209] Loss: 2.1558, average_loss: 2.2707
Epoch [24/50], Iter [40/209] Loss: 1.5894, average_loss: 2.3099
Epoch [24/50], Iter [45/209] Loss: 2.2588, average_loss: 2.2884
Epoch [24/50], Iter [50/209] Loss: 2.3633, average_loss: 2.2950
Epoch [24/50], Iter [55/209] Loss: 2.7385, average_loss: 2.2886
Epoch [24/50], Iter [60/209] Loss: 2.6678, average_loss: 2.2642
Epoch [24/50], Iter [65/209] Loss: 1.8971, average_loss: 2.2365
Epoch [24/50], Iter [70/209] Loss: 2.5604, average_loss: 2.2338

```
Epoch [24/50], Iter [75/209] Loss: 1.6161, average_loss: 2.2468
Epoch [24/50], Iter [80/209] Loss: 1.6744, average_loss: 2.2279
Epoch [24/50], Iter [85/209] Loss: 2.3639, average_loss: 2.2115
Epoch [24/50], Iter [90/209] Loss: 2.6691, average_loss: 2.2217
Epoch [24/50], Iter [95/209] Loss: 2.4148, average_loss: 2.2192
Epoch [24/50], Iter [100/209] Loss: 2.1151, average_loss: 2.1993
Epoch [24/50], Iter [105/209] Loss: 1.6706, average_loss: 2.2155
Epoch [24/50], Iter [110/209] Loss: 2.7313, average_loss: 2.2216
Epoch [24/50], Iter [115/209] Loss: 1.9352, average_loss: 2.2188
Epoch [24/50], Iter [120/209] Loss: 1.7564, average_loss: 2.2161
Epoch [24/50], Iter [125/209] Loss: 2.2056, average_loss: 2.2157
Epoch [24/50], Iter [130/209] Loss: 2.1155, average_loss: 2.2064
Epoch [24/50], Iter [135/209] Loss: 2.0031, average_loss: 2.2005
Epoch [24/50], Iter [140/209] Loss: 2.0518, average_loss: 2.2127
Epoch [24/50], Iter [145/209] Loss: 2.2538, average_loss: 2.2165
Epoch [24/50], Iter [150/209] Loss: 2.5109, average_loss: 2.2112
Epoch [24/50], Iter [155/209] Loss: 2.7610, average_loss: 2.2194
Epoch [24/50], Iter [160/209] Loss: 1.8236, average_loss: 2.2189
Epoch [24/50], Iter [165/209] Loss: 2.5945, average_loss: 2.2308
Epoch [24/50], Iter [170/209] Loss: 2.5250, average_loss: 2.2310
Epoch [24/50], Iter [175/209] Loss: 2.6340, average_loss: 2.2246
Epoch [24/50], Iter [180/209] Loss: 1.6949, average_loss: 2.2157
Epoch [24/50], Iter [185/209] Loss: 1.6874, average_loss: 2.2131
Epoch [24/50], Iter [190/209] Loss: 2.6372, average_loss: 2.2087
Epoch [24/50], Iter [195/209] Loss: 1.8663, average_loss: 2.2147
Epoch [24/50], Iter [200/209] Loss: 1.6916, average_loss: 2.2147
Epoch [24/50], Iter [205/209] Loss: 2.1210, average_loss: 2.2143
```

Starting epoch 25 / 50

Learning Rate for this epoch: 0.001

```
Epoch [25/50], Iter [5/209] Loss: 1.8528, average_loss: 2.0919
Epoch [25/50], Iter [10/209] Loss: 3.1735, average_loss: 2.2520
Epoch [25/50], Iter [15/209] Loss: 2.3706, average_loss: 2.2196
Epoch [25/50], Iter [20/209] Loss: 2.3878, average_loss: 2.2400
Epoch [25/50], Iter [25/209] Loss: 1.7401, average_loss: 2.2093
Epoch [25/50], Iter [30/209] Loss: 2.1417, average_loss: 2.2152
Epoch [25/50], Iter [35/209] Loss: 2.9984, average_loss: 2.2448
Epoch [25/50], Iter [40/209] Loss: 2.3356, average_loss: 2.2402
Epoch [25/50], Iter [45/209] Loss: 2.5351, average_loss: 2.2515
Epoch [25/50], Iter [50/209] Loss: 2.1245, average_loss: 2.2420
Epoch [25/50], Iter [55/209] Loss: 2.0885, average_loss: 2.2817
Epoch [25/50], Iter [60/209] Loss: 1.7334, average_loss: 2.2816
Epoch [25/50], Iter [65/209] Loss: 2.5428, average_loss: 2.2630
Epoch [25/50], Iter [70/209] Loss: 2.0602, average_loss: 2.2375
Epoch [25/50], Iter [75/209] Loss: 1.9741, average_loss: 2.2209
Epoch [25/50], Iter [80/209] Loss: 2.1580, average_loss: 2.2247
Epoch [25/50], Iter [85/209] Loss: 2.1941, average_loss: 2.2026
Epoch [25/50], Iter [90/209] Loss: 1.9009, average_loss: 2.2169
Epoch [25/50], Iter [95/209] Loss: 2.0969, average_loss: 2.2044
Epoch [25/50], Iter [100/209] Loss: 1.2137, average_loss: 2.1966
Epoch [25/50], Iter [105/209] Loss: 1.6575, average_loss: 2.2002
Epoch [25/50], Iter [110/209] Loss: 1.6407, average_loss: 2.1853
Epoch [25/50], Iter [115/209] Loss: 1.7952, average_loss: 2.1830
Epoch [25/50], Iter [120/209] Loss: 1.6966, average_loss: 2.1713
Epoch [25/50], Iter [125/209] Loss: 1.7449, average_loss: 2.1745
Epoch [25/50], Iter [130/209] Loss: 2.5056, average_loss: 2.1746
Epoch [25/50], Iter [135/209] Loss: 1.6817, average_loss: 2.1714
Epoch [25/50], Iter [140/209] Loss: 2.2994, average_loss: 2.1642
Epoch [25/50], Iter [145/209] Loss: 3.5985, average_loss: 2.1723
Epoch [25/50], Iter [150/209] Loss: 2.9026, average_loss: 2.1691
Epoch [25/50], Iter [155/209] Loss: 3.0298, average_loss: 2.1798
Epoch [25/50], Iter [160/209] Loss: 2.6766, average_loss: 2.1760
```

```
Epoch [25/50], Iter [165/209] Loss: 2.2727, average_loss: 2.1832
Epoch [25/50], Iter [170/209] Loss: 2.6521, average_loss: 2.1809
Epoch [25/50], Iter [175/209] Loss: 2.6842, average_loss: 2.1850
Epoch [25/50], Iter [180/209] Loss: 2.3300, average_loss: 2.1767
Epoch [25/50], Iter [185/209] Loss: 2.8006, average_loss: 2.1817
Epoch [25/50], Iter [190/209] Loss: 1.7456, average_loss: 2.1781
Epoch [25/50], Iter [195/209] Loss: 1.4475, average_loss: 2.1750
Epoch [25/50], Iter [200/209] Loss: 2.1108, average_loss: 2.1757
Epoch [25/50], Iter [205/209] Loss: 1.8383, average_loss: 2.1711
```

Starting epoch 26 / 50

Learning Rate for this epoch: 0.001

```
Epoch [26/50], Iter [5/209] Loss: 1.9662, average_loss: 2.1887
Epoch [26/50], Iter [10/209] Loss: 2.3093, average_loss: 2.1995
Epoch [26/50], Iter [15/209] Loss: 1.6914, average_loss: 2.0956
Epoch [26/50], Iter [20/209] Loss: 2.0831, average_loss: 2.1165
Epoch [26/50], Iter [25/209] Loss: 1.7070, average_loss: 2.0918
Epoch [26/50], Iter [30/209] Loss: 1.9678, average_loss: 2.0742
Epoch [26/50], Iter [35/209] Loss: 2.3755, average_loss: 2.0542
Epoch [26/50], Iter [40/209] Loss: 1.4255, average_loss: 2.0354
Epoch [26/50], Iter [45/209] Loss: 2.7358, average_loss: 2.0416
Epoch [26/50], Iter [50/209] Loss: 1.8754, average_loss: 2.0446
Epoch [26/50], Iter [55/209] Loss: 2.4268, average_loss: 2.0445
Epoch [26/50], Iter [60/209] Loss: 2.2722, average_loss: 2.0520
Epoch [26/50], Iter [65/209] Loss: 2.4211, average_loss: 2.0791
Epoch [26/50], Iter [70/209] Loss: 2.1952, average_loss: 2.0884
Epoch [26/50], Iter [75/209] Loss: 2.1914, average_loss: 2.1162
Epoch [26/50], Iter [80/209] Loss: 1.8676, average_loss: 2.1267
Epoch [26/50], Iter [85/209] Loss: 2.6516, average_loss: 2.1232
Epoch [26/50], Iter [90/209] Loss: 2.6181, average_loss: 2.1384
Epoch [26/50], Iter [95/209] Loss: 2.3702, average_loss: 2.1483
Epoch [26/50], Iter [100/209] Loss: 1.6469, average_loss: 2.1425
Epoch [26/50], Iter [105/209] Loss: 1.7252, average_loss: 2.1296
Epoch [26/50], Iter [110/209] Loss: 1.7787, average_loss: 2.1389
Epoch [26/50], Iter [115/209] Loss: 2.7799, average_loss: 2.1441
Epoch [26/50], Iter [120/209] Loss: 2.6821, average_loss: 2.1489
Epoch [26/50], Iter [125/209] Loss: 2.1919, average_loss: 2.1414
Epoch [26/50], Iter [130/209] Loss: 1.8745, average_loss: 2.1461
Epoch [26/50], Iter [135/209] Loss: 1.8193, average_loss: 2.1457
Epoch [26/50], Iter [140/209] Loss: 1.6246, average_loss: 2.1465
Epoch [26/50], Iter [145/209] Loss: 1.9649, average_loss: 2.1486
Epoch [26/50], Iter [150/209] Loss: 1.4741, average_loss: 2.1465
Epoch [26/50], Iter [155/209] Loss: 1.7045, average_loss: 2.1454
Epoch [26/50], Iter [160/209] Loss: 1.9694, average_loss: 2.1396
Epoch [26/50], Iter [165/209] Loss: 2.5894, average_loss: 2.1483
Epoch [26/50], Iter [170/209] Loss: 2.5905, average_loss: 2.1533
Epoch [26/50], Iter [175/209] Loss: 2.9423, average_loss: 2.1603
Epoch [26/50], Iter [180/209] Loss: 2.3128, average_loss: 2.1591
Epoch [26/50], Iter [185/209] Loss: 1.7071, average_loss: 2.1530
Epoch [26/50], Iter [190/209] Loss: 2.1896, average_loss: 2.1552
Epoch [26/50], Iter [195/209] Loss: 2.6751, average_loss: 2.1543
Epoch [26/50], Iter [200/209] Loss: 2.5767, average_loss: 2.1549
Epoch [26/50], Iter [205/209] Loss: 1.9780, average_loss: 2.1539
```

Starting epoch 27 / 50

Learning Rate for this epoch: 0.001

```
Epoch [27/50], Iter [5/209] Loss: 2.8630, average_loss: 2.2041
Epoch [27/50], Iter [10/209] Loss: 1.8770, average_loss: 2.1472
Epoch [27/50], Iter [15/209] Loss: 1.6552, average_loss: 2.1426
Epoch [27/50], Iter [20/209] Loss: 2.1962, average_loss: 2.1386
Epoch [27/50], Iter [25/209] Loss: 1.6632, average_loss: 2.1039
```


Epoch [27/50], Iter [30/209] Loss: 2.1132, average_loss: 2.0589
Epoch [27/50], Iter [35/209] Loss: 1.7367, average_loss: 2.0641
Epoch [27/50], Iter [40/209] Loss: 3.2781, average_loss: 2.0628
Epoch [27/50], Iter [45/209] Loss: 2.1854, average_loss: 2.0911
Epoch [27/50], Iter [50/209] Loss: 1.9733, average_loss: 2.0968
Epoch [27/50], Iter [55/209] Loss: 1.9245, average_loss: 2.0681
Epoch [27/50], Iter [60/209] Loss: 2.0556, average_loss: 2.0629
Epoch [27/50], Iter [65/209] Loss: 2.2931, average_loss: 2.0636
Epoch [27/50], Iter [70/209] Loss: 1.8927, average_loss: 2.0587
Epoch [27/50], Iter [75/209] Loss: 1.8150, average_loss: 2.0537
Epoch [27/50], Iter [80/209] Loss: 1.7164, average_loss: 2.0783
Epoch [27/50], Iter [85/209] Loss: 2.3205, average_loss: 2.0882
Epoch [27/50], Iter [90/209] Loss: 1.5922, average_loss: 2.0830
Epoch [27/50], Iter [95/209] Loss: 2.4415, average_loss: 2.0874
Epoch [27/50], Iter [100/209] Loss: 1.7394, average_loss: 2.0755
Epoch [27/50], Iter [105/209] Loss: 2.7371, average_loss: 2.0880
Epoch [27/50], Iter [110/209] Loss: 1.7089, average_loss: 2.1004
Epoch [27/50], Iter [115/209] Loss: 2.2870, average_loss: 2.1098
Epoch [27/50], Iter [120/209] Loss: 1.8087, average_loss: 2.1091
Epoch [27/50], Iter [125/209] Loss: 2.1461, average_loss: 2.1009
Epoch [27/50], Iter [130/209] Loss: 2.2006, average_loss: 2.0954
Epoch [27/50], Iter [135/209] Loss: 2.0466, average_loss: 2.0921
Epoch [27/50], Iter [140/209] Loss: 2.2038, average_loss: 2.0857
Epoch [27/50], Iter [145/209] Loss: 1.7916, average_loss: 2.0907
Epoch [27/50], Iter [150/209] Loss: 1.9096, average_loss: 2.0846
Epoch [27/50], Iter [155/209] Loss: 2.0774, average_loss: 2.0946
Epoch [27/50], Iter [160/209] Loss: 2.5376, average_loss: 2.1045
Epoch [27/50], Iter [165/209] Loss: 2.0653, average_loss: 2.1133
Epoch [27/50], Iter [170/209] Loss: 1.7024, average_loss: 2.1151
Epoch [27/50], Iter [175/209] Loss: 1.7760, average_loss: 2.1111
Epoch [27/50], Iter [180/209] Loss: 2.0299, average_loss: 2.1078
Epoch [27/50], Iter [185/209] Loss: 3.1944, average_loss: 2.1162
Epoch [27/50], Iter [190/209] Loss: 1.7339, average_loss: 2.1174
Epoch [27/50], Iter [195/209] Loss: 2.5717, average_loss: 2.1246
Epoch [27/50], Iter [200/209] Loss: 2.2058, average_loss: 2.1205
Epoch [27/50], Iter [205/209] Loss: 2.8993, average_loss: 2.1219
Updating best test loss: 2.81113

Starting epoch 28 / 50

Learning Rate for this epoch: 0.001

Epoch [28/50], Iter [5/209] Loss: 2.9179, average_loss: 2.4063
Epoch [28/50], Iter [10/209] Loss: 1.7409, average_loss: 2.1887
Epoch [28/50], Iter [15/209] Loss: 1.4646, average_loss: 2.0905
Epoch [28/50], Iter [20/209] Loss: 1.7282, average_loss: 2.0403
Epoch [28/50], Iter [25/209] Loss: 2.7535, average_loss: 2.0357
Epoch [28/50], Iter [30/209] Loss: 1.3024, average_loss: 1.9930
Epoch [28/50], Iter [35/209] Loss: 2.2975, average_loss: 2.0183
Epoch [28/50], Iter [40/209] Loss: 1.9530, average_loss: 2.0319
Epoch [28/50], Iter [45/209] Loss: 1.7364, average_loss: 2.0420
Epoch [28/50], Iter [50/209] Loss: 2.4207, average_loss: 2.0617
Epoch [28/50], Iter [55/209] Loss: 2.1284, average_loss: 2.0746
Epoch [28/50], Iter [60/209] Loss: 2.4021, average_loss: 2.0894
Epoch [28/50], Iter [65/209] Loss: 2.1372, average_loss: 2.1000
Epoch [28/50], Iter [70/209] Loss: 1.6680, average_loss: 2.1031
Epoch [28/50], Iter [75/209] Loss: 2.9567, average_loss: 2.1164
Epoch [28/50], Iter [80/209] Loss: 1.9366, average_loss: 2.0993
Epoch [28/50], Iter [85/209] Loss: 1.7276, average_loss: 2.0818
Epoch [28/50], Iter [90/209] Loss: 1.6994, average_loss: 2.0835
Epoch [28/50], Iter [95/209] Loss: 2.7023, average_loss: 2.0942
Epoch [28/50], Iter [100/209] Loss: 2.4618, average_loss: 2.1073
Epoch [28/50], Iter [105/209] Loss: 2.0286, average_loss: 2.1179
Epoch [28/50], Iter [110/209] Loss: 1.6919, average_loss: 2.1127

```
Epoch [28/50], Iter [115/209] Loss: 2.2018, average_loss: 2.1081
Epoch [28/50], Iter [120/209] Loss: 2.2653, average_loss: 2.1000
Epoch [28/50], Iter [125/209] Loss: 1.3509, average_loss: 2.0991
Epoch [28/50], Iter [130/209] Loss: 2.4218, average_loss: 2.1013
Epoch [28/50], Iter [135/209] Loss: 2.2194, average_loss: 2.1036
Epoch [28/50], Iter [140/209] Loss: 1.9881, average_loss: 2.1147
Epoch [28/50], Iter [145/209] Loss: 2.3462, average_loss: 2.1127
Epoch [28/50], Iter [150/209] Loss: 1.7252, average_loss: 2.1032
Epoch [28/50], Iter [155/209] Loss: 1.7205, average_loss: 2.0981
Epoch [28/50], Iter [160/209] Loss: 2.0956, average_loss: 2.1026
Epoch [28/50], Iter [165/209] Loss: 1.7305, average_loss: 2.1027
Epoch [28/50], Iter [170/209] Loss: 3.6381, average_loss: 2.1095
Epoch [28/50], Iter [175/209] Loss: 1.6019, average_loss: 2.1038
Epoch [28/50], Iter [180/209] Loss: 1.9743, average_loss: 2.1115
Epoch [28/50], Iter [185/209] Loss: 2.8920, average_loss: 2.1120
Epoch [28/50], Iter [190/209] Loss: 2.0544, average_loss: 2.1080
Epoch [28/50], Iter [195/209] Loss: 1.6082, average_loss: 2.1050
Epoch [28/50], Iter [200/209] Loss: 1.9945, average_loss: 2.1015
Epoch [28/50], Iter [205/209] Loss: 1.9437, average_loss: 2.1018
```

Starting epoch 29 / 50

Learning Rate for this epoch: 0.001

```
Epoch [29/50], Iter [5/209] Loss: 2.0501, average_loss: 1.8446
Epoch [29/50], Iter [10/209] Loss: 2.2381, average_loss: 2.0373
Epoch [29/50], Iter [15/209] Loss: 1.7008, average_loss: 1.9323
Epoch [29/50], Iter [20/209] Loss: 1.8983, average_loss: 1.9692
Epoch [29/50], Iter [25/209] Loss: 2.6844, average_loss: 2.0327
Epoch [29/50], Iter [30/209] Loss: 2.3546, average_loss: 2.0806
Epoch [29/50], Iter [35/209] Loss: 1.9252, average_loss: 2.0494
Epoch [29/50], Iter [40/209] Loss: 1.7125, average_loss: 2.0629
Epoch [29/50], Iter [45/209] Loss: 2.0776, average_loss: 2.0844
Epoch [29/50], Iter [50/209] Loss: 1.9116, average_loss: 2.0721
Epoch [29/50], Iter [55/209] Loss: 2.7428, average_loss: 2.0795
Epoch [29/50], Iter [60/209] Loss: 2.4951, average_loss: 2.0808
Epoch [29/50], Iter [65/209] Loss: 1.7021, average_loss: 2.0828
Epoch [29/50], Iter [70/209] Loss: 2.2140, average_loss: 2.0946
Epoch [29/50], Iter [75/209] Loss: 1.9848, average_loss: 2.0950
Epoch [29/50], Iter [80/209] Loss: 1.6580, average_loss: 2.0902
Epoch [29/50], Iter [85/209] Loss: 1.8921, average_loss: 2.0949
Epoch [29/50], Iter [90/209] Loss: 2.1094, average_loss: 2.0903
Epoch [29/50], Iter [95/209] Loss: 2.2700, average_loss: 2.1037
Epoch [29/50], Iter [100/209] Loss: 2.1660, average_loss: 2.0929
Epoch [29/50], Iter [105/209] Loss: 1.9367, average_loss: 2.0915
Epoch [29/50], Iter [110/209] Loss: 2.1960, average_loss: 2.0852
Epoch [29/50], Iter [115/209] Loss: 1.9485, average_loss: 2.0877
Epoch [29/50], Iter [120/209] Loss: 1.7567, average_loss: 2.0783
Epoch [29/50], Iter [125/209] Loss: 2.5162, average_loss: 2.0812
Epoch [29/50], Iter [130/209] Loss: 2.1516, average_loss: 2.0775
Epoch [29/50], Iter [135/209] Loss: 1.5189, average_loss: 2.0674
Epoch [29/50], Iter [140/209] Loss: 2.1288, average_loss: 2.0661
Epoch [29/50], Iter [145/209] Loss: 1.9102, average_loss: 2.0581
Epoch [29/50], Iter [150/209] Loss: 2.0807, average_loss: 2.0638
Epoch [29/50], Iter [155/209] Loss: 1.8292, average_loss: 2.0647
Epoch [29/50], Iter [160/209] Loss: 1.4234, average_loss: 2.0547
Epoch [29/50], Iter [165/209] Loss: 1.9197, average_loss: 2.0530
Epoch [29/50], Iter [170/209] Loss: 1.6883, average_loss: 2.0500
Epoch [29/50], Iter [175/209] Loss: 2.0501, average_loss: 2.0576
Epoch [29/50], Iter [180/209] Loss: 2.3376, average_loss: 2.0548
Epoch [29/50], Iter [185/209] Loss: 2.3681, average_loss: 2.0500
Epoch [29/50], Iter [190/209] Loss: 2.1533, average_loss: 2.0499
Epoch [29/50], Iter [195/209] Loss: 2.2195, average_loss: 2.0490
Epoch [29/50], Iter [200/209] Loss: 1.7406, average_loss: 2.0471
```

Epoch [29/50], Iter [205/209] Loss: 1.8444, average_loss: 2.0436
Updating best test loss: 2.80689

Starting epoch 30 / 50

Learning Rate for this epoch: 0.001

Epoch [30/50], Iter [5/209] Loss: 1.7430, average_loss: 1.7775
Epoch [30/50], Iter [10/209] Loss: 1.3024, average_loss: 1.8197
Epoch [30/50], Iter [15/209] Loss: 2.6683, average_loss: 1.9622
Epoch [30/50], Iter [20/209] Loss: 1.9456, average_loss: 1.9623
Epoch [30/50], Iter [25/209] Loss: 1.6980, average_loss: 1.9790
Epoch [30/50], Iter [30/209] Loss: 1.9126, average_loss: 1.9565
Epoch [30/50], Iter [35/209] Loss: 1.8761, average_loss: 1.9425
Epoch [30/50], Iter [40/209] Loss: 1.7040, average_loss: 1.9644
Epoch [30/50], Iter [45/209] Loss: 1.5970, average_loss: 1.9774
Epoch [30/50], Iter [50/209] Loss: 1.8337, average_loss: 1.9801
Epoch [30/50], Iter [55/209] Loss: 1.7529, average_loss: 1.9807
Epoch [30/50], Iter [60/209] Loss: 1.7310, average_loss: 1.9867
Epoch [30/50], Iter [65/209] Loss: 1.2451, average_loss: 1.9707
Epoch [30/50], Iter [70/209] Loss: 1.6595, average_loss: 1.9652
Epoch [30/50], Iter [75/209] Loss: 2.2098, average_loss: 1.9681
Epoch [30/50], Iter [80/209] Loss: 1.5836, average_loss: 1.9711
Epoch [30/50], Iter [85/209] Loss: 1.6328, average_loss: 1.9637
Epoch [30/50], Iter [90/209] Loss: 3.3202, average_loss: 1.9818
Epoch [30/50], Iter [95/209] Loss: 2.2233, average_loss: 1.9975
Epoch [30/50], Iter [100/209] Loss: 2.1397, average_loss: 1.9888
Epoch [30/50], Iter [105/209] Loss: 1.7176, average_loss: 1.9958
Epoch [30/50], Iter [110/209] Loss: 1.7920, average_loss: 1.9938
Epoch [30/50], Iter [115/209] Loss: 1.5764, average_loss: 1.9980
Epoch [30/50], Iter [120/209] Loss: 2.7950, average_loss: 2.0073
Epoch [30/50], Iter [125/209] Loss: 2.3455, average_loss: 2.0037
Epoch [30/50], Iter [130/209] Loss: 1.5568, average_loss: 1.9929
Epoch [30/50], Iter [135/209] Loss: 2.0031, average_loss: 1.9971
Epoch [30/50], Iter [140/209] Loss: 2.9267, average_loss: 2.0043
Epoch [30/50], Iter [145/209] Loss: 1.7945, average_loss: 2.0037
Epoch [30/50], Iter [150/209] Loss: 2.3263, average_loss: 2.0090
Epoch [30/50], Iter [155/209] Loss: 1.9302, average_loss: 2.0138
Epoch [30/50], Iter [160/209] Loss: 1.4198, average_loss: 2.0222
Epoch [30/50], Iter [165/209] Loss: 1.4845, average_loss: 2.0160
Epoch [30/50], Iter [170/209] Loss: 2.0565, average_loss: 2.0235
Epoch [30/50], Iter [175/209] Loss: 1.7943, average_loss: 2.0205
Epoch [30/50], Iter [180/209] Loss: 1.8124, average_loss: 2.0204
Epoch [30/50], Iter [185/209] Loss: 2.1421, average_loss: 2.0191
Epoch [30/50], Iter [190/209] Loss: 2.0844, average_loss: 2.0174
Epoch [30/50], Iter [195/209] Loss: 1.6924, average_loss: 2.0154
Epoch [30/50], Iter [200/209] Loss: 1.7474, average_loss: 2.0136
Epoch [30/50], Iter [205/209] Loss: 1.4828, average_loss: 2.0169

Starting epoch 31 / 50

Learning Rate for this epoch: 0.0001

Epoch [31/50], Iter [5/209] Loss: 1.8585, average_loss: 1.8730
Epoch [31/50], Iter [10/209] Loss: 1.4376, average_loss: 1.7751
Epoch [31/50], Iter [15/209] Loss: 1.7956, average_loss: 1.7221
Epoch [31/50], Iter [20/209] Loss: 1.4830, average_loss: 1.7170
Epoch [31/50], Iter [25/209] Loss: 1.7579, average_loss: 1.7795
Epoch [31/50], Iter [30/209] Loss: 1.8807, average_loss: 1.7981
Epoch [31/50], Iter [35/209] Loss: 2.0565, average_loss: 1.8204
Epoch [31/50], Iter [40/209] Loss: 1.9842, average_loss: 1.8655
Epoch [31/50], Iter [45/209] Loss: 2.0116, average_loss: 1.8798
Epoch [31/50], Iter [50/209] Loss: 1.9569, average_loss: 1.8711
Epoch [31/50], Iter [55/209] Loss: 1.7196, average_loss: 1.8863
Epoch [31/50], Iter [60/209] Loss: 2.0755, average_loss: 1.9014

Epoch [31/50], Iter [65/209] Loss: 1.7342, average_loss: 1.9029
Epoch [31/50], Iter [70/209] Loss: 2.6146, average_loss: 1.9233
Epoch [31/50], Iter [75/209] Loss: 1.5063, average_loss: 1.9220
Epoch [31/50], Iter [80/209] Loss: 1.7255, average_loss: 1.9190
Epoch [31/50], Iter [85/209] Loss: 1.8919, average_loss: 1.9300
Epoch [31/50], Iter [90/209] Loss: 1.7365, average_loss: 1.9326
Epoch [31/50], Iter [95/209] Loss: 1.7277, average_loss: 1.9328
Epoch [31/50], Iter [100/209] Loss: 1.6501, average_loss: 1.9198
Epoch [31/50], Iter [105/209] Loss: 1.7302, average_loss: 1.9143
Epoch [31/50], Iter [110/209] Loss: 2.1108, average_loss: 1.9206
Epoch [31/50], Iter [115/209] Loss: 2.1113, average_loss: 1.9141
Epoch [31/50], Iter [120/209] Loss: 1.8500, average_loss: 1.9257
Epoch [31/50], Iter [125/209] Loss: 2.1366, average_loss: 1.9187
Epoch [31/50], Iter [130/209] Loss: 1.7202, average_loss: 1.9190
Epoch [31/50], Iter [135/209] Loss: 1.8474, average_loss: 1.9273
Epoch [31/50], Iter [140/209] Loss: 1.6553, average_loss: 1.9210
Epoch [31/50], Iter [145/209] Loss: 1.7804, average_loss: 1.9101
Epoch [31/50], Iter [150/209] Loss: 1.9914, average_loss: 1.9058
Epoch [31/50], Iter [155/209] Loss: 2.6959, average_loss: 1.9125
Epoch [31/50], Iter [160/209] Loss: 2.6966, average_loss: 1.9220
Epoch [31/50], Iter [165/209] Loss: 2.0003, average_loss: 1.9130
Epoch [31/50], Iter [170/209] Loss: 2.0089, average_loss: 1.9145
Epoch [31/50], Iter [175/209] Loss: 1.7749, average_loss: 1.9147
Epoch [31/50], Iter [180/209] Loss: 2.6859, average_loss: 1.9185
Epoch [31/50], Iter [185/209] Loss: 1.8373, average_loss: 1.9174
Epoch [31/50], Iter [190/209] Loss: 2.2984, average_loss: 1.9200
Epoch [31/50], Iter [195/209] Loss: 1.9950, average_loss: 1.9194
Epoch [31/50], Iter [200/209] Loss: 2.4144, average_loss: 1.9316
Epoch [31/50], Iter [205/209] Loss: 1.8596, average_loss: 1.9270
Updating best test loss: 2.72753

Starting epoch 32 / 50

Learning Rate for this epoch: 0.0001

Epoch [32/50], Iter [5/209] Loss: 2.1572, average_loss: 2.2222
Epoch [32/50], Iter [10/209] Loss: 1.7889, average_loss: 2.0723
Epoch [32/50], Iter [15/209] Loss: 1.7535, average_loss: 1.9795
Epoch [32/50], Iter [20/209] Loss: 2.1918, average_loss: 1.9656
Epoch [32/50], Iter [25/209] Loss: 2.0170, average_loss: 1.9932
Epoch [32/50], Iter [30/209] Loss: 1.6375, average_loss: 1.9996
Epoch [32/50], Iter [35/209] Loss: 1.8781, average_loss: 2.0246
Epoch [32/50], Iter [40/209] Loss: 2.4907, average_loss: 1.9874
Epoch [32/50], Iter [45/209] Loss: 1.8276, average_loss: 2.0061
Epoch [32/50], Iter [50/209] Loss: 2.0384, average_loss: 2.0027
Epoch [32/50], Iter [55/209] Loss: 2.0735, average_loss: 2.0073
Epoch [32/50], Iter [60/209] Loss: 2.2938, average_loss: 1.9927
Epoch [32/50], Iter [65/209] Loss: 2.3073, average_loss: 1.9976
Epoch [32/50], Iter [70/209] Loss: 1.3765, average_loss: 1.9712
Epoch [32/50], Iter [75/209] Loss: 2.3376, average_loss: 1.9687
Epoch [32/50], Iter [80/209] Loss: 2.4026, average_loss: 1.9697
Epoch [32/50], Iter [85/209] Loss: 1.7799, average_loss: 1.9664
Epoch [32/50], Iter [90/209] Loss: 2.0607, average_loss: 1.9618
Epoch [32/50], Iter [95/209] Loss: 1.8983, average_loss: 1.9585
Epoch [32/50], Iter [100/209] Loss: 1.5613, average_loss: 1.9388
Epoch [32/50], Iter [105/209] Loss: 1.7601, average_loss: 1.9265
Epoch [32/50], Iter [110/209] Loss: 1.8164, average_loss: 1.9177
Epoch [32/50], Iter [115/209] Loss: 1.6268, average_loss: 1.9130
Epoch [32/50], Iter [120/209] Loss: 1.1661, average_loss: 1.9048
Epoch [32/50], Iter [125/209] Loss: 1.7675, average_loss: 1.9052
Epoch [32/50], Iter [130/209] Loss: 1.9708, average_loss: 1.9057
Epoch [32/50], Iter [135/209] Loss: 2.2346, average_loss: 1.9085
Epoch [32/50], Iter [140/209] Loss: 1.8689, average_loss: 1.9021
Epoch [32/50], Iter [145/209] Loss: 1.6354, average_loss: 1.9019

Epoch [32/50], Iter [150/209] Loss: 1.5029, average_loss: 1.8998
Epoch [32/50], Iter [155/209] Loss: 1.6115, average_loss: 1.9020
Epoch [32/50], Iter [160/209] Loss: 1.7357, average_loss: 1.9001
Epoch [32/50], Iter [165/209] Loss: 1.7910, average_loss: 1.9057
Epoch [32/50], Iter [170/209] Loss: 1.3533, average_loss: 1.9030
Epoch [32/50], Iter [175/209] Loss: 1.5434, average_loss: 1.8917
Epoch [32/50], Iter [180/209] Loss: 1.9532, average_loss: 1.8888
Epoch [32/50], Iter [185/209] Loss: 1.4689, average_loss: 1.8818
Epoch [32/50], Iter [190/209] Loss: 2.1990, average_loss: 1.8819
Epoch [32/50], Iter [195/209] Loss: 1.6320, average_loss: 1.8771
Epoch [32/50], Iter [200/209] Loss: 1.4914, average_loss: 1.8778
Epoch [32/50], Iter [205/209] Loss: 2.7001, average_loss: 1.8782
Updating best test loss: 2.70312

Starting epoch 33 / 50

Learning Rate for this epoch: 0.0001

Epoch [33/50], Iter [5/209] Loss: 1.8566, average_loss: 1.8694
Epoch [33/50], Iter [10/209] Loss: 1.8635, average_loss: 1.8926
Epoch [33/50], Iter [15/209] Loss: 1.5883, average_loss: 1.7733
Epoch [33/50], Iter [20/209] Loss: 1.7520, average_loss: 1.7891
Epoch [33/50], Iter [25/209] Loss: 1.8899, average_loss: 1.8325
Epoch [33/50], Iter [30/209] Loss: 1.8220, average_loss: 1.8434
Epoch [33/50], Iter [35/209] Loss: 1.5566, average_loss: 1.8545
Epoch [33/50], Iter [40/209] Loss: 2.5041, average_loss: 1.8502
Epoch [33/50], Iter [45/209] Loss: 1.5409, average_loss: 1.8156
Epoch [33/50], Iter [50/209] Loss: 1.4430, average_loss: 1.8205
Epoch [33/50], Iter [55/209] Loss: 2.2350, average_loss: 1.8385
Epoch [33/50], Iter [60/209] Loss: 1.2496, average_loss: 1.8537
Epoch [33/50], Iter [65/209] Loss: 2.3917, average_loss: 1.8925
Epoch [33/50], Iter [70/209] Loss: 2.1364, average_loss: 1.8867
Epoch [33/50], Iter [75/209] Loss: 1.9188, average_loss: 1.8992
Epoch [33/50], Iter [80/209] Loss: 1.6120, average_loss: 1.8966
Epoch [33/50], Iter [85/209] Loss: 1.7160, average_loss: 1.8947
Epoch [33/50], Iter [90/209] Loss: 1.5834, average_loss: 1.9001
Epoch [33/50], Iter [95/209] Loss: 2.1533, average_loss: 1.9250
Epoch [33/50], Iter [100/209] Loss: 1.4614, average_loss: 1.9221
Epoch [33/50], Iter [105/209] Loss: 2.2066, average_loss: 1.9222
Epoch [33/50], Iter [110/209] Loss: 2.0966, average_loss: 1.9227
Epoch [33/50], Iter [115/209] Loss: 2.0161, average_loss: 1.9258
Epoch [33/50], Iter [120/209] Loss: 2.1193, average_loss: 1.9269
Epoch [33/50], Iter [125/209] Loss: 1.9306, average_loss: 1.9219
Epoch [33/50], Iter [130/209] Loss: 1.9747, average_loss: 1.9134
Epoch [33/50], Iter [135/209] Loss: 1.2311, average_loss: 1.9032
Epoch [33/50], Iter [140/209] Loss: 2.1624, average_loss: 1.9148
Epoch [33/50], Iter [145/209] Loss: 1.7053, average_loss: 1.9165
Epoch [33/50], Iter [150/209] Loss: 3.0815, average_loss: 1.9232
Epoch [33/50], Iter [155/209] Loss: 1.5644, average_loss: 1.9123
Epoch [33/50], Iter [160/209] Loss: 1.2728, average_loss: 1.9089
Epoch [33/50], Iter [165/209] Loss: 1.6559, average_loss: 1.9010
Epoch [33/50], Iter [170/209] Loss: 2.0307, average_loss: 1.9072
Epoch [33/50], Iter [175/209] Loss: 1.7086, average_loss: 1.9016
Epoch [33/50], Iter [180/209] Loss: 2.3029, average_loss: 1.9063
Epoch [33/50], Iter [185/209] Loss: 1.7294, average_loss: 1.9052
Epoch [33/50], Iter [190/209] Loss: 1.9475, average_loss: 1.9093
Epoch [33/50], Iter [195/209] Loss: 2.9027, average_loss: 1.9085
Epoch [33/50], Iter [200/209] Loss: 1.4185, average_loss: 1.8996
Epoch [33/50], Iter [205/209] Loss: 2.1976, average_loss: 1.9027

Starting epoch 34 / 50

Learning Rate for this epoch: 0.0001

Epoch [34/50], Iter [5/209] Loss: 1.5807, average_loss: 2.0252

Epoch [34/50], Iter [10/209]	Loss: 1.8159, average_loss: 1.9115
Epoch [34/50], Iter [15/209]	Loss: 2.2873, average_loss: 1.9180
Epoch [34/50], Iter [20/209]	Loss: 1.9265, average_loss: 1.9250
Epoch [34/50], Iter [25/209]	Loss: 1.8136, average_loss: 1.9037
Epoch [34/50], Iter [30/209]	Loss: 1.7057, average_loss: 1.8735
Epoch [34/50], Iter [35/209]	Loss: 1.5153, average_loss: 1.8477
Epoch [34/50], Iter [40/209]	Loss: 1.6168, average_loss: 1.8698
Epoch [34/50], Iter [45/209]	Loss: 2.1150, average_loss: 1.8563
Epoch [34/50], Iter [50/209]	Loss: 2.0023, average_loss: 1.8552
Epoch [34/50], Iter [55/209]	Loss: 1.6310, average_loss: 1.8607
Epoch [34/50], Iter [60/209]	Loss: 2.5965, average_loss: 1.8571
Epoch [34/50], Iter [65/209]	Loss: 2.1525, average_loss: 1.8527
Epoch [34/50], Iter [70/209]	Loss: 1.6654, average_loss: 1.8665
Epoch [34/50], Iter [75/209]	Loss: 2.7512, average_loss: 1.8837
Epoch [34/50], Iter [80/209]	Loss: 2.0098, average_loss: 1.8721
Epoch [34/50], Iter [85/209]	Loss: 1.5905, average_loss: 1.8616
Epoch [34/50], Iter [90/209]	Loss: 1.7848, average_loss: 1.8681
Epoch [34/50], Iter [95/209]	Loss: 2.1921, average_loss: 1.8534
Epoch [34/50], Iter [100/209]	Loss: 2.0936, average_loss: 1.8532
Epoch [34/50], Iter [105/209]	Loss: 1.6600, average_loss: 1.8640
Epoch [34/50], Iter [110/209]	Loss: 1.9626, average_loss: 1.8632
Epoch [34/50], Iter [115/209]	Loss: 1.9410, average_loss: 1.8557
Epoch [34/50], Iter [120/209]	Loss: 2.2991, average_loss: 1.8606
Epoch [34/50], Iter [125/209]	Loss: 1.3081, average_loss: 1.8604
Epoch [34/50], Iter [130/209]	Loss: 2.0188, average_loss: 1.8631
Epoch [34/50], Iter [135/209]	Loss: 2.6707, average_loss: 1.8667
Epoch [34/50], Iter [140/209]	Loss: 1.5739, average_loss: 1.8585
Epoch [34/50], Iter [145/209]	Loss: 1.8737, average_loss: 1.8633
Epoch [34/50], Iter [150/209]	Loss: 1.8194, average_loss: 1.8602
Epoch [34/50], Iter [155/209]	Loss: 1.6567, average_loss: 1.8560
Epoch [34/50], Iter [160/209]	Loss: 1.9685, average_loss: 1.8595
Epoch [34/50], Iter [165/209]	Loss: 1.7463, average_loss: 1.8624
Epoch [34/50], Iter [170/209]	Loss: 1.8923, average_loss: 1.8567
Epoch [34/50], Iter [175/209]	Loss: 1.5904, average_loss: 1.8525
Epoch [34/50], Iter [180/209]	Loss: 1.5421, average_loss: 1.8476
Epoch [34/50], Iter [185/209]	Loss: 2.5045, average_loss: 1.8529
Epoch [34/50], Iter [190/209]	Loss: 2.1002, average_loss: 1.8636
Epoch [34/50], Iter [195/209]	Loss: 1.8325, average_loss: 1.8623
Epoch [34/50], Iter [200/209]	Loss: 1.3808, average_loss: 1.8534
Epoch [34/50], Iter [205/209]	Loss: 1.9147, average_loss: 1.8576

Starting epoch 35 / 50

Learning Rate for this epoch: 0.0001

Epoch [35/50], Iter [5/209]	Loss: 1.5244, average_loss: 1.7326
Epoch [35/50], Iter [10/209]	Loss: 1.1420, average_loss: 1.7263
Epoch [35/50], Iter [15/209]	Loss: 1.6395, average_loss: 1.7999
Epoch [35/50], Iter [20/209]	Loss: 1.8123, average_loss: 1.7850
Epoch [35/50], Iter [25/209]	Loss: 2.1486, average_loss: 1.8525
Epoch [35/50], Iter [30/209]	Loss: 1.8743, average_loss: 1.8017
Epoch [35/50], Iter [35/209]	Loss: 1.8205, average_loss: 1.8373
Epoch [35/50], Iter [40/209]	Loss: 1.5085, average_loss: 1.8344
Epoch [35/50], Iter [45/209]	Loss: 2.1747, average_loss: 1.8515
Epoch [35/50], Iter [50/209]	Loss: 1.6279, average_loss: 1.8542
Epoch [35/50], Iter [55/209]	Loss: 1.6587, average_loss: 1.8512
Epoch [35/50], Iter [60/209]	Loss: 1.6039, average_loss: 1.8172
Epoch [35/50], Iter [65/209]	Loss: 1.8261, average_loss: 1.8267
Epoch [35/50], Iter [70/209]	Loss: 2.2688, average_loss: 1.8400
Epoch [35/50], Iter [75/209]	Loss: 2.0301, average_loss: 1.8304
Epoch [35/50], Iter [80/209]	Loss: 1.5589, average_loss: 1.8417
Epoch [35/50], Iter [85/209]	Loss: 1.4420, average_loss: 1.8391
Epoch [35/50], Iter [90/209]	Loss: 2.0697, average_loss: 1.8360
Epoch [35/50], Iter [95/209]	Loss: 2.6572, average_loss: 1.8684

Epoch [35/50], Iter [100/209] Loss: 2.6266, average_loss: 1.8675
Epoch [35/50], Iter [105/209] Loss: 2.6299, average_loss: 1.8772
Epoch [35/50], Iter [110/209] Loss: 2.0449, average_loss: 1.8769
Epoch [35/50], Iter [115/209] Loss: 2.3846, average_loss: 1.8829
Epoch [35/50], Iter [120/209] Loss: 2.1376, average_loss: 1.8870
Epoch [35/50], Iter [125/209] Loss: 2.1867, average_loss: 1.8888
Epoch [35/50], Iter [130/209] Loss: 1.9261, average_loss: 1.8841
Epoch [35/50], Iter [135/209] Loss: 1.8110, average_loss: 1.8821
Epoch [35/50], Iter [140/209] Loss: 1.5642, average_loss: 1.8780
Epoch [35/50], Iter [145/209] Loss: 2.2969, average_loss: 1.8753
Epoch [35/50], Iter [150/209] Loss: 1.6318, average_loss: 1.8684
Epoch [35/50], Iter [155/209] Loss: 2.0154, average_loss: 1.8730
Epoch [35/50], Iter [160/209] Loss: 1.8489, average_loss: 1.8727
Epoch [35/50], Iter [165/209] Loss: 1.5245, average_loss: 1.8743
Epoch [35/50], Iter [170/209] Loss: 1.4213, average_loss: 1.8762
Epoch [35/50], Iter [175/209] Loss: 1.2381, average_loss: 1.8678
Epoch [35/50], Iter [180/209] Loss: 1.3692, average_loss: 1.8687
Epoch [35/50], Iter [185/209] Loss: 2.4522, average_loss: 1.8705
Epoch [35/50], Iter [190/209] Loss: 1.5242, average_loss: 1.8684
Epoch [35/50], Iter [195/209] Loss: 2.1756, average_loss: 1.8705
Epoch [35/50], Iter [200/209] Loss: 2.0228, average_loss: 1.8697
Epoch [35/50], Iter [205/209] Loss: 1.5369, average_loss: 1.8703
Updating best test loss: 2.69784

Starting epoch 36 / 50

Learning Rate for this epoch: 0.0001

Epoch [36/50], Iter [5/209] Loss: 2.2556, average_loss: 1.7486
Epoch [36/50], Iter [10/209] Loss: 1.2423, average_loss: 1.8230
Epoch [36/50], Iter [15/209] Loss: 1.9373, average_loss: 1.8300
Epoch [36/50], Iter [20/209] Loss: 1.8839, average_loss: 1.8656
Epoch [36/50], Iter [25/209] Loss: 1.5459, average_loss: 1.8565
Epoch [36/50], Iter [30/209] Loss: 2.2541, average_loss: 1.8529
Epoch [36/50], Iter [35/209] Loss: 1.3852, average_loss: 1.8323
Epoch [36/50], Iter [40/209] Loss: 2.3107, average_loss: 1.8356
Epoch [36/50], Iter [45/209] Loss: 1.8519, average_loss: 1.8370
Epoch [36/50], Iter [50/209] Loss: 2.0081, average_loss: 1.8633
Epoch [36/50], Iter [55/209] Loss: 2.6410, average_loss: 1.8913
Epoch [36/50], Iter [60/209] Loss: 1.8247, average_loss: 1.8912
Epoch [36/50], Iter [65/209] Loss: 1.6742, average_loss: 1.9018
Epoch [36/50], Iter [70/209] Loss: 1.5580, average_loss: 1.8784
Epoch [36/50], Iter [75/209] Loss: 1.9333, average_loss: 1.8756
Epoch [36/50], Iter [80/209] Loss: 2.2058, average_loss: 1.8746
Epoch [36/50], Iter [85/209] Loss: 1.7784, average_loss: 1.8736
Epoch [36/50], Iter [90/209] Loss: 1.8825, average_loss: 1.8631
Epoch [36/50], Iter [95/209] Loss: 1.8261, average_loss: 1.8472
Epoch [36/50], Iter [100/209] Loss: 1.9413, average_loss: 1.8352
Epoch [36/50], Iter [105/209] Loss: 1.6961, average_loss: 1.8308
Epoch [36/50], Iter [110/209] Loss: 2.9149, average_loss: 1.8323
Epoch [36/50], Iter [115/209] Loss: 1.8478, average_loss: 1.8356
Epoch [36/50], Iter [120/209] Loss: 1.6497, average_loss: 1.8316
Epoch [36/50], Iter [125/209] Loss: 2.2011, average_loss: 1.8300
Epoch [36/50], Iter [130/209] Loss: 2.3183, average_loss: 1.8412
Epoch [36/50], Iter [135/209] Loss: 1.5450, average_loss: 1.8332
Epoch [36/50], Iter [140/209] Loss: 1.6211, average_loss: 1.8352
Epoch [36/50], Iter [145/209] Loss: 1.7427, average_loss: 1.8331
Epoch [36/50], Iter [150/209] Loss: 2.1127, average_loss: 1.8409
Epoch [36/50], Iter [155/209] Loss: 1.5279, average_loss: 1.8401
Epoch [36/50], Iter [160/209] Loss: 1.3415, average_loss: 1.8404
Epoch [36/50], Iter [165/209] Loss: 1.7843, average_loss: 1.8365
Epoch [36/50], Iter [170/209] Loss: 1.9449, average_loss: 1.8345
Epoch [36/50], Iter [175/209] Loss: 1.7353, average_loss: 1.8336
Epoch [36/50], Iter [180/209] Loss: 1.3395, average_loss: 1.8389

Epoch [36/50], Iter [185/209] Loss: 1.3581, average_loss: 1.8353
Epoch [36/50], Iter [190/209] Loss: 2.4607, average_loss: 1.8413
Epoch [36/50], Iter [195/209] Loss: 2.3904, average_loss: 1.8369
Epoch [36/50], Iter [200/209] Loss: 1.2844, average_loss: 1.8345
Epoch [36/50], Iter [205/209] Loss: 2.4608, average_loss: 1.8455

Starting epoch 37 / 50

Learning Rate for this epoch: 0.0001

Epoch [37/50], Iter [5/209] Loss: 1.7831, average_loss: 1.8454
Epoch [37/50], Iter [10/209] Loss: 1.4300, average_loss: 1.8389
Epoch [37/50], Iter [15/209] Loss: 2.1185, average_loss: 1.7822
Epoch [37/50], Iter [20/209] Loss: 1.4289, average_loss: 1.8050
Epoch [37/50], Iter [25/209] Loss: 1.4627, average_loss: 1.8071
Epoch [37/50], Iter [30/209] Loss: 1.8892, average_loss: 1.8640
Epoch [37/50], Iter [35/209] Loss: 1.4249, average_loss: 1.8350
Epoch [37/50], Iter [40/209] Loss: 2.1775, average_loss: 1.8290
Epoch [37/50], Iter [45/209] Loss: 1.8647, average_loss: 1.8128
Epoch [37/50], Iter [50/209] Loss: 1.4202, average_loss: 1.7768
Epoch [37/50], Iter [55/209] Loss: 1.5132, average_loss: 1.7600
Epoch [37/50], Iter [60/209] Loss: 1.6387, average_loss: 1.7531
Epoch [37/50], Iter [65/209] Loss: 2.2379, average_loss: 1.7694
Epoch [37/50], Iter [70/209] Loss: 2.3741, average_loss: 1.7757
Epoch [37/50], Iter [75/209] Loss: 1.3596, average_loss: 1.7783
Epoch [37/50], Iter [80/209] Loss: 2.3528, average_loss: 1.7878
Epoch [37/50], Iter [85/209] Loss: 2.4055, average_loss: 1.8015
Epoch [37/50], Iter [90/209] Loss: 1.8164, average_loss: 1.8103
Epoch [37/50], Iter [95/209] Loss: 1.9475, average_loss: 1.8154
Epoch [37/50], Iter [100/209] Loss: 1.9773, average_loss: 1.8074
Epoch [37/50], Iter [105/209] Loss: 1.6034, average_loss: 1.8228
Epoch [37/50], Iter [110/209] Loss: 2.1376, average_loss: 1.8412
Epoch [37/50], Iter [115/209] Loss: 1.6688, average_loss: 1.8584
Epoch [37/50], Iter [120/209] Loss: 2.2061, average_loss: 1.8642
Epoch [37/50], Iter [125/209] Loss: 1.3629, average_loss: 1.8519
Epoch [37/50], Iter [130/209] Loss: 1.7103, average_loss: 1.8501
Epoch [37/50], Iter [135/209] Loss: 2.1500, average_loss: 1.8485
Epoch [37/50], Iter [140/209] Loss: 1.6447, average_loss: 1.8442
Epoch [37/50], Iter [145/209] Loss: 1.3245, average_loss: 1.8414
Epoch [37/50], Iter [150/209] Loss: 1.4072, average_loss: 1.8451
Epoch [37/50], Iter [155/209] Loss: 1.3200, average_loss: 1.8454
Epoch [37/50], Iter [160/209] Loss: 1.7690, average_loss: 1.8489
Epoch [37/50], Iter [165/209] Loss: 1.7928, average_loss: 1.8533
Epoch [37/50], Iter [170/209] Loss: 1.5903, average_loss: 1.8462
Epoch [37/50], Iter [175/209] Loss: 1.1244, average_loss: 1.8424
Epoch [37/50], Iter [180/209] Loss: 1.8571, average_loss: 1.8428
Epoch [37/50], Iter [185/209] Loss: 1.6233, average_loss: 1.8359
Epoch [37/50], Iter [190/209] Loss: 2.0706, average_loss: 1.8372
Epoch [37/50], Iter [195/209] Loss: 1.6061, average_loss: 1.8342
Epoch [37/50], Iter [200/209] Loss: 2.1082, average_loss: 1.8359
Epoch [37/50], Iter [205/209] Loss: 1.8531, average_loss: 1.8343
Updating best test loss: 2.69620

Starting epoch 38 / 50

Learning Rate for this epoch: 0.0001

Epoch [38/50], Iter [5/209] Loss: 2.4251, average_loss: 1.8662
Epoch [38/50], Iter [10/209] Loss: 1.5460, average_loss: 1.7377
Epoch [38/50], Iter [15/209] Loss: 1.8654, average_loss: 1.7076
Epoch [38/50], Iter [20/209] Loss: 1.9044, average_loss: 1.7455
Epoch [38/50], Iter [25/209] Loss: 1.9768, average_loss: 1.7661
Epoch [38/50], Iter [30/209] Loss: 1.3068, average_loss: 1.7667
Epoch [38/50], Iter [35/209] Loss: 2.7379, average_loss: 1.8087
Epoch [38/50], Iter [40/209] Loss: 1.9420, average_loss: 1.8032


```
Epoch [38/50], Iter [45/209] Loss: 1.7285, average_loss: 1.8280
Epoch [38/50], Iter [50/209] Loss: 1.7493, average_loss: 1.8412
Epoch [38/50], Iter [55/209] Loss: 2.0589, average_loss: 1.8577
Epoch [38/50], Iter [60/209] Loss: 2.5154, average_loss: 1.8662
Epoch [38/50], Iter [65/209] Loss: 1.9590, average_loss: 1.8487
Epoch [38/50], Iter [70/209] Loss: 1.2641, average_loss: 1.8504
Epoch [38/50], Iter [75/209] Loss: 1.0244, average_loss: 1.8359
Epoch [38/50], Iter [80/209] Loss: 1.4150, average_loss: 1.8132
Epoch [38/50], Iter [85/209] Loss: 2.0027, average_loss: 1.8103
Epoch [38/50], Iter [90/209] Loss: 2.3640, average_loss: 1.8266
Epoch [38/50], Iter [95/209] Loss: 1.7862, average_loss: 1.8213
Epoch [38/50], Iter [100/209] Loss: 1.6824, average_loss: 1.8228
Epoch [38/50], Iter [105/209] Loss: 1.8673, average_loss: 1.8229
Epoch [38/50], Iter [110/209] Loss: 1.7975, average_loss: 1.8169
Epoch [38/50], Iter [115/209] Loss: 2.3313, average_loss: 1.8096
Epoch [38/50], Iter [120/209] Loss: 1.6783, average_loss: 1.8010
Epoch [38/50], Iter [125/209] Loss: 1.6296, average_loss: 1.8012
Epoch [38/50], Iter [130/209] Loss: 2.7889, average_loss: 1.8151
Epoch [38/50], Iter [135/209] Loss: 1.7092, average_loss: 1.8186
Epoch [38/50], Iter [140/209] Loss: 2.0175, average_loss: 1.8198
Epoch [38/50], Iter [145/209] Loss: 2.3224, average_loss: 1.8282
Epoch [38/50], Iter [150/209] Loss: 1.4258, average_loss: 1.8290
Epoch [38/50], Iter [155/209] Loss: 1.7657, average_loss: 1.8362
Epoch [38/50], Iter [160/209] Loss: 2.0842, average_loss: 1.8403
Epoch [38/50], Iter [165/209] Loss: 1.9361, average_loss: 1.8385
Epoch [38/50], Iter [170/209] Loss: 1.8122, average_loss: 1.8400
Epoch [38/50], Iter [175/209] Loss: 2.2916, average_loss: 1.8436
Epoch [38/50], Iter [180/209] Loss: 1.5683, average_loss: 1.8410
Epoch [38/50], Iter [185/209] Loss: 1.9859, average_loss: 1.8433
Epoch [38/50], Iter [190/209] Loss: 2.0759, average_loss: 1.8441
Epoch [38/50], Iter [195/209] Loss: 1.8316, average_loss: 1.8428
Epoch [38/50], Iter [200/209] Loss: 1.5234, average_loss: 1.8418
Epoch [38/50], Iter [205/209] Loss: 2.6134, average_loss: 1.8400
```

Starting epoch 39 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [39/50], Iter [5/209] Loss: 1.8128, average_loss: 1.5132
Epoch [39/50], Iter [10/209] Loss: 2.0627, average_loss: 1.8621
Epoch [39/50], Iter [15/209] Loss: 1.6588, average_loss: 1.8086
Epoch [39/50], Iter [20/209] Loss: 1.6529, average_loss: 1.8039
Epoch [39/50], Iter [25/209] Loss: 2.9001, average_loss: 1.9017
Epoch [39/50], Iter [30/209] Loss: 2.1190, average_loss: 1.8996
Epoch [39/50], Iter [35/209] Loss: 1.4457, average_loss: 1.8894
Epoch [39/50], Iter [40/209] Loss: 1.5927, average_loss: 1.8412
Epoch [39/50], Iter [45/209] Loss: 1.4915, average_loss: 1.8350
Epoch [39/50], Iter [50/209] Loss: 1.9874, average_loss: 1.8227
Epoch [39/50], Iter [55/209] Loss: 2.1603, average_loss: 1.8081
Epoch [39/50], Iter [60/209] Loss: 1.4107, average_loss: 1.8040
Epoch [39/50], Iter [65/209] Loss: 2.0234, average_loss: 1.7961
Epoch [39/50], Iter [70/209] Loss: 1.7340, average_loss: 1.7936
Epoch [39/50], Iter [75/209] Loss: 1.7380, average_loss: 1.7943
Epoch [39/50], Iter [80/209] Loss: 1.3871, average_loss: 1.7820
Epoch [39/50], Iter [85/209] Loss: 2.3856, average_loss: 1.7936
Epoch [39/50], Iter [90/209] Loss: 2.2803, average_loss: 1.8179
Epoch [39/50], Iter [95/209] Loss: 2.4211, average_loss: 1.8215
Epoch [39/50], Iter [100/209] Loss: 2.0718, average_loss: 1.8256
Epoch [39/50], Iter [105/209] Loss: 1.5193, average_loss: 1.8216
Epoch [39/50], Iter [110/209] Loss: 1.7737, average_loss: 1.8242
Epoch [39/50], Iter [115/209] Loss: 2.3469, average_loss: 1.8489
Epoch [39/50], Iter [120/209] Loss: 1.5948, average_loss: 1.8460
Epoch [39/50], Iter [125/209] Loss: 2.2420, average_loss: 1.8524
Epoch [39/50], Iter [130/209] Loss: 1.6629, average_loss: 1.8504
```

```
Epoch [39/50], Iter [135/209] Loss: 1.3881, average_loss: 1.8398
Epoch [39/50], Iter [140/209] Loss: 2.4505, average_loss: 1.8464
Epoch [39/50], Iter [145/209] Loss: 1.6667, average_loss: 1.8380
Epoch [39/50], Iter [150/209] Loss: 1.2749, average_loss: 1.8313
Epoch [39/50], Iter [155/209] Loss: 2.0127, average_loss: 1.8354
Epoch [39/50], Iter [160/209] Loss: 2.0599, average_loss: 1.8399
Epoch [39/50], Iter [165/209] Loss: 1.6049, average_loss: 1.8315
Epoch [39/50], Iter [170/209] Loss: 1.7395, average_loss: 1.8339
Epoch [39/50], Iter [175/209] Loss: 2.1744, average_loss: 1.8322
Epoch [39/50], Iter [180/209] Loss: 2.1526, average_loss: 1.8305
Epoch [39/50], Iter [185/209] Loss: 1.4597, average_loss: 1.8268
Epoch [39/50], Iter [190/209] Loss: 1.4194, average_loss: 1.8314
Epoch [39/50], Iter [195/209] Loss: 1.9518, average_loss: 1.8288
Epoch [39/50], Iter [200/209] Loss: 2.2064, average_loss: 1.8362
Epoch [39/50], Iter [205/209] Loss: 1.4699, average_loss: 1.8351
```

Starting epoch 40 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [40/50], Iter [5/209] Loss: 1.8585, average_loss: 1.8630
Epoch [40/50], Iter [10/209] Loss: 1.4763, average_loss: 1.8460
Epoch [40/50], Iter [15/209] Loss: 2.6520, average_loss: 1.8855
Epoch [40/50], Iter [20/209] Loss: 2.0209, average_loss: 1.8821
Epoch [40/50], Iter [25/209] Loss: 1.5472, average_loss: 1.8189
Epoch [40/50], Iter [30/209] Loss: 2.1956, average_loss: 1.8354
Epoch [40/50], Iter [35/209] Loss: 2.0883, average_loss: 1.8492
Epoch [40/50], Iter [40/209] Loss: 2.0513, average_loss: 1.8258
Epoch [40/50], Iter [45/209] Loss: 2.4008, average_loss: 1.8416
Epoch [40/50], Iter [50/209] Loss: 1.4475, average_loss: 1.8368
Epoch [40/50], Iter [55/209] Loss: 1.5643, average_loss: 1.8365
Epoch [40/50], Iter [60/209] Loss: 1.5034, average_loss: 1.8494
Epoch [40/50], Iter [65/209] Loss: 1.9216, average_loss: 1.8531
Epoch [40/50], Iter [70/209] Loss: 1.8874, average_loss: 1.8406
Epoch [40/50], Iter [75/209] Loss: 2.4302, average_loss: 1.8432
Epoch [40/50], Iter [80/209] Loss: 1.8850, average_loss: 1.8484
Epoch [40/50], Iter [85/209] Loss: 1.7244, average_loss: 1.8359
Epoch [40/50], Iter [90/209] Loss: 1.5277, average_loss: 1.8716
Epoch [40/50], Iter [95/209] Loss: 2.1745, average_loss: 1.8690
Epoch [40/50], Iter [100/209] Loss: 2.2308, average_loss: 1.8645
Epoch [40/50], Iter [105/209] Loss: 1.8067, average_loss: 1.8676
Epoch [40/50], Iter [110/209] Loss: 1.6301, average_loss: 1.8584
Epoch [40/50], Iter [115/209] Loss: 1.5676, average_loss: 1.8624
Epoch [40/50], Iter [120/209] Loss: 2.0688, average_loss: 1.8653
Epoch [40/50], Iter [125/209] Loss: 2.3812, average_loss: 1.8606
Epoch [40/50], Iter [130/209] Loss: 2.0927, average_loss: 1.8609
Epoch [40/50], Iter [135/209] Loss: 1.6758, average_loss: 1.8551
Epoch [40/50], Iter [140/209] Loss: 1.9381, average_loss: 1.8612
Epoch [40/50], Iter [145/209] Loss: 1.8675, average_loss: 1.8548
Epoch [40/50], Iter [150/209] Loss: 1.1032, average_loss: 1.8510
Epoch [40/50], Iter [155/209] Loss: 1.4766, average_loss: 1.8409
Epoch [40/50], Iter [160/209] Loss: 1.1649, average_loss: 1.8333
Epoch [40/50], Iter [165/209] Loss: 1.8609, average_loss: 1.8304
Epoch [40/50], Iter [170/209] Loss: 1.8455, average_loss: 1.8249
Epoch [40/50], Iter [175/209] Loss: 2.1518, average_loss: 1.8294
Epoch [40/50], Iter [180/209] Loss: 2.5575, average_loss: 1.8203
Epoch [40/50], Iter [185/209] Loss: 1.2689, average_loss: 1.8133
Epoch [40/50], Iter [190/209] Loss: 1.9745, average_loss: 1.8172
Epoch [40/50], Iter [195/209] Loss: 2.1629, average_loss: 1.8192
Epoch [40/50], Iter [200/209] Loss: 1.4571, average_loss: 1.8132
Epoch [40/50], Iter [205/209] Loss: 2.6187, average_loss: 1.8142
Updating best test loss: 2.69175
```

Starting epoch 41 / 50

Learning Rate for this epoch: 1e-05

Epoch [41/50],	Iter [5/209]	Loss: 1.4082,	average_loss: 1.5216
Epoch [41/50],	Iter [10/209]	Loss: 1.4627,	average_loss: 1.5997
Epoch [41/50],	Iter [15/209]	Loss: 1.8022,	average_loss: 1.7174
Epoch [41/50],	Iter [20/209]	Loss: 1.9017,	average_loss: 1.7165
Epoch [41/50],	Iter [25/209]	Loss: 1.7118,	average_loss: 1.7296
Epoch [41/50],	Iter [30/209]	Loss: 2.0390,	average_loss: 1.7777
Epoch [41/50],	Iter [35/209]	Loss: 2.1629,	average_loss: 1.7764
Epoch [41/50],	Iter [40/209]	Loss: 1.4307,	average_loss: 1.7462
Epoch [41/50],	Iter [45/209]	Loss: 1.8813,	average_loss: 1.7943
Epoch [41/50],	Iter [50/209]	Loss: 1.5549,	average_loss: 1.8017
Epoch [41/50],	Iter [55/209]	Loss: 1.3481,	average_loss: 1.8035
Epoch [41/50],	Iter [60/209]	Loss: 1.8065,	average_loss: 1.8032
Epoch [41/50],	Iter [65/209]	Loss: 2.1495,	average_loss: 1.8064
Epoch [41/50],	Iter [70/209]	Loss: 2.5301,	average_loss: 1.8160
Epoch [41/50],	Iter [75/209]	Loss: 1.9646,	average_loss: 1.8108
Epoch [41/50],	Iter [80/209]	Loss: 1.8825,	average_loss: 1.8270
Epoch [41/50],	Iter [85/209]	Loss: 1.7522,	average_loss: 1.8167
Epoch [41/50],	Iter [90/209]	Loss: 1.5753,	average_loss: 1.8172
Epoch [41/50],	Iter [95/209]	Loss: 1.7438,	average_loss: 1.8219
Epoch [41/50],	Iter [100/209]	Loss: 1.5920,	average_loss: 1.8230
Epoch [41/50],	Iter [105/209]	Loss: 1.5952,	average_loss: 1.8096
Epoch [41/50],	Iter [110/209]	Loss: 1.5640,	average_loss: 1.8051
Epoch [41/50],	Iter [115/209]	Loss: 2.3140,	average_loss: 1.8094
Epoch [41/50],	Iter [120/209]	Loss: 1.7695,	average_loss: 1.8110
Epoch [41/50],	Iter [125/209]	Loss: 2.1506,	average_loss: 1.8069
Epoch [41/50],	Iter [130/209]	Loss: 1.8523,	average_loss: 1.8067
Epoch [41/50],	Iter [135/209]	Loss: 1.2045,	average_loss: 1.8071
Epoch [41/50],	Iter [140/209]	Loss: 1.7976,	average_loss: 1.8027
Epoch [41/50],	Iter [145/209]	Loss: 1.7140,	average_loss: 1.7987
Epoch [41/50],	Iter [150/209]	Loss: 1.5217,	average_loss: 1.8052
Epoch [41/50],	Iter [155/209]	Loss: 1.6123,	average_loss: 1.7996
Epoch [41/50],	Iter [160/209]	Loss: 1.6966,	average_loss: 1.8139
Epoch [41/50],	Iter [165/209]	Loss: 2.0424,	average_loss: 1.8160
Epoch [41/50],	Iter [170/209]	Loss: 1.2890,	average_loss: 1.8085
Epoch [41/50],	Iter [175/209]	Loss: 1.2128,	average_loss: 1.8034
Epoch [41/50],	Iter [180/209]	Loss: 1.1752,	average_loss: 1.8026
Epoch [41/50],	Iter [185/209]	Loss: 1.4249,	average_loss: 1.7967
Epoch [41/50],	Iter [190/209]	Loss: 2.4053,	average_loss: 1.8064
Epoch [41/50],	Iter [195/209]	Loss: 1.2227,	average_loss: 1.8031
Epoch [41/50],	Iter [200/209]	Loss: 1.8271,	average_loss: 1.8007
Epoch [41/50],	Iter [205/209]	Loss: 1.5484,	average_loss: 1.7956

Starting epoch 42 / 50

Learning Rate for this epoch: 1e-05

Epoch [42/50],	Iter [5/209]	Loss: 1.8283,	average_loss: 1.8045
Epoch [42/50],	Iter [10/209]	Loss: 1.6223,	average_loss: 1.7061
Epoch [42/50],	Iter [15/209]	Loss: 1.7964,	average_loss: 1.8277
Epoch [42/50],	Iter [20/209]	Loss: 2.5928,	average_loss: 1.8476
Epoch [42/50],	Iter [25/209]	Loss: 2.7401,	average_loss: 1.8415
Epoch [42/50],	Iter [30/209]	Loss: 1.3180,	average_loss: 1.8138
Epoch [42/50],	Iter [35/209]	Loss: 1.6744,	average_loss: 1.8192
Epoch [42/50],	Iter [40/209]	Loss: 2.0733,	average_loss: 1.8289
Epoch [42/50],	Iter [45/209]	Loss: 1.2355,	average_loss: 1.8230
Epoch [42/50],	Iter [50/209]	Loss: 1.6818,	average_loss: 1.8238
Epoch [42/50],	Iter [55/209]	Loss: 1.7466,	average_loss: 1.7967
Epoch [42/50],	Iter [60/209]	Loss: 1.6369,	average_loss: 1.8050
Epoch [42/50],	Iter [65/209]	Loss: 1.1513,	average_loss: 1.8021
Epoch [42/50],	Iter [70/209]	Loss: 1.8123,	average_loss: 1.7918
Epoch [42/50],	Iter [75/209]	Loss: 1.3930,	average_loss: 1.7821
Epoch [42/50],	Iter [80/209]	Loss: 1.5857,	average_loss: 1.7738

Epoch [42/50], Iter [85/209] Loss: 1.8168, average_loss: 1.7711
Epoch [42/50], Iter [90/209] Loss: 2.0954, average_loss: 1.7673
Epoch [42/50], Iter [95/209] Loss: 2.4004, average_loss: 1.7743
Epoch [42/50], Iter [100/209] Loss: 3.0056, average_loss: 1.8030
Epoch [42/50], Iter [105/209] Loss: 1.6820, average_loss: 1.8009
Epoch [42/50], Iter [110/209] Loss: 2.1556, average_loss: 1.8016
Epoch [42/50], Iter [115/209] Loss: 1.6835, average_loss: 1.7994
Epoch [42/50], Iter [120/209] Loss: 1.5702, average_loss: 1.8037
Epoch [42/50], Iter [125/209] Loss: 2.2841, average_loss: 1.8022
Epoch [42/50], Iter [130/209] Loss: 2.8402, average_loss: 1.8145
Epoch [42/50], Iter [135/209] Loss: 1.9534, average_loss: 1.8116
Epoch [42/50], Iter [140/209] Loss: 1.5752, average_loss: 1.8052
Epoch [42/50], Iter [145/209] Loss: 1.4798, average_loss: 1.8099
Epoch [42/50], Iter [150/209] Loss: 1.5641, average_loss: 1.8068
Epoch [42/50], Iter [155/209] Loss: 1.3553, average_loss: 1.7999
Epoch [42/50], Iter [160/209] Loss: 1.6356, average_loss: 1.7928
Epoch [42/50], Iter [165/209] Loss: 1.7108, average_loss: 1.7967
Epoch [42/50], Iter [170/209] Loss: 2.4657, average_loss: 1.7937
Epoch [42/50], Iter [175/209] Loss: 2.0536, average_loss: 1.7953
Epoch [42/50], Iter [180/209] Loss: 2.9703, average_loss: 1.8058
Epoch [42/50], Iter [185/209] Loss: 2.3317, average_loss: 1.8061
Epoch [42/50], Iter [190/209] Loss: 1.7439, average_loss: 1.8009
Epoch [42/50], Iter [195/209] Loss: 1.2294, average_loss: 1.7951
Epoch [42/50], Iter [200/209] Loss: 1.4508, average_loss: 1.7936
Epoch [42/50], Iter [205/209] Loss: 2.1071, average_loss: 1.7981

Starting epoch 43 / 50

Learning Rate for this epoch: 1e-05

Epoch [43/50], Iter [5/209] Loss: 1.2747, average_loss: 1.8917
Epoch [43/50], Iter [10/209] Loss: 1.4092, average_loss: 1.7408
Epoch [43/50], Iter [15/209] Loss: 1.9042, average_loss: 1.8049
Epoch [43/50], Iter [20/209] Loss: 2.9512, average_loss: 1.8529
Epoch [43/50], Iter [25/209] Loss: 1.5699, average_loss: 1.8191
Epoch [43/50], Iter [30/209] Loss: 1.4950, average_loss: 1.8082
Epoch [43/50], Iter [35/209] Loss: 1.5345, average_loss: 1.7659
Epoch [43/50], Iter [40/209] Loss: 1.5766, average_loss: 1.7769
Epoch [43/50], Iter [45/209] Loss: 1.8974, average_loss: 1.7776
Epoch [43/50], Iter [50/209] Loss: 1.5139, average_loss: 1.7692
Epoch [43/50], Iter [55/209] Loss: 1.5313, average_loss: 1.7662
Epoch [43/50], Iter [60/209] Loss: 1.5449, average_loss: 1.8062
Epoch [43/50], Iter [65/209] Loss: 1.6277, average_loss: 1.8032
Epoch [43/50], Iter [70/209] Loss: 1.8796, average_loss: 1.7984
Epoch [43/50], Iter [75/209] Loss: 0.8322, average_loss: 1.7864
Epoch [43/50], Iter [80/209] Loss: 1.7735, average_loss: 1.7928
Epoch [43/50], Iter [85/209] Loss: 2.1822, average_loss: 1.7878
Epoch [43/50], Iter [90/209] Loss: 2.2217, average_loss: 1.7864
Epoch [43/50], Iter [95/209] Loss: 1.7600, average_loss: 1.7778
Epoch [43/50], Iter [100/209] Loss: 2.1873, average_loss: 1.7773
Epoch [43/50], Iter [105/209] Loss: 2.2576, average_loss: 1.7688
Epoch [43/50], Iter [110/209] Loss: 1.4666, average_loss: 1.7666
Epoch [43/50], Iter [115/209] Loss: 2.0253, average_loss: 1.7664
Epoch [43/50], Iter [120/209] Loss: 1.0323, average_loss: 1.7625
Epoch [43/50], Iter [125/209] Loss: 1.5408, average_loss: 1.7600
Epoch [43/50], Iter [130/209] Loss: 1.5161, average_loss: 1.7568
Epoch [43/50], Iter [135/209] Loss: 2.5160, average_loss: 1.7647
Epoch [43/50], Iter [140/209] Loss: 2.3335, average_loss: 1.7715
Epoch [43/50], Iter [145/209] Loss: 1.7436, average_loss: 1.7748
Epoch [43/50], Iter [150/209] Loss: 1.4473, average_loss: 1.7729
Epoch [43/50], Iter [155/209] Loss: 1.9613, average_loss: 1.7735
Epoch [43/50], Iter [160/209] Loss: 1.6633, average_loss: 1.7776
Epoch [43/50], Iter [165/209] Loss: 2.0432, average_loss: 1.7792
Epoch [43/50], Iter [170/209] Loss: 2.0280, average_loss: 1.7751

```
Epoch [43/50], Iter [175/209] Loss: 2.2506, average_loss: 1.7789
Epoch [43/50], Iter [180/209] Loss: 2.2908, average_loss: 1.7875
Epoch [43/50], Iter [185/209] Loss: 2.0456, average_loss: 1.7827
Epoch [43/50], Iter [190/209] Loss: 1.5237, average_loss: 1.7812
Epoch [43/50], Iter [195/209] Loss: 1.8966, average_loss: 1.7819
Epoch [43/50], Iter [200/209] Loss: 1.7166, average_loss: 1.7802
Epoch [43/50], Iter [205/209] Loss: 1.5475, average_loss: 1.7841
```

Starting epoch 44 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [44/50], Iter [5/209] Loss: 2.5946, average_loss: 1.9250
Epoch [44/50], Iter [10/209] Loss: 1.9643, average_loss: 1.8156
Epoch [44/50], Iter [15/209] Loss: 1.9373, average_loss: 1.8486
Epoch [44/50], Iter [20/209] Loss: 2.0652, average_loss: 1.8550
Epoch [44/50], Iter [25/209] Loss: 1.8637, average_loss: 1.8436
Epoch [44/50], Iter [30/209] Loss: 1.5760, average_loss: 1.8053
Epoch [44/50], Iter [35/209] Loss: 2.3187, average_loss: 1.8013
Epoch [44/50], Iter [40/209] Loss: 1.7684, average_loss: 1.7944
Epoch [44/50], Iter [45/209] Loss: 1.4888, average_loss: 1.7978
Epoch [44/50], Iter [50/209] Loss: 1.7799, average_loss: 1.8073
Epoch [44/50], Iter [55/209] Loss: 1.9299, average_loss: 1.7900
Epoch [44/50], Iter [60/209] Loss: 1.1837, average_loss: 1.7682
Epoch [44/50], Iter [65/209] Loss: 1.8288, average_loss: 1.7633
Epoch [44/50], Iter [70/209] Loss: 1.7510, average_loss: 1.7733
Epoch [44/50], Iter [75/209] Loss: 1.6483, average_loss: 1.7774
Epoch [44/50], Iter [80/209] Loss: 2.0727, average_loss: 1.7905
Epoch [44/50], Iter [85/209] Loss: 1.9180, average_loss: 1.7911
Epoch [44/50], Iter [90/209] Loss: 2.1180, average_loss: 1.7986
Epoch [44/50], Iter [95/209] Loss: 1.5674, average_loss: 1.7947
Epoch [44/50], Iter [100/209] Loss: 2.0871, average_loss: 1.8061
Epoch [44/50], Iter [105/209] Loss: 1.3912, average_loss: 1.7953
Epoch [44/50], Iter [110/209] Loss: 1.7264, average_loss: 1.7875
Epoch [44/50], Iter [115/209] Loss: 2.0553, average_loss: 1.7830
Epoch [44/50], Iter [120/209] Loss: 1.9717, average_loss: 1.7889
Epoch [44/50], Iter [125/209] Loss: 1.5507, average_loss: 1.7956
Epoch [44/50], Iter [130/209] Loss: 1.2029, average_loss: 1.7871
Epoch [44/50], Iter [135/209] Loss: 2.7995, average_loss: 1.7916
Epoch [44/50], Iter [140/209] Loss: 1.8916, average_loss: 1.8018
Epoch [44/50], Iter [145/209] Loss: 1.7503, average_loss: 1.7997
Epoch [44/50], Iter [150/209] Loss: 2.3418, average_loss: 1.7980
Epoch [44/50], Iter [155/209] Loss: 1.9270, average_loss: 1.8022
Epoch [44/50], Iter [160/209] Loss: 2.1065, average_loss: 1.8008
Epoch [44/50], Iter [165/209] Loss: 1.3890, average_loss: 1.7987
Epoch [44/50], Iter [170/209] Loss: 1.8020, average_loss: 1.7961
Epoch [44/50], Iter [175/209] Loss: 2.2232, average_loss: 1.8034
Epoch [44/50], Iter [180/209] Loss: 1.1994, average_loss: 1.7987
Epoch [44/50], Iter [185/209] Loss: 1.6414, average_loss: 1.7977
Epoch [44/50], Iter [190/209] Loss: 2.5903, average_loss: 1.8008
Epoch [44/50], Iter [195/209] Loss: 1.5837, average_loss: 1.8028
Epoch [44/50], Iter [200/209] Loss: 1.8518, average_loss: 1.8030
Epoch [44/50], Iter [205/209] Loss: 2.8469, average_loss: 1.8070
```

Starting epoch 45 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [45/50], Iter [5/209] Loss: 1.3030, average_loss: 1.8289
Epoch [45/50], Iter [10/209] Loss: 1.8189, average_loss: 1.7351
Epoch [45/50], Iter [15/209] Loss: 2.2052, average_loss: 1.7141
Epoch [45/50], Iter [20/209] Loss: 1.9181, average_loss: 1.7714
Epoch [45/50], Iter [25/209] Loss: 1.8034, average_loss: 1.7601
Epoch [45/50], Iter [30/209] Loss: 1.3526, average_loss: 1.7998
Epoch [45/50], Iter [35/209] Loss: 3.3715, average_loss: 1.8064
```

Epoch	[45/50]	Iter	[40/209]	Loss:	1.7600	average_loss:	1.8099
Epoch	[45/50]	Iter	[45/209]	Loss:	1.5032	average_loss:	1.7920
Epoch	[45/50]	Iter	[50/209]	Loss:	1.6905	average_loss:	1.7907
Epoch	[45/50]	Iter	[55/209]	Loss:	1.8357	average_loss:	1.7953
Epoch	[45/50]	Iter	[60/209]	Loss:	1.7714	average_loss:	1.8145
Epoch	[45/50]	Iter	[65/209]	Loss:	1.1539	average_loss:	1.7889
Epoch	[45/50]	Iter	[70/209]	Loss:	1.5339	average_loss:	1.7800
Epoch	[45/50]	Iter	[75/209]	Loss:	1.7670	average_loss:	1.7730
Epoch	[45/50]	Iter	[80/209]	Loss:	2.1089	average_loss:	1.7861
Epoch	[45/50]	Iter	[85/209]	Loss:	1.4391	average_loss:	1.7767
Epoch	[45/50]	Iter	[90/209]	Loss:	1.2064	average_loss:	1.7631
Epoch	[45/50]	Iter	[95/209]	Loss:	1.2501	average_loss:	1.7479
Epoch	[45/50]	Iter	[100/209]	Loss:	2.5302	average_loss:	1.7453
Epoch	[45/50]	Iter	[105/209]	Loss:	1.1560	average_loss:	1.7385
Epoch	[45/50]	Iter	[110/209]	Loss:	1.2218	average_loss:	1.7327
Epoch	[45/50]	Iter	[115/209]	Loss:	1.7156	average_loss:	1.7413
Epoch	[45/50]	Iter	[120/209]	Loss:	1.7049	average_loss:	1.7524
Epoch	[45/50]	Iter	[125/209]	Loss:	1.7157	average_loss:	1.7538
Epoch	[45/50]	Iter	[130/209]	Loss:	2.4477	average_loss:	1.7497
Epoch	[45/50]	Iter	[135/209]	Loss:	1.2893	average_loss:	1.7569
Epoch	[45/50]	Iter	[140/209]	Loss:	1.5619	average_loss:	1.7571
Epoch	[45/50]	Iter	[145/209]	Loss:	1.2056	average_loss:	1.7585
Epoch	[45/50]	Iter	[150/209]	Loss:	1.3481	average_loss:	1.7548
Epoch	[45/50]	Iter	[155/209]	Loss:	1.9689	average_loss:	1.7508
Epoch	[45/50]	Iter	[160/209]	Loss:	1.8965	average_loss:	1.7495
Epoch	[45/50]	Iter	[165/209]	Loss:	1.4577	average_loss:	1.7478
Epoch	[45/50]	Iter	[170/209]	Loss:	2.1580	average_loss:	1.7502
Epoch	[45/50]	Iter	[175/209]	Loss:	1.5668	average_loss:	1.7537
Epoch	[45/50]	Iter	[180/209]	Loss:	1.9352	average_loss:	1.7543
Epoch	[45/50]	Iter	[185/209]	Loss:	1.6435	average_loss:	1.7638
Epoch	[45/50]	Iter	[190/209]	Loss:	1.7067	average_loss:	1.7680
Epoch	[45/50]	Iter	[195/209]	Loss:	1.2580	average_loss:	1.7688
Epoch	[45/50]	Iter	[200/209]	Loss:	2.6103	average_loss:	1.7755
Epoch	[45/50]	Iter	[205/209]	Loss:	2.1321	average_loss:	1.7760

Starting epoch 46 / 50

Learning Rate for this epoch: 1e-05

Epoch	[46/50]	Iter	[5/209]	Loss:	1.6501	average_loss:	1.8151
Epoch	[46/50]	Iter	[10/209]	Loss:	1.5016	average_loss:	1.7411
Epoch	[46/50]	Iter	[15/209]	Loss:	1.9706	average_loss:	1.7431
Epoch	[46/50]	Iter	[20/209]	Loss:	2.0627	average_loss:	1.7908
Epoch	[46/50]	Iter	[25/209]	Loss:	1.2246	average_loss:	1.7690
Epoch	[46/50]	Iter	[30/209]	Loss:	1.8959	average_loss:	1.7856
Epoch	[46/50]	Iter	[35/209]	Loss:	1.3542	average_loss:	1.7876
Epoch	[46/50]	Iter	[40/209]	Loss:	1.5808	average_loss:	1.7582
Epoch	[46/50]	Iter	[45/209]	Loss:	1.9539	average_loss:	1.7505
Epoch	[46/50]	Iter	[50/209]	Loss:	1.6758	average_loss:	1.7494
Epoch	[46/50]	Iter	[55/209]	Loss:	1.6209	average_loss:	1.7447
Epoch	[46/50]	Iter	[60/209]	Loss:	1.9007	average_loss:	1.7494
Epoch	[46/50]	Iter	[65/209]	Loss:	1.8300	average_loss:	1.7482
Epoch	[46/50]	Iter	[70/209]	Loss:	2.9476	average_loss:	1.7623
Epoch	[46/50]	Iter	[75/209]	Loss:	1.7628	average_loss:	1.7800
Epoch	[46/50]	Iter	[80/209]	Loss:	2.0380	average_loss:	1.7765
Epoch	[46/50]	Iter	[85/209]	Loss:	2.6874	average_loss:	1.7784
Epoch	[46/50]	Iter	[90/209]	Loss:	1.6849	average_loss:	1.7842
Epoch	[46/50]	Iter	[95/209]	Loss:	1.5470	average_loss:	1.7660
Epoch	[46/50]	Iter	[100/209]	Loss:	1.2786	average_loss:	1.7645
Epoch	[46/50]	Iter	[105/209]	Loss:	1.9211	average_loss:	1.7673
Epoch	[46/50]	Iter	[110/209]	Loss:	1.3435	average_loss:	1.7702
Epoch	[46/50]	Iter	[115/209]	Loss:	1.7601	average_loss:	1.7656
Epoch	[46/50]	Iter	[120/209]	Loss:	2.0474	average_loss:	1.7633
Epoch	[46/50]	Iter	[125/209]	Loss:	2.0527	average_loss:	1.7646

Epoch [46/50], Iter [130/209] Loss: 1.6326, average_loss: 1.7754
Epoch [46/50], Iter [135/209] Loss: 1.4930, average_loss: 1.7769
Epoch [46/50], Iter [140/209] Loss: 1.6717, average_loss: 1.7777
Epoch [46/50], Iter [145/209] Loss: 1.7709, average_loss: 1.7764
Epoch [46/50], Iter [150/209] Loss: 1.2565, average_loss: 1.7842
Epoch [46/50], Iter [155/209] Loss: 1.2392, average_loss: 1.7850
Epoch [46/50], Iter [160/209] Loss: 2.4521, average_loss: 1.7946
Epoch [46/50], Iter [165/209] Loss: 1.9606, average_loss: 1.8019
Epoch [46/50], Iter [170/209] Loss: 1.5927, average_loss: 1.8099
Epoch [46/50], Iter [175/209] Loss: 1.3115, average_loss: 1.8076
Epoch [46/50], Iter [180/209] Loss: 1.4621, average_loss: 1.8026
Epoch [46/50], Iter [185/209] Loss: 2.4546, average_loss: 1.8038
Epoch [46/50], Iter [190/209] Loss: 1.5671, average_loss: 1.7972
Epoch [46/50], Iter [195/209] Loss: 1.7799, average_loss: 1.8007
Epoch [46/50], Iter [200/209] Loss: 1.7385, average_loss: 1.7998
Epoch [46/50], Iter [205/209] Loss: 1.9994, average_loss: 1.8014

Starting epoch 47 / 50

Learning Rate for this epoch: 1e-05

Epoch [47/50], Iter [5/209] Loss: 2.1678, average_loss: 1.7517
Epoch [47/50], Iter [10/209] Loss: 1.9811, average_loss: 1.7407
Epoch [47/50], Iter [15/209] Loss: 1.6957, average_loss: 1.7361
Epoch [47/50], Iter [20/209] Loss: 1.3214, average_loss: 1.7346
Epoch [47/50], Iter [25/209] Loss: 2.1448, average_loss: 1.8207
Epoch [47/50], Iter [30/209] Loss: 2.4991, average_loss: 1.8416
Epoch [47/50], Iter [35/209] Loss: 1.4223, average_loss: 1.8250
Epoch [47/50], Iter [40/209] Loss: 1.5200, average_loss: 1.8175
Epoch [47/50], Iter [45/209] Loss: 1.6023, average_loss: 1.7876
Epoch [47/50], Iter [50/209] Loss: 1.3502, average_loss: 1.7933
Epoch [47/50], Iter [55/209] Loss: 1.6545, average_loss: 1.7953
Epoch [47/50], Iter [60/209] Loss: 2.5259, average_loss: 1.7959
Epoch [47/50], Iter [65/209] Loss: 1.8599, average_loss: 1.8102
Epoch [47/50], Iter [70/209] Loss: 1.7135, average_loss: 1.7939
Epoch [47/50], Iter [75/209] Loss: 1.7478, average_loss: 1.8119
Epoch [47/50], Iter [80/209] Loss: 1.6562, average_loss: 1.8024
Epoch [47/50], Iter [85/209] Loss: 1.4571, average_loss: 1.7975
Epoch [47/50], Iter [90/209] Loss: 1.7529, average_loss: 1.7878
Epoch [47/50], Iter [95/209] Loss: 2.0810, average_loss: 1.7940
Epoch [47/50], Iter [100/209] Loss: 1.4608, average_loss: 1.7814
Epoch [47/50], Iter [105/209] Loss: 1.6668, average_loss: 1.7975
Epoch [47/50], Iter [110/209] Loss: 1.6530, average_loss: 1.7950
Epoch [47/50], Iter [115/209] Loss: 1.7856, average_loss: 1.7943
Epoch [47/50], Iter [120/209] Loss: 1.3359, average_loss: 1.7908
Epoch [47/50], Iter [125/209] Loss: 1.5445, average_loss: 1.7840
Epoch [47/50], Iter [130/209] Loss: 1.2722, average_loss: 1.7716
Epoch [47/50], Iter [135/209] Loss: 2.6659, average_loss: 1.7834
Epoch [47/50], Iter [140/209] Loss: 1.6423, average_loss: 1.7776
Epoch [47/50], Iter [145/209] Loss: 1.1101, average_loss: 1.7700
Epoch [47/50], Iter [150/209] Loss: 1.3651, average_loss: 1.7611
Epoch [47/50], Iter [155/209] Loss: 2.9064, average_loss: 1.7607
Epoch [47/50], Iter [160/209] Loss: 2.0785, average_loss: 1.7644
Epoch [47/50], Iter [165/209] Loss: 1.9312, average_loss: 1.7688
Epoch [47/50], Iter [170/209] Loss: 1.9982, average_loss: 1.7778
Epoch [47/50], Iter [175/209] Loss: 2.0165, average_loss: 1.7852
Epoch [47/50], Iter [180/209] Loss: 2.0545, average_loss: 1.7828
Epoch [47/50], Iter [185/209] Loss: 2.0495, average_loss: 1.7846
Epoch [47/50], Iter [190/209] Loss: 2.2899, average_loss: 1.7869
Epoch [47/50], Iter [195/209] Loss: 1.6085, average_loss: 1.7977
Epoch [47/50], Iter [200/209] Loss: 1.8036, average_loss: 1.8008
Epoch [47/50], Iter [205/209] Loss: 2.0609, average_loss: 1.8030
Updating best test loss: 2.68888

Starting epoch 48 / 50

Learning Rate for this epoch: 1e-05

Epoch [48/50], Iter [5/209]	Loss: 1.6338, average_loss: 1.8169
Epoch [48/50], Iter [10/209]	Loss: 1.3953, average_loss: 1.9197
Epoch [48/50], Iter [15/209]	Loss: 2.2478, average_loss: 1.9240
Epoch [48/50], Iter [20/209]	Loss: 0.9396, average_loss: 1.8772
Epoch [48/50], Iter [25/209]	Loss: 2.2116, average_loss: 1.8436
Epoch [48/50], Iter [30/209]	Loss: 1.8376, average_loss: 1.8463
Epoch [48/50], Iter [35/209]	Loss: 1.8511, average_loss: 1.8598
Epoch [48/50], Iter [40/209]	Loss: 1.8471, average_loss: 1.8950
Epoch [48/50], Iter [45/209]	Loss: 2.0068, average_loss: 1.8850
Epoch [48/50], Iter [50/209]	Loss: 1.6858, average_loss: 1.8785
Epoch [48/50], Iter [55/209]	Loss: 1.8353, average_loss: 1.8554
Epoch [48/50], Iter [60/209]	Loss: 2.1060, average_loss: 1.8415
Epoch [48/50], Iter [65/209]	Loss: 1.5083, average_loss: 1.8181
Epoch [48/50], Iter [70/209]	Loss: 1.9552, average_loss: 1.8099
Epoch [48/50], Iter [75/209]	Loss: 1.3888, average_loss: 1.8008
Epoch [48/50], Iter [80/209]	Loss: 1.2401, average_loss: 1.7951
Epoch [48/50], Iter [85/209]	Loss: 1.5202, average_loss: 1.7852
Epoch [48/50], Iter [90/209]	Loss: 1.7952, average_loss: 1.7820
Epoch [48/50], Iter [95/209]	Loss: 1.7343, average_loss: 1.7879
Epoch [48/50], Iter [100/209]	Loss: 1.7332, average_loss: 1.7903
Epoch [48/50], Iter [105/209]	Loss: 1.2995, average_loss: 1.7823
Epoch [48/50], Iter [110/209]	Loss: 2.0278, average_loss: 1.7837
Epoch [48/50], Iter [115/209]	Loss: 2.0567, average_loss: 1.7862
Epoch [48/50], Iter [120/209]	Loss: 1.7396, average_loss: 1.7809
Epoch [48/50], Iter [125/209]	Loss: 2.1112, average_loss: 1.7756
Epoch [48/50], Iter [130/209]	Loss: 1.4754, average_loss: 1.7710
Epoch [48/50], Iter [135/209]	Loss: 1.7129, average_loss: 1.7769
Epoch [48/50], Iter [140/209]	Loss: 2.3067, average_loss: 1.7822
Epoch [48/50], Iter [145/209]	Loss: 2.7644, average_loss: 1.7873
Epoch [48/50], Iter [150/209]	Loss: 1.4820, average_loss: 1.7774
Epoch [48/50], Iter [155/209]	Loss: 1.5923, average_loss: 1.7738
Epoch [48/50], Iter [160/209]	Loss: 1.3685, average_loss: 1.7720
Epoch [48/50], Iter [165/209]	Loss: 1.3761, average_loss: 1.7686
Epoch [48/50], Iter [170/209]	Loss: 1.3431, average_loss: 1.7656
Epoch [48/50], Iter [175/209]	Loss: 2.1367, average_loss: 1.7669
Epoch [48/50], Iter [180/209]	Loss: 1.7610, average_loss: 1.7641
Epoch [48/50], Iter [185/209]	Loss: 1.5348, average_loss: 1.7619
Epoch [48/50], Iter [190/209]	Loss: 2.0509, average_loss: 1.7673
Epoch [48/50], Iter [195/209]	Loss: 1.4909, average_loss: 1.7600
Epoch [48/50], Iter [200/209]	Loss: 1.7759, average_loss: 1.7732
Epoch [48/50], Iter [205/209]	Loss: 2.1171, average_loss: 1.7722

Starting epoch 49 / 50

Learning Rate for this epoch: 1e-05

Epoch [49/50], Iter [5/209]	Loss: 1.6154, average_loss: 1.9328
Epoch [49/50], Iter [10/209]	Loss: 2.4427, average_loss: 1.7562
Epoch [49/50], Iter [15/209]	Loss: 1.9988, average_loss: 1.7990
Epoch [49/50], Iter [20/209]	Loss: 1.7070, average_loss: 1.8304
Epoch [49/50], Iter [25/209]	Loss: 1.2888, average_loss: 1.8268
Epoch [49/50], Iter [30/209]	Loss: 2.0985, average_loss: 1.8367
Epoch [49/50], Iter [35/209]	Loss: 1.6821, average_loss: 1.8514
Epoch [49/50], Iter [40/209]	Loss: 1.4001, average_loss: 1.8296
Epoch [49/50], Iter [45/209]	Loss: 1.9187, average_loss: 1.8105
Epoch [49/50], Iter [50/209]	Loss: 2.1693, average_loss: 1.7899
Epoch [49/50], Iter [55/209]	Loss: 2.0912, average_loss: 1.7814
Epoch [49/50], Iter [60/209]	Loss: 1.6113, average_loss: 1.7717
Epoch [49/50], Iter [65/209]	Loss: 1.7295, average_loss: 1.7821
Epoch [49/50], Iter [70/209]	Loss: 2.1454, average_loss: 1.7914
Epoch [49/50], Iter [75/209]	Loss: 1.4328, average_loss: 1.7780

Epoch [49/50], Iter [80/209] Loss: 1.3959, average_loss: 1.7546
Epoch [49/50], Iter [85/209] Loss: 1.3611, average_loss: 1.7554
Epoch [49/50], Iter [90/209] Loss: 1.5585, average_loss: 1.7547
Epoch [49/50], Iter [95/209] Loss: 1.9761, average_loss: 1.7625
Epoch [49/50], Iter [100/209] Loss: 2.1816, average_loss: 1.7659
Epoch [49/50], Iter [105/209] Loss: 1.5722, average_loss: 1.7651
Epoch [49/50], Iter [110/209] Loss: 1.2717, average_loss: 1.7791
Epoch [49/50], Iter [115/209] Loss: 2.1682, average_loss: 1.7826
Epoch [49/50], Iter [120/209] Loss: 1.7136, average_loss: 1.7873
Epoch [49/50], Iter [125/209] Loss: 1.4535, average_loss: 1.7948
Epoch [49/50], Iter [130/209] Loss: 2.0831, average_loss: 1.7907
Epoch [49/50], Iter [135/209] Loss: 1.4498, average_loss: 1.7823
Epoch [49/50], Iter [140/209] Loss: 1.5958, average_loss: 1.7762
Epoch [49/50], Iter [145/209] Loss: 2.5094, average_loss: 1.7813
Epoch [49/50], Iter [150/209] Loss: 1.5414, average_loss: 1.7856
Epoch [49/50], Iter [155/209] Loss: 2.0010, average_loss: 1.7812
Epoch [49/50], Iter [160/209] Loss: 1.6056, average_loss: 1.7831
Epoch [49/50], Iter [165/209] Loss: 1.9891, average_loss: 1.7827
Epoch [49/50], Iter [170/209] Loss: 2.3339, average_loss: 1.7877
Epoch [49/50], Iter [175/209] Loss: 1.6850, average_loss: 1.7861
Epoch [49/50], Iter [180/209] Loss: 1.9391, average_loss: 1.7892
Epoch [49/50], Iter [185/209] Loss: 1.0628, average_loss: 1.7880
Epoch [49/50], Iter [190/209] Loss: 1.8686, average_loss: 1.7809
Epoch [49/50], Iter [195/209] Loss: 1.6006, average_loss: 1.7840
Epoch [49/50], Iter [200/209] Loss: 2.5044, average_loss: 1.7928
Epoch [49/50], Iter [205/209] Loss: 1.8582, average_loss: 1.7973

Starting epoch 50 / 50

Learning Rate for this epoch: 1e-05

Epoch [50/50], Iter [5/209] Loss: 1.4934, average_loss: 1.7946
Epoch [50/50], Iter [10/209] Loss: 1.3762, average_loss: 1.6999
Epoch [50/50], Iter [15/209] Loss: 1.8213, average_loss: 1.6790
Epoch [50/50], Iter [20/209] Loss: 1.8354, average_loss: 1.7587
Epoch [50/50], Iter [25/209] Loss: 1.8859, average_loss: 1.7454
Epoch [50/50], Iter [30/209] Loss: 1.5038, average_loss: 1.7312
Epoch [50/50], Iter [35/209] Loss: 1.6541, average_loss: 1.7328
Epoch [50/50], Iter [40/209] Loss: 1.9416, average_loss: 1.7711
Epoch [50/50], Iter [45/209] Loss: 1.7386, average_loss: 1.7967
Epoch [50/50], Iter [50/209] Loss: 1.6508, average_loss: 1.7706
Epoch [50/50], Iter [55/209] Loss: 2.2674, average_loss: 1.7731
Epoch [50/50], Iter [60/209] Loss: 1.5274, average_loss: 1.7641
Epoch [50/50], Iter [65/209] Loss: 2.2142, average_loss: 1.7775
Epoch [50/50], Iter [70/209] Loss: 1.6152, average_loss: 1.7702
Epoch [50/50], Iter [75/209] Loss: 1.8727, average_loss: 1.7898
Epoch [50/50], Iter [80/209] Loss: 2.0929, average_loss: 1.7827
Epoch [50/50], Iter [85/209] Loss: 1.3644, average_loss: 1.7836
Epoch [50/50], Iter [90/209] Loss: 1.1911, average_loss: 1.7766
Epoch [50/50], Iter [95/209] Loss: 2.2225, average_loss: 1.7708
Epoch [50/50], Iter [100/209] Loss: 1.5355, average_loss: 1.7608
Epoch [50/50], Iter [105/209] Loss: 1.9257, average_loss: 1.7727
Epoch [50/50], Iter [110/209] Loss: 2.0164, average_loss: 1.7775
Epoch [50/50], Iter [115/209] Loss: 1.7056, average_loss: 1.7763
Epoch [50/50], Iter [120/209] Loss: 1.7418, average_loss: 1.7798
Epoch [50/50], Iter [125/209] Loss: 1.6007, average_loss: 1.7853
Epoch [50/50], Iter [130/209] Loss: 1.9098, average_loss: 1.7919
Epoch [50/50], Iter [135/209] Loss: 2.1010, average_loss: 1.7936
Epoch [50/50], Iter [140/209] Loss: 1.8406, average_loss: 1.7893
Epoch [50/50], Iter [145/209] Loss: 1.7104, average_loss: 1.7923
Epoch [50/50], Iter [150/209] Loss: 2.1769, average_loss: 1.7925
Epoch [50/50], Iter [155/209] Loss: 2.3309, average_loss: 1.8008
Epoch [50/50], Iter [160/209] Loss: 1.2528, average_loss: 1.7928
Epoch [50/50], Iter [165/209] Loss: 1.6374, average_loss: 1.7921

```
Epoch [50/50], Iter [170/209] Loss: 2.2828, average_loss: 1.7920
Epoch [50/50], Iter [175/209] Loss: 1.9006, average_loss: 1.7953
Epoch [50/50], Iter [180/209] Loss: 1.9253, average_loss: 1.7973
Epoch [50/50], Iter [185/209] Loss: 2.3840, average_loss: 1.7986
Epoch [50/50], Iter [190/209] Loss: 1.5074, average_loss: 1.8000
Epoch [50/50], Iter [195/209] Loss: 2.2748, average_loss: 1.8020
Epoch [50/50], Iter [200/209] Loss: 2.1630, average_loss: 1.8023
Epoch [50/50], Iter [205/209] Loss: 1.5887, average_loss: 1.8018
Updating best test loss: 2.68707
```

View example predictions

```
In [13]: 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)

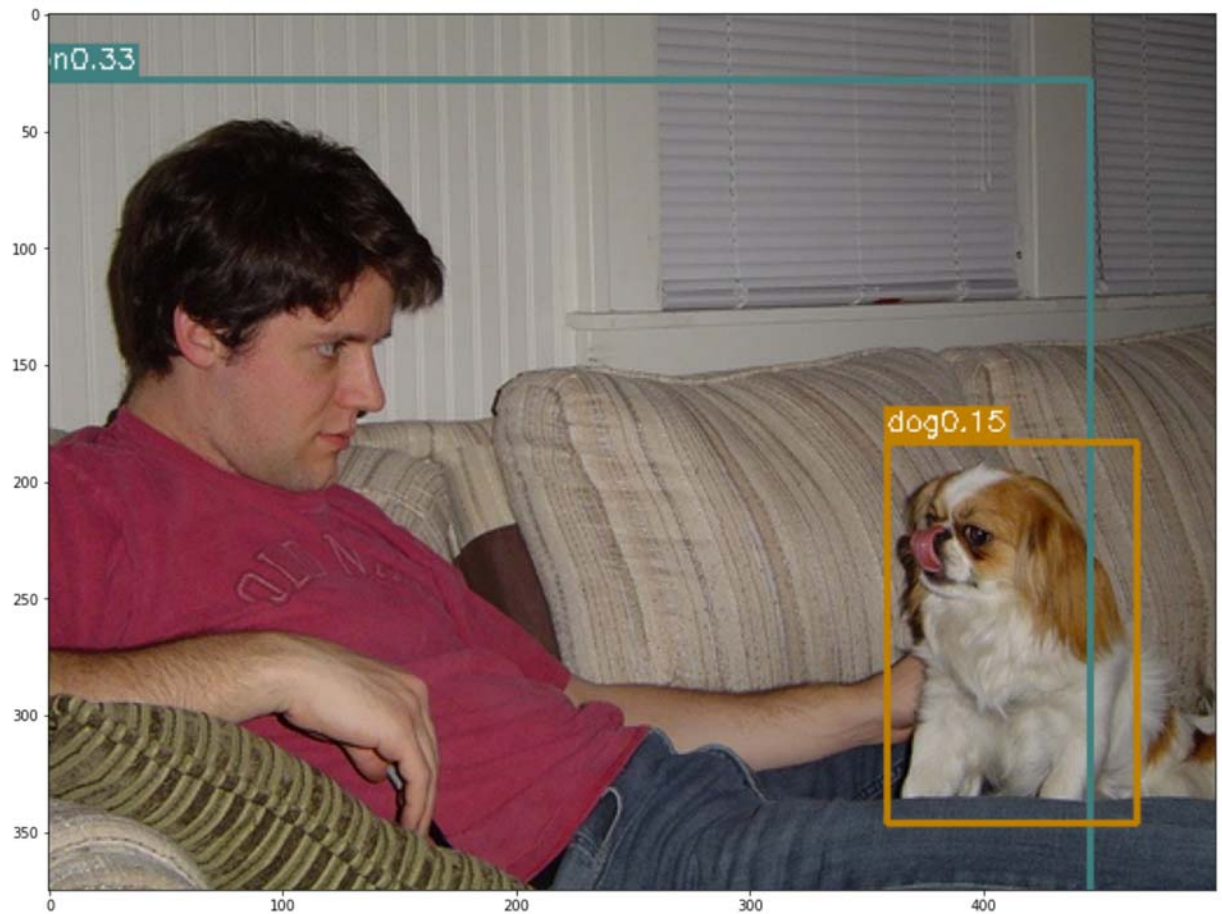
print('predicting...')
result = predict_image(net, image_name, root_img_directory=file_root_test)
for left_up, right_bottom, class_name, _, prob in result:
    color = COLORS[VOC_CLASSES.index(class_name)]
    cv2.rectangle(image, left_up, right_bottom, color, 2)
    label = class_name + str(round(prob,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

```
/content/gdrive/My Drive/Colab Notebooks/Startpkg_A3_P2/predict.py:99: UserWarning: This  
s overload of nonzero is deprecated:  
    nonzero()  
Consider using one of the following signatures instead:  
    nonzero(*, bool as_tuple) (Triggered internally at /pytorch/torch/csrc/utils/p  
ython_arg_parser.cpp:766.)  
    ids = (ovr<=threshold).nonzero().squeeze()
```

Out[13]: <matplotlib.image.AxesImage at 0x7ff18072b128>



Evaluate on Test

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

```
In [14]: test_aps = evaluate(net, test_dataset_file=annotation_file_test)
```

```
---Evaluate model on test samples---
```

```
100%|██████████| 4950/4950 [05:07<00:00, 16.11it/s]
```

```
---class aeroplane ap 0.5105135194524029---  
---class bicycle ap 0.6062137337898297---  
---class bird ap 0.47426056579025777---  
---class boat ap 0.28888942891792607---  
---class bottle ap 0.20556155442569052---  
---class bus ap 0.6319542018240447---  
---class car ap 0.6678370750839007---  
---class cat ap 0.7209403089103789---  
---class chair ap 0.2940296687075428---  
---class cow ap 0.5052405693470969---  
---class diningtable ap 0.33949080737420023---  
---class dog ap 0.6562499515290865---  
---class horse ap 0.6753197126923797---  
---class motorbike ap 0.5639379821931145---  
---class person ap 0.5178361740058215---  
---class pottedplant ap 0.17911496680303257---  
---class sheep ap 0.44120098933947566---  
---class sofa ap 0.49812533816555915---  
---class train ap 0.6233484176384051---  
---class tvmonitor ap 0.47988947967150575---  
---map 0.49399772228308264---
```

```
In [15]: output_submission_csv('my_solution_50.csv', test_aps)
```