

In [15]:

```
#!/sudo pip install tqdm
```

In [16]:

```
#!/sh download_data.sh
```

In [17]:

```
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
```

The autoreload extension is already loaded. To reload it, use:  
%reload\_ext autoreload

## Initialization

In [18]:

```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
```

In [19]:

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

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

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

In [20]:

```
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(pretrained=pretrained).to(device)
```

```

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

In [21]:

```

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 [22]:

```

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  $S \times S (B \times 5 + C)$ , 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 [23]:

```

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=1)
print('Loaded %d train images' % len(train_dataset))

```

Initializing dataset  
Loaded 5011 train images

In [24]:

```

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=1)
print('Loaded %d test images' % len(test_dataset))

```

Initializing dataset  
Loaded 4950 test images

## Train detector

In [25]:

```
in [20]:
```

```
best_test_loss = np.inf

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

    # Update learning rate late in training
    if epoch == 30 or epoch == 40:
        learning_rate /= 10.0

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

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

    total_loss = 0.
    for i, (images, target) in enumerate(train_loader):
        images, target = images.to(device), target.to(device)

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

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

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

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

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

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

Starting epoch 1 / 50

Learning Rate for this epoch: 0.001

```
Epoch [1/50], Iter [5/209] Loss: 67.2155, average_loss: 70.9981
Epoch [1/50], Iter [10/209] Loss: 47.8315, average_loss: 62.4808
Epoch [1/50], Iter [15/209] Loss: 25.7877, average_loss: 52.3723
Epoch [1/50], Iter [20/209] Loss: 19.5058, average_loss: 44.1059
Epoch [1/50], Iter [25/209] Loss: 13.0336, average_loss: 38.2586
Epoch [1/50], Iter [30/209] Loss: 9.6983, average_loss: 34.2120
Epoch [1/50], Iter [35/209] Loss: 9.6680, average_loss: 30.6841
Epoch [1/50], Iter [40/209] Loss: 6.8771, average_loss: 27.9143
Epoch [1/50], Iter [45/209] Loss: 6.7617, average_loss: 25.7489
Epoch [1/50], Iter [50/209] Loss: 10.2246, average_loss: 24.1042
Epoch [1/50], Iter [55/209] Loss: 7.3594, average_loss: 22.6484
Epoch [1/50], Iter [60/209] Loss: 7.8168, average_loss: 21.3239
Epoch [1/50], Iter [65/209] Loss: 9.7960, average_loss: 20.2466
Epoch [1/50], Iter [70/209] Loss: 5.7438, average_loss: 19.2530
Epoch [1/50], Iter [75/209] Loss: 6.9141, average_loss: 18.4133
Epoch [1/50], Iter [80/209] Loss: 6.8961, average_loss: 17.7392
Epoch [1/50], Iter [85/209] Loss: 8.9098, average_loss: 17.0794
Epoch [1/50], Iter [90/209] Loss: 5.6118, average_loss: 16.4832
Epoch [1/50], Iter [95/209] Loss: 5.6923, average_loss: 16.0000
Epoch [1/50], Iter [100/209] Loss: 7.4352, average_loss: 15.5257
Epoch [1/50], Iter [105/209] Loss: 5.5356, average_loss: 15.0413
Epoch [1/50], Iter [110/209] Loss: 5.4649, average_loss: 14.6649
```

```
Epoch [1/50], Iter [115/209] Loss: 7.8590, average_loss: 14.2810
Epoch [1/50], Iter [120/209] Loss: 6.0234, average_loss: 13.9644
Epoch [1/50], Iter [125/209] Loss: 7.4063, average_loss: 13.6373
Epoch [1/50], Iter [130/209] Loss: 5.0187, average_loss: 13.3286
Epoch [1/50], Iter [135/209] Loss: 5.5023, average_loss: 13.0457
Epoch [1/50], Iter [140/209] Loss: 5.2404, average_loss: 12.7664
Epoch [1/50], Iter [145/209] Loss: 5.2276, average_loss: 12.5144
Epoch [1/50], Iter [150/209] Loss: 5.1599, average_loss: 12.2770
Epoch [1/50], Iter [155/209] Loss: 5.3581, average_loss: 12.0667
Epoch [1/50], Iter [160/209] Loss: 4.8162, average_loss: 11.8593
Epoch [1/50], Iter [165/209] Loss: 6.6004, average_loss: 11.6576
Epoch [1/50], Iter [170/209] Loss: 5.2079, average_loss: 11.4761
Epoch [1/50], Iter [175/209] Loss: 7.1889, average_loss: 11.3192
Epoch [1/50], Iter [180/209] Loss: 4.5768, average_loss: 11.1529
Epoch [1/50], Iter [185/209] Loss: 5.7186, average_loss: 10.9947
Epoch [1/50], Iter [190/209] Loss: 5.7823, average_loss: 10.8615
Epoch [1/50], Iter [195/209] Loss: 5.3953, average_loss: 10.7284
Epoch [1/50], Iter [200/209] Loss: 5.2408, average_loss: 10.5850
Epoch [1/50], Iter [205/209] Loss: 6.5358, average_loss: 10.4643
Updating best test loss: 5.18534
```

Starting epoch 2 / 50

Learning Rate for this epoch: 0.001

```
Epoch [2/50], Iter [5/209] Loss: 3.7051, average_loss: 4.6105
Epoch [2/50], Iter [10/209] Loss: 5.2076, average_loss: 4.9877
Epoch [2/50], Iter [15/209] Loss: 4.1081, average_loss: 4.8767
Epoch [2/50], Iter [20/209] Loss: 5.0090, average_loss: 4.8079
Epoch [2/50], Iter [25/209] Loss: 4.1490, average_loss: 4.7537
Epoch [2/50], Iter [30/209] Loss: 4.5421, average_loss: 4.7540
Epoch [2/50], Iter [35/209] Loss: 6.0813, average_loss: 4.7720
Epoch [2/50], Iter [40/209] Loss: 6.7608, average_loss: 4.8783
Epoch [2/50], Iter [45/209] Loss: 5.3111, average_loss: 4.8636
Epoch [2/50], Iter [50/209] Loss: 4.7211, average_loss: 4.9016
Epoch [2/50], Iter [55/209] Loss: 5.3060, average_loss: 4.9273
Epoch [2/50], Iter [60/209] Loss: 5.0483, average_loss: 4.9359
Epoch [2/50], Iter [65/209] Loss: 3.6471, average_loss: 4.9896
Epoch [2/50], Iter [70/209] Loss: 4.2487, average_loss: 4.9730
Epoch [2/50], Iter [75/209] Loss: 4.1013, average_loss: 4.9629
Epoch [2/50], Iter [80/209] Loss: 4.7614, average_loss: 4.9650
Epoch [2/50], Iter [85/209] Loss: 3.3532, average_loss: 4.9610
Epoch [2/50], Iter [90/209] Loss: 5.8858, average_loss: 4.9664
Epoch [2/50], Iter [95/209] Loss: 4.0984, average_loss: 4.9544
Epoch [2/50], Iter [100/209] Loss: 4.9075, average_loss: 4.9683
Epoch [2/50], Iter [105/209] Loss: 5.1053, average_loss: 4.9737
Epoch [2/50], Iter [110/209] Loss: 4.8907, average_loss: 4.9702
Epoch [2/50], Iter [115/209] Loss: 4.9298, average_loss: 4.9516
Epoch [2/50], Iter [120/209] Loss: 6.8156, average_loss: 4.9562
Epoch [2/50], Iter [125/209] Loss: 5.2671, average_loss: 4.9384
Epoch [2/50], Iter [130/209] Loss: 3.2537, average_loss: 4.9198
Epoch [2/50], Iter [135/209] Loss: 4.5144, average_loss: 4.9229
Epoch [2/50], Iter [140/209] Loss: 3.9579, average_loss: 4.9017
Epoch [2/50], Iter [145/209] Loss: 4.2464, average_loss: 4.8830
Epoch [2/50], Iter [150/209] Loss: 4.7739, average_loss: 4.9006
Epoch [2/50], Iter [155/209] Loss: 4.8967, average_loss: 4.8968
Epoch [2/50], Iter [160/209] Loss: 3.8005, average_loss: 4.8895
Epoch [2/50], Iter [165/209] Loss: 4.1203, average_loss: 4.8794
Epoch [2/50], Iter [170/209] Loss: 4.7620, average_loss: 4.8801
Epoch [2/50], Iter [175/209] Loss: 4.5938, average_loss: 4.8589
Epoch [2/50], Iter [180/209] Loss: 4.8908, average_loss: 4.8457
Epoch [2/50], Iter [185/209] Loss: 4.9191, average_loss: 4.8267
Epoch [2/50], Iter [190/209] Loss: 5.5852, average_loss: 4.8294
Epoch [2/50], Iter [195/209] Loss: 4.5926, average_loss: 4.8536
Epoch [2/50], Iter [200/209] Loss: 3.4260, average_loss: 4.8488
Epoch [2/50], Iter [205/209] Loss: 4.4515, average_loss: 4.8312
Updating best test loss: 4.66971
```

Starting epoch 3 / 50

Learning Rate for this epoch: 0.001

```
Epoch [3/50], Iter [5/209] Loss: 4.1219, average_loss: 4.1749
Epoch [3/50], Iter [10/209] Loss: 3.6168, average_loss: 4.3138
Epoch [3/50], Iter [15/209] Loss: 4.4659, average_loss: 4.1912
Epoch [3/50], Iter [20/209] Loss: 4.2843, average_loss: 4.3950
Epoch [3/50], Iter [25/209] Loss: 3.9747, average_loss: 4.3119
Epoch [3/50], Iter [30/209] Loss: 4.4362, average_loss: 4.3433
Epoch [3/50], Iter [35/209] Loss: 3.8152, average_loss: 4.3777
```

Epoch [3/50], Iter [40/209] Loss: 5.1790, average\_loss: 4.4235  
Epoch [3/50], Iter [45/209] Loss: 4.1042, average\_loss: 4.3947  
Epoch [3/50], Iter [50/209] Loss: 4.4401, average\_loss: 4.4185  
Epoch [3/50], Iter [55/209] Loss: 5.3325, average\_loss: 4.4774  
Epoch [3/50], Iter [60/209] Loss: 4.1682, average\_loss: 4.4783  
Epoch [3/50], Iter [65/209] Loss: 4.4795, average\_loss: 4.4501  
Epoch [3/50], Iter [70/209] Loss: 4.4269, average\_loss: 4.4687  
Epoch [3/50], Iter [75/209] Loss: 3.2028, average\_loss: 4.4255  
Epoch [3/50], Iter [80/209] Loss: 3.7564, average\_loss: 4.4098  
Epoch [3/50], Iter [85/209] Loss: 3.5700, average\_loss: 4.4014  
Epoch [3/50], Iter [90/209] Loss: 3.7580, average\_loss: 4.3862  
Epoch [3/50], Iter [95/209] Loss: 3.7456, average\_loss: 4.3535  
Epoch [3/50], Iter [100/209] Loss: 4.7513, average\_loss: 4.3503  
Epoch [3/50], Iter [105/209] Loss: 3.9606, average\_loss: 4.3472  
Epoch [3/50], Iter [110/209] Loss: 6.2783, average\_loss: 4.3684  
Epoch [3/50], Iter [115/209] Loss: 4.9911, average\_loss: 4.3762  
Epoch [3/50], Iter [120/209] Loss: 4.8602, average\_loss: 4.3980  
Epoch [3/50], Iter [125/209] Loss: 5.5872, average\_loss: 4.3861  
Epoch [3/50], Iter [130/209] Loss: 3.7357, average\_loss: 4.3729  
Epoch [3/50], Iter [135/209] Loss: 3.4892, average\_loss: 4.3598  
Epoch [3/50], Iter [140/209] Loss: 3.9317, average\_loss: 4.3680  
Epoch [3/50], Iter [145/209] Loss: 3.5757, average\_loss: 4.3656  
Epoch [3/50], Iter [150/209] Loss: 4.6670, average\_loss: 4.3612  
Epoch [3/50], Iter [155/209] Loss: 3.5130, average\_loss: 4.3559  
Epoch [3/50], Iter [160/209] Loss: 5.3386, average\_loss: 4.3770  
Epoch [3/50], Iter [165/209] Loss: 4.4135, average\_loss: 4.3802  
Epoch [3/50], Iter [170/209] Loss: 4.0339, average\_loss: 4.3641  
Epoch [3/50], Iter [175/209] Loss: 4.1289, average\_loss: 4.3524  
Epoch [3/50], Iter [180/209] Loss: 4.9960, average\_loss: 4.3512  
Epoch [3/50], Iter [185/209] Loss: 3.4532, average\_loss: 4.3549  
Epoch [3/50], Iter [190/209] Loss: 4.1467, average\_loss: 4.3359  
Epoch [3/50], Iter [195/209] Loss: 4.3771, average\_loss: 4.3176  
Epoch [3/50], Iter [200/209] Loss: 3.8478, average\_loss: 4.2991  
Epoch [3/50], Iter [205/209] Loss: 4.7060, average\_loss: 4.3064  
Updating best test loss: 4.30669

Starting epoch 4 / 50

Learning Rate for this epoch: 0.001

Epoch [4/50], Iter [5/209] Loss: 2.9256, average\_loss: 3.8211  
Epoch [4/50], Iter [10/209] Loss: 3.9854, average\_loss: 3.7124  
Epoch [4/50], Iter [15/209] Loss: 5.1942, average\_loss: 4.0564  
Epoch [4/50], Iter [20/209] Loss: 3.1558, average\_loss: 3.9536  
Epoch [4/50], Iter [25/209] Loss: 4.1251, average\_loss: 3.9167  
Epoch [4/50], Iter [30/209] Loss: 3.4701, average\_loss: 3.8944  
Epoch [4/50], Iter [35/209] Loss: 3.3816, average\_loss: 3.9495  
Epoch [4/50], Iter [40/209] Loss: 4.2888, average\_loss: 3.9879  
Epoch [4/50], Iter [45/209] Loss: 4.0315, average\_loss: 3.9800  
Epoch [4/50], Iter [50/209] Loss: 3.5850, average\_loss: 3.9616  
Epoch [4/50], Iter [55/209] Loss: 4.1171, average\_loss: 3.9840  
Epoch [4/50], Iter [60/209] Loss: 5.9520, average\_loss: 4.0433  
Epoch [4/50], Iter [65/209] Loss: 3.5985, average\_loss: 4.0302  
Epoch [4/50], Iter [70/209] Loss: 3.1343, average\_loss: 4.0662  
Epoch [4/50], Iter [75/209] Loss: 4.9873, average\_loss: 4.0894  
Epoch [4/50], Iter [80/209] Loss: 3.0223, average\_loss: 4.0677  
Epoch [4/50], Iter [85/209] Loss: 4.6022, average\_loss: 4.0331  
Epoch [4/50], Iter [90/209] Loss: 3.5235, average\_loss: 4.0489  
Epoch [4/50], Iter [95/209] Loss: 3.9815, average\_loss: 4.0468  
Epoch [4/50], Iter [100/209] Loss: 4.2837, average\_loss: 4.0500  
Epoch [4/50], Iter [105/209] Loss: 3.4243, average\_loss: 4.0457  
Epoch [4/50], Iter [110/209] Loss: 5.6630, average\_loss: 4.0703  
Epoch [4/50], Iter [115/209] Loss: 4.2717, average\_loss: 4.0656  
Epoch [4/50], Iter [120/209] Loss: 4.0109, average\_loss: 4.0581  
Epoch [4/50], Iter [125/209] Loss: 3.2950, average\_loss: 4.0421  
Epoch [4/50], Iter [130/209] Loss: 3.8418, average\_loss: 4.0210  
Epoch [4/50], Iter [135/209] Loss: 4.7885, average\_loss: 4.0389  
Epoch [4/50], Iter [140/209] Loss: 3.8187, average\_loss: 4.0352  
Epoch [4/50], Iter [145/209] Loss: 2.9081, average\_loss: 4.0186  
Epoch [4/50], Iter [150/209] Loss: 3.7400, average\_loss: 4.0373  
Epoch [4/50], Iter [155/209] Loss: 3.0481, average\_loss: 4.0255  
Epoch [4/50], Iter [160/209] Loss: 3.3191, average\_loss: 4.0203  
Epoch [4/50], Iter [165/209] Loss: 4.0002, average\_loss: 4.0197  
Epoch [4/50], Iter [170/209] Loss: 4.1137, average\_loss: 4.0252  
Epoch [4/50], Iter [175/209] Loss: 3.8313, average\_loss: 4.0182  
Epoch [4/50], Iter [180/209] Loss: 3.4243, average\_loss: 4.0304  
Epoch [4/50], Iter [185/209] Loss: 3.5421, average\_loss: 4.0190  
Epoch [4/50], Iter [190/209] Loss: 5.3018, average\_loss: 4.0195

Epoch [4/50], Iter [195/209] Loss: 3.7680, average\_loss: 4.0019  
Epoch [4/50], Iter [200/209] Loss: 3.6118, average\_loss: 3.9940  
Epoch [4/50], Iter [205/209] Loss: 4.2968, average\_loss: 3.9996  
Updating best test loss: 3.99479

Starting epoch 5 / 50

Learning Rate for this epoch: 0.001

Epoch [5/50], Iter [5/209] Loss: 4.9093, average\_loss: 3.8492  
Epoch [5/50], Iter [10/209] Loss: 4.8398, average\_loss: 3.8710  
Epoch [5/50], Iter [15/209] Loss: 4.2983, average\_loss: 3.8012  
Epoch [5/50], Iter [20/209] Loss: 3.5634, average\_loss: 3.7437  
Epoch [5/50], Iter [25/209] Loss: 3.0908, average\_loss: 3.7853  
Epoch [5/50], Iter [30/209] Loss: 2.8906, average\_loss: 3.7015  
Epoch [5/50], Iter [35/209] Loss: 4.0678, average\_loss: 3.7085  
Epoch [5/50], Iter [40/209] Loss: 3.1959, average\_loss: 3.7151  
Epoch [5/50], Iter [45/209] Loss: 3.7396, average\_loss: 3.6940  
Epoch [5/50], Iter [50/209] Loss: 4.8018, average\_loss: 3.7175  
Epoch [5/50], Iter [55/209] Loss: 3.6711, average\_loss: 3.7158  
Epoch [5/50], Iter [60/209] Loss: 3.6906, average\_loss: 3.6988  
Epoch [5/50], Iter [65/209] Loss: 3.3759, average\_loss: 3.6787  
Epoch [5/50], Iter [70/209] Loss: 4.8172, average\_loss: 3.6999  
Epoch [5/50], Iter [75/209] Loss: 2.5364, average\_loss: 3.7005  
Epoch [5/50], Iter [80/209] Loss: 3.9013, average\_loss: 3.7536  
Epoch [5/50], Iter [85/209] Loss: 3.9330, average\_loss: 3.7437  
Epoch [5/50], Iter [90/209] Loss: 4.6021, average\_loss: 3.7771  
Epoch [5/50], Iter [95/209] Loss: 4.0140, average\_loss: 3.7677  
Epoch [5/50], Iter [100/209] Loss: 4.0219, average\_loss: 3.7713  
Epoch [5/50], Iter [105/209] Loss: 4.6416, average\_loss: 3.7912  
Epoch [5/50], Iter [110/209] Loss: 2.8990, average\_loss: 3.7832  
Epoch [5/50], Iter [115/209] Loss: 3.1294, average\_loss: 3.7736  
Epoch [5/50], Iter [120/209] Loss: 4.8867, average\_loss: 3.7763  
Epoch [5/50], Iter [125/209] Loss: 2.6872, average\_loss: 3.7719  
Epoch [5/50], Iter [130/209] Loss: 4.6054, average\_loss: 3.7822  
Epoch [5/50], Iter [135/209] Loss: 3.2993, average\_loss: 3.7761  
Epoch [5/50], Iter [140/209] Loss: 4.4443, average\_loss: 3.7866  
Epoch [5/50], Iter [145/209] Loss: 3.7756, average\_loss: 3.7911  
Epoch [5/50], Iter [150/209] Loss: 3.4134, average\_loss: 3.7774  
Epoch [5/50], Iter [155/209] Loss: 4.9532, average\_loss: 3.7760  
Epoch [5/50], Iter [160/209] Loss: 5.2077, average\_loss: 3.7941  
Epoch [5/50], Iter [165/209] Loss: 4.0409, average\_loss: 3.7933  
Epoch [5/50], Iter [170/209] Loss: 4.1896, average\_loss: 3.7902  
Epoch [5/50], Iter [175/209] Loss: 4.2680, average\_loss: 3.7876  
Epoch [5/50], Iter [180/209] Loss: 3.5392, average\_loss: 3.7782  
Epoch [5/50], Iter [185/209] Loss: 3.9026, average\_loss: 3.7762  
Epoch [5/50], Iter [190/209] Loss: 3.5079, average\_loss: 3.7805  
Epoch [5/50], Iter [195/209] Loss: 3.3106, average\_loss: 3.7783  
Epoch [5/50], Iter [200/209] Loss: 5.0551, average\_loss: 3.7848  
Epoch [5/50], Iter [205/209] Loss: 3.0308, average\_loss: 3.7697  
Updating best test loss: 3.82717

Starting epoch 6 / 50

Learning Rate for this epoch: 0.001

Epoch [6/50], Iter [5/209] Loss: 3.1949, average\_loss: 3.5618  
Epoch [6/50], Iter [10/209] Loss: 3.4558, average\_loss: 3.4673  
Epoch [6/50], Iter [15/209] Loss: 4.0123, average\_loss: 3.5352  
Epoch [6/50], Iter [20/209] Loss: 3.7137, average\_loss: 3.5618  
Epoch [6/50], Iter [25/209] Loss: 3.0074, average\_loss: 3.5091  
Epoch [6/50], Iter [30/209] Loss: 4.2111, average\_loss: 3.6398  
Epoch [6/50], Iter [35/209] Loss: 4.8524, average\_loss: 3.6843  
Epoch [6/50], Iter [40/209] Loss: 4.7263, average\_loss: 3.6814  
Epoch [6/50], Iter [45/209] Loss: 3.5394, average\_loss: 3.6525  
Epoch [6/50], Iter [50/209] Loss: 4.0405, average\_loss: 3.6995  
Epoch [6/50], Iter [55/209] Loss: 3.0033, average\_loss: 3.6405  
Epoch [6/50], Iter [60/209] Loss: 3.1812, average\_loss: 3.6530  
Epoch [6/50], Iter [65/209] Loss: 2.6107, average\_loss: 3.6518  
Epoch [6/50], Iter [70/209] Loss: 3.7675, average\_loss: 3.6389  
Epoch [6/50], Iter [75/209] Loss: 3.4904, average\_loss: 3.6103  
Epoch [6/50], Iter [80/209] Loss: 3.7156, average\_loss: 3.6275  
Epoch [6/50], Iter [85/209] Loss: 2.9519, average\_loss: 3.6290  
Epoch [6/50], Iter [90/209] Loss: 2.8776, average\_loss: 3.6134  
Epoch [6/50], Iter [95/209] Loss: 3.4198, average\_loss: 3.6253  
Epoch [6/50], Iter [100/209] Loss: 3.7636, average\_loss: 3.6215  
Epoch [6/50], Iter [105/209] Loss: 2.7553, average\_loss: 3.6018  
Epoch [6/50], Iter [110/209] Loss: 3.5255, average\_loss: 3.6008  
Epoch [6/50], Iter [115/209] Loss: 2.7771, average\_loss: 3.5875

Epoch [6/50], Iter [120/209] Loss: 2.9143, average\_loss: 3.6021  
Epoch [6/50], Iter [125/209] Loss: 5.2027, average\_loss: 3.6129  
Epoch [6/50], Iter [130/209] Loss: 2.8404, average\_loss: 3.6114  
Epoch [6/50], Iter [135/209] Loss: 3.8802, average\_loss: 3.6054  
Epoch [6/50], Iter [140/209] Loss: 2.9439, average\_loss: 3.5983  
Epoch [6/50], Iter [145/209] Loss: 3.4003, average\_loss: 3.5987  
Epoch [6/50], Iter [150/209] Loss: 3.2317, average\_loss: 3.5843  
Epoch [6/50], Iter [155/209] Loss: 2.5803, average\_loss: 3.5612  
Epoch [6/50], Iter [160/209] Loss: 4.5560, average\_loss: 3.5688  
Epoch [6/50], Iter [165/209] Loss: 3.1190, average\_loss: 3.5615  
Epoch [6/50], Iter [170/209] Loss: 4.3530, average\_loss: 3.5652  
Epoch [6/50], Iter [175/209] Loss: 3.9592, average\_loss: 3.5718  
Epoch [6/50], Iter [180/209] Loss: 4.1419, average\_loss: 3.5709  
Epoch [6/50], Iter [185/209] Loss: 2.9049, average\_loss: 3.5636  
Epoch [6/50], Iter [190/209] Loss: 3.9691, average\_loss: 3.5707  
Epoch [6/50], Iter [195/209] Loss: 3.3918, average\_loss: 3.5640  
Epoch [6/50], Iter [200/209] Loss: 3.1093, average\_loss: 3.5650  
Epoch [6/50], Iter [205/209] Loss: 2.3676, average\_loss: 3.5508  
Updating best test loss: 3.66931

Starting epoch 7 / 50

Learning Rate for this epoch: 0.001

Epoch [7/50], Iter [5/209] Loss: 2.9539, average\_loss: 3.4946  
Epoch [7/50], Iter [10/209] Loss: 3.0983, average\_loss: 3.3095  
Epoch [7/50], Iter [15/209] Loss: 2.8131, average\_loss: 3.2867  
Epoch [7/50], Iter [20/209] Loss: 3.1884, average\_loss: 3.3771  
Epoch [7/50], Iter [25/209] Loss: 4.8906, average\_loss: 3.5435  
Epoch [7/50], Iter [30/209] Loss: 3.8812, average\_loss: 3.4990  
Epoch [7/50], Iter [35/209] Loss: 2.9758, average\_loss: 3.4898  
Epoch [7/50], Iter [40/209] Loss: 2.7845, average\_loss: 3.4730  
Epoch [7/50], Iter [45/209] Loss: 2.8341, average\_loss: 3.4504  
Epoch [7/50], Iter [50/209] Loss: 3.2725, average\_loss: 3.4305  
Epoch [7/50], Iter [55/209] Loss: 3.2172, average\_loss: 3.4263  
Epoch [7/50], Iter [60/209] Loss: 2.1739, average\_loss: 3.4521  
Epoch [7/50], Iter [65/209] Loss: 3.7462, average\_loss: 3.5134  
Epoch [7/50], Iter [70/209] Loss: 3.2440, average\_loss: 3.4941  
Epoch [7/50], Iter [75/209] Loss: 3.0149, average\_loss: 3.4983  
Epoch [7/50], Iter [80/209] Loss: 3.8867, average\_loss: 3.4953  
Epoch [7/50], Iter [85/209] Loss: 4.1400, average\_loss: 3.5219  
Epoch [7/50], Iter [90/209] Loss: 3.1964, average\_loss: 3.5126  
Epoch [7/50], Iter [95/209] Loss: 3.5377, average\_loss: 3.4932  
Epoch [7/50], Iter [100/209] Loss: 3.3425, average\_loss: 3.4936  
Epoch [7/50], Iter [105/209] Loss: 3.1833, average\_loss: 3.4931  
Epoch [7/50], Iter [110/209] Loss: 2.4320, average\_loss: 3.4692  
Epoch [7/50], Iter [115/209] Loss: 3.3961, average\_loss: 3.4686  
Epoch [7/50], Iter [120/209] Loss: 4.5229, average\_loss: 3.4654  
Epoch [7/50], Iter [125/209] Loss: 3.7381, average\_loss: 3.4562  
Epoch [7/50], Iter [130/209] Loss: 3.8360, average\_loss: 3.4515  
Epoch [7/50], Iter [135/209] Loss: 2.8167, average\_loss: 3.4471  
Epoch [7/50], Iter [140/209] Loss: 3.2660, average\_loss: 3.4464  
Epoch [7/50], Iter [145/209] Loss: 3.0373, average\_loss: 3.4353  
Epoch [7/50], Iter [150/209] Loss: 4.0798, average\_loss: 3.4363  
Epoch [7/50], Iter [155/209] Loss: 3.3405, average\_loss: 3.4454  
Epoch [7/50], Iter [160/209] Loss: 4.8864, average\_loss: 3.4417  
Epoch [7/50], Iter [165/209] Loss: 3.3275, average\_loss: 3.4272  
Epoch [7/50], Iter [170/209] Loss: 3.2449, average\_loss: 3.4314  
Epoch [7/50], Iter [175/209] Loss: 3.3245, average\_loss: 3.4316  
Epoch [7/50], Iter [180/209] Loss: 3.7692, average\_loss: 3.4391  
Epoch [7/50], Iter [185/209] Loss: 3.3807, average\_loss: 3.4235  
Epoch [7/50], Iter [190/209] Loss: 2.8933, average\_loss: 3.4136  
Epoch [7/50], Iter [195/209] Loss: 4.0844, average\_loss: 3.4113  
Epoch [7/50], Iter [200/209] Loss: 4.5169, average\_loss: 3.4275  
Epoch [7/50], Iter [205/209] Loss: 2.7075, average\_loss: 3.4211  
Updating best test loss: 3.53606

Starting epoch 8 / 50

Learning Rate for this epoch: 0.001

Epoch [8/50], Iter [5/209] Loss: 2.8297, average\_loss: 3.1174  
Epoch [8/50], Iter [10/209] Loss: 2.7011, average\_loss: 3.2254  
Epoch [8/50], Iter [15/209] Loss: 3.3240, average\_loss: 3.1804  
Epoch [8/50], Iter [20/209] Loss: 2.6306, average\_loss: 3.1890  
Epoch [8/50], Iter [25/209] Loss: 4.4191, average\_loss: 3.1933  
Epoch [8/50], Iter [30/209] Loss: 3.6864, average\_loss: 3.1774  
Epoch [8/50], Iter [35/209] Loss: 2.9113, average\_loss: 3.1940  
Epoch [8/50], Iter [40/209] Loss: 3.5398, average\_loss: 3.2426

Epoch [8/50], Iter [45/209] Loss: 2.5734, average\_loss: 3.2309  
Epoch [8/50], Iter [50/209] Loss: 4.1232, average\_loss: 3.2985  
Epoch [8/50], Iter [55/209] Loss: 3.4858, average\_loss: 3.3036  
Epoch [8/50], Iter [60/209] Loss: 3.5014, average\_loss: 3.3307  
Epoch [8/50], Iter [65/209] Loss: 3.6871, average\_loss: 3.3325  
Epoch [8/50], Iter [70/209] Loss: 2.8852, average\_loss: 3.3260  
Epoch [8/50], Iter [75/209] Loss: 2.7670, average\_loss: 3.3148  
Epoch [8/50], Iter [80/209] Loss: 3.6882, average\_loss: 3.3094  
Epoch [8/50], Iter [85/209] Loss: 2.8069, average\_loss: 3.3082  
Epoch [8/50], Iter [90/209] Loss: 2.3375, average\_loss: 3.2815  
Epoch [8/50], Iter [95/209] Loss: 3.6171, average\_loss: 3.2991  
Epoch [8/50], Iter [100/209] Loss: 3.9020, average\_loss: 3.3135  
Epoch [8/50], Iter [105/209] Loss: 3.1951, average\_loss: 3.3243  
Epoch [8/50], Iter [110/209] Loss: 3.2139, average\_loss: 3.3406  
Epoch [8/50], Iter [115/209] Loss: 3.6433, average\_loss: 3.3385  
Epoch [8/50], Iter [120/209] Loss: 3.7039, average\_loss: 3.3473  
Epoch [8/50], Iter [125/209] Loss: 3.3525, average\_loss: 3.3368  
Epoch [8/50], Iter [130/209] Loss: 3.6055, average\_loss: 3.3449  
Epoch [8/50], Iter [135/209] Loss: 2.6034, average\_loss: 3.3392  
Epoch [8/50], Iter [140/209] Loss: 2.9283, average\_loss: 3.3407  
Epoch [8/50], Iter [145/209] Loss: 2.9053, average\_loss: 3.3405  
Epoch [8/50], Iter [150/209] Loss: 3.2509, average\_loss: 3.3330  
Epoch [8/50], Iter [155/209] Loss: 3.8586, average\_loss: 3.3317  
Epoch [8/50], Iter [160/209] Loss: 2.6644, average\_loss: 3.3272  
Epoch [8/50], Iter [165/209] Loss: 2.4908, average\_loss: 3.3254  
Epoch [8/50], Iter [170/209] Loss: 3.8125, average\_loss: 3.3320  
Epoch [8/50], Iter [175/209] Loss: 3.3735, average\_loss: 3.3267  
Epoch [8/50], Iter [180/209] Loss: 2.9221, average\_loss: 3.3148  
Epoch [8/50], Iter [185/209] Loss: 3.6902, average\_loss: 3.3157  
Epoch [8/50], Iter [190/209] Loss: 3.2269, average\_loss: 3.3064  
Epoch [8/50], Iter [195/209] Loss: 2.6874, average\_loss: 3.3101  
Epoch [8/50], Iter [200/209] Loss: 2.7161, average\_loss: 3.2994  
Epoch [8/50], Iter [205/209] Loss: 4.0657, average\_loss: 3.3078  
Updating best test loss: 3.48600

Starting epoch 9 / 50

Learning Rate for this epoch: 0.001

Epoch [9/50], Iter [5/209] Loss: 3.8589, average\_loss: 3.1501  
Epoch [9/50], Iter [10/209] Loss: 2.8384, average\_loss: 3.3463  
Epoch [9/50], Iter [15/209] Loss: 2.6296, average\_loss: 3.2350  
Epoch [9/50], Iter [20/209] Loss: 2.8270, average\_loss: 3.3004  
Epoch [9/50], Iter [25/209] Loss: 3.7670, average\_loss: 3.2587  
Epoch [9/50], Iter [30/209] Loss: 3.0547, average\_loss: 3.2743  
Epoch [9/50], Iter [35/209] Loss: 3.1097, average\_loss: 3.2750  
Epoch [9/50], Iter [40/209] Loss: 3.5746, average\_loss: 3.2240  
Epoch [9/50], Iter [45/209] Loss: 3.4480, average\_loss: 3.2203  
Epoch [9/50], Iter [50/209] Loss: 3.3080, average\_loss: 3.2402  
Epoch [9/50], Iter [55/209] Loss: 2.4373, average\_loss: 3.2227  
Epoch [9/50], Iter [60/209] Loss: 4.4472, average\_loss: 3.2509  
Epoch [9/50], Iter [65/209] Loss: 2.5204, average\_loss: 3.2543  
Epoch [9/50], Iter [70/209] Loss: 3.9498, average\_loss: 3.2626  
Epoch [9/50], Iter [75/209] Loss: 2.8759, average\_loss: 3.2514  
Epoch [9/50], Iter [80/209] Loss: 3.8395, average\_loss: 3.2589  
Epoch [9/50], Iter [85/209] Loss: 4.6155, average\_loss: 3.2770  
Epoch [9/50], Iter [90/209] Loss: 3.4758, average\_loss: 3.2694  
Epoch [9/50], Iter [95/209] Loss: 3.1387, average\_loss: 3.2801  
Epoch [9/50], Iter [100/209] Loss: 2.2998, average\_loss: 3.2736  
Epoch [9/50], Iter [105/209] Loss: 3.5059, average\_loss: 3.2786  
Epoch [9/50], Iter [110/209] Loss: 3.2743, average\_loss: 3.2740  
Epoch [9/50], Iter [115/209] Loss: 2.8661, average\_loss: 3.2704  
Epoch [9/50], Iter [120/209] Loss: 3.2700, average\_loss: 3.2753  
Epoch [9/50], Iter [125/209] Loss: 2.7200, average\_loss: 3.2818  
Epoch [9/50], Iter [130/209] Loss: 2.8467, average\_loss: 3.2635  
Epoch [9/50], Iter [135/209] Loss: 3.5602, average\_loss: 3.2616  
Epoch [9/50], Iter [140/209] Loss: 2.7040, average\_loss: 3.2654  
Epoch [9/50], Iter [145/209] Loss: 3.4533, average\_loss: 3.2664  
Epoch [9/50], Iter [150/209] Loss: 2.9612, average\_loss: 3.2579  
Epoch [9/50], Iter [155/209] Loss: 2.3662, average\_loss: 3.2486  
Epoch [9/50], Iter [160/209] Loss: 3.0009, average\_loss: 3.2531  
Epoch [9/50], Iter [165/209] Loss: 4.0655, average\_loss: 3.2455  
Epoch [9/50], Iter [170/209] Loss: 3.6506, average\_loss: 3.2396  
Epoch [9/50], Iter [175/209] Loss: 3.8927, average\_loss: 3.2404  
Epoch [9/50], Iter [180/209] Loss: 3.1483, average\_loss: 3.2509  
Epoch [9/50], Iter [185/209] Loss: 3.3069, average\_loss: 3.2509  
Epoch [9/50], Iter [190/209] Loss: 2.8538, average\_loss: 3.2509  
Epoch [9/50], Iter [195/209] Loss: 2.5249, average\_loss: 3.2529



Epoch [9/50], Iter [200/209] Loss: 3.5234, average\_loss: 3.2482  
Epoch [9/50], Iter [205/209] Loss: 2.6341, average\_loss: 3.2359  
Updating best test loss: 3.44835

Starting epoch 10 / 50

Learning Rate for this epoch: 0.001

Epoch [10/50], Iter [5/209] Loss: 2.9669, average\_loss: 3.1071  
Epoch [10/50], Iter [10/209] Loss: 3.4713, average\_loss: 3.1083  
Epoch [10/50], Iter [15/209] Loss: 3.8886, average\_loss: 3.1638  
Epoch [10/50], Iter [20/209] Loss: 2.7994, average\_loss: 3.2074  
Epoch [10/50], Iter [25/209] Loss: 3.3316, average\_loss: 3.1553  
Epoch [10/50], Iter [30/209] Loss: 3.5169, average\_loss: 3.2054  
Epoch [10/50], Iter [35/209] Loss: 3.1507, average\_loss: 3.2052  
Epoch [10/50], Iter [40/209] Loss: 3.0455, average\_loss: 3.2154  
Epoch [10/50], Iter [45/209] Loss: 2.9561, average\_loss: 3.1634  
Epoch [10/50], Iter [50/209] Loss: 3.6855, average\_loss: 3.1585  
Epoch [10/50], Iter [55/209] Loss: 2.4322, average\_loss: 3.1246  
Epoch [10/50], Iter [60/209] Loss: 2.8076, average\_loss: 3.1063  
Epoch [10/50], Iter [65/209] Loss: 3.0912, average\_loss: 3.1044  
Epoch [10/50], Iter [70/209] Loss: 2.5360, average\_loss: 3.0825  
Epoch [10/50], Iter [75/209] Loss: 2.4581, average\_loss: 3.0666  
Epoch [10/50], Iter [80/209] Loss: 2.8664, average\_loss: 3.0685  
Epoch [10/50], Iter [85/209] Loss: 2.7829, average\_loss: 3.0677  
Epoch [10/50], Iter [90/209] Loss: 3.1125, average\_loss: 3.0789  
Epoch [10/50], Iter [95/209] Loss: 3.3736, average\_loss: 3.0868  
Epoch [10/50], Iter [100/209] Loss: 2.7270, average\_loss: 3.0766  
Epoch [10/50], Iter [105/209] Loss: 3.0568, average\_loss: 3.1059  
Epoch [10/50], Iter [110/209] Loss: 3.3538, average\_loss: 3.1055  
Epoch [10/50], Iter [115/209] Loss: 3.2577, average\_loss: 3.1163  
Epoch [10/50], Iter [120/209] Loss: 3.9014, average\_loss: 3.1341  
Epoch [10/50], Iter [125/209] Loss: 3.4055, average\_loss: 3.1301  
Epoch [10/50], Iter [130/209] Loss: 2.8679, average\_loss: 3.1307  
Epoch [10/50], Iter [135/209] Loss: 2.9782, average\_loss: 3.1273  
Epoch [10/50], Iter [140/209] Loss: 2.7129, average\_loss: 3.1215  
Epoch [10/50], Iter [145/209] Loss: 3.1800, average\_loss: 3.1222  
Epoch [10/50], Iter [150/209] Loss: 3.5205, average\_loss: 3.1194  
Epoch [10/50], Iter [155/209] Loss: 3.0524, average\_loss: 3.1145  
Epoch [10/50], Iter [160/209] Loss: 2.4423, average\_loss: 3.1074  
Epoch [10/50], Iter [165/209] Loss: 3.6061, average\_loss: 3.1085  
Epoch [10/50], Iter [170/209] Loss: 3.0742, average\_loss: 3.1103  
Epoch [10/50], Iter [175/209] Loss: 2.7553, average\_loss: 3.1085  
Epoch [10/50], Iter [180/209] Loss: 3.4080, average\_loss: 3.1268  
Epoch [10/50], Iter [185/209] Loss: 2.6144, average\_loss: 3.1165  
Epoch [10/50], Iter [190/209] Loss: 3.8242, average\_loss: 3.1165  
Epoch [10/50], Iter [195/209] Loss: 3.0894, average\_loss: 3.1103  
Epoch [10/50], Iter [200/209] Loss: 3.5199, average\_loss: 3.1095  
Epoch [10/50], Iter [205/209] Loss: 3.3505, average\_loss: 3.1114  
Updating best test loss: 3.41420

Starting epoch 11 / 50

Learning Rate for this epoch: 0.001

Epoch [11/50], Iter [5/209] Loss: 2.8100, average\_loss: 2.9733  
Epoch [11/50], Iter [10/209] Loss: 4.5109, average\_loss: 3.1399  
Epoch [11/50], Iter [15/209] Loss: 2.7625, average\_loss: 3.0441  
Epoch [11/50], Iter [20/209] Loss: 2.8384, average\_loss: 2.9891  
Epoch [11/50], Iter [25/209] Loss: 4.4524, average\_loss: 3.0480  
Epoch [11/50], Iter [30/209] Loss: 2.3795, average\_loss: 3.0378  
Epoch [11/50], Iter [35/209] Loss: 2.7351, average\_loss: 3.0991  
Epoch [11/50], Iter [40/209] Loss: 2.1226, average\_loss: 3.0985  
Epoch [11/50], Iter [45/209] Loss: 3.5946, average\_loss: 3.0917  
Epoch [11/50], Iter [50/209] Loss: 2.3668, average\_loss: 3.1221  
Epoch [11/50], Iter [55/209] Loss: 3.0856, average\_loss: 3.0989  
Epoch [11/50], Iter [60/209] Loss: 3.0361, average\_loss: 3.1185  
Epoch [11/50], Iter [65/209] Loss: 3.2131, average\_loss: 3.1210  
Epoch [11/50], Iter [70/209] Loss: 3.2361, average\_loss: 3.1219  
Epoch [11/50], Iter [75/209] Loss: 3.5859, average\_loss: 3.1173  
Epoch [11/50], Iter [80/209] Loss: 3.0444, average\_loss: 3.1218  
Epoch [11/50], Iter [85/209] Loss: 2.9755, average\_loss: 3.1336  
Epoch [11/50], Iter [90/209] Loss: 2.8336, average\_loss: 3.1055  
Epoch [11/50], Iter [95/209] Loss: 2.6970, average\_loss: 3.0900  
Epoch [11/50], Iter [100/209] Loss: 3.2630, average\_loss: 3.0840  
Epoch [11/50], Iter [105/209] Loss: 3.0942, average\_loss: 3.0791  
Epoch [11/50], Iter [110/209] Loss: 3.1080, average\_loss: 3.0766  
Epoch [11/50], Iter [115/209] Loss: 3.0184, average\_loss: 3.0791  
Epoch [11/50], Iter [120/209] Loss: 3.7803, average\_loss: 3.0812

```
Epoch [11/50], Iter [125/209] Loss: 3.5486, average_loss: 3.0768
Epoch [11/50], Iter [130/209] Loss: 2.4563, average_loss: 3.0580
Epoch [11/50], Iter [135/209] Loss: 3.5805, average_loss: 3.0776
Epoch [11/50], Iter [140/209] Loss: 2.7842, average_loss: 3.0714
Epoch [11/50], Iter [145/209] Loss: 3.1855, average_loss: 3.0647
Epoch [11/50], Iter [150/209] Loss: 2.4194, average_loss: 3.0648
Epoch [11/50], Iter [155/209] Loss: 3.0675, average_loss: 3.0605
Epoch [11/50], Iter [160/209] Loss: 2.6476, average_loss: 3.0513
Epoch [11/50], Iter [165/209] Loss: 3.0439, average_loss: 3.0430
Epoch [11/50], Iter [170/209] Loss: 3.5861, average_loss: 3.0470
Epoch [11/50], Iter [175/209] Loss: 3.9347, average_loss: 3.0531
Epoch [11/50], Iter [180/209] Loss: 2.8395, average_loss: 3.0530
Epoch [11/50], Iter [185/209] Loss: 3.1958, average_loss: 3.0550
Epoch [11/50], Iter [190/209] Loss: 2.9970, average_loss: 3.0705
Epoch [11/50], Iter [195/209] Loss: 3.0692, average_loss: 3.0785
Epoch [11/50], Iter [200/209] Loss: 3.6902, average_loss: 3.0844
Epoch [11/50], Iter [205/209] Loss: 3.5670, average_loss: 3.0815
Updating best test loss: 3.33909
```

Starting epoch 12 / 50

Learning Rate for this epoch: 0.001

```
Epoch [12/50], Iter [5/209] Loss: 3.6166, average_loss: 3.0257
Epoch [12/50], Iter [10/209] Loss: 4.3265, average_loss: 3.1396
Epoch [12/50], Iter [15/209] Loss: 2.0008, average_loss: 2.9842
Epoch [12/50], Iter [20/209] Loss: 2.9185, average_loss: 2.9781
Epoch [12/50], Iter [25/209] Loss: 2.7536, average_loss: 2.9210
Epoch [12/50], Iter [30/209] Loss: 3.8087, average_loss: 2.9830
Epoch [12/50], Iter [35/209] Loss: 2.6570, average_loss: 2.9955
Epoch [12/50], Iter [40/209] Loss: 3.0956, average_loss: 3.0172
Epoch [12/50], Iter [45/209] Loss: 2.9163, average_loss: 3.0549
Epoch [12/50], Iter [50/209] Loss: 2.7668, average_loss: 3.0290
Epoch [12/50], Iter [55/209] Loss: 2.2955, average_loss: 3.0326
Epoch [12/50], Iter [60/209] Loss: 3.2096, average_loss: 3.0314
Epoch [12/50], Iter [65/209] Loss: 3.7447, average_loss: 3.0457
Epoch [12/50], Iter [70/209] Loss: 2.6540, average_loss: 3.0334
Epoch [12/50], Iter [75/209] Loss: 2.8110, average_loss: 3.0392
Epoch [12/50], Iter [80/209] Loss: 3.6144, average_loss: 3.0342
Epoch [12/50], Iter [85/209] Loss: 2.0522, average_loss: 3.0082
Epoch [12/50], Iter [90/209] Loss: 3.6507, average_loss: 3.0037
Epoch [12/50], Iter [95/209] Loss: 2.1520, average_loss: 3.0031
Epoch [12/50], Iter [100/209] Loss: 2.7188, average_loss: 3.0118
Epoch [12/50], Iter [105/209] Loss: 2.7421, average_loss: 3.0024
Epoch [12/50], Iter [110/209] Loss: 2.8281, average_loss: 2.9976
Epoch [12/50], Iter [115/209] Loss: 2.8320, average_loss: 3.0107
Epoch [12/50], Iter [120/209] Loss: 2.9530, average_loss: 3.0035
Epoch [12/50], Iter [125/209] Loss: 4.3725, average_loss: 3.0033
Epoch [12/50], Iter [130/209] Loss: 3.1790, average_loss: 2.9930
Epoch [12/50], Iter [135/209] Loss: 2.3338, average_loss: 2.9792
Epoch [12/50], Iter [140/209] Loss: 2.4902, average_loss: 2.9710
Epoch [12/50], Iter [145/209] Loss: 2.5415, average_loss: 2.9676
Epoch [12/50], Iter [150/209] Loss: 2.8551, average_loss: 2.9687
Epoch [12/50], Iter [155/209] Loss: 3.3067, average_loss: 2.9765
Epoch [12/50], Iter [160/209] Loss: 3.0830, average_loss: 2.9802
Epoch [12/50], Iter [165/209] Loss: 2.9757, average_loss: 2.9743
Epoch [12/50], Iter [170/209] Loss: 3.9013, average_loss: 2.9844
Epoch [12/50], Iter [175/209] Loss: 2.2646, average_loss: 2.9801
Epoch [12/50], Iter [180/209] Loss: 1.8969, average_loss: 2.9710
Epoch [12/50], Iter [185/209] Loss: 2.9016, average_loss: 2.9688
Epoch [12/50], Iter [190/209] Loss: 3.6738, average_loss: 2.9670
Epoch [12/50], Iter [195/209] Loss: 3.4558, average_loss: 2.9667
Epoch [12/50], Iter [200/209] Loss: 2.2754, average_loss: 2.9677
Epoch [12/50], Iter [205/209] Loss: 4.0474, average_loss: 2.9727
Updating best test loss: 3.32759
```

Starting epoch 13 / 50

Learning Rate for this epoch: 0.001

```
Epoch [13/50], Iter [5/209] Loss: 3.0983, average_loss: 2.9671
Epoch [13/50], Iter [10/209] Loss: 2.7923, average_loss: 3.0336
Epoch [13/50], Iter [15/209] Loss: 1.9171, average_loss: 2.8242
Epoch [13/50], Iter [20/209] Loss: 3.6699, average_loss: 2.8848
Epoch [13/50], Iter [25/209] Loss: 2.6508, average_loss: 2.8165
Epoch [13/50], Iter [30/209] Loss: 1.8319, average_loss: 2.7947
Epoch [13/50], Iter [35/209] Loss: 2.8845, average_loss: 2.8497
Epoch [13/50], Iter [40/209] Loss: 2.9478, average_loss: 2.8377
Epoch [13/50], Iter [45/209] Loss: 2.5697, average_loss: 2.8076
```

```
Epoch [13/50], Iter [50/209] Loss: 2.6103, average_loss: 2.8284
Epoch [13/50], Iter [55/209] Loss: 2.6000, average_loss: 2.8110
Epoch [13/50], Iter [60/209] Loss: 3.1527, average_loss: 2.8314
Epoch [13/50], Iter [65/209] Loss: 2.3943, average_loss: 2.8469
Epoch [13/50], Iter [70/209] Loss: 2.2074, average_loss: 2.8429
Epoch [13/50], Iter [75/209] Loss: 3.1378, average_loss: 2.8347
Epoch [13/50], Iter [80/209] Loss: 3.1432, average_loss: 2.8341
Epoch [13/50], Iter [85/209] Loss: 3.0858, average_loss: 2.8702
Epoch [13/50], Iter [90/209] Loss: 3.0274, average_loss: 2.8789
Epoch [13/50], Iter [95/209] Loss: 3.5664, average_loss: 2.8875
Epoch [13/50], Iter [100/209] Loss: 3.4926, average_loss: 2.8809
Epoch [13/50], Iter [105/209] Loss: 3.2107, average_loss: 2.9019
Epoch [13/50], Iter [110/209] Loss: 3.0295, average_loss: 2.9173
Epoch [13/50], Iter [115/209] Loss: 3.0779, average_loss: 2.9213
Epoch [13/50], Iter [120/209] Loss: 3.3822, average_loss: 2.9154
Epoch [13/50], Iter [125/209] Loss: 2.1573, average_loss: 2.8992
Epoch [13/50], Iter [130/209] Loss: 3.8068, average_loss: 2.8938
Epoch [13/50], Iter [135/209] Loss: 1.6907, average_loss: 2.8978
Epoch [13/50], Iter [140/209] Loss: 3.6057, average_loss: 2.9049
Epoch [13/50], Iter [145/209] Loss: 2.1136, average_loss: 2.8944
Epoch [13/50], Iter [150/209] Loss: 2.8593, average_loss: 2.8897
Epoch [13/50], Iter [155/209] Loss: 4.0715, average_loss: 2.8891
Epoch [13/50], Iter [160/209] Loss: 2.8334, average_loss: 2.9021
Epoch [13/50], Iter [165/209] Loss: 3.2573, average_loss: 2.9036
Epoch [13/50], Iter [170/209] Loss: 2.7221, average_loss: 2.9131
Epoch [13/50], Iter [175/209] Loss: 3.8712, average_loss: 2.9151
Epoch [13/50], Iter [180/209] Loss: 2.9748, average_loss: 2.9111
Epoch [13/50], Iter [185/209] Loss: 2.4345, average_loss: 2.9066
Epoch [13/50], Iter [190/209] Loss: 3.0772, average_loss: 2.9064
Epoch [13/50], Iter [195/209] Loss: 2.0243, average_loss: 2.9048
Epoch [13/50], Iter [200/209] Loss: 4.2372, average_loss: 2.9056
Epoch [13/50], Iter [205/209] Loss: 2.5789, average_loss: 2.9099
Updating best test loss: 3.25156
```

Starting epoch 14 / 50

Learning Rate for this epoch: 0.001

```
Epoch [14/50], Iter [5/209] Loss: 3.3553, average_loss: 2.9973
Epoch [14/50], Iter [10/209] Loss: 2.6060, average_loss: 3.0115
Epoch [14/50], Iter [15/209] Loss: 3.0580, average_loss: 2.8636
Epoch [14/50], Iter [20/209] Loss: 3.8225, average_loss: 2.9248
Epoch [14/50], Iter [25/209] Loss: 3.0486, average_loss: 2.9073
Epoch [14/50], Iter [30/209] Loss: 2.4221, average_loss: 2.8472
Epoch [14/50], Iter [35/209] Loss: 2.9568, average_loss: 2.8731
Epoch [14/50], Iter [40/209] Loss: 2.8268, average_loss: 2.8288
Epoch [14/50], Iter [45/209] Loss: 3.2032, average_loss: 2.8450
Epoch [14/50], Iter [50/209] Loss: 3.0414, average_loss: 2.8652
Epoch [14/50], Iter [55/209] Loss: 4.5998, average_loss: 2.8677
Epoch [14/50], Iter [60/209] Loss: 2.8647, average_loss: 2.8841
Epoch [14/50], Iter [65/209] Loss: 2.3716, average_loss: 2.8658
Epoch [14/50], Iter [70/209] Loss: 2.6879, average_loss: 2.8522
Epoch [14/50], Iter [75/209] Loss: 2.2880, average_loss: 2.8158
Epoch [14/50], Iter [80/209] Loss: 3.1199, average_loss: 2.8494
Epoch [14/50], Iter [85/209] Loss: 4.4009, average_loss: 2.8847
Epoch [14/50], Iter [90/209] Loss: 2.7982, average_loss: 2.8730
Epoch [14/50], Iter [95/209] Loss: 2.4839, average_loss: 2.8613
Epoch [14/50], Iter [100/209] Loss: 3.2747, average_loss: 2.8699
Epoch [14/50], Iter [105/209] Loss: 3.6616, average_loss: 2.8763
Epoch [14/50], Iter [110/209] Loss: 2.7584, average_loss: 2.8647
Epoch [14/50], Iter [115/209] Loss: 1.6808, average_loss: 2.8658
Epoch [14/50], Iter [120/209] Loss: 1.8004, average_loss: 2.8491
Epoch [14/50], Iter [125/209] Loss: 2.9512, average_loss: 2.8335
Epoch [14/50], Iter [130/209] Loss: 1.9865, average_loss: 2.8259
Epoch [14/50], Iter [135/209] Loss: 3.5320, average_loss: 2.8347
Epoch [14/50], Iter [140/209] Loss: 2.7935, average_loss: 2.8305
Epoch [14/50], Iter [145/209] Loss: 2.6165, average_loss: 2.8311
Epoch [14/50], Iter [150/209] Loss: 2.5229, average_loss: 2.8224
Epoch [14/50], Iter [155/209] Loss: 2.5720, average_loss: 2.8278
Epoch [14/50], Iter [160/209] Loss: 2.9110, average_loss: 2.8417
Epoch [14/50], Iter [165/209] Loss: 2.4071, average_loss: 2.8369
Epoch [14/50], Iter [170/209] Loss: 2.4391, average_loss: 2.8335
Epoch [14/50], Iter [175/209] Loss: 2.6253, average_loss: 2.8235
Epoch [14/50], Iter [180/209] Loss: 2.6897, average_loss: 2.8290
Epoch [14/50], Iter [185/209] Loss: 2.6418, average_loss: 2.8258
Epoch [14/50], Iter [190/209] Loss: 2.0564, average_loss: 2.8195
Epoch [14/50], Iter [195/209] Loss: 2.0951, average_loss: 2.8151
Epoch [14/50], Iter [200/209] Loss: 2.4359, average_loss: 2.8169
```

Epoch [14/50], Iter [205/209] Loss: 2.6918, average\_loss: 2.8142  
Updating best test loss: 3.23831

Starting epoch 15 / 50

Learning Rate for this epoch: 0.001

Epoch [15/50], Iter [5/209] Loss: 3.0768, average\_loss: 2.6853  
Epoch [15/50], Iter [10/209] Loss: 2.3967, average\_loss: 2.6481  
Epoch [15/50], Iter [15/209] Loss: 4.1003, average\_loss: 2.7727  
Epoch [15/50], Iter [20/209] Loss: 2.6373, average\_loss: 2.8400  
Epoch [15/50], Iter [25/209] Loss: 2.5671, average\_loss: 2.7456  
Epoch [15/50], Iter [30/209] Loss: 3.0122, average\_loss: 2.7644  
Epoch [15/50], Iter [35/209] Loss: 3.0305, average\_loss: 2.7630  
Epoch [15/50], Iter [40/209] Loss: 3.5359, average\_loss: 2.7459  
Epoch [15/50], Iter [45/209] Loss: 3.1848, average\_loss: 2.7574  
Epoch [15/50], Iter [50/209] Loss: 1.8800, average\_loss: 2.7300  
Epoch [15/50], Iter [55/209] Loss: 2.5657, average\_loss: 2.7546  
Epoch [15/50], Iter [60/209] Loss: 2.9636, average\_loss: 2.7513  
Epoch [15/50], Iter [65/209] Loss: 2.4269, average\_loss: 2.7690  
Epoch [15/50], Iter [70/209] Loss: 2.4306, average\_loss: 2.7598  
Epoch [15/50], Iter [75/209] Loss: 2.4452, average\_loss: 2.7651  
Epoch [15/50], Iter [80/209] Loss: 2.5555, average\_loss: 2.7475  
Epoch [15/50], Iter [85/209] Loss: 1.8282, average\_loss: 2.7472  
Epoch [15/50], Iter [90/209] Loss: 2.3179, average\_loss: 2.7349  
Epoch [15/50], Iter [95/209] Loss: 3.2566, average\_loss: 2.7393  
Epoch [15/50], Iter [100/209] Loss: 2.6469, average\_loss: 2.7320  
Epoch [15/50], Iter [105/209] Loss: 2.0177, average\_loss: 2.7114  
Epoch [15/50], Iter [110/209] Loss: 2.9103, average\_loss: 2.7099  
Epoch [15/50], Iter [115/209] Loss: 2.8542, average\_loss: 2.7100  
Epoch [15/50], Iter [120/209] Loss: 3.4933, average\_loss: 2.7062  
Epoch [15/50], Iter [125/209] Loss: 2.7567, average\_loss: 2.7107  
Epoch [15/50], Iter [130/209] Loss: 2.4328, average\_loss: 2.7212  
Epoch [15/50], Iter [135/209] Loss: 4.2289, average\_loss: 2.7256  
Epoch [15/50], Iter [140/209] Loss: 2.5625, average\_loss: 2.7247  
Epoch [15/50], Iter [145/209] Loss: 2.2451, average\_loss: 2.7201  
Epoch [15/50], Iter [150/209] Loss: 2.4044, average\_loss: 2.7229  
Epoch [15/50], Iter [155/209] Loss: 3.3092, average\_loss: 2.7153  
Epoch [15/50], Iter [160/209] Loss: 2.4746, average\_loss: 2.7119  
Epoch [15/50], Iter [165/209] Loss: 3.7573, average\_loss: 2.7323  
Epoch [15/50], Iter [170/209] Loss: 2.6740, average\_loss: 2.7376  
Epoch [15/50], Iter [175/209] Loss: 3.6260, average\_loss: 2.7459  
Epoch [15/50], Iter [180/209] Loss: 2.5318, average\_loss: 2.7560  
Epoch [15/50], Iter [185/209] Loss: 4.0763, average\_loss: 2.7602  
Epoch [15/50], Iter [190/209] Loss: 2.8040, average\_loss: 2.7582  
Epoch [15/50], Iter [195/209] Loss: 3.2990, average\_loss: 2.7544  
Epoch [15/50], Iter [200/209] Loss: 2.1252, average\_loss: 2.7471  
Epoch [15/50], Iter [205/209] Loss: 2.0839, average\_loss: 2.7364  
Updating best test loss: 3.17568

Starting epoch 16 / 50

Learning Rate for this epoch: 0.001

Epoch [16/50], Iter [5/209] Loss: 1.7833, average\_loss: 2.3575  
Epoch [16/50], Iter [10/209] Loss: 2.9346, average\_loss: 2.6213  
Epoch [16/50], Iter [15/209] Loss: 2.5711, average\_loss: 2.7156  
Epoch [16/50], Iter [20/209] Loss: 3.0433, average\_loss: 2.7788  
Epoch [16/50], Iter [25/209] Loss: 3.0734, average\_loss: 2.6821  
Epoch [16/50], Iter [30/209] Loss: 2.8742, average\_loss: 2.7389  
Epoch [16/50], Iter [35/209] Loss: 2.2711, average\_loss: 2.6833  
Epoch [16/50], Iter [40/209] Loss: 2.8359, average\_loss: 2.7201  
Epoch [16/50], Iter [45/209] Loss: 2.2382, average\_loss: 2.6853  
Epoch [16/50], Iter [50/209] Loss: 2.7864, average\_loss: 2.6551  
Epoch [16/50], Iter [55/209] Loss: 3.3577, average\_loss: 2.6742  
Epoch [16/50], Iter [60/209] Loss: 3.2871, average\_loss: 2.6991  
Epoch [16/50], Iter [65/209] Loss: 2.4207, average\_loss: 2.7036  
Epoch [16/50], Iter [70/209] Loss: 2.9670, average\_loss: 2.6953  
Epoch [16/50], Iter [75/209] Loss: 2.1974, average\_loss: 2.6938  
Epoch [16/50], Iter [80/209] Loss: 2.9918, average\_loss: 2.6755  
Epoch [16/50], Iter [85/209] Loss: 2.7775, average\_loss: 2.6669  
Epoch [16/50], Iter [90/209] Loss: 2.1525, average\_loss: 2.6549  
Epoch [16/50], Iter [95/209] Loss: 1.7899, average\_loss: 2.6485  
Epoch [16/50], Iter [100/209] Loss: 2.7827, average\_loss: 2.6454  
Epoch [16/50], Iter [105/209] Loss: 3.1887, average\_loss: 2.6673  
Epoch [16/50], Iter [110/209] Loss: 2.5008, average\_loss: 2.6728  
Epoch [16/50], Iter [115/209] Loss: 2.7909, average\_loss: 2.6725  
Epoch [16/50], Iter [120/209] Loss: 2.7321, average\_loss: 2.6747  
Epoch [16/50], Iter [125/209] Loss: 2.6906, average\_loss: 2.6732

```
Epoch [16/50], Iter [130/209] Loss: 2.1509, average_loss: 2.6695
Epoch [16/50], Iter [135/209] Loss: 2.8660, average_loss: 2.6634
Epoch [16/50], Iter [140/209] Loss: 1.9014, average_loss: 2.6509
Epoch [16/50], Iter [145/209] Loss: 2.5218, average_loss: 2.6624
Epoch [16/50], Iter [150/209] Loss: 3.2992, average_loss: 2.6714
Epoch [16/50], Iter [155/209] Loss: 2.3957, average_loss: 2.6692
Epoch [16/50], Iter [160/209] Loss: 2.6073, average_loss: 2.6749
Epoch [16/50], Iter [165/209] Loss: 2.4345, average_loss: 2.6660
Epoch [16/50], Iter [170/209] Loss: 3.0314, average_loss: 2.6723
Epoch [16/50], Iter [175/209] Loss: 2.4063, average_loss: 2.6762
Epoch [16/50], Iter [180/209] Loss: 4.5798, average_loss: 2.6876
Epoch [16/50], Iter [185/209] Loss: 3.7949, average_loss: 2.7061
Epoch [16/50], Iter [190/209] Loss: 2.5911, average_loss: 2.6954
Epoch [16/50], Iter [195/209] Loss: 2.3119, average_loss: 2.6870
Epoch [16/50], Iter [200/209] Loss: 3.2607, average_loss: 2.6830
Epoch [16/50], Iter [205/209] Loss: 4.4079, average_loss: 2.6911
```

Starting epoch 17 / 50

Learning Rate for this epoch: 0.001

```
Epoch [17/50], Iter [5/209] Loss: 2.8549, average_loss: 2.5270
Epoch [17/50], Iter [10/209] Loss: 2.3762, average_loss: 2.4651
Epoch [17/50], Iter [15/209] Loss: 2.1813, average_loss: 2.5769
Epoch [17/50], Iter [20/209] Loss: 2.1070, average_loss: 2.5734
Epoch [17/50], Iter [25/209] Loss: 2.0932, average_loss: 2.5674
Epoch [17/50], Iter [30/209] Loss: 2.6213, average_loss: 2.5310
Epoch [17/50], Iter [35/209] Loss: 2.4301, average_loss: 2.5756
Epoch [17/50], Iter [40/209] Loss: 2.9655, average_loss: 2.5898
Epoch [17/50], Iter [45/209] Loss: 2.5959, average_loss: 2.6120
Epoch [17/50], Iter [50/209] Loss: 2.6353, average_loss: 2.6174
Epoch [17/50], Iter [55/209] Loss: 2.5025, average_loss: 2.6437
Epoch [17/50], Iter [60/209] Loss: 3.0470, average_loss: 2.6468
Epoch [17/50], Iter [65/209] Loss: 2.5023, average_loss: 2.6220
Epoch [17/50], Iter [70/209] Loss: 2.2063, average_loss: 2.6074
Epoch [17/50], Iter [75/209] Loss: 1.9985, average_loss: 2.5925
Epoch [17/50], Iter [80/209] Loss: 2.6259, average_loss: 2.5968
Epoch [17/50], Iter [85/209] Loss: 2.7656, average_loss: 2.5956
Epoch [17/50], Iter [90/209] Loss: 3.6201, average_loss: 2.6064
Epoch [17/50], Iter [95/209] Loss: 2.1961, average_loss: 2.5927
Epoch [17/50], Iter [100/209] Loss: 1.9711, average_loss: 2.5986
Epoch [17/50], Iter [105/209] Loss: 2.7509, average_loss: 2.6035
Epoch [17/50], Iter [110/209] Loss: 2.2288, average_loss: 2.6176
Epoch [17/50], Iter [115/209] Loss: 1.8851, average_loss: 2.6200
Epoch [17/50], Iter [120/209] Loss: 3.1038, average_loss: 2.6146
Epoch [17/50], Iter [125/209] Loss: 3.1145, average_loss: 2.6139
Epoch [17/50], Iter [130/209] Loss: 2.7421, average_loss: 2.6259
Epoch [17/50], Iter [135/209] Loss: 3.5596, average_loss: 2.6504
Epoch [17/50], Iter [140/209] Loss: 2.7111, average_loss: 2.6540
Epoch [17/50], Iter [145/209] Loss: 3.1800, average_loss: 2.6532
Epoch [17/50], Iter [150/209] Loss: 2.6085, average_loss: 2.6548
Epoch [17/50], Iter [155/209] Loss: 2.5174, average_loss: 2.6487
Epoch [17/50], Iter [160/209] Loss: 2.5213, average_loss: 2.6435
Epoch [17/50], Iter [165/209] Loss: 2.0285, average_loss: 2.6317
Epoch [17/50], Iter [170/209] Loss: 2.6461, average_loss: 2.6220
Epoch [17/50], Iter [175/209] Loss: 3.2048, average_loss: 2.6366
Epoch [17/50], Iter [180/209] Loss: 2.3758, average_loss: 2.6328
Epoch [17/50], Iter [185/209] Loss: 3.2927, average_loss: 2.6469
Epoch [17/50], Iter [190/209] Loss: 2.1966, average_loss: 2.6436
Epoch [17/50], Iter [195/209] Loss: 2.2400, average_loss: 2.6445
Epoch [17/50], Iter [200/209] Loss: 2.3394, average_loss: 2.6525
Epoch [17/50], Iter [205/209] Loss: 2.2987, average_loss: 2.6585
Updating best test loss: 3.08590
```

Starting epoch 18 / 50

Learning Rate for this epoch: 0.001

```
Epoch [18/50], Iter [5/209] Loss: 2.5457, average_loss: 2.5649
Epoch [18/50], Iter [10/209] Loss: 2.9599, average_loss: 2.6799
Epoch [18/50], Iter [15/209] Loss: 1.7996, average_loss: 2.6544
Epoch [18/50], Iter [20/209] Loss: 3.3227, average_loss: 2.8041
Epoch [18/50], Iter [25/209] Loss: 2.9352, average_loss: 2.8065
Epoch [18/50], Iter [30/209] Loss: 2.2356, average_loss: 2.7864
Epoch [18/50], Iter [35/209] Loss: 2.1279, average_loss: 2.7360
Epoch [18/50], Iter [40/209] Loss: 2.1353, average_loss: 2.7397
Epoch [18/50], Iter [45/209] Loss: 3.4735, average_loss: 2.7057
Epoch [18/50], Iter [50/209] Loss: 2.4423, average_loss: 2.7039
Epoch [18/50], Iter [55/209] Loss: 2.5610, average_loss: 2.6914
```

```
Epoch [18/50], Iter [60/209] Loss: 2.9218, average_loss: 2.6944
Epoch [18/50], Iter [65/209] Loss: 2.7508, average_loss: 2.7083
Epoch [18/50], Iter [70/209] Loss: 2.1931, average_loss: 2.7059
Epoch [18/50], Iter [75/209] Loss: 1.8963, average_loss: 2.6847
Epoch [18/50], Iter [80/209] Loss: 2.1541, average_loss: 2.6638
Epoch [18/50], Iter [85/209] Loss: 2.3183, average_loss: 2.6585
Epoch [18/50], Iter [90/209] Loss: 2.8264, average_loss: 2.6610
Epoch [18/50], Iter [95/209] Loss: 1.9778, average_loss: 2.6476
Epoch [18/50], Iter [100/209] Loss: 2.7527, average_loss: 2.6391
Epoch [18/50], Iter [105/209] Loss: 2.7150, average_loss: 2.6388
Epoch [18/50], Iter [110/209] Loss: 2.9463, average_loss: 2.6350
Epoch [18/50], Iter [115/209] Loss: 2.7287, average_loss: 2.6292
Epoch [18/50], Iter [120/209] Loss: 2.9952, average_loss: 2.6233
Epoch [18/50], Iter [125/209] Loss: 2.2510, average_loss: 2.6282
Epoch [18/50], Iter [130/209] Loss: 2.7409, average_loss: 2.6497
Epoch [18/50], Iter [135/209] Loss: 2.5904, average_loss: 2.6357
Epoch [18/50], Iter [140/209] Loss: 2.3839, average_loss: 2.6352
Epoch [18/50], Iter [145/209] Loss: 3.0008, average_loss: 2.6312
Epoch [18/50], Iter [150/209] Loss: 2.4839, average_loss: 2.6123
Epoch [18/50], Iter [155/209] Loss: 2.5758, average_loss: 2.6124
Epoch [18/50], Iter [160/209] Loss: 2.9985, average_loss: 2.6056
Epoch [18/50], Iter [165/209] Loss: 2.4460, average_loss: 2.5982
Epoch [18/50], Iter [170/209] Loss: 2.4841, average_loss: 2.5981
Epoch [18/50], Iter [175/209] Loss: 3.8872, average_loss: 2.6057
Epoch [18/50], Iter [180/209] Loss: 1.6323, average_loss: 2.6038
Epoch [18/50], Iter [185/209] Loss: 2.4967, average_loss: 2.6053
Epoch [18/50], Iter [190/209] Loss: 2.0845, average_loss: 2.5936
Epoch [18/50], Iter [195/209] Loss: 2.9655, average_loss: 2.5996
Epoch [18/50], Iter [200/209] Loss: 2.7418, average_loss: 2.5962
Epoch [18/50], Iter [205/209] Loss: 1.9926, average_loss: 2.6009
```

Starting epoch 19 / 50

Learning Rate for this epoch: 0.001

```
Epoch [19/50], Iter [5/209] Loss: 2.7320, average_loss: 2.6679
Epoch [19/50], Iter [10/209] Loss: 2.8217, average_loss: 2.5261
Epoch [19/50], Iter [15/209] Loss: 2.6826, average_loss: 2.5276
Epoch [19/50], Iter [20/209] Loss: 3.4736, average_loss: 2.5282
Epoch [19/50], Iter [25/209] Loss: 2.6286, average_loss: 2.5059
Epoch [19/50], Iter [30/209] Loss: 2.3015, average_loss: 2.5128
Epoch [19/50], Iter [35/209] Loss: 2.7340, average_loss: 2.5052
Epoch [19/50], Iter [40/209] Loss: 2.3589, average_loss: 2.5373
Epoch [19/50], Iter [45/209] Loss: 2.3795, average_loss: 2.5555
Epoch [19/50], Iter [50/209] Loss: 2.4027, average_loss: 2.5627
Epoch [19/50], Iter [55/209] Loss: 2.2143, average_loss: 2.5861
Epoch [19/50], Iter [60/209] Loss: 2.4848, average_loss: 2.5731
Epoch [19/50], Iter [65/209] Loss: 3.0881, average_loss: 2.5738
Epoch [19/50], Iter [70/209] Loss: 1.7512, average_loss: 2.5766
Epoch [19/50], Iter [75/209] Loss: 2.8833, average_loss: 2.5590
Epoch [19/50], Iter [80/209] Loss: 2.5866, average_loss: 2.5460
Epoch [19/50], Iter [85/209] Loss: 3.2650, average_loss: 2.5581
Epoch [19/50], Iter [90/209] Loss: 2.3028, average_loss: 2.5543
Epoch [19/50], Iter [95/209] Loss: 2.7387, average_loss: 2.5574
Epoch [19/50], Iter [100/209] Loss: 3.7091, average_loss: 2.5688
Epoch [19/50], Iter [105/209] Loss: 2.4419, average_loss: 2.5623
Epoch [19/50], Iter [110/209] Loss: 2.2654, average_loss: 2.5667
Epoch [19/50], Iter [115/209] Loss: 2.8159, average_loss: 2.5639
Epoch [19/50], Iter [120/209] Loss: 2.4819, average_loss: 2.5562
Epoch [19/50], Iter [125/209] Loss: 2.4544, average_loss: 2.5685
Epoch [19/50], Iter [130/209] Loss: 2.4892, average_loss: 2.5631
Epoch [19/50], Iter [135/209] Loss: 2.2219, average_loss: 2.5530
Epoch [19/50], Iter [140/209] Loss: 2.3623, average_loss: 2.5622
Epoch [19/50], Iter [145/209] Loss: 2.2143, average_loss: 2.5565
Epoch [19/50], Iter [150/209] Loss: 2.8615, average_loss: 2.5609
Epoch [19/50], Iter [155/209] Loss: 2.3228, average_loss: 2.5576
Epoch [19/50], Iter [160/209] Loss: 1.9906, average_loss: 2.5459
Epoch [19/50], Iter [165/209] Loss: 1.9729, average_loss: 2.5471
Epoch [19/50], Iter [170/209] Loss: 2.6436, average_loss: 2.5478
Epoch [19/50], Iter [175/209] Loss: 3.6927, average_loss: 2.5525
Epoch [19/50], Iter [180/209] Loss: 1.7203, average_loss: 2.5441
Epoch [19/50], Iter [185/209] Loss: 2.0527, average_loss: 2.5412
Epoch [19/50], Iter [190/209] Loss: 2.6176, average_loss: 2.5388
Epoch [19/50], Iter [195/209] Loss: 2.4505, average_loss: 2.5324
Epoch [19/50], Iter [200/209] Loss: 2.9427, average_loss: 2.5401
Epoch [19/50], Iter [205/209] Loss: 3.4207, average_loss: 2.5363
Updating best test loss: 3.03954
```

Starting epoch 20 / 50

Learning Rate for this epoch: 0.001

```
Epoch [20/50], Iter [5/209] Loss: 2.3847, average_loss: 2.3144
Epoch [20/50], Iter [10/209] Loss: 2.5408, average_loss: 2.2931
Epoch [20/50], Iter [15/209] Loss: 2.0410, average_loss: 2.3306
Epoch [20/50], Iter [20/209] Loss: 3.2037, average_loss: 2.4100
Epoch [20/50], Iter [25/209] Loss: 2.7749, average_loss: 2.4063
Epoch [20/50], Iter [30/209] Loss: 2.3344, average_loss: 2.4694
Epoch [20/50], Iter [35/209] Loss: 2.0163, average_loss: 2.4010
Epoch [20/50], Iter [40/209] Loss: 2.6224, average_loss: 2.4637
Epoch [20/50], Iter [45/209] Loss: 2.2385, average_loss: 2.4565
Epoch [20/50], Iter [50/209] Loss: 2.4894, average_loss: 2.4465
Epoch [20/50], Iter [55/209] Loss: 2.6221, average_loss: 2.4852
Epoch [20/50], Iter [60/209] Loss: 2.1351, average_loss: 2.4873
Epoch [20/50], Iter [65/209] Loss: 2.3010, average_loss: 2.4813
Epoch [20/50], Iter [70/209] Loss: 3.0580, average_loss: 2.4912
Epoch [20/50], Iter [75/209] Loss: 2.5324, average_loss: 2.4767
Epoch [20/50], Iter [80/209] Loss: 2.1850, average_loss: 2.4770
Epoch [20/50], Iter [85/209] Loss: 2.1957, average_loss: 2.4757
Epoch [20/50], Iter [90/209] Loss: 2.7379, average_loss: 2.4814
Epoch [20/50], Iter [95/209] Loss: 3.0294, average_loss: 2.4841
Epoch [20/50], Iter [100/209] Loss: 1.7110, average_loss: 2.4897
Epoch [20/50], Iter [105/209] Loss: 1.7556, average_loss: 2.4818
Epoch [20/50], Iter [110/209] Loss: 3.5010, average_loss: 2.4758
Epoch [20/50], Iter [115/209] Loss: 1.9844, average_loss: 2.4893
Epoch [20/50], Iter [120/209] Loss: 2.4224, average_loss: 2.5022
Epoch [20/50], Iter [125/209] Loss: 2.2823, average_loss: 2.4947
Epoch [20/50], Iter [130/209] Loss: 2.2029, average_loss: 2.4988
Epoch [20/50], Iter [135/209] Loss: 1.7920, average_loss: 2.4997
Epoch [20/50], Iter [140/209] Loss: 2.4698, average_loss: 2.5096
Epoch [20/50], Iter [145/209] Loss: 2.0576, average_loss: 2.5142
Epoch [20/50], Iter [150/209] Loss: 2.2711, average_loss: 2.5164
Epoch [20/50], Iter [155/209] Loss: 2.3685, average_loss: 2.5284
Epoch [20/50], Iter [160/209] Loss: 2.1589, average_loss: 2.5253
Epoch [20/50], Iter [165/209] Loss: 3.1636, average_loss: 2.5285
Epoch [20/50], Iter [170/209] Loss: 1.9936, average_loss: 2.5298
Epoch [20/50], Iter [175/209] Loss: 2.2455, average_loss: 2.5272
Epoch [20/50], Iter [180/209] Loss: 2.6121, average_loss: 2.5183
Epoch [20/50], Iter [185/209] Loss: 1.9069, average_loss: 2.5243
Epoch [20/50], Iter [190/209] Loss: 2.2947, average_loss: 2.5286
Epoch [20/50], Iter [195/209] Loss: 2.0341, average_loss: 2.5279
Epoch [20/50], Iter [200/209] Loss: 1.7257, average_loss: 2.5232
Epoch [20/50], Iter [205/209] Loss: 2.0866, average_loss: 2.5214
```

Starting epoch 21 / 50

Learning Rate for this epoch: 0.001

```
Epoch [21/50], Iter [5/209] Loss: 2.8111, average_loss: 2.6504
Epoch [21/50], Iter [10/209] Loss: 2.7881, average_loss: 2.7150
Epoch [21/50], Iter [15/209] Loss: 2.0411, average_loss: 2.6309
Epoch [21/50], Iter [20/209] Loss: 2.7656, average_loss: 2.6010
Epoch [21/50], Iter [25/209] Loss: 3.0535, average_loss: 2.5606
Epoch [21/50], Iter [30/209] Loss: 2.5766, average_loss: 2.5918
Epoch [21/50], Iter [35/209] Loss: 2.9352, average_loss: 2.5839
Epoch [21/50], Iter [40/209] Loss: 2.5700, average_loss: 2.5652
Epoch [21/50], Iter [45/209] Loss: 2.6793, average_loss: 2.5558
Epoch [21/50], Iter [50/209] Loss: 2.2887, average_loss: 2.5513
Epoch [21/50], Iter [55/209] Loss: 1.8895, average_loss: 2.5351
Epoch [21/50], Iter [60/209] Loss: 2.2190, average_loss: 2.5376
Epoch [21/50], Iter [65/209] Loss: 2.5495, average_loss: 2.5098
Epoch [21/50], Iter [70/209] Loss: 2.3036, average_loss: 2.4811
Epoch [21/50], Iter [75/209] Loss: 2.6751, average_loss: 2.4795
Epoch [21/50], Iter [80/209] Loss: 2.1643, average_loss: 2.4695
Epoch [21/50], Iter [85/209] Loss: 2.1333, average_loss: 2.4713
Epoch [21/50], Iter [90/209] Loss: 1.7872, average_loss: 2.4610
Epoch [21/50], Iter [95/209] Loss: 2.5278, average_loss: 2.4599
Epoch [21/50], Iter [100/209] Loss: 2.9925, average_loss: 2.4641
Epoch [21/50], Iter [105/209] Loss: 1.8594, average_loss: 2.4592
Epoch [21/50], Iter [110/209] Loss: 2.8878, average_loss: 2.4547
Epoch [21/50], Iter [115/209] Loss: 3.2508, average_loss: 2.4634
Epoch [21/50], Iter [120/209] Loss: 1.6373, average_loss: 2.4577
Epoch [21/50], Iter [125/209] Loss: 2.0356, average_loss: 2.4569
Epoch [21/50], Iter [130/209] Loss: 2.7062, average_loss: 2.4649
Epoch [21/50], Iter [135/209] Loss: 3.5450, average_loss: 2.4712
Epoch [21/50], Iter [140/209] Loss: 1.9517, average_loss: 2.4596
Epoch [21/50], Iter [145/209] Loss: 2.5247, average_loss: 2.4733
```

```
Epoch [21/50], Iter [145/209] Loss: 2.3247, average_loss: 2.4733
Epoch [21/50], Iter [150/209] Loss: 2.7565, average_loss: 2.4904
Epoch [21/50], Iter [155/209] Loss: 2.3966, average_loss: 2.4920
Epoch [21/50], Iter [160/209] Loss: 1.8703, average_loss: 2.4866
Epoch [21/50], Iter [165/209] Loss: 2.2947, average_loss: 2.4935
Epoch [21/50], Iter [170/209] Loss: 2.1295, average_loss: 2.4878
Epoch [21/50], Iter [175/209] Loss: 1.7495, average_loss: 2.4760
Epoch [21/50], Iter [180/209] Loss: 2.5034, average_loss: 2.4804
Epoch [21/50], Iter [185/209] Loss: 3.0262, average_loss: 2.4885
Epoch [21/50], Iter [190/209] Loss: 1.8233, average_loss: 2.4954
Epoch [21/50], Iter [195/209] Loss: 2.9909, average_loss: 2.4949
Epoch [21/50], Iter [200/209] Loss: 2.3679, average_loss: 2.4954
Epoch [21/50], Iter [205/209] Loss: 2.2204, average_loss: 2.5003
Updating best test loss: 3.01751
```

Starting epoch 22 / 50

Learning Rate for this epoch: 0.001

```
Epoch [22/50], Iter [5/209] Loss: 1.8833, average_loss: 2.3923
Epoch [22/50], Iter [10/209] Loss: 2.8493, average_loss: 2.4923
Epoch [22/50], Iter [15/209] Loss: 1.9220, average_loss: 2.3479
Epoch [22/50], Iter [20/209] Loss: 2.5546, average_loss: 2.4121
Epoch [22/50], Iter [25/209] Loss: 3.1991, average_loss: 2.4366
Epoch [22/50], Iter [30/209] Loss: 2.4196, average_loss: 2.4914
Epoch [22/50], Iter [35/209] Loss: 2.4733, average_loss: 2.5117
Epoch [22/50], Iter [40/209] Loss: 2.5496, average_loss: 2.5226
Epoch [22/50], Iter [45/209] Loss: 2.8737, average_loss: 2.5440
Epoch [22/50], Iter [50/209] Loss: 2.3283, average_loss: 2.5205
Epoch [22/50], Iter [55/209] Loss: 2.0374, average_loss: 2.5125
Epoch [22/50], Iter [60/209] Loss: 1.8350, average_loss: 2.4936
Epoch [22/50], Iter [65/209] Loss: 2.0663, average_loss: 2.4710
Epoch [22/50], Iter [70/209] Loss: 2.6308, average_loss: 2.4757
Epoch [22/50], Iter [75/209] Loss: 2.4046, average_loss: 2.4594
Epoch [22/50], Iter [80/209] Loss: 3.5463, average_loss: 2.4713
Epoch [22/50], Iter [85/209] Loss: 2.0162, average_loss: 2.4553
Epoch [22/50], Iter [90/209] Loss: 2.9149, average_loss: 2.4514
Epoch [22/50], Iter [95/209] Loss: 2.8394, average_loss: 2.4533
Epoch [22/50], Iter [100/209] Loss: 3.1709, average_loss: 2.4533
Epoch [22/50], Iter [105/209] Loss: 2.6559, average_loss: 2.4539
Epoch [22/50], Iter [110/209] Loss: 2.2632, average_loss: 2.4427
Epoch [22/50], Iter [115/209] Loss: 2.3969, average_loss: 2.4429
Epoch [22/50], Iter [120/209] Loss: 2.2193, average_loss: 2.4328
Epoch [22/50], Iter [125/209] Loss: 3.1285, average_loss: 2.4443
Epoch [22/50], Iter [130/209] Loss: 2.0543, average_loss: 2.4378
Epoch [22/50], Iter [135/209] Loss: 2.8961, average_loss: 2.4420
Epoch [22/50], Iter [140/209] Loss: 2.4958, average_loss: 2.4420
Epoch [22/50], Iter [145/209] Loss: 2.3647, average_loss: 2.4411
Epoch [22/50], Iter [150/209] Loss: 2.6443, average_loss: 2.4334
Epoch [22/50], Iter [155/209] Loss: 1.9094, average_loss: 2.4269
Epoch [22/50], Iter [160/209] Loss: 2.7389, average_loss: 2.4282
Epoch [22/50], Iter [165/209] Loss: 2.2541, average_loss: 2.4229
Epoch [22/50], Iter [170/209] Loss: 1.8983, average_loss: 2.4337
Epoch [22/50], Iter [175/209] Loss: 2.6426, average_loss: 2.4515
Epoch [22/50], Iter [180/209] Loss: 2.5527, average_loss: 2.4503
Epoch [22/50], Iter [185/209] Loss: 1.8671, average_loss: 2.4461
Epoch [22/50], Iter [190/209] Loss: 2.4843, average_loss: 2.4431
Epoch [22/50], Iter [195/209] Loss: 2.2840, average_loss: 2.4379
Epoch [22/50], Iter [200/209] Loss: 2.8170, average_loss: 2.4394
Epoch [22/50], Iter [205/209] Loss: 2.0197, average_loss: 2.4316
Updating best test loss: 2.99965
```

Starting epoch 23 / 50

Learning Rate for this epoch: 0.001

```
Epoch [23/50], Iter [5/209] Loss: 1.6881, average_loss: 2.2733
Epoch [23/50], Iter [10/209] Loss: 2.0378, average_loss: 2.4490
Epoch [23/50], Iter [15/209] Loss: 2.0688, average_loss: 2.4761
Epoch [23/50], Iter [20/209] Loss: 2.5144, average_loss: 2.4641
Epoch [23/50], Iter [25/209] Loss: 2.5878, average_loss: 2.4275
Epoch [23/50], Iter [30/209] Loss: 2.6009, average_loss: 2.4556
Epoch [23/50], Iter [35/209] Loss: 1.8006, average_loss: 2.4218
Epoch [23/50], Iter [40/209] Loss: 2.0226, average_loss: 2.3851
Epoch [23/50], Iter [45/209] Loss: 1.8288, average_loss: 2.4037
Epoch [23/50], Iter [50/209] Loss: 2.9999, average_loss: 2.3870
Epoch [23/50], Iter [55/209] Loss: 3.7238, average_loss: 2.4066
Epoch [23/50], Iter [60/209] Loss: 2.2566, average_loss: 2.3898
Epoch [23/50], Iter [65/209] Loss: 2.0544, average_loss: 2.3951
Epoch [23/50], Iter [70/209] Loss: 2.3888, average_loss: 2.3865
```



```
Epoch [23/50], Iter [70/209] Loss: 2.3099, average_loss: 2.3000
Epoch [23/50], Iter [75/209] Loss: 1.8216, average_loss: 2.3647
Epoch [23/50], Iter [80/209] Loss: 1.9794, average_loss: 2.3524
Epoch [23/50], Iter [85/209] Loss: 1.7624, average_loss: 2.3456
Epoch [23/50], Iter [90/209] Loss: 2.3501, average_loss: 2.3350
Epoch [23/50], Iter [95/209] Loss: 2.6960, average_loss: 2.3423
Epoch [23/50], Iter [100/209] Loss: 2.5026, average_loss: 2.3511
Epoch [23/50], Iter [105/209] Loss: 2.8718, average_loss: 2.3593
Epoch [23/50], Iter [110/209] Loss: 1.9587, average_loss: 2.3502
Epoch [23/50], Iter [115/209] Loss: 2.6718, average_loss: 2.3744
Epoch [23/50], Iter [120/209] Loss: 1.9134, average_loss: 2.3759
Epoch [23/50], Iter [125/209] Loss: 2.0238, average_loss: 2.3834
Epoch [23/50], Iter [130/209] Loss: 2.9201, average_loss: 2.3921
Epoch [23/50], Iter [135/209] Loss: 2.8556, average_loss: 2.3960
Epoch [23/50], Iter [140/209] Loss: 2.7214, average_loss: 2.3989
Epoch [23/50], Iter [145/209] Loss: 1.7782, average_loss: 2.3988
Epoch [23/50], Iter [150/209] Loss: 2.9431, average_loss: 2.3970
Epoch [23/50], Iter [155/209] Loss: 2.9369, average_loss: 2.3951
Epoch [23/50], Iter [160/209] Loss: 2.5260, average_loss: 2.4075
Epoch [23/50], Iter [165/209] Loss: 3.0165, average_loss: 2.4131
Epoch [23/50], Iter [170/209] Loss: 1.3752, average_loss: 2.4038
Epoch [23/50], Iter [175/209] Loss: 2.3708, average_loss: 2.4012
Epoch [23/50], Iter [180/209] Loss: 2.6144, average_loss: 2.4056
Epoch [23/50], Iter [185/209] Loss: 2.6483, average_loss: 2.4099
Epoch [23/50], Iter [190/209] Loss: 2.8482, average_loss: 2.4131
Epoch [23/50], Iter [195/209] Loss: 2.5242, average_loss: 2.4211
Epoch [23/50], Iter [200/209] Loss: 2.7369, average_loss: 2.4248
Epoch [23/50], Iter [205/209] Loss: 2.0276, average_loss: 2.4296
```

Starting epoch 24 / 50

Learning Rate for this epoch: 0.001

```
Epoch [24/50], Iter [5/209] Loss: 2.3155, average_loss: 2.7357
Epoch [24/50], Iter [10/209] Loss: 2.4855, average_loss: 2.4140
Epoch [24/50], Iter [15/209] Loss: 2.0355, average_loss: 2.4530
Epoch [24/50], Iter [20/209] Loss: 1.9093, average_loss: 2.4280
Epoch [24/50], Iter [25/209] Loss: 2.8278, average_loss: 2.4486
Epoch [24/50], Iter [30/209] Loss: 2.2138, average_loss: 2.4179
Epoch [24/50], Iter [35/209] Loss: 1.8984, average_loss: 2.4145
Epoch [24/50], Iter [40/209] Loss: 3.2052, average_loss: 2.4228
Epoch [24/50], Iter [45/209] Loss: 1.5742, average_loss: 2.3669
Epoch [24/50], Iter [50/209] Loss: 2.1287, average_loss: 2.3585
Epoch [24/50], Iter [55/209] Loss: 1.8156, average_loss: 2.3585
Epoch [24/50], Iter [60/209] Loss: 3.0756, average_loss: 2.3690
Epoch [24/50], Iter [65/209] Loss: 2.5453, average_loss: 2.3828
Epoch [24/50], Iter [70/209] Loss: 2.9922, average_loss: 2.3717
Epoch [24/50], Iter [75/209] Loss: 2.3797, average_loss: 2.3674
Epoch [24/50], Iter [80/209] Loss: 2.6711, average_loss: 2.3829
Epoch [24/50], Iter [85/209] Loss: 1.3855, average_loss: 2.3513
Epoch [24/50], Iter [90/209] Loss: 2.1417, average_loss: 2.3509
Epoch [24/50], Iter [95/209] Loss: 2.9317, average_loss: 2.3452
Epoch [24/50], Iter [100/209] Loss: 2.1515, average_loss: 2.3515
Epoch [24/50], Iter [105/209] Loss: 2.4227, average_loss: 2.3586
Epoch [24/50], Iter [110/209] Loss: 2.2702, average_loss: 2.3626
Epoch [24/50], Iter [115/209] Loss: 1.8652, average_loss: 2.3741
Epoch [24/50], Iter [120/209] Loss: 2.7346, average_loss: 2.3775
Epoch [24/50], Iter [125/209] Loss: 2.7972, average_loss: 2.4016
Epoch [24/50], Iter [130/209] Loss: 2.5040, average_loss: 2.4095
Epoch [24/50], Iter [135/209] Loss: 2.1178, average_loss: 2.4090
Epoch [24/50], Iter [140/209] Loss: 2.0916, average_loss: 2.3927
Epoch [24/50], Iter [145/209] Loss: 2.2662, average_loss: 2.3932
Epoch [24/50], Iter [150/209] Loss: 2.2912, average_loss: 2.3879
Epoch [24/50], Iter [155/209] Loss: 2.2747, average_loss: 2.3823
Epoch [24/50], Iter [160/209] Loss: 1.9028, average_loss: 2.3769
Epoch [24/50], Iter [165/209] Loss: 3.0714, average_loss: 2.3725
Epoch [24/50], Iter [170/209] Loss: 2.7360, average_loss: 2.3647
Epoch [24/50], Iter [175/209] Loss: 1.8536, average_loss: 2.3642
Epoch [24/50], Iter [180/209] Loss: 1.8067, average_loss: 2.3598
Epoch [24/50], Iter [185/209] Loss: 2.9922, average_loss: 2.3645
Epoch [24/50], Iter [190/209] Loss: 1.9452, average_loss: 2.3614
Epoch [24/50], Iter [195/209] Loss: 2.2790, average_loss: 2.3599
Epoch [24/50], Iter [200/209] Loss: 2.1293, average_loss: 2.3652
Epoch [24/50], Iter [205/209] Loss: 2.4087, average_loss: 2.3617
```

Updating best test loss: 2.99032

Starting epoch 25 / 50

Learning Rate for this epoch: 0.001

Learning rate for this epoch: 0.001

```
Epoch [25/50], Iter [5/209] Loss: 1.9675, average_loss: 2.4458
Epoch [25/50], Iter [10/209] Loss: 2.6585, average_loss: 2.3748
Epoch [25/50], Iter [15/209] Loss: 2.1480, average_loss: 2.3897
Epoch [25/50], Iter [20/209] Loss: 2.3968, average_loss: 2.3987
Epoch [25/50], Iter [25/209] Loss: 2.3288, average_loss: 2.3567
Epoch [25/50], Iter [30/209] Loss: 2.7369, average_loss: 2.4271
Epoch [25/50], Iter [35/209] Loss: 2.3113, average_loss: 2.3988
Epoch [25/50], Iter [40/209] Loss: 2.6483, average_loss: 2.4077
Epoch [25/50], Iter [45/209] Loss: 1.9101, average_loss: 2.3940
Epoch [25/50], Iter [50/209] Loss: 3.0438, average_loss: 2.4032
Epoch [25/50], Iter [55/209] Loss: 2.2006, average_loss: 2.4151
Epoch [25/50], Iter [60/209] Loss: 2.3852, average_loss: 2.4058
Epoch [25/50], Iter [65/209] Loss: 1.8788, average_loss: 2.3854
Epoch [25/50], Iter [70/209] Loss: 2.2790, average_loss: 2.3757
Epoch [25/50], Iter [75/209] Loss: 1.7795, average_loss: 2.3801
Epoch [25/50], Iter [80/209] Loss: 1.9406, average_loss: 2.3586
Epoch [25/50], Iter [85/209] Loss: 2.5346, average_loss: 2.3763
Epoch [25/50], Iter [90/209] Loss: 1.8053, average_loss: 2.3745
Epoch [25/50], Iter [95/209] Loss: 1.7580, average_loss: 2.3691
Epoch [25/50], Iter [100/209] Loss: 2.4630, average_loss: 2.3566
Epoch [25/50], Iter [105/209] Loss: 1.9513, average_loss: 2.3639
Epoch [25/50], Iter [110/209] Loss: 1.7877, average_loss: 2.3719
Epoch [25/50], Iter [115/209] Loss: 3.1874, average_loss: 2.3725
Epoch [25/50], Iter [120/209] Loss: 1.8538, average_loss: 2.3653
Epoch [25/50], Iter [125/209] Loss: 2.1105, average_loss: 2.3576
Epoch [25/50], Iter [130/209] Loss: 2.3115, average_loss: 2.3738
Epoch [25/50], Iter [135/209] Loss: 2.9691, average_loss: 2.3716
Epoch [25/50], Iter [140/209] Loss: 3.0722, average_loss: 2.3796
Epoch [25/50], Iter [145/209] Loss: 2.0984, average_loss: 2.3829
Epoch [25/50], Iter [150/209] Loss: 2.1826, average_loss: 2.3757
Epoch [25/50], Iter [155/209] Loss: 3.1621, average_loss: 2.3791
Epoch [25/50], Iter [160/209] Loss: 2.3965, average_loss: 2.3763
Epoch [25/50], Iter [165/209] Loss: 1.7126, average_loss: 2.3675
Epoch [25/50], Iter [170/209] Loss: 1.9521, average_loss: 2.3582
Epoch [25/50], Iter [175/209] Loss: 2.4049, average_loss: 2.3537
Epoch [25/50], Iter [180/209] Loss: 2.4004, average_loss: 2.3610
Epoch [25/50], Iter [185/209] Loss: 2.4266, average_loss: 2.3585
Epoch [25/50], Iter [190/209] Loss: 2.2342, average_loss: 2.3564
Epoch [25/50], Iter [195/209] Loss: 1.9398, average_loss: 2.3587
Epoch [25/50], Iter [200/209] Loss: 2.3665, average_loss: 2.3555
Epoch [25/50], Iter [205/209] Loss: 3.9795, average_loss: 2.3611
Updating best test loss: 2.94294
```

Starting epoch 26 / 50

Learning Rate for this epoch: 0.001

```
Epoch [26/50], Iter [5/209] Loss: 2.4987, average_loss: 2.2013
Epoch [26/50], Iter [10/209] Loss: 2.0550, average_loss: 2.2020
Epoch [26/50], Iter [15/209] Loss: 2.0011, average_loss: 2.1461
Epoch [26/50], Iter [20/209] Loss: 2.3088, average_loss: 2.1335
Epoch [26/50], Iter [25/209] Loss: 1.5958, average_loss: 2.1746
Epoch [26/50], Iter [30/209] Loss: 2.3197, average_loss: 2.2038
Epoch [26/50], Iter [35/209] Loss: 1.7529, average_loss: 2.1738
Epoch [26/50], Iter [40/209] Loss: 2.6926, average_loss: 2.2068
Epoch [26/50], Iter [45/209] Loss: 2.2542, average_loss: 2.1965
Epoch [26/50], Iter [50/209] Loss: 1.7302, average_loss: 2.2102
Epoch [26/50], Iter [55/209] Loss: 2.4633, average_loss: 2.2090
Epoch [26/50], Iter [60/209] Loss: 2.4739, average_loss: 2.2098
Epoch [26/50], Iter [65/209] Loss: 2.3968, average_loss: 2.2198
Epoch [26/50], Iter [70/209] Loss: 2.6280, average_loss: 2.2190
Epoch [26/50], Iter [75/209] Loss: 2.1189, average_loss: 2.2473
Epoch [26/50], Iter [80/209] Loss: 2.1908, average_loss: 2.2488
Epoch [26/50], Iter [85/209] Loss: 1.9431, average_loss: 2.2392
Epoch [26/50], Iter [90/209] Loss: 2.0561, average_loss: 2.2324
Epoch [26/50], Iter [95/209] Loss: 1.7975, average_loss: 2.2162
Epoch [26/50], Iter [100/209] Loss: 2.7106, average_loss: 2.2213
Epoch [26/50], Iter [105/209] Loss: 2.0000, average_loss: 2.2411
Epoch [26/50], Iter [110/209] Loss: 2.8700, average_loss: 2.2480
Epoch [26/50], Iter [115/209] Loss: 2.3045, average_loss: 2.2614
Epoch [26/50], Iter [120/209] Loss: 2.3792, average_loss: 2.2786
Epoch [26/50], Iter [125/209] Loss: 2.0706, average_loss: 2.2818
Epoch [26/50], Iter [130/209] Loss: 1.9448, average_loss: 2.2786
Epoch [26/50], Iter [135/209] Loss: 3.2232, average_loss: 2.2970
Epoch [26/50], Iter [140/209] Loss: 2.0703, average_loss: 2.3009
Epoch [26/50], Iter [145/209] Loss: 2.4002, average_loss: 2.3028
Epoch [26/50], Iter [150/209] Loss: 2.4579, average_loss: 2.3008
```

```
Epoch [26/50], Iter [155/209] Loss: 2.1752, average_loss: 2.2914
Epoch [26/50], Iter [160/209] Loss: 2.1779, average_loss: 2.2993
Epoch [26/50], Iter [165/209] Loss: 2.5129, average_loss: 2.3019
Epoch [26/50], Iter [170/209] Loss: 1.6495, average_loss: 2.3077
Epoch [26/50], Iter [175/209] Loss: 2.4352, average_loss: 2.3011
Epoch [26/50], Iter [180/209] Loss: 2.7572, average_loss: 2.3070
Epoch [26/50], Iter [185/209] Loss: 2.3397, average_loss: 2.3074
Epoch [26/50], Iter [190/209] Loss: 2.6813, average_loss: 2.3128
Epoch [26/50], Iter [195/209] Loss: 2.0274, average_loss: 2.3133
Epoch [26/50], Iter [200/209] Loss: 2.0270, average_loss: 2.3098
Epoch [26/50], Iter [205/209] Loss: 2.5789, average_loss: 2.3128
```

Starting epoch 27 / 50

Learning Rate for this epoch: 0.001

```
Epoch [27/50], Iter [5/209] Loss: 2.8513, average_loss: 2.4130
Epoch [27/50], Iter [10/209] Loss: 2.6961, average_loss: 2.3147
Epoch [27/50], Iter [15/209] Loss: 2.4837, average_loss: 2.3986
Epoch [27/50], Iter [20/209] Loss: 2.5319, average_loss: 2.3671
Epoch [27/50], Iter [25/209] Loss: 2.5111, average_loss: 2.3767
Epoch [27/50], Iter [30/209] Loss: 2.9149, average_loss: 2.3763
Epoch [27/50], Iter [35/209] Loss: 1.3976, average_loss: 2.2761
Epoch [27/50], Iter [40/209] Loss: 2.2349, average_loss: 2.2858
Epoch [27/50], Iter [45/209] Loss: 1.8947, average_loss: 2.2936
Epoch [27/50], Iter [50/209] Loss: 1.8537, average_loss: 2.3007
Epoch [27/50], Iter [55/209] Loss: 2.0839, average_loss: 2.3048
Epoch [27/50], Iter [60/209] Loss: 2.2404, average_loss: 2.2903
Epoch [27/50], Iter [65/209] Loss: 2.7074, average_loss: 2.2745
Epoch [27/50], Iter [70/209] Loss: 2.2367, average_loss: 2.2916
Epoch [27/50], Iter [75/209] Loss: 2.5284, average_loss: 2.2851
Epoch [27/50], Iter [80/209] Loss: 2.1342, average_loss: 2.2908
Epoch [27/50], Iter [85/209] Loss: 2.4293, average_loss: 2.2960
Epoch [27/50], Iter [90/209] Loss: 2.6980, average_loss: 2.2939
Epoch [27/50], Iter [95/209] Loss: 1.4205, average_loss: 2.2903
Epoch [27/50], Iter [100/209] Loss: 1.8526, average_loss: 2.2814
Epoch [27/50], Iter [105/209] Loss: 1.9227, average_loss: 2.2875
Epoch [27/50], Iter [110/209] Loss: 2.4141, average_loss: 2.2813
Epoch [27/50], Iter [115/209] Loss: 2.2001, average_loss: 2.2853
Epoch [27/50], Iter [120/209] Loss: 2.5382, average_loss: 2.2870
Epoch [27/50], Iter [125/209] Loss: 2.1292, average_loss: 2.3013
Epoch [27/50], Iter [130/209] Loss: 2.1296, average_loss: 2.2970
Epoch [27/50], Iter [135/209] Loss: 2.6842, average_loss: 2.2994
Epoch [27/50], Iter [140/209] Loss: 2.1326, average_loss: 2.2960
Epoch [27/50], Iter [145/209] Loss: 2.6345, average_loss: 2.2916
Epoch [27/50], Iter [150/209] Loss: 2.8255, average_loss: 2.2967
Epoch [27/50], Iter [155/209] Loss: 2.1575, average_loss: 2.3000
Epoch [27/50], Iter [160/209] Loss: 1.9950, average_loss: 2.2957
Epoch [27/50], Iter [165/209] Loss: 2.0607, average_loss: 2.2911
Epoch [27/50], Iter [170/209] Loss: 2.2649, average_loss: 2.2947
Epoch [27/50], Iter [175/209] Loss: 2.2224, average_loss: 2.2945
Epoch [27/50], Iter [180/209] Loss: 2.1467, average_loss: 2.2936
Epoch [27/50], Iter [185/209] Loss: 1.8412, average_loss: 2.2966
Epoch [27/50], Iter [190/209] Loss: 2.2895, average_loss: 2.3019
Epoch [27/50], Iter [195/209] Loss: 2.2999, average_loss: 2.3023
Epoch [27/50], Iter [200/209] Loss: 2.0489, average_loss: 2.2954
Epoch [27/50], Iter [205/209] Loss: 2.6174, average_loss: 2.2963
```

Starting epoch 28 / 50

Learning Rate for this epoch: 0.001

```
Epoch [28/50], Iter [5/209] Loss: 1.7006, average_loss: 1.9502
Epoch [28/50], Iter [10/209] Loss: 1.6921, average_loss: 2.0535
Epoch [28/50], Iter [15/209] Loss: 1.3473, average_loss: 2.1266
Epoch [28/50], Iter [20/209] Loss: 1.7768, average_loss: 2.1339
Epoch [28/50], Iter [25/209] Loss: 2.0646, average_loss: 2.1559
Epoch [28/50], Iter [30/209] Loss: 1.7880, average_loss: 2.1784
Epoch [28/50], Iter [35/209] Loss: 2.0071, average_loss: 2.1619
Epoch [28/50], Iter [40/209] Loss: 2.0621, average_loss: 2.1799
Epoch [28/50], Iter [45/209] Loss: 2.0201, average_loss: 2.1763
Epoch [28/50], Iter [50/209] Loss: 2.5940, average_loss: 2.2253
Epoch [28/50], Iter [55/209] Loss: 2.1740, average_loss: 2.2444
Epoch [28/50], Iter [60/209] Loss: 1.8814, average_loss: 2.2545
Epoch [28/50], Iter [65/209] Loss: 2.9420, average_loss: 2.2679
Epoch [28/50], Iter [70/209] Loss: 1.6531, average_loss: 2.2474
Epoch [28/50], Iter [75/209] Loss: 2.0264, average_loss: 2.2462
Epoch [28/50], Iter [80/209] Loss: 2.3376, average_loss: 2.2629
Epoch [28/50], Iter [85/209] Loss: 2.2524, average_loss: 2.2673
```

```
Epoch [28/50], Iter [90/209] Loss: 1.7587, average_loss: 2.2613
Epoch [28/50], Iter [95/209] Loss: 2.4282, average_loss: 2.2631
Epoch [28/50], Iter [100/209] Loss: 1.6834, average_loss: 2.2552
Epoch [28/50], Iter [105/209] Loss: 2.1766, average_loss: 2.2521
Epoch [28/50], Iter [110/209] Loss: 2.0160, average_loss: 2.2445
Epoch [28/50], Iter [115/209] Loss: 1.9969, average_loss: 2.2340
Epoch [28/50], Iter [120/209] Loss: 2.6170, average_loss: 2.2274
Epoch [28/50], Iter [125/209] Loss: 2.5192, average_loss: 2.2366
Epoch [28/50], Iter [130/209] Loss: 2.1063, average_loss: 2.2390
Epoch [28/50], Iter [135/209] Loss: 2.0329, average_loss: 2.2366
Epoch [28/50], Iter [140/209] Loss: 2.2258, average_loss: 2.2434
Epoch [28/50], Iter [145/209] Loss: 1.8907, average_loss: 2.2543
Epoch [28/50], Iter [150/209] Loss: 1.9374, average_loss: 2.2500
Epoch [28/50], Iter [155/209] Loss: 1.7827, average_loss: 2.2452
Epoch [28/50], Iter [160/209] Loss: 2.1332, average_loss: 2.2490
Epoch [28/50], Iter [165/209] Loss: 2.3582, average_loss: 2.2559
Epoch [28/50], Iter [170/209] Loss: 2.0432, average_loss: 2.2487
Epoch [28/50], Iter [175/209] Loss: 2.8113, average_loss: 2.2474
Epoch [28/50], Iter [180/209] Loss: 2.1662, average_loss: 2.2487
Epoch [28/50], Iter [185/209] Loss: 2.4941, average_loss: 2.2484
Epoch [28/50], Iter [190/209] Loss: 2.1579, average_loss: 2.2471
Epoch [28/50], Iter [195/209] Loss: 1.7338, average_loss: 2.2573
Epoch [28/50], Iter [200/209] Loss: 1.7401, average_loss: 2.2567
Epoch [28/50], Iter [205/209] Loss: 1.9038, average_loss: 2.2616
```

Starting epoch 29 / 50

Learning Rate for this epoch: 0.001

```
Epoch [29/50], Iter [5/209] Loss: 2.5963, average_loss: 2.4057
Epoch [29/50], Iter [10/209] Loss: 1.7192, average_loss: 2.2708
Epoch [29/50], Iter [15/209] Loss: 2.5968, average_loss: 2.2634
Epoch [29/50], Iter [20/209] Loss: 2.3768, average_loss: 2.2320
Epoch [29/50], Iter [25/209] Loss: 2.8786, average_loss: 2.2117
Epoch [29/50], Iter [30/209] Loss: 2.0761, average_loss: 2.2077
Epoch [29/50], Iter [35/209] Loss: 2.3877, average_loss: 2.1816
Epoch [29/50], Iter [40/209] Loss: 2.6635, average_loss: 2.1983
Epoch [29/50], Iter [45/209] Loss: 1.9706, average_loss: 2.2224
Epoch [29/50], Iter [50/209] Loss: 2.2284, average_loss: 2.2469
Epoch [29/50], Iter [55/209] Loss: 2.3338, average_loss: 2.2455
Epoch [29/50], Iter [60/209] Loss: 2.0295, average_loss: 2.2455
Epoch [29/50], Iter [65/209] Loss: 1.9757, average_loss: 2.2433
Epoch [29/50], Iter [70/209] Loss: 1.6870, average_loss: 2.2356
Epoch [29/50], Iter [75/209] Loss: 2.1205, average_loss: 2.2444
Epoch [29/50], Iter [80/209] Loss: 1.9136, average_loss: 2.2550
Epoch [29/50], Iter [85/209] Loss: 2.7897, average_loss: 2.2505
Epoch [29/50], Iter [90/209] Loss: 1.8967, average_loss: 2.2514
Epoch [29/50], Iter [95/209] Loss: 1.8072, average_loss: 2.2523
Epoch [29/50], Iter [100/209] Loss: 2.7350, average_loss: 2.2379
Epoch [29/50], Iter [105/209] Loss: 2.2076, average_loss: 2.2419
Epoch [29/50], Iter [110/209] Loss: 2.4798, average_loss: 2.2408
Epoch [29/50], Iter [115/209] Loss: 2.8115, average_loss: 2.2370
Epoch [29/50], Iter [120/209] Loss: 2.0758, average_loss: 2.2251
Epoch [29/50], Iter [125/209] Loss: 2.3541, average_loss: 2.2302
Epoch [29/50], Iter [130/209] Loss: 2.0404, average_loss: 2.2343
Epoch [29/50], Iter [135/209] Loss: 2.6371, average_loss: 2.2393
Epoch [29/50], Iter [140/209] Loss: 2.7607, average_loss: 2.2382
Epoch [29/50], Iter [145/209] Loss: 1.6023, average_loss: 2.2360
Epoch [29/50], Iter [150/209] Loss: 3.0561, average_loss: 2.2531
Epoch [29/50], Iter [155/209] Loss: 2.8781, average_loss: 2.2741
Epoch [29/50], Iter [160/209] Loss: 1.9628, average_loss: 2.2688
Epoch [29/50], Iter [165/209] Loss: 2.2913, average_loss: 2.2686
Epoch [29/50], Iter [170/209] Loss: 2.3515, average_loss: 2.2703
Epoch [29/50], Iter [175/209] Loss: 1.8056, average_loss: 2.2649
Epoch [29/50], Iter [180/209] Loss: 2.5302, average_loss: 2.2662
Epoch [29/50], Iter [185/209] Loss: 2.8916, average_loss: 2.2702
Epoch [29/50], Iter [190/209] Loss: 2.6291, average_loss: 2.2704
Epoch [29/50], Iter [195/209] Loss: 2.0403, average_loss: 2.2682
Epoch [29/50], Iter [200/209] Loss: 1.9371, average_loss: 2.2592
Epoch [29/50], Iter [205/209] Loss: 1.8032, average_loss: 2.2529
```

Updating best test loss: 2.93577

Starting epoch 30 / 50

Learning Rate for this epoch: 0.001

```
Epoch [30/50], Iter [5/209] Loss: 2.2646, average_loss: 2.1905
Epoch [30/50], Iter [10/209] Loss: 2.6141, average_loss: 2.4176
Epoch [30/50], Iter [15/209] Loss: 2.0498, average_loss: 2.3654
```

Epoch [30/50], Iter [20/209] Loss: 1.9581, average\_loss: 2.3720  
Epoch [30/50], Iter [25/209] Loss: 2.1982, average\_loss: 2.3028  
Epoch [30/50], Iter [30/209] Loss: 1.8762, average\_loss: 2.2635  
Epoch [30/50], Iter [35/209] Loss: 2.2704, average\_loss: 2.2649  
Epoch [30/50], Iter [40/209] Loss: 1.9034, average\_loss: 2.2995  
Epoch [30/50], Iter [45/209] Loss: 2.5486, average\_loss: 2.2875  
Epoch [30/50], Iter [50/209] Loss: 2.5945, average\_loss: 2.2640  
Epoch [30/50], Iter [55/209] Loss: 2.3937, average\_loss: 2.2571  
Epoch [30/50], Iter [60/209] Loss: 2.4312, average\_loss: 2.2398  
Epoch [30/50], Iter [65/209] Loss: 2.0667, average\_loss: 2.2239  
Epoch [30/50], Iter [70/209] Loss: 1.8160, average\_loss: 2.2063  
Epoch [30/50], Iter [75/209] Loss: 1.8952, average\_loss: 2.2162  
Epoch [30/50], Iter [80/209] Loss: 1.9190, average\_loss: 2.2153  
Epoch [30/50], Iter [85/209] Loss: 1.8144, average\_loss: 2.2014  
Epoch [30/50], Iter [90/209] Loss: 1.8167, average\_loss: 2.1873  
Epoch [30/50], Iter [95/209] Loss: 2.2019, average\_loss: 2.1962  
Epoch [30/50], Iter [100/209] Loss: 2.3786, average\_loss: 2.1842  
Epoch [30/50], Iter [105/209] Loss: 3.2320, average\_loss: 2.2014  
Epoch [30/50], Iter [110/209] Loss: 2.7662, average\_loss: 2.2016  
Epoch [30/50], Iter [115/209] Loss: 2.1785, average\_loss: 2.2091  
Epoch [30/50], Iter [120/209] Loss: 2.0292, average\_loss: 2.2231  
Epoch [30/50], Iter [125/209] Loss: 1.9707, average\_loss: 2.2189  
Epoch [30/50], Iter [130/209] Loss: 2.1836, average\_loss: 2.2190  
Epoch [30/50], Iter [135/209] Loss: 2.5809, average\_loss: 2.2187  
Epoch [30/50], Iter [140/209] Loss: 1.7145, average\_loss: 2.2170  
Epoch [30/50], Iter [145/209] Loss: 2.2266, average\_loss: 2.2146  
Epoch [30/50], Iter [150/209] Loss: 2.0222, average\_loss: 2.2085  
Epoch [30/50], Iter [155/209] Loss: 1.8246, average\_loss: 2.2051  
Epoch [30/50], Iter [160/209] Loss: 1.9883, average\_loss: 2.2145  
Epoch [30/50], Iter [165/209] Loss: 2.2831, average\_loss: 2.2251  
Epoch [30/50], Iter [170/209] Loss: 2.1647, average\_loss: 2.2301  
Epoch [30/50], Iter [175/209] Loss: 1.8226, average\_loss: 2.2267  
Epoch [30/50], Iter [180/209] Loss: 2.3370, average\_loss: 2.2296  
Epoch [30/50], Iter [185/209] Loss: 2.2499, average\_loss: 2.2271  
Epoch [30/50], Iter [190/209] Loss: 1.9844, average\_loss: 2.2328  
Epoch [30/50], Iter [195/209] Loss: 2.9427, average\_loss: 2.2407  
Epoch [30/50], Iter [200/209] Loss: 1.6753, average\_loss: 2.2331  
Epoch [30/50], Iter [205/209] Loss: 1.8459, average\_loss: 2.2255

Starting epoch 31 / 50

Learning Rate for this epoch: 0.0001

Epoch [31/50], Iter [5/209] Loss: 1.7654, average\_loss: 2.0786  
Epoch [31/50], Iter [10/209] Loss: 2.0082, average\_loss: 2.1416  
Epoch [31/50], Iter [15/209] Loss: 2.4367, average\_loss: 2.1302  
Epoch [31/50], Iter [20/209] Loss: 2.1894, average\_loss: 2.1270  
Epoch [31/50], Iter [25/209] Loss: 1.8564, average\_loss: 2.0937  
Epoch [31/50], Iter [30/209] Loss: 1.9390, average\_loss: 2.0936  
Epoch [31/50], Iter [35/209] Loss: 2.1369, average\_loss: 2.0902  
Epoch [31/50], Iter [40/209] Loss: 2.6131, average\_loss: 2.1130  
Epoch [31/50], Iter [45/209] Loss: 1.4678, average\_loss: 2.0696  
Epoch [31/50], Iter [50/209] Loss: 1.8050, average\_loss: 2.0775  
Epoch [31/50], Iter [55/209] Loss: 2.2454, average\_loss: 2.1079  
Epoch [31/50], Iter [60/209] Loss: 2.0411, average\_loss: 2.1052  
Epoch [31/50], Iter [65/209] Loss: 1.8307, average\_loss: 2.0913  
Epoch [31/50], Iter [70/209] Loss: 2.7028, average\_loss: 2.1176  
Epoch [31/50], Iter [75/209] Loss: 1.8082, average\_loss: 2.1144  
Epoch [31/50], Iter [80/209] Loss: 2.1967, average\_loss: 2.1126  
Epoch [31/50], Iter [85/209] Loss: 1.8024, average\_loss: 2.1182  
Epoch [31/50], Iter [90/209] Loss: 2.1133, average\_loss: 2.1213  
Epoch [31/50], Iter [95/209] Loss: 1.9745, average\_loss: 2.1193  
Epoch [31/50], Iter [100/209] Loss: 2.0858, average\_loss: 2.1391  
Epoch [31/50], Iter [105/209] Loss: 1.7917, average\_loss: 2.1460  
Epoch [31/50], Iter [110/209] Loss: 2.5852, average\_loss: 2.1720  
Epoch [31/50], Iter [115/209] Loss: 3.0470, average\_loss: 2.1885  
Epoch [31/50], Iter [120/209] Loss: 2.0139, average\_loss: 2.1771  
Epoch [31/50], Iter [125/209] Loss: 1.9711, average\_loss: 2.1733  
Epoch [31/50], Iter [130/209] Loss: 2.3980, average\_loss: 2.1705  
Epoch [31/50], Iter [135/209] Loss: 3.2029, average\_loss: 2.1821  
Epoch [31/50], Iter [140/209] Loss: 2.3515, average\_loss: 2.1778  
Epoch [31/50], Iter [145/209] Loss: 1.7668, average\_loss: 2.1719  
Epoch [31/50], Iter [150/209] Loss: 1.8389, average\_loss: 2.1695  
Epoch [31/50], Iter [155/209] Loss: 2.4048, average\_loss: 2.1774  
Epoch [31/50], Iter [160/209] Loss: 1.6922, average\_loss: 2.1680  
Epoch [31/50], Iter [165/209] Loss: 1.7227, average\_loss: 2.1508  
Epoch [31/50], Iter [170/209] Loss: 2.3085, average\_loss: 2.1482  
Epoch [31/50], Iter [175/209] Loss: 1.8352, average\_loss: 2.1486

Epoch [31/50], Iter [180/209] Loss: 2.7550, average\_loss: 2.1483  
Epoch [31/50], Iter [185/209] Loss: 2.6233, average\_loss: 2.1516  
Epoch [31/50], Iter [190/209] Loss: 2.2798, average\_loss: 2.1439  
Epoch [31/50], Iter [195/209] Loss: 1.5553, average\_loss: 2.1411  
Epoch [31/50], Iter [200/209] Loss: 1.5411, average\_loss: 2.1459  
Epoch [31/50], Iter [205/209] Loss: 2.1694, average\_loss: 2.1422  
Updating best test loss: 2.90299

Starting epoch 32 / 50

Learning Rate for this epoch: 0.0001

Epoch [32/50], Iter [5/209] Loss: 1.9068, average\_loss: 2.0161  
Epoch [32/50], Iter [10/209] Loss: 2.3870, average\_loss: 2.1175  
Epoch [32/50], Iter [15/209] Loss: 1.9354, average\_loss: 2.1246  
Epoch [32/50], Iter [20/209] Loss: 1.8327, average\_loss: 2.0587  
Epoch [32/50], Iter [25/209] Loss: 2.0598, average\_loss: 2.0363  
Epoch [32/50], Iter [30/209] Loss: 2.0723, average\_loss: 2.1088  
Epoch [32/50], Iter [35/209] Loss: 1.5933, average\_loss: 2.0948  
Epoch [32/50], Iter [40/209] Loss: 2.4358, average\_loss: 2.0969  
Epoch [32/50], Iter [45/209] Loss: 1.8337, average\_loss: 2.0970  
Epoch [32/50], Iter [50/209] Loss: 2.0594, average\_loss: 2.1291  
Epoch [32/50], Iter [55/209] Loss: 2.5886, average\_loss: 2.1562  
Epoch [32/50], Iter [60/209] Loss: 2.1367, average\_loss: 2.1512  
Epoch [32/50], Iter [65/209] Loss: 2.3156, average\_loss: 2.1281  
Epoch [32/50], Iter [70/209] Loss: 2.3716, average\_loss: 2.1313  
Epoch [32/50], Iter [75/209] Loss: 2.0228, average\_loss: 2.1570  
Epoch [32/50], Iter [80/209] Loss: 1.8169, average\_loss: 2.1378  
Epoch [32/50], Iter [85/209] Loss: 1.6773, average\_loss: 2.1136  
Epoch [32/50], Iter [90/209] Loss: 2.4156, average\_loss: 2.1199  
Epoch [32/50], Iter [95/209] Loss: 2.2365, average\_loss: 2.1231  
Epoch [32/50], Iter [100/209] Loss: 2.6976, average\_loss: 2.1131  
Epoch [32/50], Iter [105/209] Loss: 2.0840, average\_loss: 2.1241  
Epoch [32/50], Iter [110/209] Loss: 2.0464, average\_loss: 2.1367  
Epoch [32/50], Iter [115/209] Loss: 1.7496, average\_loss: 2.1271  
Epoch [32/50], Iter [120/209] Loss: 2.0715, average\_loss: 2.1266  
Epoch [32/50], Iter [125/209] Loss: 2.3613, average\_loss: 2.1209  
Epoch [32/50], Iter [130/209] Loss: 2.0246, average\_loss: 2.1222  
Epoch [32/50], Iter [135/209] Loss: 2.0258, average\_loss: 2.1193  
Epoch [32/50], Iter [140/209] Loss: 2.4361, average\_loss: 2.1225  
Epoch [32/50], Iter [145/209] Loss: 1.8023, average\_loss: 2.1211  
Epoch [32/50], Iter [150/209] Loss: 2.1777, average\_loss: 2.1250  
Epoch [32/50], Iter [155/209] Loss: 1.7113, average\_loss: 2.1272  
Epoch [32/50], Iter [160/209] Loss: 2.3175, average\_loss: 2.1327  
Epoch [32/50], Iter [165/209] Loss: 2.0853, average\_loss: 2.1367  
Epoch [32/50], Iter [170/209] Loss: 1.4221, average\_loss: 2.1336  
Epoch [32/50], Iter [175/209] Loss: 2.1759, average\_loss: 2.1319  
Epoch [32/50], Iter [180/209] Loss: 2.3380, average\_loss: 2.1286  
Epoch [32/50], Iter [185/209] Loss: 1.3697, average\_loss: 2.1213  
Epoch [32/50], Iter [190/209] Loss: 1.4624, average\_loss: 2.1234  
Epoch [32/50], Iter [195/209] Loss: 2.5705, average\_loss: 2.1243  
Epoch [32/50], Iter [200/209] Loss: 2.8545, average\_loss: 2.1372  
Epoch [32/50], Iter [205/209] Loss: 2.0821, average\_loss: 2.1341  
Updating best test loss: 2.89198

Starting epoch 33 / 50

Learning Rate for this epoch: 0.0001

Epoch [33/50], Iter [5/209] Loss: 1.9154, average\_loss: 2.0637  
Epoch [33/50], Iter [10/209] Loss: 2.0434, average\_loss: 2.0354  
Epoch [33/50], Iter [15/209] Loss: 2.4373, average\_loss: 1.9909  
Epoch [33/50], Iter [20/209] Loss: 1.6247, average\_loss: 1.9353  
Epoch [33/50], Iter [25/209] Loss: 2.7913, average\_loss: 1.9842  
Epoch [33/50], Iter [30/209] Loss: 1.5652, average\_loss: 1.9640  
Epoch [33/50], Iter [35/209] Loss: 1.9266, average\_loss: 2.0296  
Epoch [33/50], Iter [40/209] Loss: 1.5056, average\_loss: 1.9989  
Epoch [33/50], Iter [45/209] Loss: 1.5059, average\_loss: 1.9793  
Epoch [33/50], Iter [50/209] Loss: 2.1310, average\_loss: 1.9893  
Epoch [33/50], Iter [55/209] Loss: 1.9941, average\_loss: 1.9843  
Epoch [33/50], Iter [60/209] Loss: 2.6028, average\_loss: 1.9963  
Epoch [33/50], Iter [65/209] Loss: 1.7823, average\_loss: 2.0078  
Epoch [33/50], Iter [70/209] Loss: 2.5044, average\_loss: 2.0279  
Epoch [33/50], Iter [75/209] Loss: 1.6454, average\_loss: 2.0130  
Epoch [33/50], Iter [80/209] Loss: 2.2387, average\_loss: 2.0139  
Epoch [33/50], Iter [85/209] Loss: 2.2379, average\_loss: 2.0348  
Epoch [33/50], Iter [90/209] Loss: 2.0684, average\_loss: 2.0394  
Epoch [33/50], Iter [95/209] Loss: 2.3258, average\_loss: 2.0380  
Epoch [33/50], Iter [100/209] Loss: 2.2828, average\_loss: 2.0607

```
Epoch [33/50], Iter [105/209] Loss: 1.9784, average_loss: 2.0567
Epoch [33/50], Iter [110/209] Loss: 1.4214, average_loss: 2.0462
Epoch [33/50], Iter [115/209] Loss: 1.9384, average_loss: 2.0441
Epoch [33/50], Iter [120/209] Loss: 1.8580, average_loss: 2.0452
Epoch [33/50], Iter [125/209] Loss: 1.7091, average_loss: 2.0454
Epoch [33/50], Iter [130/209] Loss: 2.0696, average_loss: 2.0618
Epoch [33/50], Iter [135/209] Loss: 2.4684, average_loss: 2.0805
Epoch [33/50], Iter [140/209] Loss: 1.7852, average_loss: 2.0804
Epoch [33/50], Iter [145/209] Loss: 1.8721, average_loss: 2.0791
Epoch [33/50], Iter [150/209] Loss: 2.1532, average_loss: 2.0853
Epoch [33/50], Iter [155/209] Loss: 2.0531, average_loss: 2.0798
Epoch [33/50], Iter [160/209] Loss: 1.6673, average_loss: 2.0850
Epoch [33/50], Iter [165/209] Loss: 1.7935, average_loss: 2.0827
Epoch [33/50], Iter [170/209] Loss: 1.9760, average_loss: 2.0865
Epoch [33/50], Iter [175/209] Loss: 2.7375, average_loss: 2.0859
Epoch [33/50], Iter [180/209] Loss: 1.5682, average_loss: 2.0753
Epoch [33/50], Iter [185/209] Loss: 2.4175, average_loss: 2.0757
Epoch [33/50], Iter [190/209] Loss: 2.0056, average_loss: 2.0768
Epoch [33/50], Iter [195/209] Loss: 1.7768, average_loss: 2.0792
Epoch [33/50], Iter [200/209] Loss: 2.5192, average_loss: 2.0813
Epoch [33/50], Iter [205/209] Loss: 2.3799, average_loss: 2.0844
```

Starting epoch 34 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [34/50], Iter [5/209] Loss: 2.2957, average_loss: 2.1409
Epoch [34/50], Iter [10/209] Loss: 2.1565, average_loss: 2.1959
Epoch [34/50], Iter [15/209] Loss: 1.4326, average_loss: 2.0551
Epoch [34/50], Iter [20/209] Loss: 1.7310, average_loss: 2.0874
Epoch [34/50], Iter [25/209] Loss: 2.0085, average_loss: 2.0558
Epoch [34/50], Iter [30/209] Loss: 2.5336, average_loss: 2.0874
Epoch [34/50], Iter [35/209] Loss: 2.1162, average_loss: 2.0412
Epoch [34/50], Iter [40/209] Loss: 2.0002, average_loss: 2.0375
Epoch [34/50], Iter [45/209] Loss: 2.1071, average_loss: 2.0437
Epoch [34/50], Iter [50/209] Loss: 1.9291, average_loss: 2.0423
Epoch [34/50], Iter [55/209] Loss: 2.0168, average_loss: 2.0450
Epoch [34/50], Iter [60/209] Loss: 2.0676, average_loss: 2.0289
Epoch [34/50], Iter [65/209] Loss: 2.0104, average_loss: 2.0229
Epoch [34/50], Iter [70/209] Loss: 1.9409, average_loss: 2.0375
Epoch [34/50], Iter [75/209] Loss: 2.3111, average_loss: 2.0296
Epoch [34/50], Iter [80/209] Loss: 3.1143, average_loss: 2.0429
Epoch [34/50], Iter [85/209] Loss: 2.1074, average_loss: 2.0581
Epoch [34/50], Iter [90/209] Loss: 2.3697, average_loss: 2.0548
Epoch [34/50], Iter [95/209] Loss: 2.0951, average_loss: 2.0572
Epoch [34/50], Iter [100/209] Loss: 1.8377, average_loss: 2.0453
Epoch [34/50], Iter [105/209] Loss: 2.2246, average_loss: 2.0535
Epoch [34/50], Iter [110/209] Loss: 2.3318, average_loss: 2.0492
Epoch [34/50], Iter [115/209] Loss: 1.8598, average_loss: 2.0500
Epoch [34/50], Iter [120/209] Loss: 2.5656, average_loss: 2.0550
Epoch [34/50], Iter [125/209] Loss: 1.7976, average_loss: 2.0647
Epoch [34/50], Iter [130/209] Loss: 1.9300, average_loss: 2.0590
Epoch [34/50], Iter [135/209] Loss: 2.3420, average_loss: 2.0566
Epoch [34/50], Iter [140/209] Loss: 2.6941, average_loss: 2.0658
Epoch [34/50], Iter [145/209] Loss: 2.7345, average_loss: 2.0736
Epoch [34/50], Iter [150/209] Loss: 1.6816, average_loss: 2.0742
Epoch [34/50], Iter [155/209] Loss: 1.8602, average_loss: 2.0763
Epoch [34/50], Iter [160/209] Loss: 2.1189, average_loss: 2.0767
Epoch [34/50], Iter [165/209] Loss: 2.6434, average_loss: 2.0797
Epoch [34/50], Iter [170/209] Loss: 2.4171, average_loss: 2.0812
Epoch [34/50], Iter [175/209] Loss: 2.8375, average_loss: 2.0856
Epoch [34/50], Iter [180/209] Loss: 1.8728, average_loss: 2.0830
Epoch [34/50], Iter [185/209] Loss: 2.1317, average_loss: 2.0928
Epoch [34/50], Iter [190/209] Loss: 2.0422, average_loss: 2.0932
Epoch [34/50], Iter [195/209] Loss: 1.7344, average_loss: 2.0942
Epoch [34/50], Iter [200/209] Loss: 1.8619, average_loss: 2.0956
Epoch [34/50], Iter [205/209] Loss: 1.9669, average_loss: 2.0971
```

Starting epoch 35 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [35/50], Iter [5/209] Loss: 1.3474, average_loss: 1.8047
Epoch [35/50], Iter [10/209] Loss: 3.3103, average_loss: 1.9735
Epoch [35/50], Iter [15/209] Loss: 2.1882, average_loss: 2.0389
Epoch [35/50], Iter [20/209] Loss: 1.6929, average_loss: 2.0631
Epoch [35/50], Iter [25/209] Loss: 1.9304, average_loss: 2.1043
Epoch [35/50], Iter [30/209] Loss: 2.6694, average_loss: 2.1031
Epoch [35/50], Iter [35/209] Loss: 2.2364, average_loss: 2.0977
```

```
Epoch [35/50], Iter [40/209] Loss: 3.2933, average_loss: 2.0899
Epoch [35/50], Iter [45/209] Loss: 2.0512, average_loss: 2.0856
Epoch [35/50], Iter [50/209] Loss: 1.9113, average_loss: 2.0764
Epoch [35/50], Iter [55/209] Loss: 1.9007, average_loss: 2.0646
Epoch [35/50], Iter [60/209] Loss: 1.7000, average_loss: 2.0415
Epoch [35/50], Iter [65/209] Loss: 1.1422, average_loss: 2.0152
Epoch [35/50], Iter [70/209] Loss: 2.6155, average_loss: 2.0170
Epoch [35/50], Iter [75/209] Loss: 2.7287, average_loss: 2.0397
Epoch [35/50], Iter [80/209] Loss: 2.9053, average_loss: 2.0568
Epoch [35/50], Iter [85/209] Loss: 2.0526, average_loss: 2.0616
Epoch [35/50], Iter [90/209] Loss: 2.2659, average_loss: 2.0528
Epoch [35/50], Iter [95/209] Loss: 2.0718, average_loss: 2.0457
Epoch [35/50], Iter [100/209] Loss: 2.0131, average_loss: 2.0441
Epoch [35/50], Iter [105/209] Loss: 2.4260, average_loss: 2.0480
Epoch [35/50], Iter [110/209] Loss: 2.0931, average_loss: 2.0547
Epoch [35/50], Iter [115/209] Loss: 2.0086, average_loss: 2.0545
Epoch [35/50], Iter [120/209] Loss: 2.6571, average_loss: 2.0679
Epoch [35/50], Iter [125/209] Loss: 1.6621, average_loss: 2.0630
Epoch [35/50], Iter [130/209] Loss: 2.0193, average_loss: 2.0621
Epoch [35/50], Iter [135/209] Loss: 2.2678, average_loss: 2.0719
Epoch [35/50], Iter [140/209] Loss: 1.4201, average_loss: 2.0705
Epoch [35/50], Iter [145/209] Loss: 1.5008, average_loss: 2.0748
Epoch [35/50], Iter [150/209] Loss: 2.2228, average_loss: 2.0814
Epoch [35/50], Iter [155/209] Loss: 2.2736, average_loss: 2.0819
Epoch [35/50], Iter [160/209] Loss: 1.7674, average_loss: 2.0764
Epoch [35/50], Iter [165/209] Loss: 1.8862, average_loss: 2.0725
Epoch [35/50], Iter [170/209] Loss: 2.2505, average_loss: 2.0649
Epoch [35/50], Iter [175/209] Loss: 1.8945, average_loss: 2.0679
Epoch [35/50], Iter [180/209] Loss: 1.9468, average_loss: 2.0625
Epoch [35/50], Iter [185/209] Loss: 1.6699, average_loss: 2.0518
Epoch [35/50], Iter [190/209] Loss: 1.5452, average_loss: 2.0488
Epoch [35/50], Iter [195/209] Loss: 1.9821, average_loss: 2.0466
Epoch [35/50], Iter [200/209] Loss: 2.2876, average_loss: 2.0431
Epoch [35/50], Iter [205/209] Loss: 2.4539, average_loss: 2.0493
```

Starting epoch 36 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [36/50], Iter [5/209] Loss: 1.4438, average_loss: 2.0761
Epoch [36/50], Iter [10/209] Loss: 2.7177, average_loss: 2.0751
Epoch [36/50], Iter [15/209] Loss: 1.7394, average_loss: 2.0052
Epoch [36/50], Iter [20/209] Loss: 2.4332, average_loss: 2.0262
Epoch [36/50], Iter [25/209] Loss: 1.9824, average_loss: 2.0027
Epoch [36/50], Iter [30/209] Loss: 2.6196, average_loss: 2.0468
Epoch [36/50], Iter [35/209] Loss: 2.0204, average_loss: 2.0683
Epoch [36/50], Iter [40/209] Loss: 1.7632, average_loss: 2.0336
Epoch [36/50], Iter [45/209] Loss: 2.7518, average_loss: 2.0875
Epoch [36/50], Iter [50/209] Loss: 2.8165, average_loss: 2.1089
Epoch [36/50], Iter [55/209] Loss: 2.1748, average_loss: 2.1191
Epoch [36/50], Iter [60/209] Loss: 2.3795, average_loss: 2.1353
Epoch [36/50], Iter [65/209] Loss: 1.9204, average_loss: 2.1383
Epoch [36/50], Iter [70/209] Loss: 2.3303, average_loss: 2.1313
Epoch [36/50], Iter [75/209] Loss: 2.2529, average_loss: 2.1498
Epoch [36/50], Iter [80/209] Loss: 2.4298, average_loss: 2.1458
Epoch [36/50], Iter [85/209] Loss: 1.5077, average_loss: 2.1236
Epoch [36/50], Iter [90/209] Loss: 2.3810, average_loss: 2.1328
Epoch [36/50], Iter [95/209] Loss: 2.3329, average_loss: 2.1246
Epoch [36/50], Iter [100/209] Loss: 1.1724, average_loss: 2.1077
Epoch [36/50], Iter [105/209] Loss: 2.2563, average_loss: 2.1039
Epoch [36/50], Iter [110/209] Loss: 1.9508, average_loss: 2.1009
Epoch [36/50], Iter [115/209] Loss: 1.8008, average_loss: 2.0901
Epoch [36/50], Iter [120/209] Loss: 2.4685, average_loss: 2.0829
Epoch [36/50], Iter [125/209] Loss: 2.0332, average_loss: 2.0896
Epoch [36/50], Iter [130/209] Loss: 1.8297, average_loss: 2.0898
Epoch [36/50], Iter [135/209] Loss: 2.3389, average_loss: 2.0822
Epoch [36/50], Iter [140/209] Loss: 1.4618, average_loss: 2.0725
Epoch [36/50], Iter [145/209] Loss: 2.0617, average_loss: 2.0684
Epoch [36/50], Iter [150/209] Loss: 1.9442, average_loss: 2.0726
Epoch [36/50], Iter [155/209] Loss: 2.3193, average_loss: 2.0707
Epoch [36/50], Iter [160/209] Loss: 1.6599, average_loss: 2.0677
Epoch [36/50], Iter [165/209] Loss: 1.7970, average_loss: 2.0695
Epoch [36/50], Iter [170/209] Loss: 2.2769, average_loss: 2.0710
Epoch [36/50], Iter [175/209] Loss: 1.8927, average_loss: 2.0666
Epoch [36/50], Iter [180/209] Loss: 1.4104, average_loss: 2.0698
Epoch [36/50], Iter [185/209] Loss: 1.9614, average_loss: 2.0722
Epoch [36/50], Iter [190/209] Loss: 3.1734, average_loss: 2.0833
Epoch [36/50], Iter [195/209] Loss: 2.4698, average_loss: 2.0789
```



Epoch [36/50], Iter [200/209] Loss: 2.6640, average\_loss: 2.0838  
Epoch [36/50], Iter [205/209] Loss: 2.0221, average\_loss: 2.0837  
Updating best test loss: 2.86414

Starting epoch 37 / 50

Learning Rate for this epoch: 0.0001

Epoch [37/50], Iter [5/209] Loss: 1.8529, average\_loss: 2.1966  
Epoch [37/50], Iter [10/209] Loss: 1.8957, average\_loss: 2.1804  
Epoch [37/50], Iter [15/209] Loss: 1.9664, average\_loss: 2.1393  
Epoch [37/50], Iter [20/209] Loss: 1.7825, average\_loss: 2.0134  
Epoch [37/50], Iter [25/209] Loss: 1.7269, average\_loss: 1.9736  
Epoch [37/50], Iter [30/209] Loss: 2.0274, average\_loss: 1.9854  
Epoch [37/50], Iter [35/209] Loss: 1.4510, average\_loss: 1.9919  
Epoch [37/50], Iter [40/209] Loss: 1.8674, average\_loss: 2.0247  
Epoch [37/50], Iter [45/209] Loss: 2.3288, average\_loss: 2.0208  
Epoch [37/50], Iter [50/209] Loss: 2.2106, average\_loss: 2.0099  
Epoch [37/50], Iter [55/209] Loss: 2.2279, average\_loss: 2.0231  
Epoch [37/50], Iter [60/209] Loss: 2.5395, average\_loss: 2.0475  
Epoch [37/50], Iter [65/209] Loss: 1.7490, average\_loss: 2.0504  
Epoch [37/50], Iter [70/209] Loss: 1.5165, average\_loss: 2.0399  
Epoch [37/50], Iter [75/209] Loss: 1.9973, average\_loss: 2.0306  
Epoch [37/50], Iter [80/209] Loss: 2.1442, average\_loss: 2.0459  
Epoch [37/50], Iter [85/209] Loss: 1.8076, average\_loss: 2.0480  
Epoch [37/50], Iter [90/209] Loss: 2.0445, average\_loss: 2.0330  
Epoch [37/50], Iter [95/209] Loss: 1.5832, average\_loss: 2.0342  
Epoch [37/50], Iter [100/209] Loss: 1.7913, average\_loss: 2.0279  
Epoch [37/50], Iter [105/209] Loss: 2.4967, average\_loss: 2.0338  
Epoch [37/50], Iter [110/209] Loss: 1.9177, average\_loss: 2.0450  
Epoch [37/50], Iter [115/209] Loss: 2.2489, average\_loss: 2.0436  
Epoch [37/50], Iter [120/209] Loss: 2.1846, average\_loss: 2.0481  
Epoch [37/50], Iter [125/209] Loss: 1.9424, average\_loss: 2.0433  
Epoch [37/50], Iter [130/209] Loss: 1.7483, average\_loss: 2.0489  
Epoch [37/50], Iter [135/209] Loss: 1.7692, average\_loss: 2.0583  
Epoch [37/50], Iter [140/209] Loss: 1.9448, average\_loss: 2.0598  
Epoch [37/50], Iter [145/209] Loss: 1.9620, average\_loss: 2.0555  
Epoch [37/50], Iter [150/209] Loss: 1.5652, average\_loss: 2.0441  
Epoch [37/50], Iter [155/209] Loss: 2.2011, average\_loss: 2.0393  
Epoch [37/50], Iter [160/209] Loss: 1.7401, average\_loss: 2.0385  
Epoch [37/50], Iter [165/209] Loss: 1.8397, average\_loss: 2.0451  
Epoch [37/50], Iter [170/209] Loss: 2.0263, average\_loss: 2.0441  
Epoch [37/50], Iter [175/209] Loss: 3.1551, average\_loss: 2.0577  
Epoch [37/50], Iter [180/209] Loss: 2.0169, average\_loss: 2.0614  
Epoch [37/50], Iter [185/209] Loss: 1.9541, average\_loss: 2.0657  
Epoch [37/50], Iter [190/209] Loss: 1.9041, average\_loss: 2.0678  
Epoch [37/50], Iter [195/209] Loss: 1.6566, average\_loss: 2.0667  
Epoch [37/50], Iter [200/209] Loss: 2.0198, average\_loss: 2.0680  
Epoch [37/50], Iter [205/209] Loss: 2.0470, average\_loss: 2.0657

Starting epoch 38 / 50

Learning Rate for this epoch: 0.0001

Epoch [38/50], Iter [5/209] Loss: 1.9628, average\_loss: 2.1073  
Epoch [38/50], Iter [10/209] Loss: 1.8882, average\_loss: 2.1129  
Epoch [38/50], Iter [15/209] Loss: 2.4522, average\_loss: 2.1200  
Epoch [38/50], Iter [20/209] Loss: 1.4329, average\_loss: 2.0813  
Epoch [38/50], Iter [25/209] Loss: 1.7461, average\_loss: 2.0439  
Epoch [38/50], Iter [30/209] Loss: 2.6247, average\_loss: 1.9845  
Epoch [38/50], Iter [35/209] Loss: 2.0725, average\_loss: 1.9529  
Epoch [38/50], Iter [40/209] Loss: 2.4087, average\_loss: 1.9538  
Epoch [38/50], Iter [45/209] Loss: 2.7510, average\_loss: 1.9874  
Epoch [38/50], Iter [50/209] Loss: 1.8633, average\_loss: 1.9906  
Epoch [38/50], Iter [55/209] Loss: 1.7843, average\_loss: 1.9951  
Epoch [38/50], Iter [60/209] Loss: 2.1077, average\_loss: 2.0066  
Epoch [38/50], Iter [65/209] Loss: 2.0797, average\_loss: 2.0065  
Epoch [38/50], Iter [70/209] Loss: 1.6418, average\_loss: 1.9991  
Epoch [38/50], Iter [75/209] Loss: 1.3841, average\_loss: 2.0056  
Epoch [38/50], Iter [80/209] Loss: 1.8386, average\_loss: 2.0154  
Epoch [38/50], Iter [85/209] Loss: 3.1184, average\_loss: 2.0143  
Epoch [38/50], Iter [90/209] Loss: 2.1474, average\_loss: 2.0128  
Epoch [38/50], Iter [95/209] Loss: 1.6321, average\_loss: 2.0096  
Epoch [38/50], Iter [100/209] Loss: 2.0181, average\_loss: 2.0026  
Epoch [38/50], Iter [105/209] Loss: 1.9527, average\_loss: 2.0026  
Epoch [38/50], Iter [110/209] Loss: 2.3059, average\_loss: 2.0093  
Epoch [38/50], Iter [115/209] Loss: 2.2359, average\_loss: 2.0050  
Epoch [38/50], Iter [120/209] Loss: 1.8374, average\_loss: 2.0040  
Epoch [38/50], Iter [125/209] Loss: 1.8853, average\_loss: 2.0053

```
Epoch [38/50], Iter [130/209] Loss: 1.8623, average_loss: 1.9964
Epoch [38/50], Iter [135/209] Loss: 2.1821, average_loss: 2.0074
Epoch [38/50], Iter [140/209] Loss: 2.2425, average_loss: 2.0089
Epoch [38/50], Iter [145/209] Loss: 2.0169, average_loss: 2.0086
Epoch [38/50], Iter [150/209] Loss: 1.9404, average_loss: 2.0149
Epoch [38/50], Iter [155/209] Loss: 2.3724, average_loss: 2.0111
Epoch [38/50], Iter [160/209] Loss: 2.0595, average_loss: 2.0094
Epoch [38/50], Iter [165/209] Loss: 3.2745, average_loss: 2.0188
Epoch [38/50], Iter [170/209] Loss: 1.9745, average_loss: 2.0166
Epoch [38/50], Iter [175/209] Loss: 1.8853, average_loss: 2.0157
Epoch [38/50], Iter [180/209] Loss: 2.3174, average_loss: 2.0249
Epoch [38/50], Iter [185/209] Loss: 1.7871, average_loss: 2.0269
Epoch [38/50], Iter [190/209] Loss: 1.7619, average_loss: 2.0313
Epoch [38/50], Iter [195/209] Loss: 1.9357, average_loss: 2.0336
Epoch [38/50], Iter [200/209] Loss: 2.5753, average_loss: 2.0409
Epoch [38/50], Iter [205/209] Loss: 1.8259, average_loss: 2.0412
```

Starting epoch 39 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [39/50], Iter [5/209] Loss: 2.6085, average_loss: 2.1147
Epoch [39/50], Iter [10/209] Loss: 2.2291, average_loss: 2.0477
Epoch [39/50], Iter [15/209] Loss: 2.8346, average_loss: 2.1268
Epoch [39/50], Iter [20/209] Loss: 3.2766, average_loss: 2.1317
Epoch [39/50], Iter [25/209] Loss: 2.3252, average_loss: 2.1220
Epoch [39/50], Iter [30/209] Loss: 2.5097, average_loss: 2.1414
Epoch [39/50], Iter [35/209] Loss: 2.2452, average_loss: 2.1316
Epoch [39/50], Iter [40/209] Loss: 2.0006, average_loss: 2.1113
Epoch [39/50], Iter [45/209] Loss: 3.2619, average_loss: 2.1036
Epoch [39/50], Iter [50/209] Loss: 2.3461, average_loss: 2.1165
Epoch [39/50], Iter [55/209] Loss: 2.0936, average_loss: 2.0873
Epoch [39/50], Iter [60/209] Loss: 1.5666, average_loss: 2.0871
Epoch [39/50], Iter [65/209] Loss: 2.6624, average_loss: 2.0870
Epoch [39/50], Iter [70/209] Loss: 1.8124, average_loss: 2.0745
Epoch [39/50], Iter [75/209] Loss: 2.0156, average_loss: 2.0581
Epoch [39/50], Iter [80/209] Loss: 1.7606, average_loss: 2.0554
Epoch [39/50], Iter [85/209] Loss: 2.3305, average_loss: 2.0673
Epoch [39/50], Iter [90/209] Loss: 1.7692, average_loss: 2.0524
Epoch [39/50], Iter [95/209] Loss: 1.9683, average_loss: 2.0461
Epoch [39/50], Iter [100/209] Loss: 2.2091, average_loss: 2.0480
Epoch [39/50], Iter [105/209] Loss: 2.3759, average_loss: 2.0479
Epoch [39/50], Iter [110/209] Loss: 1.5898, average_loss: 2.0373
Epoch [39/50], Iter [115/209] Loss: 1.8269, average_loss: 2.0432
Epoch [39/50], Iter [120/209] Loss: 2.2163, average_loss: 2.0462
Epoch [39/50], Iter [125/209] Loss: 2.0142, average_loss: 2.0411
Epoch [39/50], Iter [130/209] Loss: 3.4391, average_loss: 2.0571
Epoch [39/50], Iter [135/209] Loss: 2.8561, average_loss: 2.0647
Epoch [39/50], Iter [140/209] Loss: 2.1526, average_loss: 2.0626
Epoch [39/50], Iter [145/209] Loss: 2.2751, average_loss: 2.0552
Epoch [39/50], Iter [150/209] Loss: 1.7730, average_loss: 2.0478
Epoch [39/50], Iter [155/209] Loss: 1.9414, average_loss: 2.0435
Epoch [39/50], Iter [160/209] Loss: 2.2704, average_loss: 2.0390
Epoch [39/50], Iter [165/209] Loss: 2.1423, average_loss: 2.0484
Epoch [39/50], Iter [170/209] Loss: 1.9885, average_loss: 2.0413
Epoch [39/50], Iter [175/209] Loss: 3.1481, average_loss: 2.0486
Epoch [39/50], Iter [180/209] Loss: 2.3048, average_loss: 2.0491
Epoch [39/50], Iter [185/209] Loss: 2.4621, average_loss: 2.0607
Epoch [39/50], Iter [190/209] Loss: 1.7551, average_loss: 2.0676
Epoch [39/50], Iter [195/209] Loss: 1.7667, average_loss: 2.0671
Epoch [39/50], Iter [200/209] Loss: 2.2382, average_loss: 2.0695
Epoch [39/50], Iter [205/209] Loss: 2.8338, average_loss: 2.0685
```

Starting epoch 40 / 50

Learning Rate for this epoch: 0.0001

```
Epoch [40/50], Iter [5/209] Loss: 3.0143, average_loss: 2.1432
Epoch [40/50], Iter [10/209] Loss: 1.8480, average_loss: 2.1092
Epoch [40/50], Iter [15/209] Loss: 2.4179, average_loss: 2.0771
Epoch [40/50], Iter [20/209] Loss: 2.0629, average_loss: 2.0236
Epoch [40/50], Iter [25/209] Loss: 1.9850, average_loss: 2.0439
Epoch [40/50], Iter [30/209] Loss: 3.2912, average_loss: 2.1162
Epoch [40/50], Iter [35/209] Loss: 2.5247, average_loss: 2.1169
Epoch [40/50], Iter [40/209] Loss: 2.6571, average_loss: 2.1244
Epoch [40/50], Iter [45/209] Loss: 2.3149, average_loss: 2.1208
Epoch [40/50], Iter [50/209] Loss: 2.4668, average_loss: 2.1507
Epoch [40/50], Iter [55/209] Loss: 2.1725, average_loss: 2.1634
Epoch [40/50], Iter [60/209] Loss: 1.4344, average_loss: 2.1245
```

```
Epoch [40/50], Iter [65/209] Loss: 2.4843, average_loss: 2.1115
Epoch [40/50], Iter [70/209] Loss: 1.2939, average_loss: 2.1035
Epoch [40/50], Iter [75/209] Loss: 1.7695, average_loss: 2.0927
Epoch [40/50], Iter [80/209] Loss: 2.5667, average_loss: 2.0926
Epoch [40/50], Iter [85/209] Loss: 2.5220, average_loss: 2.0980
Epoch [40/50], Iter [90/209] Loss: 2.1827, average_loss: 2.0915
Epoch [40/50], Iter [95/209] Loss: 2.0760, average_loss: 2.0875
Epoch [40/50], Iter [100/209] Loss: 1.8692, average_loss: 2.0647
Epoch [40/50], Iter [105/209] Loss: 2.0817, average_loss: 2.0633
Epoch [40/50], Iter [110/209] Loss: 1.9773, average_loss: 2.0640
Epoch [40/50], Iter [115/209] Loss: 2.4802, average_loss: 2.0671
Epoch [40/50], Iter [120/209] Loss: 1.7293, average_loss: 2.0619
Epoch [40/50], Iter [125/209] Loss: 2.2631, average_loss: 2.0628
Epoch [40/50], Iter [130/209] Loss: 2.9610, average_loss: 2.0703
Epoch [40/50], Iter [135/209] Loss: 2.6340, average_loss: 2.0689
Epoch [40/50], Iter [140/209] Loss: 2.6091, average_loss: 2.0789
Epoch [40/50], Iter [145/209] Loss: 1.9401, average_loss: 2.0807
Epoch [40/50], Iter [150/209] Loss: 2.1997, average_loss: 2.0760
Epoch [40/50], Iter [155/209] Loss: 1.6029, average_loss: 2.0645
Epoch [40/50], Iter [160/209] Loss: 2.3668, average_loss: 2.0711
Epoch [40/50], Iter [165/209] Loss: 2.1760, average_loss: 2.0659
Epoch [40/50], Iter [170/209] Loss: 2.3214, average_loss: 2.0709
Epoch [40/50], Iter [175/209] Loss: 1.6748, average_loss: 2.0654
Epoch [40/50], Iter [180/209] Loss: 1.8488, average_loss: 2.0711
Epoch [40/50], Iter [185/209] Loss: 2.3859, average_loss: 2.0660
Epoch [40/50], Iter [190/209] Loss: 1.8737, average_loss: 2.0664
Epoch [40/50], Iter [195/209] Loss: 1.6428, average_loss: 2.0661
Epoch [40/50], Iter [200/209] Loss: 1.5434, average_loss: 2.0676
Epoch [40/50], Iter [205/209] Loss: 1.9897, average_loss: 2.0710
```

Starting epoch 41 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [41/50], Iter [5/209] Loss: 2.2920, average_loss: 1.9073
Epoch [41/50], Iter [10/209] Loss: 1.9092, average_loss: 2.0234
Epoch [41/50], Iter [15/209] Loss: 2.7854, average_loss: 2.1925
Epoch [41/50], Iter [20/209] Loss: 1.5645, average_loss: 2.1081
Epoch [41/50], Iter [25/209] Loss: 1.7769, average_loss: 2.0563
Epoch [41/50], Iter [30/209] Loss: 2.7598, average_loss: 2.0669
Epoch [41/50], Iter [35/209] Loss: 1.8639, average_loss: 2.0788
Epoch [41/50], Iter [40/209] Loss: 2.2493, average_loss: 2.0992
Epoch [41/50], Iter [45/209] Loss: 1.6419, average_loss: 2.0837
Epoch [41/50], Iter [50/209] Loss: 1.4085, average_loss: 2.0421
Epoch [41/50], Iter [55/209] Loss: 2.5141, average_loss: 2.0459
Epoch [41/50], Iter [60/209] Loss: 2.0425, average_loss: 2.0308
Epoch [41/50], Iter [65/209] Loss: 1.7518, average_loss: 2.0307
Epoch [41/50], Iter [70/209] Loss: 1.8298, average_loss: 2.0418
Epoch [41/50], Iter [75/209] Loss: 1.8801, average_loss: 2.0560
Epoch [41/50], Iter [80/209] Loss: 1.8877, average_loss: 2.0610
Epoch [41/50], Iter [85/209] Loss: 2.3623, average_loss: 2.0534
Epoch [41/50], Iter [90/209] Loss: 2.6326, average_loss: 2.0522
Epoch [41/50], Iter [95/209] Loss: 1.8939, average_loss: 2.0377
Epoch [41/50], Iter [100/209] Loss: 1.9424, average_loss: 2.0358
Epoch [41/50], Iter [105/209] Loss: 1.8751, average_loss: 2.0395
Epoch [41/50], Iter [110/209] Loss: 1.7427, average_loss: 2.0369
Epoch [41/50], Iter [115/209] Loss: 1.7144, average_loss: 2.0234
Epoch [41/50], Iter [120/209] Loss: 1.7038, average_loss: 2.0282
Epoch [41/50], Iter [125/209] Loss: 1.7160, average_loss: 2.0285
Epoch [41/50], Iter [130/209] Loss: 1.7192, average_loss: 2.0313
Epoch [41/50], Iter [135/209] Loss: 2.5061, average_loss: 2.0259
Epoch [41/50], Iter [140/209] Loss: 1.5245, average_loss: 2.0275
Epoch [41/50], Iter [145/209] Loss: 1.8912, average_loss: 2.0260
Epoch [41/50], Iter [150/209] Loss: 2.0516, average_loss: 2.0192
Epoch [41/50], Iter [155/209] Loss: 1.8105, average_loss: 2.0239
Epoch [41/50], Iter [160/209] Loss: 1.5710, average_loss: 2.0208
Epoch [41/50], Iter [165/209] Loss: 1.6221, average_loss: 2.0165
Epoch [41/50], Iter [170/209] Loss: 1.5258, average_loss: 2.0167
Epoch [41/50], Iter [175/209] Loss: 1.7424, average_loss: 2.0217
Epoch [41/50], Iter [180/209] Loss: 2.9155, average_loss: 2.0347
Epoch [41/50], Iter [185/209] Loss: 2.3809, average_loss: 2.0361
Epoch [41/50], Iter [190/209] Loss: 1.8521, average_loss: 2.0372
Epoch [41/50], Iter [195/209] Loss: 1.4110, average_loss: 2.0394
Epoch [41/50], Iter [200/209] Loss: 2.1962, average_loss: 2.0414
Epoch [41/50], Iter [205/209] Loss: 1.8491, average_loss: 2.0432
```

Starting epoch 42 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [42/50], Iter [5/209] Loss: 3.2581, average_loss: 2.3248
Epoch [42/50], Iter [10/209] Loss: 2.1484, average_loss: 2.2038
Epoch [42/50], Iter [15/209] Loss: 1.6584, average_loss: 2.1281
Epoch [42/50], Iter [20/209] Loss: 2.6343, average_loss: 2.1941
Epoch [42/50], Iter [25/209] Loss: 1.6913, average_loss: 2.2067
Epoch [42/50], Iter [30/209] Loss: 1.6282, average_loss: 2.1522
Epoch [42/50], Iter [35/209] Loss: 2.2251, average_loss: 2.1248
Epoch [42/50], Iter [40/209] Loss: 1.8640, average_loss: 2.0914
Epoch [42/50], Iter [45/209] Loss: 1.9823, average_loss: 2.0953
Epoch [42/50], Iter [50/209] Loss: 1.9214, average_loss: 2.0775
Epoch [42/50], Iter [55/209] Loss: 2.2891, average_loss: 2.0905
Epoch [42/50], Iter [60/209] Loss: 1.6994, average_loss: 2.0655
Epoch [42/50], Iter [65/209] Loss: 1.6458, average_loss: 2.0457
Epoch [42/50], Iter [70/209] Loss: 1.9404, average_loss: 2.0576
Epoch [42/50], Iter [75/209] Loss: 1.5577, average_loss: 2.0378
Epoch [42/50], Iter [80/209] Loss: 1.9284, average_loss: 2.0276
Epoch [42/50], Iter [85/209] Loss: 2.5822, average_loss: 2.0360
Epoch [42/50], Iter [90/209] Loss: 2.2794, average_loss: 2.0428
Epoch [42/50], Iter [95/209] Loss: 1.3601, average_loss: 2.0335
Epoch [42/50], Iter [100/209] Loss: 1.9438, average_loss: 2.0268
Epoch [42/50], Iter [105/209] Loss: 1.5653, average_loss: 2.0258
Epoch [42/50], Iter [110/209] Loss: 2.6627, average_loss: 2.0338
Epoch [42/50], Iter [115/209] Loss: 1.8614, average_loss: 2.0374
Epoch [42/50], Iter [120/209] Loss: 2.2044, average_loss: 2.0448
Epoch [42/50], Iter [125/209] Loss: 1.8735, average_loss: 2.0342
Epoch [42/50], Iter [130/209] Loss: 2.3937, average_loss: 2.0345
Epoch [42/50], Iter [135/209] Loss: 2.5626, average_loss: 2.0416
Epoch [42/50], Iter [140/209] Loss: 1.5847, average_loss: 2.0340
Epoch [42/50], Iter [145/209] Loss: 2.2034, average_loss: 2.0357
Epoch [42/50], Iter [150/209] Loss: 1.5724, average_loss: 2.0413
Epoch [42/50], Iter [155/209] Loss: 2.7884, average_loss: 2.0468
Epoch [42/50], Iter [160/209] Loss: 2.2219, average_loss: 2.0474
Epoch [42/50], Iter [165/209] Loss: 1.8895, average_loss: 2.0503
Epoch [42/50], Iter [170/209] Loss: 2.1402, average_loss: 2.0528
Epoch [42/50], Iter [175/209] Loss: 1.7192, average_loss: 2.0550
Epoch [42/50], Iter [180/209] Loss: 2.0140, average_loss: 2.0511
Epoch [42/50], Iter [185/209] Loss: 1.9960, average_loss: 2.0545
Epoch [42/50], Iter [190/209] Loss: 1.4089, average_loss: 2.0472
Epoch [42/50], Iter [195/209] Loss: 1.8761, average_loss: 2.0416
Epoch [42/50], Iter [200/209] Loss: 1.3370, average_loss: 2.0391
Epoch [42/50], Iter [205/209] Loss: 1.4255, average_loss: 2.0367
```

Starting epoch 43 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [43/50], Iter [5/209] Loss: 1.7953, average_loss: 1.9654
Epoch [43/50], Iter [10/209] Loss: 1.5952, average_loss: 1.9406
Epoch [43/50], Iter [15/209] Loss: 1.8187, average_loss: 1.9317
Epoch [43/50], Iter [20/209] Loss: 2.8846, average_loss: 2.0211
Epoch [43/50], Iter [25/209] Loss: 2.3540, average_loss: 2.0345
Epoch [43/50], Iter [30/209] Loss: 1.9422, average_loss: 2.0739
Epoch [43/50], Iter [35/209] Loss: 2.9439, average_loss: 2.1002
Epoch [43/50], Iter [40/209] Loss: 2.1111, average_loss: 2.0997
Epoch [43/50], Iter [45/209] Loss: 1.5105, average_loss: 2.0809
Epoch [43/50], Iter [50/209] Loss: 2.1203, average_loss: 2.1016
Epoch [43/50], Iter [55/209] Loss: 1.8415, average_loss: 2.0890
Epoch [43/50], Iter [60/209] Loss: 2.5215, average_loss: 2.1178
Epoch [43/50], Iter [65/209] Loss: 2.2639, average_loss: 2.1096
Epoch [43/50], Iter [70/209] Loss: 2.4886, average_loss: 2.0853
Epoch [43/50], Iter [75/209] Loss: 2.1252, average_loss: 2.0736
Epoch [43/50], Iter [80/209] Loss: 2.1824, average_loss: 2.0670
Epoch [43/50], Iter [85/209] Loss: 2.7122, average_loss: 2.0677
Epoch [43/50], Iter [90/209] Loss: 2.4670, average_loss: 2.0667
Epoch [43/50], Iter [95/209] Loss: 2.0218, average_loss: 2.0752
Epoch [43/50], Iter [100/209] Loss: 1.6927, average_loss: 2.0713
Epoch [43/50], Iter [105/209] Loss: 1.3322, average_loss: 2.0650
Epoch [43/50], Iter [110/209] Loss: 2.6791, average_loss: 2.0627
Epoch [43/50], Iter [115/209] Loss: 2.1253, average_loss: 2.0655
Epoch [43/50], Iter [120/209] Loss: 1.5466, average_loss: 2.0535
Epoch [43/50], Iter [125/209] Loss: 1.9298, average_loss: 2.0488
Epoch [43/50], Iter [130/209] Loss: 2.6409, average_loss: 2.0561
Epoch [43/50], Iter [135/209] Loss: 1.5121, average_loss: 2.0557
Epoch [43/50], Iter [140/209] Loss: 2.4728, average_loss: 2.0498
Epoch [43/50], Iter [145/209] Loss: 1.9126, average_loss: 2.0511
Epoch [43/50], Iter [150/209] Loss: 1.7616, average_loss: 2.0500
Epoch [43/50], Iter [155/209] Loss: 1.6927, average_loss: 2.0501
```

```
Epoch [43/50], Iter [160/209] Loss: 2.0613, average_loss: 2.0514
Epoch [43/50], Iter [165/209] Loss: 1.7188, average_loss: 2.0589
Epoch [43/50], Iter [170/209] Loss: 2.1630, average_loss: 2.0642
Epoch [43/50], Iter [175/209] Loss: 2.4031, average_loss: 2.0583
Epoch [43/50], Iter [180/209] Loss: 1.8756, average_loss: 2.0534
Epoch [43/50], Iter [185/209] Loss: 1.7505, average_loss: 2.0513
Epoch [43/50], Iter [190/209] Loss: 2.0429, average_loss: 2.0518
Epoch [43/50], Iter [195/209] Loss: 1.5919, average_loss: 2.0437
Epoch [43/50], Iter [200/209] Loss: 2.3658, average_loss: 2.0400
Epoch [43/50], Iter [205/209] Loss: 2.9166, average_loss: 2.0432
```

Starting epoch 44 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [44/50], Iter [5/209] Loss: 2.9021, average_loss: 1.9120
Epoch [44/50], Iter [10/209] Loss: 2.0133, average_loss: 1.9665
Epoch [44/50], Iter [15/209] Loss: 1.5939, average_loss: 2.0175
Epoch [44/50], Iter [20/209] Loss: 2.4390, average_loss: 2.0102
Epoch [44/50], Iter [25/209] Loss: 1.6088, average_loss: 2.0033
Epoch [44/50], Iter [30/209] Loss: 1.4249, average_loss: 2.0081
Epoch [44/50], Iter [35/209] Loss: 1.6803, average_loss: 1.9735
Epoch [44/50], Iter [40/209] Loss: 1.9748, average_loss: 1.9874
Epoch [44/50], Iter [45/209] Loss: 2.2233, average_loss: 1.9949
Epoch [44/50], Iter [50/209] Loss: 1.9439, average_loss: 1.9817
Epoch [44/50], Iter [55/209] Loss: 1.7880, average_loss: 1.9749
Epoch [44/50], Iter [60/209] Loss: 2.1410, average_loss: 1.9791
Epoch [44/50], Iter [65/209] Loss: 3.2939, average_loss: 1.9983
Epoch [44/50], Iter [70/209] Loss: 1.2522, average_loss: 1.9794
Epoch [44/50], Iter [75/209] Loss: 2.2215, average_loss: 1.9880
Epoch [44/50], Iter [80/209] Loss: 1.8716, average_loss: 1.9800
Epoch [44/50], Iter [85/209] Loss: 1.9775, average_loss: 1.9903
Epoch [44/50], Iter [90/209] Loss: 2.0492, average_loss: 2.0003
Epoch [44/50], Iter [95/209] Loss: 1.8722, average_loss: 1.9950
Epoch [44/50], Iter [100/209] Loss: 1.5245, average_loss: 1.9886
Epoch [44/50], Iter [105/209] Loss: 2.0327, average_loss: 1.9850
Epoch [44/50], Iter [110/209] Loss: 2.1218, average_loss: 2.0038
Epoch [44/50], Iter [115/209] Loss: 2.7537, average_loss: 2.0029
Epoch [44/50], Iter [120/209] Loss: 1.6765, average_loss: 1.9990
Epoch [44/50], Iter [125/209] Loss: 1.6291, average_loss: 1.9987
Epoch [44/50], Iter [130/209] Loss: 2.1363, average_loss: 1.9968
Epoch [44/50], Iter [135/209] Loss: 1.7018, average_loss: 1.9929
Epoch [44/50], Iter [140/209] Loss: 1.7569, average_loss: 1.9883
Epoch [44/50], Iter [145/209] Loss: 1.5657, average_loss: 1.9838
Epoch [44/50], Iter [150/209] Loss: 2.1683, average_loss: 1.9847
Epoch [44/50], Iter [155/209] Loss: 1.7339, average_loss: 1.9838
Epoch [44/50], Iter [160/209] Loss: 1.7963, average_loss: 1.9893
Epoch [44/50], Iter [165/209] Loss: 1.9386, average_loss: 1.9881
Epoch [44/50], Iter [170/209] Loss: 1.9235, average_loss: 1.9861
Epoch [44/50], Iter [175/209] Loss: 2.8602, average_loss: 1.9942
Epoch [44/50], Iter [180/209] Loss: 2.2865, average_loss: 1.9946
Epoch [44/50], Iter [185/209] Loss: 1.8043, average_loss: 1.9922
Epoch [44/50], Iter [190/209] Loss: 2.1351, average_loss: 1.9984
Epoch [44/50], Iter [195/209] Loss: 1.4134, average_loss: 1.9926
Epoch [44/50], Iter [200/209] Loss: 2.1803, average_loss: 1.9954
Epoch [44/50], Iter [205/209] Loss: 2.0864, average_loss: 1.9954
```

Updating best test loss: 2.86280

Starting epoch 45 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [45/50], Iter [5/209] Loss: 1.7546, average_loss: 2.2085
Epoch [45/50], Iter [10/209] Loss: 1.6724, average_loss: 2.0586
Epoch [45/50], Iter [15/209] Loss: 2.1127, average_loss: 2.0037
Epoch [45/50], Iter [20/209] Loss: 2.2204, average_loss: 2.0131
Epoch [45/50], Iter [25/209] Loss: 2.4268, average_loss: 1.9969
Epoch [45/50], Iter [30/209] Loss: 2.9243, average_loss: 2.0500
Epoch [45/50], Iter [35/209] Loss: 2.5914, average_loss: 2.0889
Epoch [45/50], Iter [40/209] Loss: 1.4293, average_loss: 2.0654
Epoch [45/50], Iter [45/209] Loss: 2.1157, average_loss: 2.0620
Epoch [45/50], Iter [50/209] Loss: 1.9490, average_loss: 2.0606
Epoch [45/50], Iter [55/209] Loss: 1.6795, average_loss: 2.0375
Epoch [45/50], Iter [60/209] Loss: 2.1091, average_loss: 2.0177
Epoch [45/50], Iter [65/209] Loss: 1.8081, average_loss: 2.0186
Epoch [45/50], Iter [70/209] Loss: 1.8377, average_loss: 2.0222
Epoch [45/50], Iter [75/209] Loss: 1.8491, average_loss: 2.0081
Epoch [45/50], Iter [80/209] Loss: 1.9526, average_loss: 2.0234
Epoch [45/50], Iter [85/209] Loss: 1.8738, average_loss: 2.0104
```

```
Epoch [45/50], Iter [90/209] Loss: 2.3813, average_loss: 2.0099
Epoch [45/50], Iter [95/209] Loss: 2.0316, average_loss: 2.0082
Epoch [45/50], Iter [100/209] Loss: 2.0848, average_loss: 2.0131
Epoch [45/50], Iter [105/209] Loss: 2.3299, average_loss: 2.0147
Epoch [45/50], Iter [110/209] Loss: 2.6815, average_loss: 2.0031
Epoch [45/50], Iter [115/209] Loss: 2.0072, average_loss: 2.0031
Epoch [45/50], Iter [120/209] Loss: 1.8221, average_loss: 2.0102
Epoch [45/50], Iter [125/209] Loss: 1.9276, average_loss: 2.0148
Epoch [45/50], Iter [130/209] Loss: 1.5727, average_loss: 2.0073
Epoch [45/50], Iter [135/209] Loss: 2.0783, average_loss: 2.0173
Epoch [45/50], Iter [140/209] Loss: 3.1002, average_loss: 2.0329
Epoch [45/50], Iter [145/209] Loss: 1.3991, average_loss: 2.0298
Epoch [45/50], Iter [150/209] Loss: 2.3961, average_loss: 2.0379
Epoch [45/50], Iter [155/209] Loss: 1.8123, average_loss: 2.0318
Epoch [45/50], Iter [160/209] Loss: 1.8813, average_loss: 2.0285
Epoch [45/50], Iter [165/209] Loss: 2.0425, average_loss: 2.0385
Epoch [45/50], Iter [170/209] Loss: 1.8702, average_loss: 2.0363
Epoch [45/50], Iter [175/209] Loss: 1.6134, average_loss: 2.0351
Epoch [45/50], Iter [180/209] Loss: 1.9414, average_loss: 2.0319
Epoch [45/50], Iter [185/209] Loss: 1.2402, average_loss: 2.0277
Epoch [45/50], Iter [190/209] Loss: 1.7024, average_loss: 2.0218
Epoch [45/50], Iter [195/209] Loss: 2.2017, average_loss: 2.0292
Epoch [45/50], Iter [200/209] Loss: 1.7938, average_loss: 2.0287
Epoch [45/50], Iter [205/209] Loss: 2.2380, average_loss: 2.0313
```

Starting epoch 46 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [46/50], Iter [5/209] Loss: 1.6565, average_loss: 1.8960
Epoch [46/50], Iter [10/209] Loss: 1.9783, average_loss: 1.8194
Epoch [46/50], Iter [15/209] Loss: 1.6210, average_loss: 1.8168
Epoch [46/50], Iter [20/209] Loss: 2.4777, average_loss: 1.9062
Epoch [46/50], Iter [25/209] Loss: 2.1310, average_loss: 1.9012
Epoch [46/50], Iter [30/209] Loss: 1.7180, average_loss: 1.9370
Epoch [46/50], Iter [35/209] Loss: 2.3480, average_loss: 1.9739
Epoch [46/50], Iter [40/209] Loss: 2.3012, average_loss: 1.9943
Epoch [46/50], Iter [45/209] Loss: 1.6317, average_loss: 1.9669
Epoch [46/50], Iter [50/209] Loss: 1.3580, average_loss: 1.9506
Epoch [46/50], Iter [55/209] Loss: 2.6850, average_loss: 1.9775
Epoch [46/50], Iter [60/209] Loss: 2.3995, average_loss: 1.9887
Epoch [46/50], Iter [65/209] Loss: 2.3744, average_loss: 1.9666
Epoch [46/50], Iter [70/209] Loss: 1.7875, average_loss: 1.9754
Epoch [46/50], Iter [75/209] Loss: 2.2078, average_loss: 1.9703
Epoch [46/50], Iter [80/209] Loss: 2.4315, average_loss: 1.9729
Epoch [46/50], Iter [85/209] Loss: 2.2081, average_loss: 1.9835
Epoch [46/50], Iter [90/209] Loss: 2.2837, average_loss: 1.9871
Epoch [46/50], Iter [95/209] Loss: 1.8061, average_loss: 1.9967
Epoch [46/50], Iter [100/209] Loss: 1.6105, average_loss: 2.0060
Epoch [46/50], Iter [105/209] Loss: 1.9149, average_loss: 2.0007
Epoch [46/50], Iter [110/209] Loss: 1.9881, average_loss: 1.9943
Epoch [46/50], Iter [115/209] Loss: 2.6455, average_loss: 1.9918
Epoch [46/50], Iter [120/209] Loss: 1.7324, average_loss: 1.9906
Epoch [46/50], Iter [125/209] Loss: 3.4269, average_loss: 2.0049
Epoch [46/50], Iter [130/209] Loss: 2.7895, average_loss: 2.0042
Epoch [46/50], Iter [135/209] Loss: 2.4986, average_loss: 2.0103
Epoch [46/50], Iter [140/209] Loss: 2.6293, average_loss: 2.0216
Epoch [46/50], Iter [145/209] Loss: 2.6584, average_loss: 2.0273
Epoch [46/50], Iter [150/209] Loss: 2.8960, average_loss: 2.0409
Epoch [46/50], Iter [155/209] Loss: 1.8679, average_loss: 2.0348
Epoch [46/50], Iter [160/209] Loss: 2.6250, average_loss: 2.0291
Epoch [46/50], Iter [165/209] Loss: 2.4341, average_loss: 2.0263
Epoch [46/50], Iter [170/209] Loss: 1.7778, average_loss: 2.0250
Epoch [46/50], Iter [175/209] Loss: 1.4359, average_loss: 2.0124
Epoch [46/50], Iter [180/209] Loss: 2.3756, average_loss: 2.0101
Epoch [46/50], Iter [185/209] Loss: 2.4563, average_loss: 2.0150
Epoch [46/50], Iter [190/209] Loss: 2.3001, average_loss: 2.0169
Epoch [46/50], Iter [195/209] Loss: 1.9468, average_loss: 2.0225
Epoch [46/50], Iter [200/209] Loss: 2.3358, average_loss: 2.0234
Epoch [46/50], Iter [205/209] Loss: 2.9097, average_loss: 2.0291
```

Starting epoch 47 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [47/50], Iter [5/209] Loss: 1.8481, average_loss: 1.8315
Epoch [47/50], Iter [10/209] Loss: 2.3261, average_loss: 2.0847
Epoch [47/50], Iter [15/209] Loss: 2.0559, average_loss: 2.0860
Epoch [47/50], Iter [20/209] Loss: 2.0046, average_loss: 2.0413
```

```
Epoch [47/50], Iter [20/209] Loss: 2.8916, average_loss: 2.0110
Epoch [47/50], Iter [25/209] Loss: 3.0157, average_loss: 2.0374
Epoch [47/50], Iter [30/209] Loss: 2.8232, average_loss: 2.0450
Epoch [47/50], Iter [35/209] Loss: 1.9141, average_loss: 2.0296
Epoch [47/50], Iter [40/209] Loss: 1.7929, average_loss: 2.0126
Epoch [47/50], Iter [45/209] Loss: 2.0580, average_loss: 2.0155
Epoch [47/50], Iter [50/209] Loss: 2.3628, average_loss: 2.0303
Epoch [47/50], Iter [55/209] Loss: 2.6065, average_loss: 2.0347
Epoch [47/50], Iter [60/209] Loss: 1.9494, average_loss: 2.0348
Epoch [47/50], Iter [65/209] Loss: 2.3122, average_loss: 2.0322
Epoch [47/50], Iter [70/209] Loss: 2.9367, average_loss: 2.0581
Epoch [47/50], Iter [75/209] Loss: 1.7315, average_loss: 2.0822
Epoch [47/50], Iter [80/209] Loss: 1.5145, average_loss: 2.0624
Epoch [47/50], Iter [85/209] Loss: 1.7494, average_loss: 2.0393
Epoch [47/50], Iter [90/209] Loss: 1.9693, average_loss: 2.0306
Epoch [47/50], Iter [95/209] Loss: 1.8865, average_loss: 2.0167
Epoch [47/50], Iter [100/209] Loss: 2.1829, average_loss: 2.0123
Epoch [47/50], Iter [105/209] Loss: 1.6094, average_loss: 2.0147
Epoch [47/50], Iter [110/209] Loss: 2.5761, average_loss: 2.0165
Epoch [47/50], Iter [115/209] Loss: 2.0740, average_loss: 2.0079
Epoch [47/50], Iter [120/209] Loss: 1.7544, average_loss: 2.0000
Epoch [47/50], Iter [125/209] Loss: 1.7287, average_loss: 1.9945
Epoch [47/50], Iter [130/209] Loss: 2.7067, average_loss: 2.0010
Epoch [47/50], Iter [135/209] Loss: 1.9280, average_loss: 1.9995
Epoch [47/50], Iter [140/209] Loss: 2.5722, average_loss: 2.0070
Epoch [47/50], Iter [145/209] Loss: 1.8857, average_loss: 2.0112
Epoch [47/50], Iter [150/209] Loss: 1.3483, average_loss: 2.0143
Epoch [47/50], Iter [155/209] Loss: 1.7263, average_loss: 2.0123
Epoch [47/50], Iter [160/209] Loss: 1.4601, average_loss: 2.0116
Epoch [47/50], Iter [165/209] Loss: 1.8389, average_loss: 2.0180
Epoch [47/50], Iter [170/209] Loss: 1.5502, average_loss: 2.0196
Epoch [47/50], Iter [175/209] Loss: 3.0111, average_loss: 2.0221
Epoch [47/50], Iter [180/209] Loss: 1.6594, average_loss: 2.0188
Epoch [47/50], Iter [185/209] Loss: 1.8553, average_loss: 2.0219
Epoch [47/50], Iter [190/209] Loss: 1.9808, average_loss: 2.0239
Epoch [47/50], Iter [195/209] Loss: 1.6848, average_loss: 2.0187
Epoch [47/50], Iter [200/209] Loss: 1.6175, average_loss: 2.0212
Epoch [47/50], Iter [205/209] Loss: 2.6714, average_loss: 2.0279
```

Starting epoch 48 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [48/50], Iter [5/209] Loss: 2.0685, average_loss: 2.0235
Epoch [48/50], Iter [10/209] Loss: 1.9458, average_loss: 2.0344
Epoch [48/50], Iter [15/209] Loss: 1.9820, average_loss: 2.0773
Epoch [48/50], Iter [20/209] Loss: 2.8316, average_loss: 2.1571
Epoch [48/50], Iter [25/209] Loss: 1.7635, average_loss: 2.0925
Epoch [48/50], Iter [30/209] Loss: 2.6740, average_loss: 2.0703
Epoch [48/50], Iter [35/209] Loss: 2.1098, average_loss: 2.0730
Epoch [48/50], Iter [40/209] Loss: 1.8757, average_loss: 2.0532
Epoch [48/50], Iter [45/209] Loss: 2.2121, average_loss: 2.0503
Epoch [48/50], Iter [50/209] Loss: 1.4945, average_loss: 2.0398
Epoch [48/50], Iter [55/209] Loss: 2.1425, average_loss: 2.0225
Epoch [48/50], Iter [60/209] Loss: 1.7300, average_loss: 2.0494
Epoch [48/50], Iter [65/209] Loss: 2.7230, average_loss: 2.0626
Epoch [48/50], Iter [70/209] Loss: 2.0507, average_loss: 2.0628
Epoch [48/50], Iter [75/209] Loss: 1.8542, average_loss: 2.0506
Epoch [48/50], Iter [80/209] Loss: 1.4133, average_loss: 2.0256
Epoch [48/50], Iter [85/209] Loss: 2.3148, average_loss: 2.0248
Epoch [48/50], Iter [90/209] Loss: 1.4128, average_loss: 2.0285
Epoch [48/50], Iter [95/209] Loss: 1.6149, average_loss: 2.0309
Epoch [48/50], Iter [100/209] Loss: 2.0044, average_loss: 2.0188
Epoch [48/50], Iter [105/209] Loss: 1.4043, average_loss: 2.0004
Epoch [48/50], Iter [110/209] Loss: 1.3690, average_loss: 1.9908
Epoch [48/50], Iter [115/209] Loss: 2.2975, average_loss: 1.9913
Epoch [48/50], Iter [120/209] Loss: 2.0338, average_loss: 1.9986
Epoch [48/50], Iter [125/209] Loss: 2.3100, average_loss: 1.9998
Epoch [48/50], Iter [130/209] Loss: 1.3750, average_loss: 1.9968
Epoch [48/50], Iter [135/209] Loss: 1.3990, average_loss: 1.9870
Epoch [48/50], Iter [140/209] Loss: 1.9276, average_loss: 1.9885
Epoch [48/50], Iter [145/209] Loss: 1.8123, average_loss: 1.9842
Epoch [48/50], Iter [150/209] Loss: 2.4742, average_loss: 1.9933
Epoch [48/50], Iter [155/209] Loss: 1.6610, average_loss: 1.9908
Epoch [48/50], Iter [160/209] Loss: 1.7568, average_loss: 1.9985
Epoch [48/50], Iter [165/209] Loss: 2.8734, average_loss: 2.0037
Epoch [48/50], Iter [170/209] Loss: 1.8323, average_loss: 1.9940
Epoch [48/50], Iter [175/209] Loss: 2.0306, average_loss: 1.9924
Epoch [48/50], Iter [180/209] Loss: 1.7283, average_loss: 1.9950
```

```
Epoch [48/50], Iter [180/209] Loss: 1.7203, average_loss: 1.9930
Epoch [48/50], Iter [185/209] Loss: 1.8044, average_loss: 1.9935
Epoch [48/50], Iter [190/209] Loss: 1.8113, average_loss: 1.9947
Epoch [48/50], Iter [195/209] Loss: 1.5038, average_loss: 1.9884
Epoch [48/50], Iter [200/209] Loss: 2.7127, average_loss: 1.9925
Epoch [48/50], Iter [205/209] Loss: 2.2556, average_loss: 2.0013
```

Starting epoch 49 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [49/50], Iter [5/209] Loss: 1.9572, average_loss: 1.8969
Epoch [49/50], Iter [10/209] Loss: 1.8501, average_loss: 1.8074
Epoch [49/50], Iter [15/209] Loss: 2.3728, average_loss: 1.8699
Epoch [49/50], Iter [20/209] Loss: 2.1012, average_loss: 1.9124
Epoch [49/50], Iter [25/209] Loss: 2.1796, average_loss: 1.9087
Epoch [49/50], Iter [30/209] Loss: 1.5403, average_loss: 1.9153
Epoch [49/50], Iter [35/209] Loss: 1.4073, average_loss: 1.8879
Epoch [49/50], Iter [40/209] Loss: 2.5950, average_loss: 1.8833
Epoch [49/50], Iter [45/209] Loss: 2.1381, average_loss: 1.9048
Epoch [49/50], Iter [50/209] Loss: 1.8677, average_loss: 1.8927
Epoch [49/50], Iter [55/209] Loss: 2.0012, average_loss: 1.9255
Epoch [49/50], Iter [60/209] Loss: 1.5688, average_loss: 1.9290
Epoch [49/50], Iter [65/209] Loss: 1.6133, average_loss: 1.9198
Epoch [49/50], Iter [70/209] Loss: 2.1171, average_loss: 1.9111
Epoch [49/50], Iter [75/209] Loss: 1.6036, average_loss: 1.9268
Epoch [49/50], Iter [80/209] Loss: 2.0232, average_loss: 1.9532
Epoch [49/50], Iter [85/209] Loss: 2.5093, average_loss: 1.9652
Epoch [49/50], Iter [90/209] Loss: 1.6614, average_loss: 1.9808
Epoch [49/50], Iter [95/209] Loss: 2.0191, average_loss: 1.9919
Epoch [49/50], Iter [100/209] Loss: 3.2153, average_loss: 2.0032
Epoch [49/50], Iter [105/209] Loss: 2.1082, average_loss: 2.0058
Epoch [49/50], Iter [110/209] Loss: 1.8142, average_loss: 2.0095
Epoch [49/50], Iter [115/209] Loss: 1.9738, average_loss: 1.9973
Epoch [49/50], Iter [120/209] Loss: 2.3006, average_loss: 1.9943
Epoch [49/50], Iter [125/209] Loss: 1.5983, average_loss: 1.9889
Epoch [49/50], Iter [130/209] Loss: 1.9779, average_loss: 1.9825
Epoch [49/50], Iter [135/209] Loss: 1.8782, average_loss: 1.9722
Epoch [49/50], Iter [140/209] Loss: 2.7216, average_loss: 1.9858
Epoch [49/50], Iter [145/209] Loss: 1.3351, average_loss: 1.9801
Epoch [49/50], Iter [150/209] Loss: 2.2004, average_loss: 1.9844
Epoch [49/50], Iter [155/209] Loss: 2.3584, average_loss: 1.9876
Epoch [49/50], Iter [160/209] Loss: 2.0511, average_loss: 1.9863
Epoch [49/50], Iter [165/209] Loss: 1.8826, average_loss: 1.9892
Epoch [49/50], Iter [170/209] Loss: 2.0341, average_loss: 1.9878
Epoch [49/50], Iter [175/209] Loss: 1.8938, average_loss: 1.9856
Epoch [49/50], Iter [180/209] Loss: 1.9815, average_loss: 1.9888
Epoch [49/50], Iter [185/209] Loss: 1.4412, average_loss: 1.9846
Epoch [49/50], Iter [190/209] Loss: 2.5751, average_loss: 1.9839
Epoch [49/50], Iter [195/209] Loss: 1.4749, average_loss: 1.9848
Epoch [49/50], Iter [200/209] Loss: 2.2630, average_loss: 1.9932
Epoch [49/50], Iter [205/209] Loss: 1.5762, average_loss: 1.9952
```

Starting epoch 50 / 50

Learning Rate for this epoch: 1e-05

```
Epoch [50/50], Iter [5/209] Loss: 1.5599, average_loss: 2.0217
Epoch [50/50], Iter [10/209] Loss: 2.0195, average_loss: 1.9915
Epoch [50/50], Iter [15/209] Loss: 2.5350, average_loss: 2.0531
Epoch [50/50], Iter [20/209] Loss: 2.2445, average_loss: 2.0485
Epoch [50/50], Iter [25/209] Loss: 1.4464, average_loss: 2.0255
Epoch [50/50], Iter [30/209] Loss: 1.6955, average_loss: 2.0018
Epoch [50/50], Iter [35/209] Loss: 1.8594, average_loss: 2.0144
Epoch [50/50], Iter [40/209] Loss: 1.6972, average_loss: 1.9807
Epoch [50/50], Iter [45/209] Loss: 1.6318, average_loss: 1.9634
Epoch [50/50], Iter [50/209] Loss: 1.8386, average_loss: 1.9404
Epoch [50/50], Iter [55/209] Loss: 2.7126, average_loss: 1.9737
Epoch [50/50], Iter [60/209] Loss: 1.7756, average_loss: 1.9906
Epoch [50/50], Iter [65/209] Loss: 2.2641, average_loss: 1.9889
Epoch [50/50], Iter [70/209] Loss: 2.1277, average_loss: 1.9994
Epoch [50/50], Iter [75/209] Loss: 1.5106, average_loss: 1.9914
Epoch [50/50], Iter [80/209] Loss: 1.8552, average_loss: 1.9896
Epoch [50/50], Iter [85/209] Loss: 2.4454, average_loss: 1.9882
Epoch [50/50], Iter [90/209] Loss: 3.9792, average_loss: 2.0224
Epoch [50/50], Iter [95/209] Loss: 1.7621, average_loss: 2.0069
Epoch [50/50], Iter [100/209] Loss: 1.9898, average_loss: 2.0183
Epoch [50/50], Iter [105/209] Loss: 1.8132, average_loss: 2.0082
Epoch [50/50], Iter [110/209] Loss: 2.0780, average_loss: 2.0014
Epoch [50/50], Iter [115/209] Loss: 2.7000, average_loss: 2.0005
```



```
Epoch [50/50], Iter [110/209] Loss: 2.7000, average_loss: 2.0000
Epoch [50/50], Iter [120/209] Loss: 2.2674, average_loss: 2.0057
Epoch [50/50], Iter [125/209] Loss: 1.5452, average_loss: 2.0005
Epoch [50/50], Iter [130/209] Loss: 2.2107, average_loss: 1.9975
Epoch [50/50], Iter [135/209] Loss: 2.6105, average_loss: 2.0058
Epoch [50/50], Iter [140/209] Loss: 2.2156, average_loss: 2.0149
Epoch [50/50], Iter [145/209] Loss: 1.8247, average_loss: 2.0131
Epoch [50/50], Iter [150/209] Loss: 1.6112, average_loss: 2.0056
Epoch [50/50], Iter [155/209] Loss: 1.3598, average_loss: 2.0069
Epoch [50/50], Iter [160/209] Loss: 2.4584, average_loss: 2.0074
Epoch [50/50], Iter [165/209] Loss: 2.1562, average_loss: 2.0072
Epoch [50/50], Iter [170/209] Loss: 1.7782, average_loss: 2.0024
Epoch [50/50], Iter [175/209] Loss: 1.8421, average_loss: 1.9982
Epoch [50/50], Iter [180/209] Loss: 1.5608, average_loss: 1.9972
Epoch [50/50], Iter [185/209] Loss: 2.0288, average_loss: 1.9950
Epoch [50/50], Iter [190/209] Loss: 2.3986, average_loss: 1.9918
Epoch [50/50], Iter [195/209] Loss: 2.6568, average_loss: 1.9954
Epoch [50/50], Iter [200/209] Loss: 1.7769, average_loss: 1.9857
Epoch [50/50], Iter [205/209] Loss: 2.3632, average_loss: 1.9910
```

## View example predictions

In [26]:

```
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)

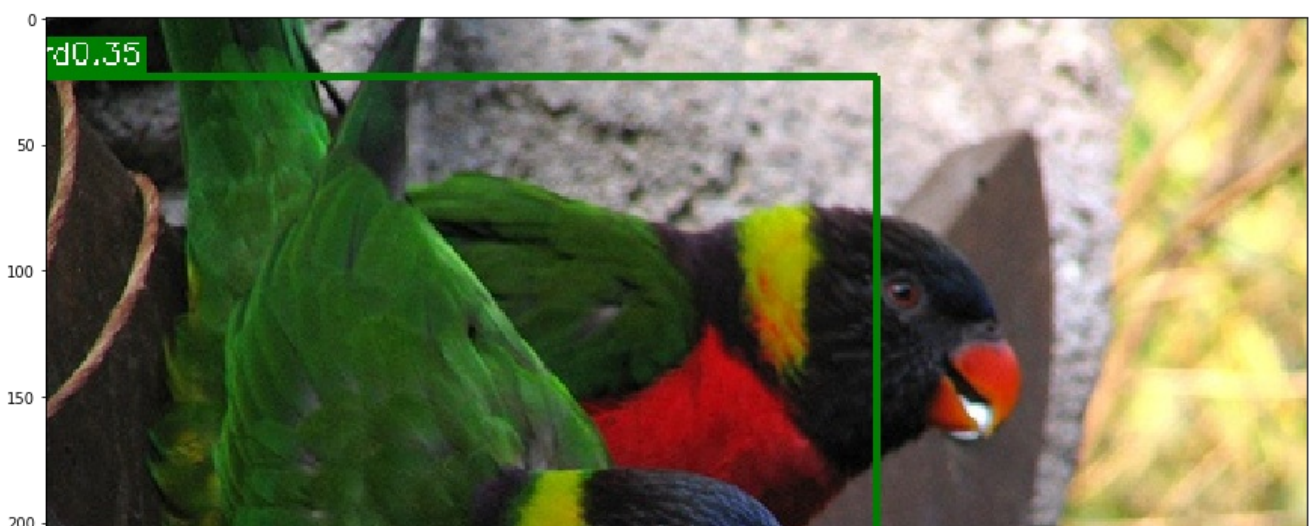
print('predicting...')
result = predict_image(net, image_name, root_img_directory=file_root_test)
for left_up, right_bottom, class_name, _, prob in result:
    color = COLORS[VOC_CLASSES.index(class_name)]
    cv2.rectangle(image, left_up, right_bottom, color, 2)
    label = class_name + str(round(prob,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

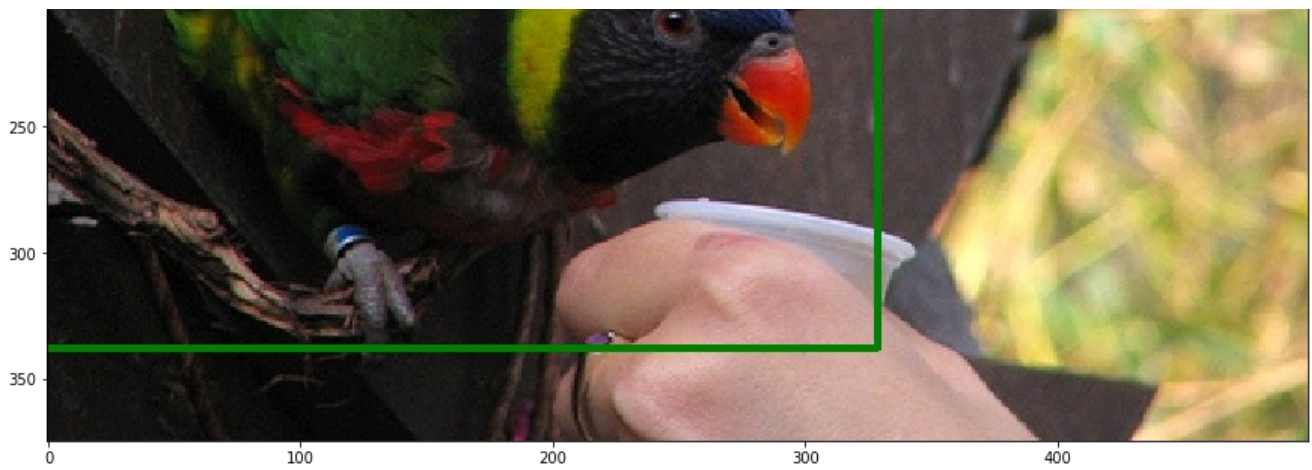
plt.figure(figsize = (15,15))
plt.imshow(image)
```

predicting...

Out [26]:

<matplotlib.image.AxesImage at 0x7fe8220e14e0>





## Evaluate on Test

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

In [27]:

```
test_aps = evaluate(net, test_dataset_file=annotation_file_test)
```

---Evaluate model on test samples---

100%|██████████| 4950/4950 [01:20<00:00, 61.56it/s]

```
---class aeroplane ap 0.5219431666198617---
---class bicycle ap 0.5531239051111997---
---class bird ap 0.4687586039012494---
---class boat ap 0.29974215858188114---
---class bottle ap 0.17845519373584734---
---class bus ap 0.6069752156819173---
---class car ap 0.6402728196694963---
---class cat ap 0.6659224839746425---
---class chair ap 0.28632598656343594---
---class cow ap 0.4631157860369395---
---class diningtable ap 0.33702565324480327---
---class dog ap 0.5798445290243366---
---class horse ap 0.6719423663676491---
---class motorbike ap 0.5579638123008808---
---class person ap 0.4989258659594752---
---class pottedplant ap 0.2093356809038186---
---class sheep ap 0.47632316107997574---
---class sofa ap 0.4530998925955493---
---class train ap 0.6474984791245608---
---class tvmonitor ap 0.46491123033716253---
---map 0.4790752995407342---
```

In [28]:

```
output_submission_csv('my_solution.csv', test_aps)
```