

In [1]:

```
!bash download_data.sh
```

```
--2018-11-11 20:55:19-- http://host.robots.ox.ac.uk/pascal/VOC/voc2007/VOCTrainval_06-Nov-2007.tar
```

```
Resolving host.robots.ox.ac.uk (host.robots.ox.ac.uk)... 129.67.94.152
```

```
Connecting to host.robots.ox.ac.uk (host.robots.ox.ac.uk)|129.67.94.152|:80... connected.
```

```
HTTP request sent, awaiting response... 200 OK
```

```
Length: 460032000 (439M) [application/x-tar]
```

```
Saving to: 'VOCTrainval_06-Nov-2007.tar'
```

```
VOCTrainval_06-Nov- 100%[=====>] 438.72M  4.92MB/s    in 70s
```

```
2018-11-11 20:56:30 (6.25 MB/s) - 'VOCTrainval_06-Nov-2007.tar' saved [460032000/460032000]
```

```
--2018-11-11 20:56:31-- http://host.robots.ox.ac.uk/pascal/VOC/voc2007/VOCTest_06-Nov-2007.tar
```

```
Resolving host.robots.ox.ac.uk (host.robots.ox.ac.uk)... 129.67.94.152
```

```
Connecting to host.robots.ox.ac.uk (host.robots.ox.ac.uk)|129.67.94.152|:80... connected.
```

```
HTTP request sent, awaiting response... 200 OK
```

```
Length: 451020800 (430M) [application/x-tar]
```

```
Saving to: 'VOCTest_06-Nov-2007.tar'
```

```
VOCTest_06-Nov-2007 100%[=====>] 430.13M  1.55MB/s    in 3m 5s
```

```
2018-11-11 20:59:36 (2.33 MB/s) - 'VOCTest_06-Nov-2007.tar' saved [451020800/451020800]
```

In [3]:

```
import sys
sys.path.append('.local/lib/python3.5/site-packages')
```

In [1]:

```
import os
import random

import cv2
import numpy as np

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

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

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

## Initialization

In [2]:

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

In [3]:

```
# YOLO network hyperparameters
B = 2 # number of bounding box predictions per cell
```

```
S = 14 # width/height of network output grid (larger than 7x7 from paper since we use a different network)
```

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

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

In [4]:

```
load_network_path = None
pretrained = True

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

Load pre-trained model

In [5]:

```
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 [6]:

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

## Reading Pascal Data

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

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

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

In [7]:

```
file_root_train = 'VOCdevkit_2007/VOC2007/JPEGImages/'
annotation_file_train = 'voc2007.txt'

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

Initializing dataset

Loaded 5011 train images

In [8]:

```
file_root_test = 'VOCdevkit_2007/VOC2007test/JPEGImages/'
annotation_file_test = 'voc2007test.txt'

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

Initializing dataset  
Loaded 4950 test images

## Train detector

In [9]:

```
best_test_loss = np.inf

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

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

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

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

    total_loss = 0.

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

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

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

    # 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

24  
24  
24  
24  
24

Epoch [1/50], Iter [5/209] Loss: 65.6879, average\_loss: 69.4378

24  
24  
24  
24  
24

Epoch [1/50], Iter [10/209] Loss: 42.2458, average\_loss: 60.6060

24  
24  
24  
24  
24

Epoch [1/50], Iter [15/209] Loss: 24.0236, average\_loss: 51.3178

24  
24  
24  
24  
24

Epoch [1/50], Iter [20/209] Loss: 17.3097, average\_loss: 43.3452

24  
24  
24  
24  
24

Epoch [1/50], Iter [25/209] Loss: 13.3500, average\_loss: 37.6687

24  
24  
24  
24  
24

Epoch [1/50], Iter [30/209] Loss: 10.7645, average\_loss: 33.9599

24  
24  
24  
24  
24

Epoch [1/50], Iter [35/209] Loss: 12.0206, average\_loss: 30.7107

24  
24  
24  
24  
24

Epoch [1/50], Iter [40/209] Loss: 9.7506, average\_loss: 28.0264

24  
24  
24  
24  
24

Epoch [1/50], Iter [45/209] Loss: 8.0579, average\_loss: 25.9670

24  
24  
24  
24  
24

Epoch [1/50], Iter [50/209] Loss: 7.0559, average\_loss: 24.2046

24  
24  
24  
24  
24

Epoch [1/50], Iter [55/209] Loss: 7.2270, average\_loss: 22.6728

24  
24  
24  
24  
24

Epoch [1/50], Iter [60/209] Loss: 8.7898, average\_loss: 21.4282

24  
24  
24  
24

24  
Epoch [1/50], Iter [65/209] Loss: 8.3291, average\_loss: 20.4513  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [70/209] Loss: 4.9328, average\_loss: 19.6052  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [75/209] Loss: 9.5404, average\_loss: 18.8071  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [80/209] Loss: 7.2301, average\_loss: 18.1086  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [85/209] Loss: 9.0094, average\_loss: 17.5191  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [90/209] Loss: 7.0734, average\_loss: 16.9620  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [95/209] Loss: 7.7621, average\_loss: 16.4114  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [100/209] Loss: 5.2301, average\_loss: 15.9004  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [105/209] Loss: 7.7581, average\_loss: 15.4347  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [110/209] Loss: 6.8917, average\_loss: 15.1137  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [115/209] Loss: 6.3885, average\_loss: 14.7135  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [120/209] Loss: 6.5114, average\_loss: 14.3364  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [125/209] Loss: 6.1557, average\_loss: 14.0156  
24  
24  
24

24  
24  
Epoch [1/50], Iter [130/209] Loss: 5.0821, average\_loss: 13.7464  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [135/209] Loss: 5.1985, average\_loss: 13.4681  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [140/209] Loss: 5.8316, average\_loss: 13.1959  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [145/209] Loss: 5.0564, average\_loss: 12.9427  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [150/209] Loss: 5.5757, average\_loss: 12.7279  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [155/209] Loss: 5.7552, average\_loss: 12.5095  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [160/209] Loss: 5.1231, average\_loss: 12.2796  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [165/209] Loss: 6.3750, average\_loss: 12.0772  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [170/209] Loss: 5.2222, average\_loss: 11.8876  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [175/209] Loss: 6.1586, average\_loss: 11.7038  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [180/209] Loss: 4.1181, average\_loss: 11.5243  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [185/209] Loss: 6.1509, average\_loss: 11.3598  
24  
24  
24  
24  
24  
Epoch [1/50], Iter [190/209] Loss: 6.2490, average\_loss: 11.2087  
24  
24

```

Epoch [1/50], Iter [195/209] Loss: 6.3707, average_loss: 11.0738
Epoch [1/50], Iter [200/209] Loss: 6.6295, average_loss: 10.9376
Epoch [1/50], Iter [205/209] Loss: 5.8793, average_loss: 10.8073

```

[illegible]



Updating best test loss: 5.34364

Starting epoch 2 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [2/50], Iter [5/209] Loss: 3.7070, average\_loss: 4.9735

24

24

24

24

24

Epoch [2/50], Iter [10/209] Loss: 6.6279, average\_loss: 5.5379

24

24

24

24

24

Epoch [2/50], Iter [15/209] Loss: 5.0977, average\_loss: 5.3622

24

24

24

24

24

Epoch [2/50], Iter [20/209] Loss: 5.2979, average\_loss: 5.4347

24

24

24

24

24

Epoch [2/50], Iter [25/209] Loss: 5.1754, average\_loss: 5.3554

24

24

24

24

24

Epoch [2/50], Iter [30/209] Loss: 3.9401, average\_loss: 5.2879

24

24

24

24

24

Epoch [2/50], Iter [35/209] Loss: 4.2849, average\_loss: 5.2364

24

24

24

24

24

Epoch [2/50], Iter [40/209] Loss: 5.2242, average\_loss: 5.2336

24

24

24

24

24

Epoch [2/50], Iter [45/209] Loss: 3.6850, average\_loss: 5.2567

24

24

24

24

24

Epoch [2/50], Iter [50/209] Loss: 4.3417, average\_loss: 5.2386

24

24

24

24

24

Epoch [2/50], Iter [55/209] Loss: 4.8755, average\_loss: 5.2468

24

24

24

24

24

Epoch [2/50], Iter [60/209] Loss: 5.6357, average\_loss: 5.1881

24

24

24

24

24  
Epoch [2/50], Iter [65/209] Loss: 4.8417, average\_loss: 5.1680  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [70/209] Loss: 4.6432, average\_loss: 5.1493  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [75/209] Loss: 4.9919, average\_loss: 5.1421  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [80/209] Loss: 5.2130, average\_loss: 5.1207  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [85/209] Loss: 3.9593, average\_loss: 5.0962  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [90/209] Loss: 6.3142, average\_loss: 5.0992  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [95/209] Loss: 5.0256, average\_loss: 5.0598  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [100/209] Loss: 5.4860, average\_loss: 5.0358  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [105/209] Loss: 5.6590, average\_loss: 5.0657  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [110/209] Loss: 2.8855, average\_loss: 5.0331  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [115/209] Loss: 5.5264, average\_loss: 5.0128  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [120/209] Loss: 4.4664, average\_loss: 4.9932  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [125/209] Loss: 4.4721, average\_loss: 5.0081  
24  
24  
24

24  
24  
Epoch [2/50], Iter [130/209] Loss: 6.1990, average\_loss: 5.0140  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [135/209] Loss: 4.9573, average\_loss: 4.9947  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [140/209] Loss: 5.3036, average\_loss: 4.9854  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [145/209] Loss: 6.1484, average\_loss: 5.0043  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [150/209] Loss: 3.5341, average\_loss: 5.0016  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [155/209] Loss: 6.6815, average\_loss: 5.0096  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [160/209] Loss: 5.0581, average\_loss: 5.0125  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [165/209] Loss: 4.5173, average\_loss: 5.0055  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [170/209] Loss: 4.5801, average\_loss: 4.9885  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [175/209] Loss: 4.3289, average\_loss: 4.9663  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [180/209] Loss: 4.0634, average\_loss: 4.9701  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [185/209] Loss: 5.2439, average\_loss: 4.9622  
24  
24  
24  
24  
24  
Epoch [2/50], Iter [190/209] Loss: 7.1000, average\_loss: 4.9691  
24  
24

[illegible]

[illegible]

[illegible]

Up

Starting epoch 3 / 50

Starting epoch 2 / 50

```
Starting epoch 3 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [3/50], Iter [5/209] Loss: 3.7995, average_loss: 4.1504
24
24
24
24
24
Epoch [3/50], Iter [10/209] Loss: 4.5488, average_loss: 4.2438
24
24
24
24
24
Epoch [3/50], Iter [15/209] Loss: 4.8742, average_loss: 4.4054
24
24
24
24
24
Epoch [3/50], Iter [20/209] Loss: 4.0483, average_loss: 4.3858
24
24
24
24
24
Epoch [3/50], Iter [25/209] Loss: 3.4506, average_loss: 4.3882
24
24
24
24
24
Epoch [3/50], Iter [30/209] Loss: 3.9365, average_loss: 4.4589
24
24
24
24
24
Epoch [3/50], Iter [35/209] Loss: 6.7469, average_loss: 4.5520
24
24
24
24
24
Epoch [3/50], Iter [40/209] Loss: 2.7027, average_loss: 4.4753
24
24
24
24
24
Epoch [3/50], Iter [45/209] Loss: 5.3099, average_loss: 4.4499
24
24
24
24
24
Epoch [3/50], Iter [50/209] Loss: 4.2617, average_loss: 4.4288
24
24
24
24
24
Epoch [3/50], Iter [55/209] Loss: 5.3999, average_loss: 4.4736
24
24
24
24
24
Epoch [3/50], Iter [60/209] Loss: 5.3047, average_loss: 4.5154
24
24
24
24
```



```
24
24
Epoch [3/50], Iter [65/209] Loss: 3.7512, average_loss: 4.4645
24
24
24
24
24
Epoch [3/50], Iter [70/209] Loss: 3.0732, average_loss: 4.4606
24
24
24
24
24
Epoch [3/50], Iter [75/209] Loss: 3.8658, average_loss: 4.4490
24
24
24
24
24
Epoch [3/50], Iter [80/209] Loss: 4.0771, average_loss: 4.4496
24
24
24
24
24
Epoch [3/50], Iter [85/209] Loss: 5.3176, average_loss: 4.4744
24
24
24
24
24
Epoch [3/50], Iter [90/209] Loss: 4.6098, average_loss: 4.4701
24
24
24
24
24
Epoch [3/50], Iter [95/209] Loss: 3.7071, average_loss: 4.4679
24
24
24
24
24
Epoch [3/50], Iter [100/209] Loss: 4.4994, average_loss: 4.4556
24
24
24
24
24
Epoch [3/50], Iter [105/209] Loss: 2.7554, average_loss: 4.4319
24
24
24
24
24
Epoch [3/50], Iter [110/209] Loss: 3.6555, average_loss: 4.4306
24
24
24
24
24
Epoch [3/50], Iter [115/209] Loss: 5.8730, average_loss: 4.4362
24
24
24
24
24
Epoch [3/50], Iter [120/209] Loss: 3.5010, average_loss: 4.4325
24
24
24
24
24
Epoch [3/50], Iter [125/209] Loss: 4.0485, average_loss: 4.4282
24
24
24
```

24  
24  
24  
Epoch [3/50], Iter [130/209] Loss: 3.5390, average\_loss: 4.4045  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [135/209] Loss: 3.2579, average\_loss: 4.3786  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [140/209] Loss: 3.7992, average\_loss: 4.3709  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [145/209] Loss: 3.1344, average\_loss: 4.3836  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [150/209] Loss: 3.7174, average\_loss: 4.3675  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [155/209] Loss: 6.8318, average\_loss: 4.3923  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [160/209] Loss: 4.8927, average\_loss: 4.3908  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [165/209] Loss: 5.2323, average\_loss: 4.4054  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [170/209] Loss: 3.4412, average\_loss: 4.4025  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [175/209] Loss: 3.9667, average\_loss: 4.3912  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [180/209] Loss: 3.3973, average\_loss: 4.3799  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [185/209] Loss: 3.3198, average\_loss: 4.3656  
24  
24  
24  
24  
24  
Epoch [3/50], Iter [190/209] Loss: 3.3825, average\_loss: 4.3520  
24  
24

```
Epoch [3/50], Iter [195/209] Loss: 3.4944, average_loss: 4.3594
Epoch [3/50], Iter [200/209] Loss: 3.2583, average_loss: 4.3608
Epoch [3/50], Iter [205/209] Loss: 3.2172, average_loss: 4.3521
```



```
Updating best test loss: 4.22631
```

```
Starting epoch 4 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [4/50], Iter [5/209] Loss: 6.1938, average_loss: 4.5734
24
24
24
24
24
Epoch [4/50], Iter [10/209] Loss: 5.6334, average_loss: 4.5352
24
24
24
24
24
Epoch [4/50], Iter [15/209] Loss: 3.9680, average_loss: 4.3825
24
24
24
24
24
Epoch [4/50], Iter [20/209] Loss: 4.0072, average_loss: 4.3749
24
24
24
24
24
Epoch [4/50], Iter [25/209] Loss: 3.7596, average_loss: 4.4316
24
24
24
24
24
Epoch [4/50], Iter [30/209] Loss: 3.7393, average_loss: 4.3543
24
24
24
24
24
Epoch [4/50], Iter [35/209] Loss: 3.4678, average_loss: 4.2427
24
24
24
24
24
Epoch [4/50], Iter [40/209] Loss: 5.1092, average_loss: 4.2427
24
24
24
24
24
Epoch [4/50], Iter [45/209] Loss: 5.2433, average_loss: 4.2223
24
24
24
24
24
Epoch [4/50], Iter [50/209] Loss: 3.9187, average_loss: 4.1810
24
24
24
24
24
Epoch [4/50], Iter [55/209] Loss: 3.5941, average_loss: 4.1084
24
24
24
24
24
Epoch [4/50], Iter [60/209] Loss: 3.1660, average_loss: 4.0757
24
24
24
```

24  
24  
Epoch [4/50], Iter [65/209] Loss: 3.7277, average\_loss: 4.0441  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [70/209] Loss: 2.9779, average\_loss: 4.0239  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [75/209] Loss: 3.6883, average\_loss: 4.0204  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [80/209] Loss: 2.6603, average\_loss: 4.0113  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [85/209] Loss: 4.3872, average\_loss: 4.0296  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [90/209] Loss: 5.6760, average\_loss: 4.0492  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [95/209] Loss: 4.7031, average\_loss: 4.0283  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [100/209] Loss: 3.9303, average\_loss: 4.0316  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [105/209] Loss: 3.8416, average\_loss: 4.0585  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [110/209] Loss: 4.5316, average\_loss: 4.0508  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [115/209] Loss: 3.6520, average\_loss: 4.0640  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [120/209] Loss: 3.2733, average\_loss: 4.0619  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [125/209] Loss: 4.1447, average\_loss: 4.0627  
24  
24

24  
24  
24  
Epoch [4/50], Iter [130/209] Loss: 3.5189, average\_loss: 4.0491  
24  
24  
24  
24  
24  
Epoch [4/50], Iter [135/209] Loss: 3.7099, average\_loss: 4.0521  
24  
24  
24  
24  
Epoch [4/50], Iter [140/209] Loss: 3.7486, average\_loss: 4.0509  
24  
24  
24  
24  
Epoch [4/50], Iter [145/209] Loss: 3.1039, average\_loss: 4.0465  
24  
24  
24  
24  
Epoch [4/50], Iter [150/209] Loss: 4.2577, average\_loss: 4.0400  
24  
24  
24  
24  
Epoch [4/50], Iter [155/209] Loss: 4.4062, average\_loss: 4.0371  
24  
24  
24  
24  
Epoch [4/50], Iter [160/209] Loss: 4.4969, average\_loss: 4.0457  
24  
24  
24  
24  
Epoch [4/50], Iter [165/209] Loss: 6.0970, average\_loss: 4.0643  
24  
24  
24  
24  
Epoch [4/50], Iter [170/209] Loss: 3.7901, average\_loss: 4.0500  
24  
24  
24  
24  
Epoch [4/50], Iter [175/209] Loss: 4.3490, average\_loss: 4.0441  
24  
24  
24  
24  
Epoch [4/50], Iter [180/209] Loss: 5.3790, average\_loss: 4.0471  
24  
24  
24  
24  
Epoch [4/50], Iter [185/209] Loss: 3.5411, average\_loss: 4.0411  
24  
24  
24  
24  
Epoch [4/50], Iter [190/209] Loss: 3.7845, average\_loss: 4.0379  
24



```
Epoch [4/50], Iter [195/209] Loss: 4.0263, average_loss: 4.0378
Epoch [4/50], Iter [200/209] Loss: 3.5197, average_loss: 4.0287
Epoch [4/50], Iter [205/209] Loss: 2.9175, average_loss: 4.0217
```

[illegible]

```
Updating best test loss: 3.99619
```

```
Starting epoch 5 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [5/50], Iter [5/209] Loss: 2.8624, average_loss: 3.3163
24
24
24
24
24
Epoch [5/50], Iter [10/209] Loss: 3.7067, average_loss: 3.4266
24
24
24
24
24
Epoch [5/50], Iter [15/209] Loss: 3.8950, average_loss: 3.3943
24
24
24
24
24
Epoch [5/50], Iter [20/209] Loss: 3.2925, average_loss: 3.5462
24
24
24
24
24
Epoch [5/50], Iter [25/209] Loss: 3.3668, average_loss: 3.6229
24
24
24
24
24
Epoch [5/50], Iter [30/209] Loss: 3.1047, average_loss: 3.6679
24
24
24
24
24
Epoch [5/50], Iter [35/209] Loss: 4.4713, average_loss: 3.6650
24
24
24
24
24
Epoch [5/50], Iter [40/209] Loss: 4.9048, average_loss: 3.6856
24
24
24
24
24
Epoch [5/50], Iter [45/209] Loss: 2.9042, average_loss: 3.7100
24
24
24
24
24
Epoch [5/50], Iter [50/209] Loss: 3.2386, average_loss: 3.7165
24
24
24
24
24
Epoch [5/50], Iter [55/209] Loss: 4.0024, average_loss: 3.7324
24
24
24
24
24
Epoch [5/50], Iter [60/209] Loss: 3.5139, average_loss: 3.7355
24
24
24
```

24  
24  
Epoch [5/50], Iter [65/209] Loss: 3.4816, average\_loss: 3.7322  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [70/209] Loss: 3.4435, average\_loss: 3.7661  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [75/209] Loss: 3.6645, average\_loss: 3.7317  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [80/209] Loss: 3.1871, average\_loss: 3.7103  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [85/209] Loss: 2.8627, average\_loss: 3.7456  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [90/209] Loss: 3.9650, average\_loss: 3.7315  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [95/209] Loss: 4.7629, average\_loss: 3.7509  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [100/209] Loss: 3.6966, average\_loss: 3.7639  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [105/209] Loss: 2.9025, average\_loss: 3.7869  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [110/209] Loss: 3.2128, average\_loss: 3.7857  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [115/209] Loss: 3.2857, average\_loss: 3.7795  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [120/209] Loss: 2.6292, average\_loss: 3.7635  
24  
24  
24  
24  
24  
Epoch [5/50], Iter [125/209] Loss: 3.1828, average\_loss: 3.7594  
24  
24

```
24
24
24
Epoch [5/50], Iter [130/209] Loss: 3.7927, average_loss: 3.7580
24
24
24
24
Epoch [5/50], Iter [135/209] Loss: 4.0606, average_loss: 3.7555
24
24
24
24
Epoch [5/50], Iter [140/209] Loss: 4.9441, average_loss: 3.7744
24
24
24
24
Epoch [5/50], Iter [145/209] Loss: 3.2868, average_loss: 3.7750
24
24
24
24
Epoch [5/50], Iter [150/209] Loss: 2.7588, average_loss: 3.7573
24
24
24
24
Epoch [5/50], Iter [155/209] Loss: 3.8585, average_loss: 3.7481
24
24
24
24
Epoch [5/50], Iter [160/209] Loss: 5.1383, average_loss: 3.7599
24
24
24
24
Epoch [5/50], Iter [165/209] Loss: 4.4953, average_loss: 3.7667
24
24
24
24
Epoch [5/50], Iter [170/209] Loss: 4.2162, average_loss: 3.7643
24
24
24
24
Epoch [5/50], Iter [175/209] Loss: 3.4078, average_loss: 3.7584
24
24
24
24
Epoch [5/50], Iter [180/209] Loss: 3.3428, average_loss: 3.7489
24
24
24
24
Epoch [5/50], Iter [185/209] Loss: 3.1512, average_loss: 3.7596
24
24
24
24
Epoch [5/50], Iter [190/209] Loss: 3.8247, average_loss: 3.7604
24
```

[illegible]

[illegible]



[illegible]

```
Updating best test loss: 3.78137
```

```
Updating best test loss: 3.78137
```

```
Starting epoch 6 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [6/50], Iter [5/209] Loss: 3.9614, average_loss: 3.6328
24
24
24
24
24
Epoch [6/50], Iter [10/209] Loss: 3.1428, average_loss: 3.3375
24
24
24
24
24
Epoch [6/50], Iter [15/209] Loss: 4.3734, average_loss: 3.6719
24
24
24
24
24
Epoch [6/50], Iter [20/209] Loss: 3.9429, average_loss: 3.6902
24
24
24
24
24
Epoch [6/50], Iter [25/209] Loss: 3.0221, average_loss: 3.5886
24
24
24
24
24
Epoch [6/50], Iter [30/209] Loss: 3.9099, average_loss: 3.6458
24
24
24
24
24
Epoch [6/50], Iter [35/209] Loss: 2.8920, average_loss: 3.6147
24
24
24
24
24
Epoch [6/50], Iter [40/209] Loss: 3.9659, average_loss: 3.6276
24
24
24
24
24
Epoch [6/50], Iter [45/209] Loss: 2.7823, average_loss: 3.5854
24
24
24
24
24
Epoch [6/50], Iter [50/209] Loss: 3.5496, average_loss: 3.5880
24
24
24
24
24
Epoch [6/50], Iter [55/209] Loss: 4.7267, average_loss: 3.5988
24
24
24
24
24
Epoch [6/50], Iter [60/209] Loss: 2.8758, average_loss: 3.5844
24
24
24
```

24  
24  
24  
Epoch [6/50], Iter [65/209] Loss: 3.2286, average\_loss: 3.5714  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [70/209] Loss: 3.4518, average\_loss: 3.5975  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [75/209] Loss: 2.7728, average\_loss: 3.5871  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [80/209] Loss: 2.9207, average\_loss: 3.5677  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [85/209] Loss: 4.1377, average\_loss: 3.5717  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [90/209] Loss: 4.2948, average\_loss: 3.5868  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [95/209] Loss: 3.5801, average\_loss: 3.5679  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [100/209] Loss: 2.8802, average\_loss: 3.5604  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [105/209] Loss: 3.6507, average\_loss: 3.5601  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [110/209] Loss: 3.2016, average\_loss: 3.5714  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [115/209] Loss: 3.3055, average\_loss: 3.5568  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [120/209] Loss: 3.3134, average\_loss: 3.5465  
24  
24  
24  
24  
24  
Epoch [6/50], Iter [125/209] Loss: 3.1580, average\_loss: 3.5529  
24  
24

```
24
24
24
24
Epoch [6/50], Iter [130/209] Loss: 3.0637, average_loss: 3.5441
24
24
24
24
24
Epoch [6/50], Iter [135/209] Loss: 3.7625, average_loss: 3.5585
24
24
24
24
24
Epoch [6/50], Iter [140/209] Loss: 3.2268, average_loss: 3.5575
24
24
24
24
24
Epoch [6/50], Iter [145/209] Loss: 3.1717, average_loss: 3.5632
24
24
24
24
24
Epoch [6/50], Iter [150/209] Loss: 3.3420, average_loss: 3.5740
24
24
24
24
24
Epoch [6/50], Iter [155/209] Loss: 3.2765, average_loss: 3.5601
24
24
24
24
24
Epoch [6/50], Iter [160/209] Loss: 3.5484, average_loss: 3.5726
24
24
24
24
24
Epoch [6/50], Iter [165/209] Loss: 4.5785, average_loss: 3.5754
24
24
24
24
24
Epoch [6/50], Iter [170/209] Loss: 3.4448, average_loss: 3.5654
24
24
24
24
24
Epoch [6/50], Iter [175/209] Loss: 3.4284, average_loss: 3.5575
24
24
24
24
24
Epoch [6/50], Iter [180/209] Loss: 2.8841, average_loss: 3.5562
24
24
24
24
24
Epoch [6/50], Iter [185/209] Loss: 3.3212, average_loss: 3.5544
24
24
24
24
24
Epoch [6/50], Iter [190/209] Loss: 4.1389, average_loss: 3.5550
24
```



[illegible]

```
Updating best test loss: 3.72864
```

```
Starting epoch 7 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [7/50], Iter [5/209] Loss: 3.5111, average_loss: 3.4172
24
24
24
24
24
Epoch [7/50], Iter [10/209] Loss: 3.4899, average_loss: 3.5757
24
24
24
24
24
Epoch [7/50], Iter [15/209] Loss: 3.7131, average_loss: 3.4839
24
24
24
24
24
Epoch [7/50], Iter [20/209] Loss: 2.7615, average_loss: 3.4143
24
24
24
24
24
Epoch [7/50], Iter [25/209] Loss: 3.0926, average_loss: 3.4500
24
24
24
24
24
Epoch [7/50], Iter [30/209] Loss: 3.2588, average_loss: 3.4185
24
24
24
24
24
Epoch [7/50], Iter [35/209] Loss: 2.6887, average_loss: 3.3512
24
24
24
24
24
Epoch [7/50], Iter [40/209] Loss: 4.1505, average_loss: 3.3978
24
24
24
24
24
Epoch [7/50], Iter [45/209] Loss: 3.8068, average_loss: 3.4144
24
24
24
24
24
Epoch [7/50], Iter [50/209] Loss: 3.9174, average_loss: 3.4287
24
24
24
24
24
Epoch [7/50], Iter [55/209] Loss: 2.9487, average_loss: 3.4792
24
24
24
24
24
Epoch [7/50], Iter [60/209] Loss: 3.0494, average_loss: 3.4581
24
24
```



24  
24  
24  
Epoch [7/50], Iter [65/209] Loss: 3.3601, average\_loss: 3.4697  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [70/209] Loss: 3.0626, average\_loss: 3.4520  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [75/209] Loss: 3.3094, average\_loss: 3.4721  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [80/209] Loss: 2.9356, average\_loss: 3.4669  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [85/209] Loss: 2.4737, average\_loss: 3.4307  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [90/209] Loss: 3.2779, average\_loss: 3.4213  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [95/209] Loss: 4.8212, average\_loss: 3.4448  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [100/209] Loss: 3.4014, average\_loss: 3.4331  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [105/209] Loss: 2.9614, average\_loss: 3.4355  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [110/209] Loss: 2.6687, average\_loss: 3.4104  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [115/209] Loss: 3.9936, average\_loss: 3.4139  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [120/209] Loss: 3.5564, average\_loss: 3.4097  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [125/209] Loss: 3.7382, average\_loss: 3.4111  
24

24  
24  
24  
24  
Epoch [7/50], Iter [130/209] Loss: 2.8198, average\_loss: 3.3994  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [135/209] Loss: 2.7217, average\_loss: 3.3841  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [140/209] Loss: 2.8873, average\_loss: 3.3972  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [145/209] Loss: 3.9569, average\_loss: 3.4103  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [150/209] Loss: 2.9137, average\_loss: 3.4160  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [155/209] Loss: 3.6381, average\_loss: 3.4151  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [160/209] Loss: 2.7871, average\_loss: 3.4035  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [165/209] Loss: 3.5182, average\_loss: 3.4066  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [170/209] Loss: 2.4056, average\_loss: 3.4022  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [175/209] Loss: 4.4970, average\_loss: 3.4143  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [180/209] Loss: 2.0813, average\_loss: 3.4047  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [185/209] Loss: 4.3312, average\_loss: 3.4025  
24  
24  
24  
24  
24  
Epoch [7/50], Iter [190/209] Loss: 3.0337, average\_loss: 3.3995

```

Epoch [7/50], Iter [195/209] Loss: 3.2382, average_loss: 3.3939
Epoch [7/50], Iter [200/209] Loss: 3.1497, average_loss: 3.3989
Epoch [7/50], Iter [205/209] Loss: 4.7812, average_loss: 3.4006
```

[illegible]

[illegible]

```
Updating best test loss: 3.57215
```

```
Starting epoch 8 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [8/50], Iter [5/209] Loss: 2.8217, average_loss: 3.1461
24
24
24
24
24
Epoch [8/50], Iter [10/209] Loss: 2.8462, average_loss: 3.1944
24
24
24
24
24
Epoch [8/50], Iter [15/209] Loss: 3.4291, average_loss: 3.1987
24
24
24
24
24
Epoch [8/50], Iter [20/209] Loss: 3.0952, average_loss: 3.2578
24
24
24
24
24
Epoch [8/50], Iter [25/209] Loss: 2.1436, average_loss: 3.2514
24
24
24
24
24
Epoch [8/50], Iter [30/209] Loss: 3.0145, average_loss: 3.2076
24
24
24
24
24
Epoch [8/50], Iter [35/209] Loss: 3.6992, average_loss: 3.1860
24
24
24
24
24
Epoch [8/50], Iter [40/209] Loss: 4.4053, average_loss: 3.2338
24
24
24
24
24
Epoch [8/50], Iter [45/209] Loss: 2.7568, average_loss: 3.2495
24
24
24
24
24
Epoch [8/50], Iter [50/209] Loss: 2.8449, average_loss: 3.2222
24
24
24
24
24
Epoch [8/50], Iter [55/209] Loss: 2.9893, average_loss: 3.2292
24
24
24
24
24
Epoch [8/50], Iter [60/209] Loss: 3.6053, average_loss: 3.2426
24
24
```

```
24
24
24
Epoch [8/50], Iter [65/209] Loss: 3.1449, average_loss: 3.2663
24
24
24
24
Epoch [8/50], Iter [70/209] Loss: 3.6756, average_loss: 3.2961
24
24
24
24
Epoch [8/50], Iter [75/209] Loss: 3.7494, average_loss: 3.2930
24
24
24
24
Epoch [8/50], Iter [80/209] Loss: 2.6129, average_loss: 3.2791
24
24
24
24
Epoch [8/50], Iter [85/209] Loss: 3.1356, average_loss: 3.2745
24
24
24
24
Epoch [8/50], Iter [90/209] Loss: 3.3504, average_loss: 3.2633
24
24
24
24
Epoch [8/50], Iter [95/209] Loss: 3.8009, average_loss: 3.2641
24
24
24
24
Epoch [8/50], Iter [100/209] Loss: 4.0530, average_loss: 3.2580
24
24
24
24
Epoch [8/50], Iter [105/209] Loss: 2.6376, average_loss: 3.2638
24
24
24
24
Epoch [8/50], Iter [110/209] Loss: 4.8088, average_loss: 3.2718
24
24
24
24
Epoch [8/50], Iter [115/209] Loss: 4.7426, average_loss: 3.2861
24
24
24
24
Epoch [8/50], Iter [120/209] Loss: 3.4807, average_loss: 3.2793
24
24
24
24
Epoch [8/50], Iter [125/209] Loss: 3.0709, average_loss: 3.2836
24
```

24  
24  
24  
24  
Epoch [8/50], Iter [130/209] Loss: 2.6636, average\_loss: 3.2779  
24  
24  
24  
24  
Epoch [8/50], Iter [135/209] Loss: 3.7302, average\_loss: 3.2727  
24  
24  
24  
24  
Epoch [8/50], Iter [140/209] Loss: 3.4381, average\_loss: 3.2671  
24  
24  
24  
24  
Epoch [8/50], Iter [145/209] Loss: 2.3112, average\_loss: 3.2629  
24  
24  
24  
24  
Epoch [8/50], Iter [150/209] Loss: 4.3144, average\_loss: 3.2787  
24  
24  
24  
24  
Epoch [8/50], Iter [155/209] Loss: 3.2192, average\_loss: 3.2731  
24  
24  
24  
24  
Epoch [8/50], Iter [160/209] Loss: 4.2948, average\_loss: 3.2701  
24  
24  
24  
24  
Epoch [8/50], Iter [165/209] Loss: 3.1089, average\_loss: 3.2722  
24  
24  
24  
24  
Epoch [8/50], Iter [170/209] Loss: 3.6082, average\_loss: 3.2727  
24  
24  
24  
24  
Epoch [8/50], Iter [175/209] Loss: 3.0718, average\_loss: 3.2644  
24  
24  
24  
24  
Epoch [8/50], Iter [180/209] Loss: 3.2292, average\_loss: 3.2720  
24  
24  
24  
24  
Epoch [8/50], Iter [185/209] Loss: 2.3881, average\_loss: 3.2628  
24  
24  
24  
24  
Epoch [8/50], Iter [190/209] Loss: 3.2006, average\_loss: 3.2652



[illegible]

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24



```
Starting epoch 9 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [9/50], Iter [5/209] Loss: 3.3986, average_loss: 3.5310
24
24
24
24
24
Epoch [9/50], Iter [10/209] Loss: 4.3252, average_loss: 3.6467
24
24
24
24
24
Epoch [9/50], Iter [15/209] Loss: 3.0646, average_loss: 3.4837
24
24
24
24
24
Epoch [9/50], Iter [20/209] Loss: 2.8027, average_loss: 3.3231
24
24
24
24
24
Epoch [9/50], Iter [25/209] Loss: 2.7750, average_loss: 3.2809
24
24
24
24
24
Epoch [9/50], Iter [30/209] Loss: 3.1260, average_loss: 3.2875
24
24
24
24
24
Epoch [9/50], Iter [35/209] Loss: 3.1738, average_loss: 3.2989
24
24
24
24
24
Epoch [9/50], Iter [40/209] Loss: 3.9459, average_loss: 3.3346
24
24
24
24
24
Epoch [9/50], Iter [45/209] Loss: 2.7834, average_loss: 3.3782
24
24
24
24
24
Epoch [9/50], Iter [50/209] Loss: 2.2543, average_loss: 3.3185
24
24
24
24
24
Epoch [9/50], Iter [55/209] Loss: 4.0197, average_loss: 3.3249
24
24
24
24
24
Epoch [9/50], Iter [60/209] Loss: 2.9891, average_loss: 3.2972
24
24
```

24  
24  
24  
24  
Epoch [9/50], Iter [65/209] Loss: 3.3193, average\_loss: 3.3332  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [70/209] Loss: 3.1991, average\_loss: 3.3023  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [75/209] Loss: 3.1380, average\_loss: 3.2928  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [80/209] Loss: 2.7819, average\_loss: 3.2882  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [85/209] Loss: 2.8595, average\_loss: 3.2815  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [90/209] Loss: 2.5790, average\_loss: 3.2521  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [95/209] Loss: 2.9190, average\_loss: 3.2445  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [100/209] Loss: 4.4403, average\_loss: 3.2400  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [105/209] Loss: 2.9356, average\_loss: 3.2505  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [110/209] Loss: 2.3555, average\_loss: 3.2559  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [115/209] Loss: 3.0824, average\_loss: 3.2419  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [120/209] Loss: 3.1236, average\_loss: 3.2324  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [125/209] Loss: 3.8954, average\_loss: 3.2480  
24

24  
24  
24  
24  
24  
Epoch [9/50], Iter [130/209] Loss: 3.3017, average\_loss: 3.2397  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [135/209] Loss: 3.3366, average\_loss: 3.2358  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [140/209] Loss: 3.1656, average\_loss: 3.2299  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [145/209] Loss: 3.2083, average\_loss: 3.2150  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [150/209] Loss: 2.7026, average\_loss: 3.2067  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [155/209] Loss: 3.6061, average\_loss: 3.2096  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [160/209] Loss: 3.1057, average\_loss: 3.2050  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [165/209] Loss: 3.8114, average\_loss: 3.2099  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [170/209] Loss: 2.8770, average\_loss: 3.2086  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [175/209] Loss: 2.9458, average\_loss: 3.2066  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [180/209] Loss: 3.4295, average\_loss: 3.2174  
24  
24  
24  
24  
24  
Epoch [9/50], Iter [185/209] Loss: 3.0732, average\_loss: 3.2148  
24  
24  
24  
24  
24

[illegible]

[illegible]



[illegible]

```
Updating best test loss: 3.41507
```

```
Starting epoch 10 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [10/50], Iter [5/209] Loss: 3.7180, average_loss: 2.8048
24
24
24
24
24
Epoch [10/50], Iter [10/209] Loss: 3.0616, average_loss: 2.8217
24
24
24
24
24
Epoch [10/50], Iter [15/209] Loss: 2.4428, average_loss: 2.9278
24
24
24
24
24
Epoch [10/50], Iter [20/209] Loss: 2.9647, average_loss: 2.9973
24
24
24
24
24
Epoch [10/50], Iter [25/209] Loss: 3.7027, average_loss: 3.0936
24
24
24
24
24
Epoch [10/50], Iter [30/209] Loss: 4.6254, average_loss: 3.1319
24
24
24
24
24
Epoch [10/50], Iter [35/209] Loss: 2.6339, average_loss: 3.0578
24
24
24
24
24
Epoch [10/50], Iter [40/209] Loss: 3.1153, average_loss: 3.0214
24
24
24
24
24
Epoch [10/50], Iter [45/209] Loss: 3.7266, average_loss: 3.0611
24
24
24
24
24
Epoch [10/50], Iter [50/209] Loss: 3.0565, average_loss: 3.0359
24
24
24
24
24
Epoch [10/50], Iter [55/209] Loss: 3.6896, average_loss: 3.0167
24
24
24
24
24
Epoch [10/50], Iter [60/209] Loss: 2.3739, average_loss: 3.0080
24
```

24  
24  
24  
24  
Epoch [10/50], Iter [65/209] Loss: 3.1096, average\_loss: 3.0286  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [70/209] Loss: 3.3006, average\_loss: 3.0540  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [75/209] Loss: 2.1517, average\_loss: 3.0381  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [80/209] Loss: 3.6926, average\_loss: 3.0406  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [85/209] Loss: 3.1160, average\_loss: 3.0486  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [90/209] Loss: 2.8997, average\_loss: 3.0627  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [95/209] Loss: 3.7923, average\_loss: 3.0883  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [100/209] Loss: 2.5237, average\_loss: 3.0925  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [105/209] Loss: 3.2370, average\_loss: 3.0804  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [110/209] Loss: 2.7039, average\_loss: 3.0711  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [115/209] Loss: 3.5300, average\_loss: 3.0829  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [120/209] Loss: 3.0871, average\_loss: 3.0989  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [125/209] Loss: 2.8821, average\_loss: 3.1039

24  
24  
24  
24  
24  
Epoch [10/50], Iter [130/209] Loss: 3.6862, average\_loss: 3.1074  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [135/209] Loss: 2.9399, average\_loss: 3.1069  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [140/209] Loss: 3.3100, average\_loss: 3.0992  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [145/209] Loss: 3.9816, average\_loss: 3.1142  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [150/209] Loss: 2.4569, average\_loss: 3.1111  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [155/209] Loss: 3.6729, average\_loss: 3.1099  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [160/209] Loss: 2.4884, average\_loss: 3.0970  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [165/209] Loss: 2.6508, average\_loss: 3.1078  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [170/209] Loss: 3.7373, average\_loss: 3.1069  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [175/209] Loss: 3.2572, average\_loss: 3.1075  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [180/209] Loss: 4.2059, average\_loss: 3.1134  
24  
24  
24  
24  
24  
Epoch [10/50], Iter [185/209] Loss: 2.2888, average\_loss: 3.1045  
24  
24  
24  
24  
24

[illegible]

[illegible]

[illegible]

Updating best test loss: 3.41056

Spawning 2000 2000 2000: 0.12000

Starting epoch 11 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [11/50], Iter [5/209] Loss: 2.4928, average\_loss: 3.3151

24

24

24

24

24

Epoch [11/50], Iter [10/209] Loss: 2.3638, average\_loss: 3.2500

24

24

24

24

24

Epoch [11/50], Iter [15/209] Loss: 2.1495, average\_loss: 3.1544

24

24

24

24

24

Epoch [11/50], Iter [20/209] Loss: 2.7976, average\_loss: 3.1210

24

24

24

24

24

Epoch [11/50], Iter [25/209] Loss: 2.9706, average\_loss: 3.1448

24

24

24

24

24

Epoch [11/50], Iter [30/209] Loss: 2.9133, average\_loss: 3.2023

24

24

24

24

24

Epoch [11/50], Iter [35/209] Loss: 2.4415, average\_loss: 3.1611

24

24

24

24

24

Epoch [11/50], Iter [40/209] Loss: 3.4171, average\_loss: 3.1191

24

24

24

24

24

Epoch [11/50], Iter [45/209] Loss: 3.2152, average\_loss: 3.1031

24

24

24

24

24

Epoch [11/50], Iter [50/209] Loss: 2.0572, average\_loss: 3.1089

24

24

24

24

24

Epoch [11/50], Iter [55/209] Loss: 2.5963, average\_loss: 3.1128

24

24

24

24

24

Epoch [11/50], Iter [60/209] Loss: 4.1205, average\_loss: 3.1423

24



```
24
24
24
24
Epoch [11/50], Iter [65/209] Loss: 3.6280, average_loss: 3.1383
24
24
24
24
Epoch [11/50], Iter [70/209] Loss: 3.1176, average_loss: 3.1264
24
24
24
24
Epoch [11/50], Iter [75/209] Loss: 2.7329, average_loss: 3.1172
24
24
24
24
Epoch [11/50], Iter [80/209] Loss: 3.4714, average_loss: 3.1093
24
24
24
24
Epoch [11/50], Iter [85/209] Loss: 3.6061, average_loss: 3.1088
24
24
24
24
Epoch [11/50], Iter [90/209] Loss: 4.1457, average_loss: 3.1218
24
24
24
24
Epoch [11/50], Iter [95/209] Loss: 3.6419, average_loss: 3.1224
24
24
24
24
Epoch [11/50], Iter [100/209] Loss: 2.6771, average_loss: 3.1018
24
24
24
24
Epoch [11/50], Iter [105/209] Loss: 3.3813, average_loss: 3.1028
24
24
24
24
Epoch [11/50], Iter [110/209] Loss: 3.5851, average_loss: 3.0963
24
24
24
24
Epoch [11/50], Iter [115/209] Loss: 2.6084, average_loss: 3.0815
24
24
24
24
Epoch [11/50], Iter [120/209] Loss: 3.1873, average_loss: 3.0768
24
24
24
24
Epoch [11/50], Iter [125/209] Loss: 3.6093, average_loss: 3.0804
```

```
Epoch [11/50], Iter [120/209] Loss: 3.0000, average_loss: 3.0000
24
24
24
24
24
Epoch [11/50], Iter [130/209] Loss: 2.7349, average_loss: 3.0706
24
24
24
24
24
Epoch [11/50], Iter [135/209] Loss: 2.2342, average_loss: 3.0679
24
24
24
24
24
Epoch [11/50], Iter [140/209] Loss: 3.9570, average_loss: 3.0786
24
24
24
24
24
Epoch [11/50], Iter [145/209] Loss: 3.1413, average_loss: 3.0729
24
24
24
24
24
Epoch [11/50], Iter [150/209] Loss: 3.6168, average_loss: 3.0730
24
24
24
24
24
Epoch [11/50], Iter [155/209] Loss: 2.9622, average_loss: 3.0726
24
24
24
24
24
Epoch [11/50], Iter [160/209] Loss: 3.4471, average_loss: 3.0774
24
24
24
24
24
Epoch [11/50], Iter [165/209] Loss: 4.0007, average_loss: 3.0671
24
24
24
24
24
Epoch [11/50], Iter [170/209] Loss: 3.7031, average_loss: 3.0674
24
24
24
24
24
Epoch [11/50], Iter [175/209] Loss: 3.4891, average_loss: 3.0743
24
24
24
24
24
Epoch [11/50], Iter [180/209] Loss: 2.8571, average_loss: 3.0758
24
24
24
24
24
Epoch [11/50], Iter [185/209] Loss: 3.5833, average_loss: 3.0834
24
24
24
24
24
```

[illegible]





Updating best test loss: 3.34524

Starting epoch 12 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [12/50], Iter [5/209] Loss: 4.2544, average\_loss: 2.9280

24

24

24

24

24

Epoch [12/50], Iter [10/209] Loss: 2.9997, average\_loss: 2.9021

24

24

24

24

24

Epoch [12/50], Iter [15/209] Loss: 2.3894, average\_loss: 2.8571

24

24

24

24

24

Epoch [12/50], Iter [20/209] Loss: 2.5015, average\_loss: 2.8676

24

24

24

24

24

Epoch [12/50], Iter [25/209] Loss: 3.0953, average\_loss: 2.9684

24

24

24

24

24

Epoch [12/50], Iter [30/209] Loss: 2.8560, average\_loss: 2.9776

24

24

24

24

24

Epoch [12/50], Iter [35/209] Loss: 4.5545, average\_loss: 3.0308

24

24

24

24

24

Epoch [12/50], Iter [40/209] Loss: 3.2127, average\_loss: 3.0530

24

24

24

24

24

Epoch [12/50], Iter [45/209] Loss: 2.5375, average\_loss: 3.0738

24

24

24

24

24

Epoch [12/50], Iter [50/209] Loss: 2.8359, average\_loss: 3.0761

24

24

24

24

24

Epoch [12/50], Iter [55/209] Loss: 2.2518, average\_loss: 3.0537

24

24

24

24

24

Epoch [12/50], Iter [60/209] Loss: 2.9807, average\_loss: 3.0135

24

24  
24  
24  
24  
24  
Epoch [12/50], Iter [65/209] Loss: 2.2224, average\_loss: 2.9949  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [70/209] Loss: 1.9405, average\_loss: 2.9728  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [75/209] Loss: 2.6718, average\_loss: 2.9798  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [80/209] Loss: 2.8412, average\_loss: 2.9776  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [85/209] Loss: 2.5689, average\_loss: 2.9708  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [90/209] Loss: 2.6952, average\_loss: 2.9617  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [95/209] Loss: 2.7455, average\_loss: 2.9520  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [100/209] Loss: 2.5256, average\_loss: 2.9458  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [105/209] Loss: 3.2616, average\_loss: 2.9536  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [110/209] Loss: 2.8169, average\_loss: 2.9557  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [115/209] Loss: 2.3929, average\_loss: 2.9426  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [120/209] Loss: 2.6200, average\_loss: 2.9325  
24  
24  
24  
24  
24

Epoch [12/50], Iter [125/209] Loss: 3.0278, average\_loss: 2.9299  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [130/209] Loss: 2.9119, average\_loss: 2.9402  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [135/209] Loss: 2.3743, average\_loss: 2.9241  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [140/209] Loss: 3.6015, average\_loss: 2.9367  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [145/209] Loss: 3.3486, average\_loss: 2.9464  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [150/209] Loss: 2.7371, average\_loss: 2.9420  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [155/209] Loss: 2.9489, average\_loss: 2.9512  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [160/209] Loss: 2.6913, average\_loss: 2.9584  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [165/209] Loss: 3.2033, average\_loss: 2.9566  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [170/209] Loss: 2.0515, average\_loss: 2.9614  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [175/209] Loss: 2.5297, average\_loss: 2.9525  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [180/209] Loss: 5.0646, average\_loss: 2.9656  
24  
24  
24  
24  
24  
Epoch [12/50], Iter [185/209] Loss: 4.4210, average\_loss: 2.9787  
24  
24  
24  
24



```
Epoch [12/50], Iter [190/209] Loss: 2.4912, average_loss: 2.9844
Epoch [12/50], Iter [195/209] Loss: 3.0638, average_loss: 2.9746
Epoch [12/50], Iter [200/209] Loss: 2.3229, average_loss: 2.9667
Epoch [12/50], Iter [205/209] Loss: 3.0952, average_loss: 2.9577
```



[illegible]

Updating best test loss: 3.32490

Starting epoch 13 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [13/50], Iter [5/209] Loss: 3.1045, average\_loss: 2.9439

24

24

24

24

24

Epoch [13/50], Iter [10/209] Loss: 2.2885, average\_loss: 2.8379

24

24

24

24

24

Epoch [13/50], Iter [15/209] Loss: 2.8043, average\_loss: 2.8517

24

24

24

24

24

Epoch [13/50], Iter [20/209] Loss: 2.7415, average\_loss: 2.8927

24

24

24

24

24

Epoch [13/50], Iter [25/209] Loss: 2.3061, average\_loss: 2.8870

24

24

24

24

24

Epoch [13/50], Iter [30/209] Loss: 2.9069, average\_loss: 2.8873

24

24

24

24

24

Epoch [13/50], Iter [35/209] Loss: 2.4177, average\_loss: 2.9419

24

24

24

24

24

Epoch [13/50], Iter [40/209] Loss: 2.2741, average\_loss: 2.9300

24

24

24

24

24

Epoch [13/50], Iter [45/209] Loss: 2.1753, average\_loss: 2.8909

24

24

24

24

24

Epoch [13/50], Iter [50/209] Loss: 2.9879, average\_loss: 2.8885

24

24

24

24

24

Epoch [13/50], Iter [55/209] Loss: 3.2609, average\_loss: 2.8868

24

24

24

24

24

Epoch [13/50], Iter [60/209] Loss: 3.9233, average\_loss: 2.8899

24  
24  
24  
24  
24  
Epoch [13/50], Iter [65/209] Loss: 2.4621, average\_loss: 2.8621  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [70/209] Loss: 3.2589, average\_loss: 2.8807  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [75/209] Loss: 2.2209, average\_loss: 2.8584  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [80/209] Loss: 3.2434, average\_loss: 2.8573  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [85/209] Loss: 2.7011, average\_loss: 2.8626  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [90/209] Loss: 2.8880, average\_loss: 2.8732  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [95/209] Loss: 2.3421, average\_loss: 2.8685  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [100/209] Loss: 2.9887, average\_loss: 2.8693  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [105/209] Loss: 2.6406, average\_loss: 2.8812  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [110/209] Loss: 2.3458, average\_loss: 2.8747  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [115/209] Loss: 2.4211, average\_loss: 2.8658  
24  
24  
24  
24  
24  
Epoch [13/50], Iter [120/209] Loss: 3.3650, average\_loss: 2.8782  
24  
24  
24  
24  
24

```
Epoch [13/50], Iter [125/209] Loss: 3.3487, average_loss: 2.8867
24
24
24
24
24
Epoch [13/50], Iter [130/209] Loss: 2.3967, average_loss: 2.8823
24
24
24
24
24
Epoch [13/50], Iter [135/209] Loss: 3.0153, average_loss: 2.8878
24
24
24
24
24
Epoch [13/50], Iter [140/209] Loss: 2.7578, average_loss: 2.8850
24
24
24
24
24
Epoch [13/50], Iter [145/209] Loss: 4.4813, average_loss: 2.8982
24
24
24
24
24
Epoch [13/50], Iter [150/209] Loss: 2.4786, average_loss: 2.8829
24
24
24
24
24
Epoch [13/50], Iter [155/209] Loss: 3.4943, average_loss: 2.8890
24
24
24
24
24
Epoch [13/50], Iter [160/209] Loss: 2.9268, average_loss: 2.8928
24
24
24
24
24
Epoch [13/50], Iter [165/209] Loss: 3.4803, average_loss: 2.8946
24
24
24
24
24
Epoch [13/50], Iter [170/209] Loss: 2.3681, average_loss: 2.8827
24
24
24
24
24
Epoch [13/50], Iter [175/209] Loss: 2.8272, average_loss: 2.8842
24
24
24
24
24
Epoch [13/50], Iter [180/209] Loss: 3.1320, average_loss: 2.8958
24
24
24
24
24
Epoch [13/50], Iter [185/209] Loss: 3.8245, average_loss: 2.9090
24
24
24
24
```

```
Epoch [13/50], Iter [190/209] Loss: 3.1463, average_loss: 2.9170
Epoch [13/50], Iter [195/209] Loss: 4.2682, average_loss: 2.9173
Epoch [13/50], Iter [200/209] Loss: 2.4774, average_loss: 2.9056
Epoch [13/50], Iter [205/209] Loss: 1.7580, average_loss: 2.9012
```

[illegible]





~  
Updating best test loss: 3.21093

Starting epoch 14 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24  
Epoch [14/50], Iter [5/209] Loss: 2.3988, average\_loss: 2.8795  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [10/209] Loss: 2.7498, average\_loss: 2.8104  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [15/209] Loss: 2.8132, average\_loss: 2.6797  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [20/209] Loss: 2.7158, average\_loss: 2.7451  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [25/209] Loss: 2.5445, average\_loss: 2.7261  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [30/209] Loss: 2.9500, average\_loss: 2.7564  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [35/209] Loss: 3.9366, average\_loss: 2.8028  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [40/209] Loss: 3.3015, average\_loss: 2.8232  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [45/209] Loss: 2.2894, average\_loss: 2.8729  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [50/209] Loss: 3.4442, average\_loss: 2.9098  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [55/209] Loss: 2.5922, average\_loss: 2.9035  
24  
24  
24  
24  
24  
Epoch [14/50], Iter [60/209] Loss: 2.6326, average\_loss: 2.9277

```
Epoch [14/50], Iter [60/209] Loss: 2.9320, average_loss: 2.9277
24
24
24
24
24
Epoch [14/50], Iter [65/209] Loss: 2.4123, average_loss: 2.9121
24
24
24
24
24
Epoch [14/50], Iter [70/209] Loss: 2.7277, average_loss: 2.8796
24
24
24
24
24
Epoch [14/50], Iter [75/209] Loss: 3.1556, average_loss: 2.9251
24
24
24
24
24
Epoch [14/50], Iter [80/209] Loss: 2.2143, average_loss: 2.9108
24
24
24
24
24
Epoch [14/50], Iter [85/209] Loss: 3.0070, average_loss: 2.9141
24
24
24
24
24
Epoch [14/50], Iter [90/209] Loss: 2.8989, average_loss: 2.9331
24
24
24
24
24
Epoch [14/50], Iter [95/209] Loss: 2.6332, average_loss: 2.9289
24
24
24
24
24
Epoch [14/50], Iter [100/209] Loss: 2.4954, average_loss: 2.9152
24
24
24
24
24
Epoch [14/50], Iter [105/209] Loss: 2.6846, average_loss: 2.9090
24
24
24
24
24
Epoch [14/50], Iter [110/209] Loss: 2.6756, average_loss: 2.9032
24
24
24
24
24
Epoch [14/50], Iter [115/209] Loss: 2.4793, average_loss: 2.9230
24
24
24
24
24
Epoch [14/50], Iter [120/209] Loss: 2.4414, average_loss: 2.8998
24
24
24
24
24
```

```
Epoch [14/50], Iter [125/209] Loss: 2.6782, average_loss: 2.8837
24
24
24
24
24
Epoch [14/50], Iter [130/209] Loss: 4.1014, average_loss: 2.8965
24
24
24
24
24
Epoch [14/50], Iter [135/209] Loss: 2.4546, average_loss: 2.8843
24
24
24
24
24
Epoch [14/50], Iter [140/209] Loss: 3.3075, average_loss: 2.8871
24
24
24
24
24
Epoch [14/50], Iter [145/209] Loss: 4.4242, average_loss: 2.8887
24
24
24
24
24
Epoch [14/50], Iter [150/209] Loss: 2.0922, average_loss: 2.8830
24
24
24
24
24
Epoch [14/50], Iter [155/209] Loss: 2.5689, average_loss: 2.8813
24
24
24
24
24
Epoch [14/50], Iter [160/209] Loss: 2.6134, average_loss: 2.8710
24
24
24
24
24
Epoch [14/50], Iter [165/209] Loss: 3.0187, average_loss: 2.8777
24
24
24
24
24
Epoch [14/50], Iter [170/209] Loss: 2.7598, average_loss: 2.8768
24
24
24
24
24
Epoch [14/50], Iter [175/209] Loss: 2.9635, average_loss: 2.8717
24
24
24
24
24
Epoch [14/50], Iter [180/209] Loss: 2.1970, average_loss: 2.8617
24
24
24
24
24
Epoch [14/50], Iter [185/209] Loss: 3.3428, average_loss: 2.8602
24
24
24
24
```

```
Epoch [14/50], Iter [190/209] Loss: 2.8721, average_loss: 2.8584
Epoch [14/50], Iter [195/209] Loss: 2.2388, average_loss: 2.8529
Epoch [14/50], Iter [200/209] Loss: 2.6388, average_loss: 2.8493
Epoch [14/50], Iter [205/209] Loss: 2.3575, average_loss: 2.8456
```

[illegible]

[illegible]

6

Updating best test loss: 3.14737

Starting epoch 15 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [15/50], Iter [5/209] Loss: 2.9484, average\_loss: 2.9745

24

24

24

24

24

Epoch [15/50], Iter [10/209] Loss: 2.4897, average\_loss: 2.7728

24

24

24

24

24

Epoch [15/50], Iter [15/209] Loss: 3.5342, average\_loss: 2.8072

24

24

24

24

24

Epoch [15/50], Iter [20/209] Loss: 2.5742, average\_loss: 2.8020

24

24

24

24

24

Epoch [15/50], Iter [25/209] Loss: 3.4395, average\_loss: 2.7829

24

24

24

24

24

Epoch [15/50], Iter [30/209] Loss: 3.9820, average\_loss: 2.8078

24

24

24

24

24

Epoch [15/50], Iter [35/209] Loss: 2.2474, average\_loss: 2.8742

24

24

24

24

24

Epoch [15/50], Iter [40/209] Loss: 3.0036, average\_loss: 2.8555

24

24

24

24

24

Epoch [15/50], Iter [45/209] Loss: 3.1921, average\_loss: 2.8498

24

24

24

24

24

Epoch [15/50], Iter [50/209] Loss: 2.8587, average\_loss: 2.8547

24

24

24

24

24

Epoch [15/50], Iter [55/209] Loss: 3.2955, average\_loss: 2.8662

24

24

24

24

24



Epoch [15/50], Iter [60/209] Loss: 2.8173, average\_loss: 2.8455  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [65/209] Loss: 2.6470, average\_loss: 2.8464  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [70/209] Loss: 3.0192, average\_loss: 2.8430  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [75/209] Loss: 3.8303, average\_loss: 2.8329  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [80/209] Loss: 3.5009, average\_loss: 2.8425  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [85/209] Loss: 2.9101, average\_loss: 2.8665  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [90/209] Loss: 2.3252, average\_loss: 2.8552  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [95/209] Loss: 2.9350, average\_loss: 2.8510  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [100/209] Loss: 2.4989, average\_loss: 2.8325  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [105/209] Loss: 2.5761, average\_loss: 2.8093  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [110/209] Loss: 2.7467, average\_loss: 2.8296  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [115/209] Loss: 2.4372, average\_loss: 2.8116  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [120/209] Loss: 2.4000, average\_loss: 2.7906  
24  
24  
24  
24

24  
Epoch [15/50], Iter [125/209] Loss: 3.3146, average\_loss: 2.7900  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [130/209] Loss: 2.2953, average\_loss: 2.7798  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [135/209] Loss: 3.0976, average\_loss: 2.7788  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [140/209] Loss: 2.7111, average\_loss: 2.7958  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [145/209] Loss: 2.3943, average\_loss: 2.7845  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [150/209] Loss: 2.6323, average\_loss: 2.7865  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [155/209] Loss: 2.2197, average\_loss: 2.7768  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [160/209] Loss: 3.4035, average\_loss: 2.7763  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [165/209] Loss: 2.5832, average\_loss: 2.7683  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [170/209] Loss: 2.3483, average\_loss: 2.7646  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [175/209] Loss: 2.3757, average\_loss: 2.7596  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [180/209] Loss: 2.8260, average\_loss: 2.7648  
24  
24  
24  
24  
24  
Epoch [15/50], Iter [185/209] Loss: 3.0466, average\_loss: 2.7654  
24  
24  
24

```
Epoch [15/50], Iter [190/209] Loss: 2.8157, average_loss: 2.7618
Epoch [15/50], Iter [195/209] Loss: 1.9871, average_loss: 2.7562
Epoch [15/50], Iter [200/209] Loss: 3.3896, average_loss: 2.7631
Epoch [15/50], Iter [205/209] Loss: 2.5385, average_loss: 2.7611
```

[illegible]

[illegible]

6

Updating best test loss: 3.14011

Starting epoch 16 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [16/50], Iter [5/209] Loss: 2.1496, average\_loss: 2.5106

24

24

24

24

24

Epoch [16/50], Iter [10/209] Loss: 2.5675, average\_loss: 2.4207

24

24

24

24

24

Epoch [16/50], Iter [15/209] Loss: 2.9820, average\_loss: 2.5497

24

24

24

24

24

Epoch [16/50], Iter [20/209] Loss: 2.7198, average\_loss: 2.5959

24

24

24

24

24

Epoch [16/50], Iter [25/209] Loss: 2.0677, average\_loss: 2.5798

24

24

24

24

24

Epoch [16/50], Iter [30/209] Loss: 2.5633, average\_loss: 2.6133

24

24

24

24

24

Epoch [16/50], Iter [35/209] Loss: 2.8211, average\_loss: 2.6115

24

24

24

24

24

Epoch [16/50], Iter [40/209] Loss: 2.3560, average\_loss: 2.5956

24

24

24

24

24

Epoch [16/50], Iter [45/209] Loss: 2.2610, average\_loss: 2.5849

24

24

24

24

24

Epoch [16/50], Iter [50/209] Loss: 2.3429, average\_loss: 2.5611

24

24

24

24

24

Epoch [16/50], Iter [55/209] Loss: 2.5617, average\_loss: 2.5402

24

24

24

24

24

Epoch [16/50], Iter [60/209] Loss: 2.5225, average\_loss: 2.5514  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [65/209] Loss: 3.2354, average\_loss: 2.6090  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [70/209] Loss: 2.7038, average\_loss: 2.6163  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [75/209] Loss: 2.2737, average\_loss: 2.6368  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [80/209] Loss: 2.5905, average\_loss: 2.6558  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [85/209] Loss: 2.8736, average\_loss: 2.6459  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [90/209] Loss: 3.0822, average\_loss: 2.6454  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [95/209] Loss: 2.5519, average\_loss: 2.6571  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [100/209] Loss: 2.9972, average\_loss: 2.6625  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [105/209] Loss: 3.3910, average\_loss: 2.6738  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [110/209] Loss: 3.1307, average\_loss: 2.6925  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [115/209] Loss: 2.1115, average\_loss: 2.6867  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [120/209] Loss: 2.8065, average\_loss: 2.6896  
24  
24  
24  
24

24  
Epoch [16/50], Iter [125/209] Loss: 1.9992, average\_loss: 2.6846  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [130/209] Loss: 2.1529, average\_loss: 2.6841  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [135/209] Loss: 2.4746, average\_loss: 2.6865  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [140/209] Loss: 2.3560, average\_loss: 2.6820  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [145/209] Loss: 1.9727, average\_loss: 2.6715  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [150/209] Loss: 2.7291, average\_loss: 2.6660  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [155/209] Loss: 2.6505, average\_loss: 2.6686  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [160/209] Loss: 3.4310, average\_loss: 2.6700  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [165/209] Loss: 1.8938, average\_loss: 2.6773  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [170/209] Loss: 2.9729, average\_loss: 2.6778  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [175/209] Loss: 3.0488, average\_loss: 2.6782  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [180/209] Loss: 2.5899, average\_loss: 2.6681  
24  
24  
24  
24  
24  
Epoch [16/50], Iter [185/209] Loss: 3.6033, average\_loss: 2.6737  
24  
24  
24



```
Epoch [16/50], Iter [190/209] Loss: 2.7285, average_loss: 2.6735
Epoch [16/50], Iter [195/209] Loss: 2.3597, average_loss: 2.6722
Epoch [16/50], Iter [200/209] Loss: 2.6167, average_loss: 2.6746
Epoch [16/50], Iter [205/209] Loss: 3.0473, average_loss: 2.6792
```

[illegible]

[illegible]

```
24
6
Updating best test loss: 3.06912

Starting epoch 17 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [17/50], Iter [5/209] Loss: 2.8426, average_loss: 2.6661
24
24
24
24
24
Epoch [17/50], Iter [10/209] Loss: 3.5262, average_loss: 2.6592
24
24
24
24
24
Epoch [17/50], Iter [15/209] Loss: 2.4620, average_loss: 2.7087
24
24
24
24
24
Epoch [17/50], Iter [20/209] Loss: 2.1063, average_loss: 2.6649
24
24
24
24
24
Epoch [17/50], Iter [25/209] Loss: 3.1336, average_loss: 2.6586
24
24
24
24
24
Epoch [17/50], Iter [30/209] Loss: 2.3560, average_loss: 2.6215
24
24
24
24
24
Epoch [17/50], Iter [35/209] Loss: 3.2795, average_loss: 2.6446
24
24
24
24
24
Epoch [17/50], Iter [40/209] Loss: 2.3249, average_loss: 2.6419
24
24
24
24
24
Epoch [17/50], Iter [45/209] Loss: 2.1984, average_loss: 2.6147
24
24
24
24
24
Epoch [17/50], Iter [50/209] Loss: 2.3112, average_loss: 2.5942
24
24
24
24
24
Epoch [17/50], Iter [55/209] Loss: 2.3580, average_loss: 2.5695
24
24
24
24
24
```

```
Epoch [17/50], Iter [60/209] Loss: 2.9662, average_loss: 2.5770
24
24
24
24
24
Epoch [17/50], Iter [65/209] Loss: 2.1046, average_loss: 2.5523
24
24
24
24
24
Epoch [17/50], Iter [70/209] Loss: 2.2219, average_loss: 2.5662
24
24
24
24
24
Epoch [17/50], Iter [75/209] Loss: 1.9292, average_loss: 2.5719
24
24
24
24
24
Epoch [17/50], Iter [80/209] Loss: 2.4317, average_loss: 2.5832
24
24
24
24
24
Epoch [17/50], Iter [85/209] Loss: 2.5795, average_loss: 2.5727
24
24
24
24
24
Epoch [17/50], Iter [90/209] Loss: 2.3675, average_loss: 2.5590
24
24
24
24
24
Epoch [17/50], Iter [95/209] Loss: 2.1360, average_loss: 2.5647
24
24
24
24
24
Epoch [17/50], Iter [100/209] Loss: 3.2389, average_loss: 2.5725
24
24
24
24
24
Epoch [17/50], Iter [105/209] Loss: 2.3593, average_loss: 2.5640
24
24
24
24
24
Epoch [17/50], Iter [110/209] Loss: 2.3641, average_loss: 2.5667
24
24
24
24
24
Epoch [17/50], Iter [115/209] Loss: 4.2280, average_loss: 2.5853
24
24
24
24
24
Epoch [17/50], Iter [120/209] Loss: 3.5214, average_loss: 2.5980
24
24
24
24
```

```
24
24
Epoch [17/50], Iter [125/209] Loss: 3.8621, average_loss: 2.6192
24
24
24
24
24
Epoch [17/50], Iter [130/209] Loss: 2.7366, average_loss: 2.6135
24
24
24
24
24
Epoch [17/50], Iter [135/209] Loss: 1.8694, average_loss: 2.6125
24
24
24
24
24
Epoch [17/50], Iter [140/209] Loss: 2.1753, average_loss: 2.6092
24
24
24
24
24
Epoch [17/50], Iter [145/209] Loss: 2.2025, average_loss: 2.6105
24
24
24
24
24
Epoch [17/50], Iter [150/209] Loss: 2.8031, average_loss: 2.6088
24
24
24
24
24
Epoch [17/50], Iter [155/209] Loss: 2.9590, average_loss: 2.6110
24
24
24
24
24
Epoch [17/50], Iter [160/209] Loss: 2.2043, average_loss: 2.5971
24
24
24
24
24
Epoch [17/50], Iter [165/209] Loss: 2.6360, average_loss: 2.6025
24
24
24
24
24
Epoch [17/50], Iter [170/209] Loss: 2.3742, average_loss: 2.6052
24
24
24
24
24
Epoch [17/50], Iter [175/209] Loss: 2.4985, average_loss: 2.6004
24
24
24
24
24
Epoch [17/50], Iter [180/209] Loss: 2.9127, average_loss: 2.6033
24
24
24
24
24
Epoch [17/50], Iter [185/209] Loss: 2.3320, average_loss: 2.5992
24
24
24
```

```
Epoch [17/50], Iter [190/209] Loss: 2.4082, average_loss: 2.6066
Epoch [17/50], Iter [195/209] Loss: 2.4001, average_loss: 2.6027
Epoch [17/50], Iter [200/209] Loss: 2.6865, average_loss: 2.6046
Epoch [17/50], Iter [205/209] Loss: 2.0106, average_loss: 2.5947
```

[illegible]





24  
6

Starting epoch 18 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [18/50], Iter [5/209] Loss: 2.9801, average\_loss: 3.0287

24

24

24

24

24

Epoch [18/50], Iter [10/209] Loss: 3.1698, average\_loss: 2.6930

24

24

24

24

24

Epoch [18/50], Iter [15/209] Loss: 2.1409, average\_loss: 2.6325

24

24

24

24

24

Epoch [18/50], Iter [20/209] Loss: 2.7201, average\_loss: 2.5742

24

24

24

24

24

Epoch [18/50], Iter [25/209] Loss: 2.5145, average\_loss: 2.5950

24

24

24

24

24

Epoch [18/50], Iter [30/209] Loss: 1.9672, average\_loss: 2.5824

24

24

24

24

24

Epoch [18/50], Iter [35/209] Loss: 2.8715, average\_loss: 2.5734

24

24

24

24

24

Epoch [18/50], Iter [40/209] Loss: 2.6384, average\_loss: 2.5692

24

24

24

24

24

Epoch [18/50], Iter [45/209] Loss: 3.0711, average\_loss: 2.5958

24

24

24

24

24

Epoch [18/50], Iter [50/209] Loss: 3.0607, average\_loss: 2.5710

24

24

24

24

24

Epoch [18/50], Iter [55/209] Loss: 2.4073, average\_loss: 2.5848

24

24

24

24

24

Epoch [18/50], Iter [60/209] Loss: 2.6736, average\_loss: 2.5824  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [65/209] Loss: 2.9060, average\_loss: 2.5934  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [70/209] Loss: 3.1550, average\_loss: 2.6018  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [75/209] Loss: 2.2949, average\_loss: 2.5882  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [80/209] Loss: 2.2264, average\_loss: 2.5973  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [85/209] Loss: 3.0194, average\_loss: 2.6132  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [90/209] Loss: 2.6249, average\_loss: 2.6095  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [95/209] Loss: 2.1267, average\_loss: 2.5928  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [100/209] Loss: 2.2503, average\_loss: 2.5861  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [105/209] Loss: 1.9222, average\_loss: 2.5762  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [110/209] Loss: 2.1208, average\_loss: 2.5872  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [115/209] Loss: 3.0774, average\_loss: 2.5885  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [120/209] Loss: 2.4493, average\_loss: 2.5806  
24  
24  
24  
24

24  
Epoch [18/50], Iter [125/209] Loss: 2.6164, average\_loss: 2.5872  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [130/209] Loss: 2.7415, average\_loss: 2.5934  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [135/209] Loss: 2.7320, average\_loss: 2.6111  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [140/209] Loss: 2.6161, average\_loss: 2.6044  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [145/209] Loss: 2.8269, average\_loss: 2.6053  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [150/209] Loss: 2.6597, average\_loss: 2.6007  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [155/209] Loss: 2.7704, average\_loss: 2.5894  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [160/209] Loss: 3.3055, average\_loss: 2.5888  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [165/209] Loss: 2.4308, average\_loss: 2.5805  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [170/209] Loss: 2.2562, average\_loss: 2.5742  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [175/209] Loss: 2.6376, average\_loss: 2.5618  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [180/209] Loss: 2.9713, average\_loss: 2.5658  
24  
24  
24  
24  
24  
Epoch [18/50], Iter [185/209] Loss: 2.7641, average\_loss: 2.5648  
24  
24  
24

```
Epoch [18/50], Iter [190/209] Loss: 2.7287, average_loss: 2.5650  
Epoch [18/50], Iter [195/209] Loss: 2.6376, average_loss: 2.5684  
Epoch [18/50], Iter [200/209] Loss: 2.7749, average_loss: 2.5716  
Epoch [18/50], Iter [205/209] Loss: 2.4262, average_loss: 2.5747
```

[illegible]

[illegible]

6

Updating best test loss: 3.05680

Starting epoch 19 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [19/50], Iter [5/209] Loss: 2.1771, average\_loss: 2.8049

24

24

24

24

24

Epoch [19/50], Iter [10/209] Loss: 3.3733, average\_loss: 2.8585

24

24

24

24

24

Epoch [19/50], Iter [15/209] Loss: 2.3654, average\_loss: 2.7215

24

24

24

24

24

Epoch [19/50], Iter [20/209] Loss: 2.9346, average\_loss: 2.7044

24

24

24

24

24

Epoch [19/50], Iter [25/209] Loss: 2.4315, average\_loss: 2.6723

24

24

24

24

24

Epoch [19/50], Iter [30/209] Loss: 3.0566, average\_loss: 2.6970

24

24

24

24

24

Epoch [19/50], Iter [35/209] Loss: 3.2618, average\_loss: 2.7148

24

24

24

24

24

Epoch [19/50], Iter [40/209] Loss: 2.8989, average\_loss: 2.7058

24

24

24

24

24

Epoch [19/50], Iter [45/209] Loss: 2.8335, average\_loss: 2.6859

24

24

24

24

24

Epoch [19/50], Iter [50/209] Loss: 2.0599, average\_loss: 2.6585

24

24

24

24

24

Epoch [19/50], Iter [55/209] Loss: 2.2983, average\_loss: 2.6387

24

24

24

24

24



Epoch [19/50], Iter [60/209] Loss: 1.8521, average\_loss: 2.6196  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [65/209] Loss: 2.5610, average\_loss: 2.6155  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [70/209] Loss: 2.7096, average\_loss: 2.6181  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [75/209] Loss: 2.0943, average\_loss: 2.5951  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [80/209] Loss: 2.3387, average\_loss: 2.5782  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [85/209] Loss: 2.4168, average\_loss: 2.5690  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [90/209] Loss: 2.9695, average\_loss: 2.5943  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [95/209] Loss: 2.1761, average\_loss: 2.5909  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [100/209] Loss: 1.8339, average\_loss: 2.5797  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [105/209] Loss: 2.2309, average\_loss: 2.5756  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [110/209] Loss: 2.3131, average\_loss: 2.5727  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [115/209] Loss: 3.0303, average\_loss: 2.5570  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [120/209] Loss: 3.1187, average\_loss: 2.5695  
24  
24  
24  
24

24  
Epoch [19/50], Iter [125/209] Loss: 2.2503, average\_loss: 2.5648  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [130/209] Loss: 2.7368, average\_loss: 2.5442  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [135/209] Loss: 2.2239, average\_loss: 2.5349  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [140/209] Loss: 2.3752, average\_loss: 2.5321  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [145/209] Loss: 2.5104, average\_loss: 2.5326  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [150/209] Loss: 2.8020, average\_loss: 2.5376  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [155/209] Loss: 3.0921, average\_loss: 2.5457  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [160/209] Loss: 2.3105, average\_loss: 2.5441  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [165/209] Loss: 3.2367, average\_loss: 2.5427  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [170/209] Loss: 3.0473, average\_loss: 2.5518  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [175/209] Loss: 3.2283, average\_loss: 2.5498  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [180/209] Loss: 2.3032, average\_loss: 2.5440  
24  
24  
24  
24  
24  
Epoch [19/50], Iter [185/209] Loss: 2.5562, average\_loss: 2.5449  
24  
24  
24

```

Epoch [19/50], Iter [190/209] Loss: 2.6324, average_loss: 2.5404
Epoch [19/50], Iter [195/209] Loss: 3.0833, average_loss: 2.5412
Epoch [19/50], Iter [200/209] Loss: 2.6208, average_loss: 2.5382
Epoch [19/50], Iter [205/209] Loss: 2.7984, average_loss: 2.5371

```

[illegible]

[illegible]

```
24
6
Updating best test loss: 3.02279

Starting epoch 20 / 50
Learning Rate for this epoch: 0.001
24
24
24
24
24
Epoch [20/50], Iter [5/209] Loss: 2.1882, average_loss: 2.1483
24
24
24
24
24
Epoch [20/50], Iter [10/209] Loss: 3.0167, average_loss: 2.4098
24
24
24
24
24
Epoch [20/50], Iter [15/209] Loss: 2.7063, average_loss: 2.4610
24
24
24
24
24
Epoch [20/50], Iter [20/209] Loss: 2.6710, average_loss: 2.4737
24
24
24
24
24
Epoch [20/50], Iter [25/209] Loss: 2.0672, average_loss: 2.5230
24
24
24
24
24
Epoch [20/50], Iter [30/209] Loss: 2.5752, average_loss: 2.4971
24
24
24
24
24
Epoch [20/50], Iter [35/209] Loss: 2.4039, average_loss: 2.4213
24
24
24
24
24
Epoch [20/50], Iter [40/209] Loss: 3.1024, average_loss: 2.4858
24
24
24
24
24
Epoch [20/50], Iter [45/209] Loss: 2.8506, average_loss: 2.4805
24
24
24
24
24
Epoch [20/50], Iter [50/209] Loss: 2.3322, average_loss: 2.4781
24
24
24
24
24
Epoch [20/50], Iter [55/209] Loss: 3.3283, average_loss: 2.4994
24
24
24
24
24
```

```
24
Epoch [20/50], Iter [60/209] Loss: 2.0694, average_loss: 2.4963
24
24
24
24
24
Epoch [20/50], Iter [65/209] Loss: 2.1678, average_loss: 2.4838
24
24
24
24
24
Epoch [20/50], Iter [70/209] Loss: 2.6656, average_loss: 2.4978
24
24
24
24
24
Epoch [20/50], Iter [75/209] Loss: 2.1627, average_loss: 2.4771
24
24
24
24
24
Epoch [20/50], Iter [80/209] Loss: 2.5958, average_loss: 2.4846
24
24
24
24
24
Epoch [20/50], Iter [85/209] Loss: 2.3845, average_loss: 2.4901
24
24
24
24
24
Epoch [20/50], Iter [90/209] Loss: 2.4488, average_loss: 2.5055
24
24
24
24
24
Epoch [20/50], Iter [95/209] Loss: 2.9431, average_loss: 2.5104
24
24
24
24
24
Epoch [20/50], Iter [100/209] Loss: 1.7494, average_loss: 2.4944
24
24
24
24
24
Epoch [20/50], Iter [105/209] Loss: 2.1463, average_loss: 2.4853
24
24
24
24
24
Epoch [20/50], Iter [110/209] Loss: 1.9892, average_loss: 2.4924
24
24
24
24
24
Epoch [20/50], Iter [115/209] Loss: 2.3530, average_loss: 2.4889
24
24
24
24
24
Epoch [20/50], Iter [120/209] Loss: 2.4796, average_loss: 2.4900
24
24
24
24
```

```
24
24
Epoch [20/50], Iter [125/209] Loss: 3.7892, average_loss: 2.5075
24
24
24
24
24
Epoch [20/50], Iter [130/209] Loss: 2.4251, average_loss: 2.4974
24
24
24
24
24
Epoch [20/50], Iter [135/209] Loss: 3.0786, average_loss: 2.4987
24
24
24
24
24
Epoch [20/50], Iter [140/209] Loss: 2.3206, average_loss: 2.5032
24
24
24
24
24
Epoch [20/50], Iter [145/209] Loss: 2.1637, average_loss: 2.4961
24
24
24
24
24
Epoch [20/50], Iter [150/209] Loss: 2.8992, average_loss: 2.4948
24
24
24
24
24
Epoch [20/50], Iter [155/209] Loss: 1.7479, average_loss: 2.4896
24
24
24
24
24
Epoch [20/50], Iter [160/209] Loss: 2.4077, average_loss: 2.4956
24
24
24
24
24
Epoch [20/50], Iter [165/209] Loss: 3.6955, average_loss: 2.5015
24
24
24
24
24
Epoch [20/50], Iter [170/209] Loss: 2.6222, average_loss: 2.5015
24
24
24
24
24
Epoch [20/50], Iter [175/209] Loss: 2.1500, average_loss: 2.4955
24
24
24
24
24
Epoch [20/50], Iter [180/209] Loss: 2.8427, average_loss: 2.4948
24
24
24
24
24
Epoch [20/50], Iter [185/209] Loss: 2.2853, average_loss: 2.4902
24
24
24
```



```
Epoch [20/50], Iter [190/209] Loss: 2.8109, average_loss: 2.4827
Epoch [20/50], Iter [195/209] Loss: 2.0151, average_loss: 2.4866
Epoch [20/50], Iter [200/209] Loss: 2.1684, average_loss: 2.4784
Epoch [20/50], Iter [205/209] Loss: 2.8138, average_loss: 2.4833
```



[illegible]

24  
6

Starting epoch 21 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24

Epoch [21/50], Iter [5/209] Loss: 2.3627, average\_loss: 2.4269

24  
24  
24  
24  
24

Epoch [21/50], Iter [10/209] Loss: 3.0004, average\_loss: 2.3443

24  
24  
24  
24  
24

Epoch [21/50], Iter [15/209] Loss: 2.5714, average\_loss: 2.3866

24  
24  
24  
24  
24

Epoch [21/50], Iter [20/209] Loss: 2.3082, average\_loss: 2.4369

24  
24  
24  
24  
24

Epoch [21/50], Iter [25/209] Loss: 1.9723, average\_loss: 2.3708

24  
24  
24  
24  
24

Epoch [21/50], Iter [30/209] Loss: 1.7040, average\_loss: 2.3958

24  
24  
24  
24  
24

Epoch [21/50], Iter [35/209] Loss: 2.5630, average\_loss: 2.3847

24  
24  
24  
24  
24

Epoch [21/50], Iter [40/209] Loss: 1.8088, average\_loss: 2.3812

24  
24  
24  
24  
24

Epoch [21/50], Iter [45/209] Loss: 2.1924, average\_loss: 2.3941

24  
24  
24  
24  
24

Epoch [21/50], Iter [50/209] Loss: 3.2575, average\_loss: 2.3952

24  
24  
24  
24  
24

Epoch [21/50], Iter [55/209] Loss: 1.8530, average\_loss: 2.3837

24  
24  
24  
24  
24

Epoch [21/50], Iter [60/209] Loss: 2.1886, average\_loss: 2.3839  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [65/209] Loss: 1.4840, average\_loss: 2.3677  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [70/209] Loss: 2.6069, average\_loss: 2.3542  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [75/209] Loss: 3.4185, average\_loss: 2.3671  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [80/209] Loss: 2.1427, average\_loss: 2.3562  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [85/209] Loss: 3.0071, average\_loss: 2.3654  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [90/209] Loss: 2.0836, average\_loss: 2.3808  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [95/209] Loss: 2.3929, average\_loss: 2.3839  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [100/209] Loss: 1.9445, average\_loss: 2.3933  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [105/209] Loss: 1.9625, average\_loss: 2.3878  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [110/209] Loss: 3.0361, average\_loss: 2.3828  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [115/209] Loss: 2.5540, average\_loss: 2.3991  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [120/209] Loss: 1.9278, average\_loss: 2.3964  
24  
24  
24  
24

24  
Epoch [21/50], Iter [125/209] Loss: 2.6745, average\_loss: 2.3912  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [130/209] Loss: 1.8800, average\_loss: 2.3859  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [135/209] Loss: 2.3942, average\_loss: 2.3858  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [140/209] Loss: 1.7484, average\_loss: 2.3822  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [145/209] Loss: 2.5586, average\_loss: 2.3932  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [150/209] Loss: 2.7479, average\_loss: 2.3996  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [155/209] Loss: 2.2541, average\_loss: 2.4005  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [160/209] Loss: 2.0118, average\_loss: 2.4016  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [165/209] Loss: 2.4794, average\_loss: 2.4063  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [170/209] Loss: 2.6144, average\_loss: 2.4096  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [175/209] Loss: 2.7425, average\_loss: 2.4122  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [180/209] Loss: 2.7214, average\_loss: 2.4123  
24  
24  
24  
24  
24  
Epoch [21/50], Iter [185/209] Loss: 2.5867, average\_loss: 2.4183  
24  
24  
24

```
Epoch [21/50], Iter [190/209] Loss: 2.8460, average_loss: 2.4286
Epoch [21/50], Iter [195/209] Loss: 1.9574, average_loss: 2.4321
Epoch [21/50], Iter [200/209] Loss: 2.0446, average_loss: 2.4332
Epoch [21/50], Iter [205/209] Loss: 2.5157, average_loss: 2.4327
```

[illegible]



[illegible]

6

Updating best test loss: 3.00548

Starting epoch 22 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [22/50], Iter [5/209] Loss: 2.2508, average\_loss: 2.2923

24

24

24

24

24

Epoch [22/50], Iter [10/209] Loss: 3.2409, average\_loss: 2.3930

24

24

24

24

24

Epoch [22/50], Iter [15/209] Loss: 2.1000, average\_loss: 2.3974

24

24

24

24

24

Epoch [22/50], Iter [20/209] Loss: 2.0930, average\_loss: 2.4321

24

24

24

24

24

Epoch [22/50], Iter [25/209] Loss: 2.1707, average\_loss: 2.4381

24

24

24

24

24

Epoch [22/50], Iter [30/209] Loss: 2.9067, average\_loss: 2.4440

24

24

24

24

24

Epoch [22/50], Iter [35/209] Loss: 2.4769, average\_loss: 2.4109

24

24

24

24

24

Epoch [22/50], Iter [40/209] Loss: 2.7223, average\_loss: 2.3889

24

24

24

24

24

Epoch [22/50], Iter [45/209] Loss: 2.5682, average\_loss: 2.3905

24

24

24

24

24

Epoch [22/50], Iter [50/209] Loss: 2.8850, average\_loss: 2.3973

24

24

24

24

24

Epoch [22/50], Iter [55/209] Loss: 2.1353, average\_loss: 2.4156

24

24

24

24

24

Epoch [22/50], Iter [60/209] Loss: 1.5809, average\_loss: 2.4243  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [65/209] Loss: 2.6713, average\_loss: 2.4394  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [70/209] Loss: 2.3559, average\_loss: 2.4382  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [75/209] Loss: 2.3582, average\_loss: 2.4268  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [80/209] Loss: 2.4545, average\_loss: 2.4226  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [85/209] Loss: 1.9805, average\_loss: 2.4101  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [90/209] Loss: 2.8443, average\_loss: 2.4022  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [95/209] Loss: 2.2842, average\_loss: 2.3870  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [100/209] Loss: 1.9790, average\_loss: 2.3851  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [105/209] Loss: 1.4529, average\_loss: 2.3847  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [110/209] Loss: 3.0859, average\_loss: 2.3830  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [115/209] Loss: 2.0601, average\_loss: 2.3832  
24  
24  
24  
24  
24  
Epoch [22/50], Iter [120/209] Loss: 2.8305, average\_loss: 2.3918  
24  
24  
24  
24

```
24
Epoch [22/50], Iter [125/209] Loss: 2.4937, average_loss: 2.3847
24
24
24
24
Epoch [22/50], Iter [130/209] Loss: 2.9255, average_loss: 2.3901
24
24
24
24
Epoch [22/50], Iter [135/209] Loss: 3.9095, average_loss: 2.4144
24
24
24
24
Epoch [22/50], Iter [140/209] Loss: 2.4953, average_loss: 2.4023
24
24
24
24
Epoch [22/50], Iter [145/209] Loss: 1.8762, average_loss: 2.3952
24
24
24
24
Epoch [22/50], Iter [150/209] Loss: 2.2089, average_loss: 2.3840
24
24
24
24
Epoch [22/50], Iter [155/209] Loss: 3.3121, average_loss: 2.3961
24
24
24
24
Epoch [22/50], Iter [160/209] Loss: 1.9118, average_loss: 2.4010
24
24
24
24
Epoch [22/50], Iter [165/209] Loss: 3.0484, average_loss: 2.4038
24
24
24
24
Epoch [22/50], Iter [170/209] Loss: 2.8033, average_loss: 2.4036
24
24
24
24
Epoch [22/50], Iter [175/209] Loss: 2.1330, average_loss: 2.3995
24
24
24
24
Epoch [22/50], Iter [180/209] Loss: 1.9614, average_loss: 2.3994
24
24
24
24
Epoch [22/50], Iter [185/209] Loss: 2.3118, average_loss: 2.3996
24
24
24
```

```
Epoch [22/50], Iter [190/209] Loss: 1.8703, average_loss: 2.3943
Epoch [22/50], Iter [195/209] Loss: 2.6810, average_loss: 2.3983
Epoch [22/50], Iter [200/209] Loss: 2.0526, average_loss: 2.3979
Epoch [22/50], Iter [205/209] Loss: 2.1084, average_loss: 2.3971
```

[illegible]

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24  
6  
Updating best test loss: 2.95721

Starting epoch 23 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24  
Epoch [23/50], Iter [5/209] Loss: 2.8718, average\_loss: 2.9755  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [10/209] Loss: 2.0804, average\_loss: 2.5276  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [15/209] Loss: 2.1259, average\_loss: 2.4550  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [20/209] Loss: 2.5664, average\_loss: 2.4477  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [25/209] Loss: 2.6868, average\_loss: 2.5288  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [30/209] Loss: 2.6597, average\_loss: 2.5125  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [35/209] Loss: 2.5079, average\_loss: 2.4925  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [40/209] Loss: 2.3500, average\_loss: 2.4803  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [45/209] Loss: 1.9995, average\_loss: 2.4696  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [50/209] Loss: 2.2311, average\_loss: 2.4593  
24  
24  
24  
24  
24  
Epoch [23/50], Iter [55/209] Loss: 3.4149, average\_loss: 2.4531  
24  
24  
24  
24  
24



```
24
Epoch [23/50], Iter [60/209] Loss: 2.5754, average_loss: 2.4272
24
24
24
24
24
Epoch [23/50], Iter [65/209] Loss: 2.3880, average_loss: 2.4046
24
24
24
24
24
Epoch [23/50], Iter [70/209] Loss: 2.0739, average_loss: 2.3990
24
24
24
24
24
Epoch [23/50], Iter [75/209] Loss: 1.7055, average_loss: 2.3780
24
24
24
24
24
Epoch [23/50], Iter [80/209] Loss: 2.1850, average_loss: 2.3781
24
24
24
24
24
Epoch [23/50], Iter [85/209] Loss: 2.3545, average_loss: 2.3721
24
24
24
24
24
Epoch [23/50], Iter [90/209] Loss: 2.3987, average_loss: 2.3667
24
24
24
24
24
Epoch [23/50], Iter [95/209] Loss: 2.7672, average_loss: 2.3516
24
24
24
24
24
Epoch [23/50], Iter [100/209] Loss: 2.0639, average_loss: 2.3506
24
24
24
24
24
Epoch [23/50], Iter [105/209] Loss: 2.9222, average_loss: 2.3777
24
24
24
24
24
Epoch [23/50], Iter [110/209] Loss: 2.3143, average_loss: 2.3768
24
24
24
24
24
Epoch [23/50], Iter [115/209] Loss: 1.8270, average_loss: 2.3653
24
24
24
24
24
Epoch [23/50], Iter [120/209] Loss: 1.9252, average_loss: 2.3564
24
24
24
24
```

```
24
24
Epoch [23/50], Iter [125/209] Loss: 2.6078, average_loss: 2.3617
24
24
24
24
24
Epoch [23/50], Iter [130/209] Loss: 2.4124, average_loss: 2.3542
24
24
24
24
24
Epoch [23/50], Iter [135/209] Loss: 3.2192, average_loss: 2.3546
24
24
24
24
24
Epoch [23/50], Iter [140/209] Loss: 2.2447, average_loss: 2.3469
24
24
24
24
24
Epoch [23/50], Iter [145/209] Loss: 1.7868, average_loss: 2.3368
24
24
24
24
24
Epoch [23/50], Iter [150/209] Loss: 2.2425, average_loss: 2.3399
24
24
24
24
24
Epoch [23/50], Iter [155/209] Loss: 2.3875, average_loss: 2.3450
24
24
24
24
24
Epoch [23/50], Iter [160/209] Loss: 2.7321, average_loss: 2.3428
24
24
24
24
24
Epoch [23/50], Iter [165/209] Loss: 2.2549, average_loss: 2.3537
24
24
24
24
24
Epoch [23/50], Iter [170/209] Loss: 1.8545, average_loss: 2.3527
24
24
24
24
24
Epoch [23/50], Iter [175/209] Loss: 1.7019, average_loss: 2.3519
24
24
24
24
24
Epoch [23/50], Iter [180/209] Loss: 2.7376, average_loss: 2.3551
24
24
24
24
24
Epoch [23/50], Iter [185/209] Loss: 1.5482, average_loss: 2.3545
24
24
24
```

```
Epoch [23/50], Iter [190/209] Loss: 2.2568, average_loss: 2.3552
Epoch [23/50], Iter [195/209] Loss: 2.5998, average_loss: 2.3526
Epoch [23/50], Iter [200/209] Loss: 1.9014, average_loss: 2.3588
Epoch [23/50], Iter [205/209] Loss: 2.3671, average_loss: 2.3596
```

[illegible]

[illegible]

24  
6  
Updating best test loss: 2.92904

Starting epoch 24 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24  
Epoch [24/50], Iter [5/209] Loss: 2.5978, average\_loss: 2.3074  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [10/209] Loss: 3.8046, average\_loss: 2.4705  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [15/209] Loss: 2.8075, average\_loss: 2.4805  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [20/209] Loss: 3.2408, average\_loss: 2.4732  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [25/209] Loss: 1.8770, average\_loss: 2.3814  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [30/209] Loss: 1.4208, average\_loss: 2.3653  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [35/209] Loss: 2.0250, average\_loss: 2.3435  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [40/209] Loss: 2.5475, average\_loss: 2.3266  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [45/209] Loss: 2.0284, average\_loss: 2.3035  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [50/209] Loss: 2.4298, average\_loss: 2.3185  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [55/209] Loss: 2.5169, average\_loss: 2.3225  
24  
24  
24  
24

24  
Epoch [24/50], Iter [60/209] Loss: 1.9414, average\_loss: 2.3149  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [65/209] Loss: 2.2481, average\_loss: 2.3224  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [70/209] Loss: 2.3577, average\_loss: 2.3227  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [75/209] Loss: 2.8350, average\_loss: 2.3260  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [80/209] Loss: 2.4477, average\_loss: 2.3408  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [85/209] Loss: 2.9650, average\_loss: 2.3591  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [90/209] Loss: 1.7577, average\_loss: 2.3525  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [95/209] Loss: 2.5475, average\_loss: 2.3508  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [100/209] Loss: 2.6381, average\_loss: 2.3542  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [105/209] Loss: 2.3041, average\_loss: 2.3522  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [110/209] Loss: 2.1209, average\_loss: 2.3618  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [115/209] Loss: 2.2672, average\_loss: 2.3539  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [120/209] Loss: 2.1863, average\_loss: 2.3443  
24  
24  
24

24  
24  
Epoch [24/50], Iter [125/209] Loss: 2.1687, average\_loss: 2.3414  
24  
24  
24  
24  
Epoch [24/50], Iter [130/209] Loss: 2.6821, average\_loss: 2.3404  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [135/209] Loss: 2.2932, average\_loss: 2.3347  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [140/209] Loss: 2.0593, average\_loss: 2.3331  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [145/209] Loss: 1.8666, average\_loss: 2.3254  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [150/209] Loss: 1.9252, average\_loss: 2.3269  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [155/209] Loss: 3.0765, average\_loss: 2.3310  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [160/209] Loss: 2.3117, average\_loss: 2.3207  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [165/209] Loss: 3.5323, average\_loss: 2.3235  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [170/209] Loss: 2.3582, average\_loss: 2.3303  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [175/209] Loss: 2.0752, average\_loss: 2.3223  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [180/209] Loss: 2.4054, average\_loss: 2.3172  
24  
24  
24  
24  
24  
Epoch [24/50], Iter [185/209] Loss: 2.1548, average\_loss: 2.3131  
24  
24



```

Epoch [24/50], Iter [190/209] Loss: 2.6889, average_loss: 2.3134
Epoch [24/50], Iter [195/209] Loss: 2.1810, average_loss: 2.3079
Epoch [24/50], Iter [200/209] Loss: 1.6823, average_loss: 2.2998
Epoch [24/50], Iter [205/209] Loss: 1.9013, average_loss: 2.2931

```

[illegible]

[illegible]

24

6

Updating best test loss: 2.87434

Starting epoch 25 / 50

Learning Rate for this epoch: 0.001

24

24

24

24

24

Epoch [25/50], Iter [5/209] Loss: 2.0814, average\_loss: 2.1750

24

24

24

24

24

Epoch [25/50], Iter [10/209] Loss: 2.3921, average\_loss: 2.2812

24

24

24

24

24

Epoch [25/50], Iter [15/209] Loss: 1.5302, average\_loss: 2.2801

24

24

24

24

24

Epoch [25/50], Iter [20/209] Loss: 2.3642, average\_loss: 2.2880

24

24

24

24

24

Epoch [25/50], Iter [25/209] Loss: 2.7666, average\_loss: 2.2773

24

24

24

24

24

Epoch [25/50], Iter [30/209] Loss: 2.2220, average\_loss: 2.3162

24

24

24

24

24

Epoch [25/50], Iter [35/209] Loss: 3.1320, average\_loss: 2.3123

24

24

24

24

24

Epoch [25/50], Iter [40/209] Loss: 1.7350, average\_loss: 2.2870

24

24

24

24

24

Epoch [25/50], Iter [45/209] Loss: 2.0164, average\_loss: 2.2596

24

24

24

24

24

Epoch [25/50], Iter [50/209] Loss: 1.7191, average\_loss: 2.2589

24

24

24

24

24

Epoch [25/50], Iter [55/209] Loss: 2.1130, average\_loss: 2.2731

24

24

24

24

```
24
Epoch [25/50], Iter [60/209] Loss: 2.2221, average_loss: 2.2631
24
24
24
24
24
Epoch [25/50], Iter [65/209] Loss: 2.2908, average_loss: 2.2556
24
24
24
24
24
Epoch [25/50], Iter [70/209] Loss: 2.4293, average_loss: 2.2577
24
24
24
24
24
Epoch [25/50], Iter [75/209] Loss: 2.1422, average_loss: 2.2605
24
24
24
24
24
Epoch [25/50], Iter [80/209] Loss: 2.8297, average_loss: 2.2627
24
24
24
24
24
Epoch [25/50], Iter [85/209] Loss: 2.3124, average_loss: 2.2604
24
24
24
24
24
Epoch [25/50], Iter [90/209] Loss: 2.7856, average_loss: 2.2706
24
24
24
24
24
Epoch [25/50], Iter [95/209] Loss: 1.9144, average_loss: 2.2739
24
24
24
24
24
Epoch [25/50], Iter [100/209] Loss: 1.4453, average_loss: 2.2633
24
24
24
24
24
Epoch [25/50], Iter [105/209] Loss: 1.9874, average_loss: 2.2515
24
24
24
24
24
Epoch [25/50], Iter [110/209] Loss: 1.6177, average_loss: 2.2491
24
24
24
24
24
Epoch [25/50], Iter [115/209] Loss: 2.9010, average_loss: 2.2491
24
24
24
24
24
Epoch [25/50], Iter [120/209] Loss: 1.8237, average_loss: 2.2385
24
24
24
```

```
24
24
24
Epoch [25/50], Iter [125/209] Loss: 1.8432, average_loss: 2.2377
24
24
24
24
24
Epoch [25/50], Iter [130/209] Loss: 2.0268, average_loss: 2.2569
24
24
24
24
24
Epoch [25/50], Iter [135/209] Loss: 1.6876, average_loss: 2.2541
24
24
24
24
24
Epoch [25/50], Iter [140/209] Loss: 1.7543, average_loss: 2.2531
24
24
24
24
24
Epoch [25/50], Iter [145/209] Loss: 2.7117, average_loss: 2.2548
24
24
24
24
24
Epoch [25/50], Iter [150/209] Loss: 2.3255, average_loss: 2.2580
24
24
24
24
24
Epoch [25/50], Iter [155/209] Loss: 2.9668, average_loss: 2.2704
24
24
24
24
24
Epoch [25/50], Iter [160/209] Loss: 1.9365, average_loss: 2.2680
24
24
24
24
24
Epoch [25/50], Iter [165/209] Loss: 2.0631, average_loss: 2.2579
24
24
24
24
24
Epoch [25/50], Iter [170/209] Loss: 2.3512, average_loss: 2.2568
24
24
24
24
24
Epoch [25/50], Iter [175/209] Loss: 2.6193, average_loss: 2.2629
24
24
24
24
24
Epoch [25/50], Iter [180/209] Loss: 2.0949, average_loss: 2.2555
24
24
24
24
24
Epoch [25/50], Iter [185/209] Loss: 2.3875, average_loss: 2.2625
24
24
```

```

Epoch [25/50], Iter [190/209] Loss: 1.8627, average_loss: 2.2567
Epoch [25/50], Iter [195/209] Loss: 2.0913, average_loss: 2.2508
Epoch [25/50], Iter [200/209] Loss: 2.5472, average_loss: 2.2424
Epoch [25/50], Iter [205/209] Loss: 2.2242, average_loss: 2.2406

```

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24





24  
24  
6

Starting epoch 26 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24

Epoch [26/50], Iter [5/209] Loss: 2.6401, average\_loss: 2.3217

24  
24  
24  
24  
24

Epoch [26/50], Iter [10/209] Loss: 2.0020, average\_loss: 2.3276

24  
24  
24  
24  
24

Epoch [26/50], Iter [15/209] Loss: 2.5409, average\_loss: 2.3388

24  
24  
24  
24  
24

Epoch [26/50], Iter [20/209] Loss: 2.6615, average\_loss: 2.2878

24  
24  
24  
24  
24

Epoch [26/50], Iter [25/209] Loss: 2.8319, average\_loss: 2.2608

24  
24  
24  
24  
24

Epoch [26/50], Iter [30/209] Loss: 2.0406, average\_loss: 2.2305

24  
24  
24  
24  
24

Epoch [26/50], Iter [35/209] Loss: 1.5876, average\_loss: 2.2385

24  
24  
24  
24  
24

Epoch [26/50], Iter [40/209] Loss: 2.8107, average\_loss: 2.2829

24  
24  
24  
24  
24

Epoch [26/50], Iter [45/209] Loss: 1.8349, average\_loss: 2.2652

24  
24  
24  
24  
24

Epoch [26/50], Iter [50/209] Loss: 2.0174, average\_loss: 2.2584

24  
24  
24  
24  
24

Epoch [26/50], Iter [55/209] Loss: 2.6399, average\_loss: 2.2404

24  
24  
24  
24  
24

```
24
Epoch [26/50], Iter [60/209] Loss: 2.0825, average_loss: 2.2424
24
24
24
24
24
Epoch [26/50], Iter [65/209] Loss: 2.1715, average_loss: 2.2448
24
24
24
24
24
Epoch [26/50], Iter [70/209] Loss: 1.8687, average_loss: 2.2383
24
24
24
24
24
Epoch [26/50], Iter [75/209] Loss: 1.5322, average_loss: 2.2275
24
24
24
24
24
Epoch [26/50], Iter [80/209] Loss: 1.5636, average_loss: 2.2185
24
24
24
24
24
Epoch [26/50], Iter [85/209] Loss: 3.4684, average_loss: 2.2436
24
24
24
24
24
Epoch [26/50], Iter [90/209] Loss: 2.6825, average_loss: 2.2401
24
24
24
24
24
Epoch [26/50], Iter [95/209] Loss: 2.3971, average_loss: 2.2483
24
24
24
24
24
Epoch [26/50], Iter [100/209] Loss: 2.3282, average_loss: 2.2403
24
24
24
24
24
Epoch [26/50], Iter [105/209] Loss: 2.1838, average_loss: 2.2372
24
24
24
24
24
Epoch [26/50], Iter [110/209] Loss: 1.6522, average_loss: 2.2305
24
24
24
24
24
Epoch [26/50], Iter [115/209] Loss: 1.3449, average_loss: 2.2129
24
24
24
24
24
Epoch [26/50], Iter [120/209] Loss: 1.8476, average_loss: 2.2009
24
24
24
24
24
```

```
24
24
Epoch [26/50], Iter [125/209] Loss: 2.2026, average_loss: 2.2070
24
24
24
24
24
Epoch [26/50], Iter [130/209] Loss: 2.4541, average_loss: 2.2102
24
24
24
24
24
Epoch [26/50], Iter [135/209] Loss: 2.1682, average_loss: 2.2121
24
24
24
24
24
Epoch [26/50], Iter [140/209] Loss: 2.1633, average_loss: 2.2136
24
24
24
24
24
Epoch [26/50], Iter [145/209] Loss: 2.1743, average_loss: 2.2140
24
24
24
24
24
Epoch [26/50], Iter [150/209] Loss: 2.1673, average_loss: 2.2180
24
24
24
24
24
Epoch [26/50], Iter [155/209] Loss: 2.6235, average_loss: 2.2187
24
24
24
24
24
Epoch [26/50], Iter [160/209] Loss: 1.8135, average_loss: 2.2259
24
24
24
24
24
Epoch [26/50], Iter [165/209] Loss: 2.0096, average_loss: 2.2295
24
24
24
24
24
Epoch [26/50], Iter [170/209] Loss: 1.8099, average_loss: 2.2166
24
24
24
24
24
Epoch [26/50], Iter [175/209] Loss: 1.8904, average_loss: 2.2169
24
24
24
24
24
Epoch [26/50], Iter [180/209] Loss: 2.4135, average_loss: 2.2199
24
24
24
24
24
Epoch [26/50], Iter [185/209] Loss: 2.3703, average_loss: 2.2206
24
24
-
```

```

Epoch [26/50], Iter [190/209] Loss: 2.1521, average_loss: 2.2209
Epoch [26/50], Iter [195/209] Loss: 1.9380, average_loss: 2.2172
Epoch [26/50], Iter [200/209] Loss: 2.6758, average_loss: 2.2240
Epoch [26/50], Iter [205/209] Loss: 1.6465, average_loss: 2.2123

```

[illegible]



24  
6  
Updating best test loss: 2.83513

Starting epoch 27 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24  
Epoch [27/50], Iter [5/209] Loss: 2.0715, average\_loss: 2.1206  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [10/209] Loss: 2.2720, average\_loss: 2.1654  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [15/209] Loss: 1.9232, average\_loss: 2.0722  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [20/209] Loss: 2.7830, average\_loss: 2.1839  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [25/209] Loss: 2.4521, average\_loss: 2.1911  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [30/209] Loss: 2.2804, average\_loss: 2.2134  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [35/209] Loss: 1.8026, average\_loss: 2.2125  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [40/209] Loss: 2.2309, average\_loss: 2.2017  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [45/209] Loss: 2.3614, average\_loss: 2.1969  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [50/209] Loss: 1.9726, average\_loss: 2.1821  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [55/209] Loss: 2.0583, average\_loss: 2.1552  
24  
24  
24  
24



24  
Epoch [27/50], Iter [60/209] Loss: 1.8471, average\_loss: 2.1654  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [65/209] Loss: 1.8741, average\_loss: 2.1688  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [70/209] Loss: 2.7688, average\_loss: 2.1655  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [75/209] Loss: 2.9814, average\_loss: 2.1760  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [80/209] Loss: 3.2810, average\_loss: 2.1760  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [85/209] Loss: 1.6984, average\_loss: 2.1658  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [90/209] Loss: 1.6824, average\_loss: 2.1674  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [95/209] Loss: 2.2029, average\_loss: 2.1613  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [100/209] Loss: 2.1707, average\_loss: 2.1785  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [105/209] Loss: 2.4209, average\_loss: 2.1821  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [110/209] Loss: 1.8421, average\_loss: 2.1740  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [115/209] Loss: 1.5211, average\_loss: 2.1687  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [120/209] Loss: 2.2018, average\_loss: 2.1739  
24  
24  
24

24  
24  
Epoch [27/50], Iter [125/209] Loss: 2.2894, average\_loss: 2.1735  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [130/209] Loss: 3.4564, average\_loss: 2.1789  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [135/209] Loss: 2.0427, average\_loss: 2.1821  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [140/209] Loss: 2.3982, average\_loss: 2.1787  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [145/209] Loss: 1.9765, average\_loss: 2.1679  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [150/209] Loss: 1.8571, average\_loss: 2.1695  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [155/209] Loss: 2.3527, average\_loss: 2.1758  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [160/209] Loss: 1.7460, average\_loss: 2.1671  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [165/209] Loss: 2.5256, average\_loss: 2.1731  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [170/209] Loss: 2.0849, average\_loss: 2.1778  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [175/209] Loss: 2.0511, average\_loss: 2.1787  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [180/209] Loss: 1.8274, average\_loss: 2.1850  
24  
24  
24  
24  
24  
Epoch [27/50], Iter [185/209] Loss: 3.3088, average\_loss: 2.1897  
24  
24

```

Epoch [27/50], Iter [190/209] Loss: 1.7761, average_loss: 2.1885
Epoch [27/50], Iter [195/209] Loss: 1.9476, average_loss: 2.1895
Epoch [27/50], Iter [200/209] Loss: 2.5723, average_loss: 2.1924
Epoch [27/50], Iter [205/209] Loss: 2.1216, average_loss: 2.1961

```

[illegible]

[illegible]

24  
24  
6

Starting epoch 28 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24

Epoch [28/50], Iter [5/209] Loss: 1.8766, average\_loss: 2.0698

24  
24  
24  
24  
24

Epoch [28/50], Iter [10/209] Loss: 2.1051, average\_loss: 2.1812

24  
24  
24  
24  
24

Epoch [28/50], Iter [15/209] Loss: 1.6052, average\_loss: 2.2119

24  
24  
24  
24  
24

Epoch [28/50], Iter [20/209] Loss: 1.7500, average\_loss: 2.1855

24  
24  
24  
24  
24

Epoch [28/50], Iter [25/209] Loss: 2.1216, average\_loss: 2.1790

24  
24  
24  
24  
24

Epoch [28/50], Iter [30/209] Loss: 2.0947, average\_loss: 2.1855

24  
24  
24  
24  
24

Epoch [28/50], Iter [35/209] Loss: 2.3351, average\_loss: 2.1787

24  
24  
24  
24  
24

Epoch [28/50], Iter [40/209] Loss: 1.9027, average\_loss: 2.1541

24  
24  
24  
24  
24

Epoch [28/50], Iter [45/209] Loss: 2.0632, average\_loss: 2.1412

24  
24  
24  
24  
24

Epoch [28/50], Iter [50/209] Loss: 2.1889, average\_loss: 2.1575

24  
24  
24  
24  
24

Epoch [28/50], Iter [55/209] Loss: 2.6975, average\_loss: 2.1549

24  
24  
24  
24  
24

```
Epoch [28/50], Iter [60/209] Loss: 2.9661, average_loss: 2.1778
24
24
24
24
24
Epoch [28/50], Iter [65/209] Loss: 2.2618, average_loss: 2.1691
24
24
24
24
24
Epoch [28/50], Iter [70/209] Loss: 2.2167, average_loss: 2.1633
24
24
24
24
24
Epoch [28/50], Iter [75/209] Loss: 2.2160, average_loss: 2.1843
24
24
24
24
24
Epoch [28/50], Iter [80/209] Loss: 2.3901, average_loss: 2.1890
24
24
24
24
24
Epoch [28/50], Iter [85/209] Loss: 1.9442, average_loss: 2.1856
24
24
24
24
24
Epoch [28/50], Iter [90/209] Loss: 1.5963, average_loss: 2.1873
24
24
24
24
24
Epoch [28/50], Iter [95/209] Loss: 2.1219, average_loss: 2.1901
24
24
24
24
24
Epoch [28/50], Iter [100/209] Loss: 2.0399, average_loss: 2.1903
24
24
24
24
24
Epoch [28/50], Iter [105/209] Loss: 2.6286, average_loss: 2.2007
24
24
24
24
24
Epoch [28/50], Iter [110/209] Loss: 1.9457, average_loss: 2.2085
24
24
24
24
24
Epoch [28/50], Iter [115/209] Loss: 1.8212, average_loss: 2.2055
24
24
24
24
24
Epoch [28/50], Iter [120/209] Loss: 1.7426, average_loss: 2.1914
24
24
24
24
```

```
24
Epoch [28/50], Iter [125/209] Loss: 1.8609, average_loss: 2.1848
24
24
24
24
Epoch [28/50], Iter [130/209] Loss: 1.8488, average_loss: 2.1764
24
24
24
24
24
Epoch [28/50], Iter [135/209] Loss: 1.5402, average_loss: 2.1725
24
24
24
24
24
Epoch [28/50], Iter [140/209] Loss: 2.3952, average_loss: 2.1691
24
24
24
24
24
Epoch [28/50], Iter [145/209] Loss: 1.4798, average_loss: 2.1675
24
24
24
24
24
Epoch [28/50], Iter [150/209] Loss: 1.4154, average_loss: 2.1583
24
24
24
24
24
Epoch [28/50], Iter [155/209] Loss: 1.5852, average_loss: 2.1504
24
24
24
24
24
Epoch [28/50], Iter [160/209] Loss: 2.8648, average_loss: 2.1478
24
24
24
24
24
Epoch [28/50], Iter [165/209] Loss: 1.5328, average_loss: 2.1420
24
24
24
24
24
Epoch [28/50], Iter [170/209] Loss: 1.9967, average_loss: 2.1427
24
24
24
24
24
Epoch [28/50], Iter [175/209] Loss: 2.1714, average_loss: 2.1466
24
24
24
24
24
Epoch [28/50], Iter [180/209] Loss: 1.9676, average_loss: 2.1445
24
24
24
24
24
Epoch [28/50], Iter [185/209] Loss: 1.9284, average_loss: 2.1534
24
24
24
```



```
Epoch [28/50], Iter [190/209] Loss: 1.7854, average_loss: 2.1498
Epoch [28/50], Iter [195/209] Loss: 2.5019, average_loss: 2.1498
Epoch [28/50], Iter [200/209] Loss: 1.7866, average_loss: 2.1484
Epoch [28/50], Iter [205/209] Loss: 2.3731, average_loss: 2.1540
```



[illegible]

24  
6  
Updating best test loss: 2.83324

Starting epoch 29 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24  
Epoch [29/50], Iter [5/209] Loss: 1.8547, average\_loss: 1.7719  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [10/209] Loss: 1.9584, average\_loss: 1.9386  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [15/209] Loss: 2.2179, average\_loss: 1.9576  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [20/209] Loss: 1.7660, average\_loss: 1.9493  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [25/209] Loss: 2.1137, average\_loss: 1.9685  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [30/209] Loss: 1.6930, average\_loss: 2.0444  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [35/209] Loss: 2.0610, average\_loss: 2.0505  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [40/209] Loss: 2.1268, average\_loss: 2.0754  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [45/209] Loss: 1.7012, average\_loss: 2.0656  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [50/209] Loss: 1.5808, average\_loss: 2.0506  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [55/209] Loss: 2.6504, average\_loss: 2.0859  
24  
24  
24  
24  
24

```
24
Epoch [29/50], Iter [60/209] Loss: 1.6045, average_loss: 2.0954
24
24
24
24
24
Epoch [29/50], Iter [65/209] Loss: 2.3925, average_loss: 2.0973
24
24
24
24
24
Epoch [29/50], Iter [70/209] Loss: 1.6248, average_loss: 2.0844
24
24
24
24
24
Epoch [29/50], Iter [75/209] Loss: 2.1400, average_loss: 2.1087
24
24
24
24
24
Epoch [29/50], Iter [80/209] Loss: 2.0373, average_loss: 2.1313
24
24
24
24
24
Epoch [29/50], Iter [85/209] Loss: 1.4765, average_loss: 2.1181
24
24
24
24
24
Epoch [29/50], Iter [90/209] Loss: 1.9538, average_loss: 2.1126
24
24
24
24
24
Epoch [29/50], Iter [95/209] Loss: 2.1231, average_loss: 2.0978
24
24
24
24
24
Epoch [29/50], Iter [100/209] Loss: 1.7008, average_loss: 2.0966
24
24
24
24
24
Epoch [29/50], Iter [105/209] Loss: 2.2179, average_loss: 2.1147
24
24
24
24
24
Epoch [29/50], Iter [110/209] Loss: 2.3808, average_loss: 2.1089
24
24
24
24
24
Epoch [29/50], Iter [115/209] Loss: 2.9079, average_loss: 2.1088
24
24
24
24
24
Epoch [29/50], Iter [120/209] Loss: 1.9535, average_loss: 2.1001
24
24
24
24
24
```

24  
24  
Epoch [29/50], Iter [125/209] Loss: 1.5007, average\_loss: 2.0988  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [130/209] Loss: 1.6962, average\_loss: 2.0913  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [135/209] Loss: 1.5634, average\_loss: 2.0908  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [140/209] Loss: 2.2797, average\_loss: 2.0962  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [145/209] Loss: 2.4875, average\_loss: 2.1028  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [150/209] Loss: 2.3040, average\_loss: 2.1104  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [155/209] Loss: 2.1376, average\_loss: 2.1095  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [160/209] Loss: 1.9411, average\_loss: 2.1074  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [165/209] Loss: 1.9790, average\_loss: 2.1072  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [170/209] Loss: 1.9670, average\_loss: 2.1017  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [175/209] Loss: 1.7464, average\_loss: 2.1012  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [180/209] Loss: 2.2880, average\_loss: 2.0996  
24  
24  
24  
24  
24  
Epoch [29/50], Iter [185/209] Loss: 1.7921, average\_loss: 2.1014  
24  
24

```
Epoch [29/50], Iter [190/209] Loss: 2.6159, average_loss: 2.1032
Epoch [29/50], Iter [195/209] Loss: 2.4768, average_loss: 2.1092
Epoch [29/50], Iter [200/209] Loss: 2.3771, average_loss: 2.1048
Epoch [29/50], Iter [205/209] Loss: 1.6573, average_loss: 2.1037
```





[illegible]

24  
6

Starting epoch 30 / 50  
Learning Rate for this epoch: 0.001

24  
24  
24  
24  
24

Epoch [30/50], Iter [5/209] Loss: 1.7312, average\_loss: 2.2936

24  
24  
24  
24  
24

Epoch [30/50], Iter [10/209] Loss: 1.9268, average\_loss: 2.1162

24  
24  
24  
24  
24

Epoch [30/50], Iter [15/209] Loss: 1.6597, average\_loss: 2.1885

24  
24  
24  
24  
24

Epoch [30/50], Iter [20/209] Loss: 1.8606, average\_loss: 2.0531

24  
24  
24  
24  
24

Epoch [30/50], Iter [25/209] Loss: 1.7300, average\_loss: 1.9992

24  
24  
24  
24  
24

Epoch [30/50], Iter [30/209] Loss: 1.5104, average\_loss: 2.0652

24  
24  
24  
24  
24

Epoch [30/50], Iter [35/209] Loss: 1.5253, average\_loss: 2.0613

24  
24  
24  
24  
24

Epoch [30/50], Iter [40/209] Loss: 2.1268, average\_loss: 2.0788

24  
24  
24  
24  
24

Epoch [30/50], Iter [45/209] Loss: 2.0715, average\_loss: 2.0811

24  
24  
24  
24  
24

Epoch [30/50], Iter [50/209] Loss: 1.6824, average\_loss: 2.0942

24  
24  
24  
24  
24

Epoch [30/50], Iter [55/209] Loss: 2.2896, average\_loss: 2.1209

24  
24  
24  
24  
24

Epoch [30/50], Iter [60/209] Loss: 2.0995, average\_loss: 2.1194  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [65/209] Loss: 2.4085, average\_loss: 2.1066  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [70/209] Loss: 1.6765, average\_loss: 2.0865  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [75/209] Loss: 2.5891, average\_loss: 2.1195  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [80/209] Loss: 1.9930, average\_loss: 2.1109  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [85/209] Loss: 1.6719, average\_loss: 2.0975  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [90/209] Loss: 1.5886, average\_loss: 2.1038  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [95/209] Loss: 1.5378, average\_loss: 2.0945  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [100/209] Loss: 1.9242, average\_loss: 2.0794  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [105/209] Loss: 1.8867, average\_loss: 2.0686  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [110/209] Loss: 2.4445, average\_loss: 2.0795  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [115/209] Loss: 2.6220, average\_loss: 2.0802  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [120/209] Loss: 2.5172, average\_loss: 2.0884  
24  
24  
24  
24

24  
Epoch [30/50], Iter [125/209] Loss: 1.6999, average\_loss: 2.0835  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [130/209] Loss: 1.7321, average\_loss: 2.0888  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [135/209] Loss: 1.6201, average\_loss: 2.0732  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [140/209] Loss: 1.5781, average\_loss: 2.0732  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [145/209] Loss: 1.6429, average\_loss: 2.0775  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [150/209] Loss: 1.7399, average\_loss: 2.0697  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [155/209] Loss: 2.9268, average\_loss: 2.0814  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [160/209] Loss: 1.8764, average\_loss: 2.0858  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [165/209] Loss: 2.4266, average\_loss: 2.0847  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [170/209] Loss: 2.0865, average\_loss: 2.0838  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [175/209] Loss: 1.4568, average\_loss: 2.0789  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [180/209] Loss: 1.5519, average\_loss: 2.0750  
24  
24  
24  
24  
24  
Epoch [30/50], Iter [185/209] Loss: 2.3181, average\_loss: 2.0798  
24  
24  
24

```
Epoch [30/50], Iter [190/209] Loss: 1.7061, average_loss: 2.0794
Epoch [30/50], Iter [195/209] Loss: 2.2053, average_loss: 2.0788
Epoch [30/50], Iter [200/209] Loss: 3.0662, average_loss: 2.0848
Epoch [30/50], Iter [205/209] Loss: 2.4500, average_loss: 2.0857
```

[illegible]

[illegible]

```
24
6
Updating best test loss: 2.81505

Starting epoch 31 / 50
Learning Rate for this epoch: 0.0001
24
24
24
24
24
Epoch [31/50], Iter [5/209] Loss: 2.3383, average_loss: 2.0126
24
24
24
24
24
Epoch [31/50], Iter [10/209] Loss: 1.9268, average_loss: 1.9607
24
24
24
24
24
Epoch [31/50], Iter [15/209] Loss: 1.7788, average_loss: 1.9740
24
24
24
24
24
Epoch [31/50], Iter [20/209] Loss: 1.6076, average_loss: 1.9692
24
24
24
24
24
Epoch [31/50], Iter [25/209] Loss: 2.1575, average_loss: 1.9793
24
24
24
24
24
Epoch [31/50], Iter [30/209] Loss: 1.7844, average_loss: 2.0008
24
24
24
24
24
Epoch [31/50], Iter [35/209] Loss: 1.8318, average_loss: 1.9824
24
24
24
24
24
Epoch [31/50], Iter [40/209] Loss: 2.2775, average_loss: 1.9743
24
24
24
24
24
Epoch [31/50], Iter [45/209] Loss: 1.7263, average_loss: 1.9663
24
24
24
24
24
Epoch [31/50], Iter [50/209] Loss: 1.9037, average_loss: 1.9669
24
24
24
24
24
Epoch [31/50], Iter [55/209] Loss: 1.8088, average_loss: 1.9768
24
24
24
24
24
```



```
Epoch [31/50], Iter [60/209] Loss: 1.7190, average_loss: 1.9983
24
24
24
24
24
Epoch [31/50], Iter [65/209] Loss: 2.2856, average_loss: 2.0107
24
24
24
24
24
Epoch [31/50], Iter [70/209] Loss: 2.2585, average_loss: 2.0060
24
24
24
24
24
Epoch [31/50], Iter [75/209] Loss: 2.2714, average_loss: 2.0067
24
24
24
24
24
Epoch [31/50], Iter [80/209] Loss: 1.4867, average_loss: 1.9947
24
24
24
24
24
Epoch [31/50], Iter [85/209] Loss: 1.9402, average_loss: 2.0334
24
24
24
24
24
Epoch [31/50], Iter [90/209] Loss: 1.6613, average_loss: 2.0226
24
24
24
24
24
Epoch [31/50], Iter [95/209] Loss: 2.1038, average_loss: 2.0131
24
24
24
24
24
Epoch [31/50], Iter [100/209] Loss: 1.7848, average_loss: 2.0025
24
24
24
24
24
Epoch [31/50], Iter [105/209] Loss: 2.0941, average_loss: 1.9973
24
24
24
24
24
Epoch [31/50], Iter [110/209] Loss: 2.3317, average_loss: 2.0064
24
24
24
24
24
Epoch [31/50], Iter [115/209] Loss: 2.3311, average_loss: 2.0165
24
24
24
24
24
Epoch [31/50], Iter [120/209] Loss: 2.3817, average_loss: 2.0109
24
24
24
24
```

```
24
Epoch [31/50], Iter [125/209] Loss: 2.2269, average_loss: 2.0087
24
24
24
24
24
Epoch [31/50], Iter [130/209] Loss: 2.0241, average_loss: 2.0017
24
24
24
24
24
Epoch [31/50], Iter [135/209] Loss: 1.6181, average_loss: 2.0017
24
24
24
24
24
Epoch [31/50], Iter [140/209] Loss: 1.7763, average_loss: 2.0085
24
24
24
24
24
Epoch [31/50], Iter [145/209] Loss: 1.6529, average_loss: 2.0065
24
24
24
24
24
Epoch [31/50], Iter [150/209] Loss: 1.6475, average_loss: 1.9948
24
24
24
24
24
Epoch [31/50], Iter [155/209] Loss: 1.8104, average_loss: 1.9921
24
24
24
24
24
Epoch [31/50], Iter [160/209] Loss: 2.0819, average_loss: 1.9934
24
24
24
24
24
Epoch [31/50], Iter [165/209] Loss: 1.4684, average_loss: 1.9840
24
24
24
24
24
Epoch [31/50], Iter [170/209] Loss: 1.7082, average_loss: 1.9862
24
24
24
24
24
Epoch [31/50], Iter [175/209] Loss: 1.9868, average_loss: 1.9878
24
24
24
24
24
Epoch [31/50], Iter [180/209] Loss: 2.4211, average_loss: 1.9896
24
24
24
24
24
Epoch [31/50], Iter [185/209] Loss: 1.6913, average_loss: 1.9832
24
24
24
```

```
Epoch [31/50], Iter [190/209] Loss: 1.9178, average_loss: 1.9928
Epoch [31/50], Iter [195/209] Loss: 2.7379, average_loss: 1.9964
Epoch [31/50], Iter [200/209] Loss: 1.7770, average_loss: 1.9926
Epoch [31/50], Iter [205/209] Loss: 1.7219, average_loss: 1.9914
```

[illegible]

[illegible]

24  
6  
Updating best test loss: 2.75515

Starting epoch 32 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [32/50], Iter [5/209] Loss: 2.0795, average\_loss: 1.7670  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [10/209] Loss: 1.7294, average\_loss: 1.7711  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [15/209] Loss: 2.3328, average\_loss: 1.8631  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [20/209] Loss: 2.1670, average\_loss: 1.9351  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [25/209] Loss: 1.5759, average\_loss: 1.9625  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [30/209] Loss: 1.4426, average\_loss: 1.9344  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [35/209] Loss: 1.6840, average\_loss: 1.9686  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [40/209] Loss: 2.0147, average\_loss: 1.9548  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [45/209] Loss: 1.8956, average\_loss: 1.9632  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [50/209] Loss: 1.8537, average\_loss: 1.9524  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [55/209] Loss: 1.3684, average\_loss: 1.9746  
24  
24  
24  
24  
24

24  
Epoch [32/50], Iter [60/209] Loss: 1.8116, average\_loss: 1.9746  
24  
24  
24  
24  
Epoch [32/50], Iter [65/209] Loss: 2.7242, average\_loss: 1.9875  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [70/209] Loss: 1.9628, average\_loss: 1.9909  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [75/209] Loss: 3.1996, average\_loss: 1.9865  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [80/209] Loss: 2.3834, average\_loss: 1.9837  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [85/209] Loss: 2.0652, average\_loss: 1.9699  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [90/209] Loss: 2.3313, average\_loss: 1.9732  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [95/209] Loss: 2.5226, average\_loss: 1.9680  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [100/209] Loss: 2.1365, average\_loss: 1.9720  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [105/209] Loss: 2.3515, average\_loss: 1.9778  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [110/209] Loss: 2.2553, average\_loss: 1.9734  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [115/209] Loss: 2.3876, average\_loss: 1.9817  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [120/209] Loss: 1.7422, average\_loss: 1.9812  
24  
24  
24

24  
24  
Epoch [32/50], Iter [125/209] Loss: 1.2416, average\_loss: 1.9726  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [130/209] Loss: 1.2729, average\_loss: 1.9726  
24  
24  
24  
24  
Epoch [32/50], Iter [135/209] Loss: 2.3614, average\_loss: 1.9765  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [140/209] Loss: 1.6731, average\_loss: 1.9681  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [145/209] Loss: 1.8066, average\_loss: 1.9630  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [150/209] Loss: 2.2090, average\_loss: 1.9647  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [155/209] Loss: 1.4508, average\_loss: 1.9581  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [160/209] Loss: 2.2960, average\_loss: 1.9546  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [165/209] Loss: 1.6392, average\_loss: 1.9580  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [170/209] Loss: 1.1233, average\_loss: 1.9549  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [175/209] Loss: 2.4544, average\_loss: 1.9533  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [180/209] Loss: 1.8477, average\_loss: 1.9560  
24  
24  
24  
24  
24  
Epoch [32/50], Iter [185/209] Loss: 1.8266, average\_loss: 1.9512  
24  
24



```
Epoch [32/50], Iter [190/209] Loss: 1.5171, average_loss: 1.9467
Epoch [32/50], Iter [195/209] Loss: 1.8075, average_loss: 1.9393
Epoch [32/50], Iter [200/209] Loss: 2.0401, average_loss: 1.9362
Epoch [32/50], Iter [205/209] Loss: 1.1867, average_loss: 1.9310
```

[illegible]

[illegible]

24

6

Updating best test loss: 2.74026

Starting epoch 33 / 50

Learning Rate for this epoch: 0.0001

24

24

24

24

24

Epoch [33/50], Iter [5/209] Loss: 1.6521, average\_loss: 1.7393

24

24

24

24

24

Epoch [33/50], Iter [10/209] Loss: 1.8355, average\_loss: 1.8941

24

24

24

24

24

Epoch [33/50], Iter [15/209] Loss: 1.8141, average\_loss: 1.8704

24

24

24

24

24

Epoch [33/50], Iter [20/209] Loss: 1.6488, average\_loss: 1.8193

24

24

24

24

24

Epoch [33/50], Iter [25/209] Loss: 2.1024, average\_loss: 1.8489

24

24

24

24

24

Epoch [33/50], Iter [30/209] Loss: 1.5223, average\_loss: 1.8550

24

24

24

24

24

Epoch [33/50], Iter [35/209] Loss: 1.4584, average\_loss: 1.8818

24

24

24

24

24

Epoch [33/50], Iter [40/209] Loss: 2.6248, average\_loss: 1.9046

24

24

24

24

24

Epoch [33/50], Iter [45/209] Loss: 2.5181, average\_loss: 1.9365

24

24

24

24

24

Epoch [33/50], Iter [50/209] Loss: 2.2560, average\_loss: 1.9122

24

24

24

24

24

Epoch [33/50], Iter [55/209] Loss: 1.6370, average\_loss: 1.8946

24

24

24

24

```
24
Epoch [33/50], Iter [60/209] Loss: 2.2093, average_loss: 1.9013
24
24
24
24
24
Epoch [33/50], Iter [65/209] Loss: 1.8132, average_loss: 1.9281
24
24
24
24
24
Epoch [33/50], Iter [70/209] Loss: 2.0918, average_loss: 1.9406
24
24
24
24
24
Epoch [33/50], Iter [75/209] Loss: 2.1841, average_loss: 1.9504
24
24
24
24
24
Epoch [33/50], Iter [80/209] Loss: 2.0063, average_loss: 1.9468
24
24
24
24
24
Epoch [33/50], Iter [85/209] Loss: 1.8451, average_loss: 1.9503
24
24
24
24
24
Epoch [33/50], Iter [90/209] Loss: 1.3898, average_loss: 1.9672
24
24
24
24
24
Epoch [33/50], Iter [95/209] Loss: 1.9694, average_loss: 1.9642
24
24
24
24
24
Epoch [33/50], Iter [100/209] Loss: 2.7127, average_loss: 1.9795
24
24
24
24
24
Epoch [33/50], Iter [105/209] Loss: 1.1738, average_loss: 1.9634
24
24
24
24
24
Epoch [33/50], Iter [110/209] Loss: 1.3119, average_loss: 1.9446
24
24
24
24
24
Epoch [33/50], Iter [115/209] Loss: 2.6725, average_loss: 1.9444
24
24
24
24
24
Epoch [33/50], Iter [120/209] Loss: 1.7717, average_loss: 1.9462
24
24
24
```

24  
24  
Epoch [33/50], Iter [125/209] Loss: 1.6152, average\_loss: 1.9505  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [130/209] Loss: 2.3800, average\_loss: 1.9542  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [135/209] Loss: 2.1573, average\_loss: 1.9535  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [140/209] Loss: 1.5894, average\_loss: 1.9431  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [145/209] Loss: 1.9240, average\_loss: 1.9413  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [150/209] Loss: 1.2999, average\_loss: 1.9377  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [155/209] Loss: 1.9646, average\_loss: 1.9374  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [160/209] Loss: 1.8536, average\_loss: 1.9375  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [165/209] Loss: 1.7686, average\_loss: 1.9325  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [170/209] Loss: 1.4914, average\_loss: 1.9269  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [175/209] Loss: 2.1278, average\_loss: 1.9300  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [180/209] Loss: 1.8658, average\_loss: 1.9328  
24  
24  
24  
24  
24  
Epoch [33/50], Iter [185/209] Loss: 2.5897, average\_loss: 1.9399  
24  
24

```
Epoch [33/50], Iter [190/209] Loss: 2.2628, average_loss: 1.9431
Epoch [33/50], Iter [195/209] Loss: 1.9496, average_loss: 1.9398
Epoch [33/50], Iter [200/209] Loss: 2.3326, average_loss: 1.9433
Epoch [33/50], Iter [205/209] Loss: 1.9133, average_loss: 1.9379
```

[illegible]





24  
24  
6  
Updating best test loss: 2.73536

Starting epoch 34 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [34/50], Iter [5/209] Loss: 2.0339, average\_loss: 1.7638  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [10/209] Loss: 1.5883, average\_loss: 1.7909  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [15/209] Loss: 1.8541, average\_loss: 1.9008  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [20/209] Loss: 1.6040, average\_loss: 1.8519  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [25/209] Loss: 2.1592, average\_loss: 1.8623  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [30/209] Loss: 2.1975, average\_loss: 1.8580  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [35/209] Loss: 1.4446, average\_loss: 1.8197  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [40/209] Loss: 1.8379, average\_loss: 1.7985  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [45/209] Loss: 1.6053, average\_loss: 1.8211  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [50/209] Loss: 1.8281, average\_loss: 1.8505  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [55/209] Loss: 2.4024, average\_loss: 1.8563  
24  
24  
24  
24

```
24
Epoch [34/50], Iter [60/209] Loss: 2.1427, average_loss: 1.8500
24
24
24
24
24
Epoch [34/50], Iter [65/209] Loss: 1.7266, average_loss: 1.8479
24
24
24
24
24
Epoch [34/50], Iter [70/209] Loss: 1.5983, average_loss: 1.8486
24
24
24
24
24
Epoch [34/50], Iter [75/209] Loss: 1.1869, average_loss: 1.8326
24
24
24
24
24
Epoch [34/50], Iter [80/209] Loss: 2.0187, average_loss: 1.8351
24
24
24
24
24
Epoch [34/50], Iter [85/209] Loss: 2.0561, average_loss: 1.8380
24
24
24
24
24
Epoch [34/50], Iter [90/209] Loss: 1.7433, average_loss: 1.8368
24
24
24
24
24
Epoch [34/50], Iter [95/209] Loss: 1.2628, average_loss: 1.8387
24
24
24
24
24
Epoch [34/50], Iter [100/209] Loss: 3.1893, average_loss: 1.8558
24
24
24
24
24
Epoch [34/50], Iter [105/209] Loss: 2.0065, average_loss: 1.8649
24
24
24
24
24
Epoch [34/50], Iter [110/209] Loss: 2.1253, average_loss: 1.8724
24
24
24
24
24
Epoch [34/50], Iter [115/209] Loss: 1.5566, average_loss: 1.8719
24
24
24
24
24
Epoch [34/50], Iter [120/209] Loss: 1.7107, average_loss: 1.8745
24
24
24
```

24  
24  
24  
Epoch [34/50], Iter [125/209] Loss: 1.6288, average\_loss: 1.8797  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [130/209] Loss: 2.0452, average\_loss: 1.8776  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [135/209] Loss: 2.3201, average\_loss: 1.8784  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [140/209] Loss: 2.1444, average\_loss: 1.8801  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [145/209] Loss: 1.8631, average\_loss: 1.8797  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [150/209] Loss: 1.7204, average\_loss: 1.8797  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [155/209] Loss: 2.0174, average\_loss: 1.8772  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [160/209] Loss: 1.4444, average\_loss: 1.8683  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [165/209] Loss: 2.0944, average\_loss: 1.8788  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [170/209] Loss: 2.2778, average\_loss: 1.8879  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [175/209] Loss: 1.6116, average\_loss: 1.8972  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [180/209] Loss: 2.0739, average\_loss: 1.9020  
24  
24  
24  
24  
24  
Epoch [34/50], Iter [185/209] Loss: 2.2197, average\_loss: 1.9077  
24  
24

```
Epoch [34/50], Iter [190/209] Loss: 2.2212, average_loss: 1.9050
Epoch [34/50], Iter [195/209] Loss: 2.2579, average_loss: 1.9124
Epoch [34/50], Iter [200/209] Loss: 1.5575, average_loss: 1.9156
Epoch [34/50], Iter [205/209] Loss: 1.3417, average_loss: 1.9149
```

[illegible]



24  
24  
6  
Updating best test loss: 2.73191

Starting epoch 35 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [35/50], Iter [5/209] Loss: 1.3567, average\_loss: 1.8349  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [10/209] Loss: 1.8964, average\_loss: 1.7361  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [15/209] Loss: 2.4015, average\_loss: 1.8154  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [20/209] Loss: 1.2977, average\_loss: 1.8406  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [25/209] Loss: 1.7625, average\_loss: 1.8634  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [30/209] Loss: 1.9573, average\_loss: 1.8602  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [35/209] Loss: 1.2690, average\_loss: 1.8587  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [40/209] Loss: 1.9694, average\_loss: 1.8672  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [45/209] Loss: 2.2843, average\_loss: 1.8802  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [50/209] Loss: 2.0276, average\_loss: 1.9000  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [55/209] Loss: 1.6104, average\_loss: 1.8852  
24  
24  
24



```
24
24
Epoch [35/50], Iter [60/209] Loss: 1.9376, average_loss: 1.8900
24
24
24
24
24
Epoch [35/50], Iter [65/209] Loss: 1.7253, average_loss: 1.8943
24
24
24
24
24
Epoch [35/50], Iter [70/209] Loss: 1.5594, average_loss: 1.8823
24
24
24
24
24
Epoch [35/50], Iter [75/209] Loss: 1.8779, average_loss: 1.8776
24
24
24
24
24
Epoch [35/50], Iter [80/209] Loss: 1.5721, average_loss: 1.8608
24
24
24
24
24
Epoch [35/50], Iter [85/209] Loss: 1.5125, average_loss: 1.8533
24
24
24
24
24
Epoch [35/50], Iter [90/209] Loss: 1.3014, average_loss: 1.8483
24
24
24
24
24
Epoch [35/50], Iter [95/209] Loss: 1.7920, average_loss: 1.8510
24
24
24
24
24
Epoch [35/50], Iter [100/209] Loss: 1.5534, average_loss: 1.8585
24
24
24
24
24
Epoch [35/50], Iter [105/209] Loss: 1.7834, average_loss: 1.8606
24
24
24
24
24
Epoch [35/50], Iter [110/209] Loss: 1.5957, average_loss: 1.8653
24
24
24
24
24
Epoch [35/50], Iter [115/209] Loss: 2.2103, average_loss: 1.8585
24
24
24
24
24
Epoch [35/50], Iter [120/209] Loss: 1.9878, average_loss: 1.8586
24
24
```

24  
24  
24  
Epoch [35/50], Iter [125/209] Loss: 2.1637, average\_loss: 1.8718  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [130/209] Loss: 2.1339, average\_loss: 1.8660  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [135/209] Loss: 2.3584, average\_loss: 1.8691  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [140/209] Loss: 1.8870, average\_loss: 1.8625  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [145/209] Loss: 2.5429, average\_loss: 1.8613  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [150/209] Loss: 2.0805, average\_loss: 1.8574  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [155/209] Loss: 2.1329, average\_loss: 1.8701  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [160/209] Loss: 2.2001, average\_loss: 1.8695  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [165/209] Loss: 2.9943, average\_loss: 1.8747  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [170/209] Loss: 1.7285, average\_loss: 1.8712  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [175/209] Loss: 1.5830, average\_loss: 1.8715  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [180/209] Loss: 2.3863, average\_loss: 1.8724  
24  
24  
24  
24  
24  
Epoch [35/50], Iter [185/209] Loss: 1.9654, average\_loss: 1.8747  
24

```

Epoch [35/50], Iter [190/209] Loss: 2.1508, average_loss: 1.8811
Epoch [35/50], Iter [195/209] Loss: 1.9993, average_loss: 1.8892
Epoch [35/50], Iter [200/209] Loss: 2.8517, average_loss: 1.8940
Epoch [35/50], Iter [205/209] Loss: 2.1678, average_loss: 1.8921

```

[illegible]



24  
24  
6  
Updating best test loss: 2.72813

Starting epoch 36 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [36/50], Iter [5/209] Loss: 1.9952, average\_loss: 1.7915  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [10/209] Loss: 2.6076, average\_loss: 1.9495  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [15/209] Loss: 1.6331, average\_loss: 1.8744  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [20/209] Loss: 1.5014, average\_loss: 1.9292  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [25/209] Loss: 2.0175, average\_loss: 1.8646  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [30/209] Loss: 1.4263, average\_loss: 1.8332  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [35/209] Loss: 1.8657, average\_loss: 1.8446  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [40/209] Loss: 1.3866, average\_loss: 1.8643  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [45/209] Loss: 1.4928, average\_loss: 1.8411  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [50/209] Loss: 1.6671, average\_loss: 1.8338  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [55/209] Loss: 1.5934, average\_loss: 1.8203  
24  
24  
24

24  
24  
Epoch [36/50], Iter [60/209] Loss: 1.7775, average\_loss: 1.8252  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [65/209] Loss: 2.2718, average\_loss: 1.8531  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [70/209] Loss: 2.1700, average\_loss: 1.8553  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [75/209] Loss: 1.4857, average\_loss: 1.8511  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [80/209] Loss: 2.3832, average\_loss: 1.8620  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [85/209] Loss: 1.4926, average\_loss: 1.8703  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [90/209] Loss: 2.1387, average\_loss: 1.8727  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [95/209] Loss: 1.3184, average\_loss: 1.8648  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [100/209] Loss: 1.3235, average\_loss: 1.8543  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [105/209] Loss: 1.3045, average\_loss: 1.8500  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [110/209] Loss: 2.9110, average\_loss: 1.8590  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [115/209] Loss: 1.7298, average\_loss: 1.8624  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [120/209] Loss: 2.0673, average\_loss: 1.8575  
24  
24

24  
24  
24  
Epoch [36/50], Iter [125/209] Loss: 1.8516, average\_loss: 1.8618  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [130/209] Loss: 1.7346, average\_loss: 1.8654  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [135/209] Loss: 2.3694, average\_loss: 1.8675  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [140/209] Loss: 2.0681, average\_loss: 1.8706  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [145/209] Loss: 1.8587, average\_loss: 1.8764  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [150/209] Loss: 1.7094, average\_loss: 1.8732  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [155/209] Loss: 1.6769, average\_loss: 1.8863  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [160/209] Loss: 1.8170, average\_loss: 1.8870  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [165/209] Loss: 2.2174, average\_loss: 1.8890  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [170/209] Loss: 2.2150, average\_loss: 1.8963  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [175/209] Loss: 2.1882, average\_loss: 1.8960  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [180/209] Loss: 2.7635, average\_loss: 1.9092  
24  
24  
24  
24  
24  
Epoch [36/50], Iter [185/209] Loss: 1.9235, average\_loss: 1.9043  
24



[illegible]



[illegible]

24  
24  
24  
6  
Updating best test loss: 2.72242

Starting epoch 37 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [37/50], Iter [5/209] Loss: 1.7796, average\_loss: 1.6383  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [10/209] Loss: 2.7281, average\_loss: 1.8516  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [15/209] Loss: 1.4930, average\_loss: 1.7651  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [20/209] Loss: 1.9345, average\_loss: 1.7998  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [25/209] Loss: 2.3841, average\_loss: 1.8040  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [30/209] Loss: 1.7240, average\_loss: 1.8168  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [35/209] Loss: 1.8280, average\_loss: 1.7851  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [40/209] Loss: 1.7745, average\_loss: 1.7785  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [45/209] Loss: 1.3230, average\_loss: 1.8052  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [50/209] Loss: 1.6801, average\_loss: 1.8260  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [55/209] Loss: 1.6888, average\_loss: 1.8384  
24  
24  
24

24  
24  
24  
Epoch [37/50], Iter [60/209] Loss: 1.7119, average\_loss: 1.8189  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [65/209] Loss: 1.8288, average\_loss: 1.8209  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [70/209] Loss: 1.7463, average\_loss: 1.8396  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [75/209] Loss: 1.5870, average\_loss: 1.8608  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [80/209] Loss: 1.2770, average\_loss: 1.8496  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [85/209] Loss: 1.9739, average\_loss: 1.8454  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [90/209] Loss: 1.4406, average\_loss: 1.8455  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [95/209] Loss: 2.0067, average\_loss: 1.8679  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [100/209] Loss: 1.9437, average\_loss: 1.8588  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [105/209] Loss: 2.6660, average\_loss: 1.8615  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [110/209] Loss: 2.0373, average\_loss: 1.8619  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [115/209] Loss: 2.8375, average\_loss: 1.8547  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [120/209] Loss: 1.3003, average\_loss: 1.8550  
24  
24

24  
24  
24  
24  
Epoch [37/50], Iter [125/209] Loss: 1.6752, average\_loss: 1.8614  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [130/209] Loss: 1.7517, average\_loss: 1.8706  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [135/209] Loss: 2.5111, average\_loss: 1.8728  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [140/209] Loss: 1.4650, average\_loss: 1.8750  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [145/209] Loss: 1.6263, average\_loss: 1.8694  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [150/209] Loss: 1.6547, average\_loss: 1.8667  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [155/209] Loss: 2.0349, average\_loss: 1.8604  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [160/209] Loss: 1.4911, average\_loss: 1.8583  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [165/209] Loss: 1.2872, average\_loss: 1.8506  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [170/209] Loss: 2.0023, average\_loss: 1.8489  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [175/209] Loss: 1.8709, average\_loss: 1.8538  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [180/209] Loss: 3.5003, average\_loss: 1.8652  
24  
24  
24  
24  
24  
Epoch [37/50], Iter [185/209] Loss: 1.8208, average\_loss: 1.8670  
24

```

Epoch [37/50], Iter [190/209] Loss: 2.6977, average_loss: 1.8705
Epoch [37/50], Iter [195/209] Loss: 1.8681, average_loss: 1.8696
Epoch [37/50], Iter [200/209] Loss: 1.6933, average_loss: 1.8644
Epoch [37/50], Iter [205/209] Loss: 2.5332, average_loss: 1.8685

```





[illegible]

24  
24  
24  
6

Starting epoch 38 / 50

Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24

Epoch [38/50], Iter [5/209] Loss: 2.2059, average\_loss: 2.0053

24  
24  
24  
24  
24

Epoch [38/50], Iter [10/209] Loss: 1.8817, average\_loss: 1.9935

24  
24  
24  
24  
24

Epoch [38/50], Iter [15/209] Loss: 2.9847, average\_loss: 2.0262

24  
24  
24  
24  
24

Epoch [38/50], Iter [20/209] Loss: 1.3143, average\_loss: 1.9647

24  
24  
24  
24  
24

Epoch [38/50], Iter [25/209] Loss: 1.5512, average\_loss: 1.8731

24  
24  
24  
24  
24

Epoch [38/50], Iter [30/209] Loss: 1.6345, average\_loss: 1.8849

24  
24  
24  
24  
24

Epoch [38/50], Iter [35/209] Loss: 1.7275, average\_loss: 1.8776

24  
24  
24  
24  
24

Epoch [38/50], Iter [40/209] Loss: 1.5240, average\_loss: 1.8513

24  
24  
24  
24  
24

Epoch [38/50], Iter [45/209] Loss: 1.9343, average\_loss: 1.8492

24  
24  
24  
24  
24

Epoch [38/50], Iter [50/209] Loss: 2.2218, average\_loss: 1.8552

24  
24  
24  
24  
24

Epoch [38/50], Iter [55/209] Loss: 1.7846, average\_loss: 1.8381

24  
24  
24

24  
24  
Epoch [38/50], Iter [60/209] Loss: 1.9122, average\_loss: 1.8471  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [65/209] Loss: 1.7629, average\_loss: 1.8528  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [70/209] Loss: 1.4220, average\_loss: 1.8409  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [75/209] Loss: 1.1502, average\_loss: 1.8304  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [80/209] Loss: 1.7740, average\_loss: 1.8290  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [85/209] Loss: 2.0344, average\_loss: 1.8182  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [90/209] Loss: 1.3420, average\_loss: 1.8184  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [95/209] Loss: 2.0031, average\_loss: 1.8309  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [100/209] Loss: 1.6037, average\_loss: 1.8212  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [105/209] Loss: 1.7527, average\_loss: 1.8106  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [110/209] Loss: 1.4270, average\_loss: 1.8095  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [115/209] Loss: 1.8216, average\_loss: 1.8320  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [120/209] Loss: 1.8808, average\_loss: 1.8304  
24  
24

24  
24  
24  
Epoch [38/50], Iter [125/209] Loss: 1.7596, average\_loss: 1.8324  
24  
24  
24  
24  
24  
Epoch [38/50], Iter [130/209] Loss: 1.3870, average\_loss: 1.8424  
24  
24  
24  
24  
Epoch [38/50], Iter [135/209] Loss: 1.9209, average\_loss: 1.8456  
24  
24  
24  
24  
Epoch [38/50], Iter [140/209] Loss: 1.5335, average\_loss: 1.8443  
24  
24  
24  
24  
Epoch [38/50], Iter [145/209] Loss: 2.5033, average\_loss: 1.8500  
24  
24  
24  
24  
Epoch [38/50], Iter [150/209] Loss: 1.8348, average\_loss: 1.8543  
24  
24  
24  
24  
Epoch [38/50], Iter [155/209] Loss: 1.8235, average\_loss: 1.8514  
24  
24  
24  
24  
Epoch [38/50], Iter [160/209] Loss: 1.9549, average\_loss: 1.8508  
24  
24  
24  
24  
Epoch [38/50], Iter [165/209] Loss: 1.7538, average\_loss: 1.8476  
24  
24  
24  
24  
Epoch [38/50], Iter [170/209] Loss: 1.3982, average\_loss: 1.8388  
24  
24  
24  
24  
Epoch [38/50], Iter [175/209] Loss: 1.3715, average\_loss: 1.8445  
24  
24  
24  
24  
Epoch [38/50], Iter [180/209] Loss: 1.7445, average\_loss: 1.8474  
24  
24  
24  
24  
Epoch [38/50], Iter [185/209] Loss: 1.8168, average\_loss: 1.8493  
24

```
Epoch [38/50], Iter [190/209] Loss: 2.0616, average_loss: 1.8468  
Epoch [38/50], Iter [195/209] Loss: 1.5520, average_loss: 1.8494  
Epoch [38/50], Iter [200/209] Loss: 2.2944, average_loss: 1.8502  
Epoch [38/50], Iter [205/209] Loss: 1.4087, average_loss: 1.8471
```



[illegible]

24  
24  
6

Starting epoch 39 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24

Epoch [39/50], Iter [5/209] Loss: 2.2617, average\_loss: 1.8363

24  
24  
24  
24  
24

Epoch [39/50], Iter [10/209] Loss: 1.5429, average\_loss: 1.7465

24  
24  
24  
24  
24

Epoch [39/50], Iter [15/209] Loss: 1.8844, average\_loss: 1.7616

24  
24  
24  
24  
24

Epoch [39/50], Iter [20/209] Loss: 1.9870, average\_loss: 1.7741

24  
24  
24  
24  
24

Epoch [39/50], Iter [25/209] Loss: 1.9832, average\_loss: 1.7297

24  
24  
24  
24  
24

Epoch [39/50], Iter [30/209] Loss: 2.0426, average\_loss: 1.7231

24  
24  
24  
24  
24

Epoch [39/50], Iter [35/209] Loss: 2.2759, average\_loss: 1.7669

24  
24  
24  
24  
24

Epoch [39/50], Iter [40/209] Loss: 1.1907, average\_loss: 1.7733

24  
24  
24  
24  
24

Epoch [39/50], Iter [45/209] Loss: 1.4306, average\_loss: 1.7622

24  
24  
24  
24  
24

Epoch [39/50], Iter [50/209] Loss: 2.1054, average\_loss: 1.7890

24  
24  
24  
24  
24

Epoch [39/50], Iter [55/209] Loss: 1.2750, average\_loss: 1.7822

24  
24  
24  
24



24  
Epoch [39/50], Iter [60/209] Loss: 2.8179, average\_loss: 1.7935  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [65/209] Loss: 1.7640, average\_loss: 1.7959  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [70/209] Loss: 2.1840, average\_loss: 1.8015  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [75/209] Loss: 1.9993, average\_loss: 1.8165  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [80/209] Loss: 1.7076, average\_loss: 1.8213  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [85/209] Loss: 2.6035, average\_loss: 1.8484  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [90/209] Loss: 1.5035, average\_loss: 1.8411  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [95/209] Loss: 1.6000, average\_loss: 1.8496  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [100/209] Loss: 2.1827, average\_loss: 1.8529  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [105/209] Loss: 1.6459, average\_loss: 1.8487  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [110/209] Loss: 2.0444, average\_loss: 1.8481  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [115/209] Loss: 1.5793, average\_loss: 1.8483  
24  
24  
24  
24  
24  
Epoch [39/50], Iter [120/209] Loss: 1.8267, average\_loss: 1.8507  
24  
24  
24

```
24
24
Epoch [39/50], Iter [125/209] Loss: 1.3183, average_loss: 1.8483
24
24
24
24
Epoch [39/50], Iter [130/209] Loss: 1.3506, average_loss: 1.8494
24
24
24
24
Epoch [39/50], Iter [135/209] Loss: 1.9087, average_loss: 1.8522
24
24
24
24
Epoch [39/50], Iter [140/209] Loss: 2.2993, average_loss: 1.8466
24
24
24
24
Epoch [39/50], Iter [145/209] Loss: 2.1938, average_loss: 1.8504
24
24
24
24
Epoch [39/50], Iter [150/209] Loss: 1.1870, average_loss: 1.8522
24
24
24
24
Epoch [39/50], Iter [155/209] Loss: 1.7569, average_loss: 1.8463
24
24
24
24
Epoch [39/50], Iter [160/209] Loss: 1.5854, average_loss: 1.8469
24
24
24
24
Epoch [39/50], Iter [165/209] Loss: 1.4259, average_loss: 1.8490
24
24
24
24
Epoch [39/50], Iter [170/209] Loss: 2.1014, average_loss: 1.8547
24
24
24
24
Epoch [39/50], Iter [175/209] Loss: 1.5521, average_loss: 1.8574
24
24
24
24
Epoch [39/50], Iter [180/209] Loss: 2.2804, average_loss: 1.8567
24
24
24
24
Epoch [39/50], Iter [185/209] Loss: 1.9672, average_loss: 1.8670
24
24
```

```
Epoch [39/50], Iter [190/209] Loss: 2.0938, average_loss: 1.8616
Epoch [39/50], Iter [195/209] Loss: 1.2942, average_loss: 1.8619
Epoch [39/50], Iter [200/209] Loss: 1.6535, average_loss: 1.8647
Epoch [39/50], Iter [205/209] Loss: 1.5560, average_loss: 1.8606
```

[illegible]

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24  
24  
6  
Updating best test loss: 2.72147

Starting epoch 40 / 50  
Learning Rate for this epoch: 0.0001

24  
24  
24  
24  
24  
Epoch [40/50], Iter [5/209] Loss: 2.0707, average\_loss: 1.9529  
24  
24  
24  
24  
Epoch [40/50], Iter [10/209] Loss: 1.7145, average\_loss: 2.0506  
24  
24  
24  
24  
Epoch [40/50], Iter [15/209] Loss: 1.4635, average\_loss: 1.9873  
24  
24  
24  
24  
Epoch [40/50], Iter [20/209] Loss: 2.1810, average\_loss: 2.0007  
24  
24  
24  
24  
Epoch [40/50], Iter [25/209] Loss: 1.4411, average\_loss: 1.9437  
24  
24  
24  
24  
Epoch [40/50], Iter [30/209] Loss: 1.9085, average\_loss: 1.9105  
24  
24  
24  
24  
Epoch [40/50], Iter [35/209] Loss: 1.9576, average\_loss: 1.8903  
24  
24  
24  
24  
Epoch [40/50], Iter [40/209] Loss: 1.4077, average\_loss: 1.8711  
24  
24  
24  
24  
Epoch [40/50], Iter [45/209] Loss: 1.8038, average\_loss: 1.8580  
24  
24  
24  
24  
Epoch [40/50], Iter [50/209] Loss: 1.8356, average\_loss: 1.8468  
24  
24  
24  
24  
Epoch [40/50], Iter [55/209] Loss: 1.8752, average\_loss: 1.8383  
24  
24  
24  
24

```
24
24
Epoch [40/50], Iter [60/209] Loss: 1.5632, average_loss: 1.8359
24
24
24
24
24
Epoch [40/50], Iter [65/209] Loss: 1.5515, average_loss: 1.8345
24
24
24
24
24
Epoch [40/50], Iter [70/209] Loss: 3.2809, average_loss: 1.8658
24
24
24
24
24
Epoch [40/50], Iter [75/209] Loss: 2.0513, average_loss: 1.8606
24
24
24
24
24
Epoch [40/50], Iter [80/209] Loss: 1.8802, average_loss: 1.8506
24
24
24
24
24
Epoch [40/50], Iter [85/209] Loss: 1.7012, average_loss: 1.8430
24
24
24
24
24
Epoch [40/50], Iter [90/209] Loss: 2.2023, average_loss: 1.8375
24
24
24
24
24
Epoch [40/50], Iter [95/209] Loss: 1.5282, average_loss: 1.8357
24
24
24
24
24
Epoch [40/50], Iter [100/209] Loss: 1.8329, average_loss: 1.8286
24
24
24
24
24
Epoch [40/50], Iter [105/209] Loss: 1.4973, average_loss: 1.8248
24
24
24
24
24
Epoch [40/50], Iter [110/209] Loss: 2.1630, average_loss: 1.8340
24
24
24
24
24
Epoch [40/50], Iter [115/209] Loss: 2.6828, average_loss: 1.8464
24
24
24
24
24
Epoch [40/50], Iter [120/209] Loss: 1.6209, average_loss: 1.8479
24
24
24
```

```
24
24
24
Epoch [40/50], Iter [125/209] Loss: 1.3902, average_loss: 1.8455
24
24
24
24
24
Epoch [40/50], Iter [130/209] Loss: 1.7393, average_loss: 1.8567
24
24
24
24
24
Epoch [40/50], Iter [135/209] Loss: 2.7480, average_loss: 1.8640
24
24
24
24
24
Epoch [40/50], Iter [140/209] Loss: 2.7583, average_loss: 1.8675
24
24
24
24
24
Epoch [40/50], Iter [145/209] Loss: 2.6029, average_loss: 1.8780
24
24
24
24
24
Epoch [40/50], Iter [150/209] Loss: 1.4591, average_loss: 1.8713
24
24
24
24
24
Epoch [40/50], Iter [155/209] Loss: 2.1422, average_loss: 1.8749
24
24
24
24
24
Epoch [40/50], Iter [160/209] Loss: 1.3321, average_loss: 1.8794
24
24
24
24
24
Epoch [40/50], Iter [165/209] Loss: 1.5842, average_loss: 1.8699
24
24
24
24
24
Epoch [40/50], Iter [170/209] Loss: 2.5551, average_loss: 1.8643
24
24
24
24
24
Epoch [40/50], Iter [175/209] Loss: 1.9623, average_loss: 1.8681
24
24
24
24
24
Epoch [40/50], Iter [180/209] Loss: 2.3988, average_loss: 1.8721
24
24
24
24
24
Epoch [40/50], Iter [185/209] Loss: 2.2882, average_loss: 1.8729
24
24
```



```
Epoch [40/50], Iter [190/209] Loss: 1.5883, average_loss: 1.8613
Epoch [40/50], Iter [195/209] Loss: 1.6172, average_loss: 1.8575
Epoch [40/50], Iter [200/209] Loss: 1.3017, average_loss: 1.8539
Epoch [40/50], Iter [205/209] Loss: 1.9168, average_loss: 1.8539
```



[illegible]

24  
24  
6  
Updating best test loss: 2.71717

Starting epoch 41 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24  
Epoch [41/50], Iter [5/209] Loss: 2.3760, average\_loss: 1.9491  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [10/209] Loss: 1.9776, average\_loss: 1.9728  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [15/209] Loss: 1.5143, average\_loss: 1.9168  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [20/209] Loss: 1.3352, average\_loss: 1.8757  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [25/209] Loss: 1.9224, average\_loss: 1.9025  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [30/209] Loss: 2.7155, average\_loss: 1.8832  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [35/209] Loss: 1.5286, average\_loss: 1.8202  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [40/209] Loss: 1.8036, average\_loss: 1.8235  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [45/209] Loss: 1.8137, average\_loss: 1.8267  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [50/209] Loss: 2.4646, average\_loss: 1.8495  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [55/209] Loss: 2.2613, average\_loss: 1.8409  
24  
24  
24

24  
24  
Epoch [41/50], Iter [60/209] Loss: 1.7387, average\_loss: 1.8447  
24  
24  
24  
24  
Epoch [41/50], Iter [65/209] Loss: 2.2468, average\_loss: 1.8457  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [70/209] Loss: 1.5704, average\_loss: 1.8413  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [75/209] Loss: 1.6579, average\_loss: 1.8353  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [80/209] Loss: 1.5533, average\_loss: 1.8236  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [85/209] Loss: 1.9115, average\_loss: 1.8125  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [90/209] Loss: 2.6337, average\_loss: 1.8357  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [95/209] Loss: 1.7073, average\_loss: 1.8288  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [100/209] Loss: 1.5242, average\_loss: 1.8384  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [105/209] Loss: 1.6513, average\_loss: 1.8440  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [110/209] Loss: 2.0749, average\_loss: 1.8431  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [115/209] Loss: 1.5005, average\_loss: 1.8357  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [120/209] Loss: 1.9194, average\_loss: 1.8291  
24  
24

24  
24  
24  
Epoch [41/50], Iter [125/209] Loss: 1.4145, average\_loss: 1.8262  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [130/209] Loss: 1.6937, average\_loss: 1.8360  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [135/209] Loss: 1.7523, average\_loss: 1.8355  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [140/209] Loss: 1.7164, average\_loss: 1.8322  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [145/209] Loss: 1.8625, average\_loss: 1.8377  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [150/209] Loss: 1.4010, average\_loss: 1.8475  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [155/209] Loss: 1.8339, average\_loss: 1.8447  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [160/209] Loss: 2.5662, average\_loss: 1.8479  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [165/209] Loss: 1.5481, average\_loss: 1.8394  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [170/209] Loss: 1.6344, average\_loss: 1.8468  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [175/209] Loss: 1.5404, average\_loss: 1.8415  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [180/209] Loss: 1.7494, average\_loss: 1.8387  
24  
24  
24  
24  
24  
Epoch [41/50], Iter [185/209] Loss: 1.6320, average\_loss: 1.8449  
24

```
Epoch [41/50], Iter [190/209] Loss: 1.6444, average_loss: 1.8376  
Epoch [41/50], Iter [195/209] Loss: 2.0524, average_loss: 1.8434  
Epoch [41/50], Iter [200/209] Loss: 1.6856, average_loss: 1.8468  
Epoch [41/50], Iter [205/209] Loss: 1.7966, average_loss: 1.8468
```

[illegible]



[illegible]

24  
24  
6  
Updating best test loss: 2.71407

Starting epoch 42 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24  
Epoch [42/50], Iter [5/209] Loss: 1.7378, average\_loss: 1.7179  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [10/209] Loss: 2.2445, average\_loss: 1.7559  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [15/209] Loss: 2.6331, average\_loss: 1.8489  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [20/209] Loss: 1.3525, average\_loss: 1.8102  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [25/209] Loss: 1.2563, average\_loss: 1.7841  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [30/209] Loss: 1.4457, average\_loss: 1.7690  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [35/209] Loss: 2.2005, average\_loss: 1.8042  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [40/209] Loss: 1.5520, average\_loss: 1.7804  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [45/209] Loss: 2.2508, average\_loss: 1.8155  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [50/209] Loss: 2.8921, average\_loss: 1.8426  
24  
24  
24  
24  
24  
Epoch [42/50], Iter [55/209] Loss: 2.6935, average\_loss: 1.8533  
24  
24  
24

```
24
24
Epoch [42/50], Iter [60/209] Loss: 1.9705, average_loss: 1.8364
24
24
24
24
Epoch [42/50], Iter [65/209] Loss: 1.6605, average_loss: 1.8472
24
24
24
24
Epoch [42/50], Iter [70/209] Loss: 1.7616, average_loss: 1.8641
24
24
24
24
Epoch [42/50], Iter [75/209] Loss: 1.6883, average_loss: 1.8622
24
24
24
24
Epoch [42/50], Iter [80/209] Loss: 2.0529, average_loss: 1.8753
24
24
24
24
Epoch [42/50], Iter [85/209] Loss: 1.4497, average_loss: 1.8685
24
24
24
24
Epoch [42/50], Iter [90/209] Loss: 1.9366, average_loss: 1.8739
24
24
24
24
Epoch [42/50], Iter [95/209] Loss: 1.4201, average_loss: 1.8643
24
24
24
24
Epoch [42/50], Iter [100/209] Loss: 1.4594, average_loss: 1.8660
24
24
24
24
Epoch [42/50], Iter [105/209] Loss: 1.7933, average_loss: 1.8523
24
24
24
24
Epoch [42/50], Iter [110/209] Loss: 1.5647, average_loss: 1.8516
24
24
24
24
Epoch [42/50], Iter [115/209] Loss: 1.4676, average_loss: 1.8591
24
24
24
24
Epoch [42/50], Iter [120/209] Loss: 1.5424, average_loss: 1.8679
24
24
```

```
24
24
24
Epoch [42/50], Iter [125/209] Loss: 2.0536, average_loss: 1.8598
24
24
24
24
24
Epoch [42/50], Iter [130/209] Loss: 1.3278, average_loss: 1.8429
24
24
24
24
24
Epoch [42/50], Iter [135/209] Loss: 1.7919, average_loss: 1.8440
24
24
24
24
24
Epoch [42/50], Iter [140/209] Loss: 1.8512, average_loss: 1.8382
24
24
24
24
24
Epoch [42/50], Iter [145/209] Loss: 1.2899, average_loss: 1.8423
24
24
24
24
24
Epoch [42/50], Iter [150/209] Loss: 2.1174, average_loss: 1.8490
24
24
24
24
24
Epoch [42/50], Iter [155/209] Loss: 1.6962, average_loss: 1.8517
24
24
24
24
24
Epoch [42/50], Iter [160/209] Loss: 1.7090, average_loss: 1.8444
24
24
24
24
24
Epoch [42/50], Iter [165/209] Loss: 1.5470, average_loss: 1.8444
24
24
24
24
24
Epoch [42/50], Iter [170/209] Loss: 2.0201, average_loss: 1.8375
24
24
24
24
24
Epoch [42/50], Iter [175/209] Loss: 2.1928, average_loss: 1.8397
24
24
24
24
24
Epoch [42/50], Iter [180/209] Loss: 1.7200, average_loss: 1.8410
24
24
24
24
24
Epoch [42/50], Iter [185/209] Loss: 1.5685, average_loss: 1.8364
24
```

```

Epoch [42/50], Iter [190/209] Loss: 1.7332, average_loss: 1.8347
Epoch [42/50], Iter [195/209] Loss: 2.3807, average_loss: 1.8324
Epoch [42/50], Iter [200/209] Loss: 2.4329, average_loss: 1.8336
Epoch [42/50], Iter [205/209] Loss: 1.3246, average_loss: 1.8296

```

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24

24



24  
24  
24  
6

Starting epoch 43 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24

Epoch [43/50], Iter [5/209] Loss: 1.6484, average\_loss: 1.6185

24  
24  
24  
24  
24

Epoch [43/50], Iter [10/209] Loss: 1.4961, average\_loss: 1.6159

24  
24  
24  
24  
24

Epoch [43/50], Iter [15/209] Loss: 1.2731, average\_loss: 1.6133

24  
24  
24  
24  
24

Epoch [43/50], Iter [20/209] Loss: 1.9156, average\_loss: 1.6457

24  
24  
24  
24  
24

Epoch [43/50], Iter [25/209] Loss: 1.1270, average\_loss: 1.6127

24  
24  
24  
24  
24

Epoch [43/50], Iter [30/209] Loss: 1.8634, average\_loss: 1.6665

24  
24  
24  
24  
24

Epoch [43/50], Iter [35/209] Loss: 2.1705, average\_loss: 1.7050

24  
24  
24  
24  
24

Epoch [43/50], Iter [40/209] Loss: 2.3441, average\_loss: 1.7144

24  
24  
24  
24  
24

Epoch [43/50], Iter [45/209] Loss: 1.6594, average\_loss: 1.7293

24  
24  
24  
24  
24

Epoch [43/50], Iter [50/209] Loss: 1.8430, average\_loss: 1.7369

24  
24  
24  
24  
24

Epoch [43/50], Iter [55/209] Loss: 1.9855, average\_loss: 1.7455

24  
24  
24  
24



```
24
24
Epoch [43/50], Iter [60/209] Loss: 2.1585, average_loss: 1.7628
24
24
24
24
24
Epoch [43/50], Iter [65/209] Loss: 2.0657, average_loss: 1.7655
24
24
24
24
24
Epoch [43/50], Iter [70/209] Loss: 1.4981, average_loss: 1.7685
24
24
24
24
24
Epoch [43/50], Iter [75/209] Loss: 2.8025, average_loss: 1.7831
24
24
24
24
24
Epoch [43/50], Iter [80/209] Loss: 2.4019, average_loss: 1.8005
24
24
24
24
24
Epoch [43/50], Iter [85/209] Loss: 1.9505, average_loss: 1.8153
24
24
24
24
24
Epoch [43/50], Iter [90/209] Loss: 1.2211, average_loss: 1.8006
24
24
24
24
24
Epoch [43/50], Iter [95/209] Loss: 1.6535, average_loss: 1.8053
24
24
24
24
24
Epoch [43/50], Iter [100/209] Loss: 1.9330, average_loss: 1.8160
24
24
24
24
24
Epoch [43/50], Iter [105/209] Loss: 1.8086, average_loss: 1.8074
24
24
24
24
24
Epoch [43/50], Iter [110/209] Loss: 1.8132, average_loss: 1.8031
24
24
24
24
24
Epoch [43/50], Iter [115/209] Loss: 1.9493, average_loss: 1.8116
24
24
24
24
24
Epoch [43/50], Iter [120/209] Loss: 2.0469, average_loss: 1.8216
24
24
24
```

```
24
24
24
Epoch [43/50], Iter [125/209] Loss: 2.1381, average_loss: 1.8281
24
24
24
24
24
Epoch [43/50], Iter [130/209] Loss: 1.4015, average_loss: 1.8253
24
24
24
24
24
Epoch [43/50], Iter [135/209] Loss: 1.7324, average_loss: 1.8242
24
24
24
24
24
Epoch [43/50], Iter [140/209] Loss: 1.8273, average_loss: 1.8338
24
24
24
24
24
Epoch [43/50], Iter [145/209] Loss: 1.3026, average_loss: 1.8389
24
24
24
24
24
Epoch [43/50], Iter [150/209] Loss: 1.9660, average_loss: 1.8385
24
24
24
24
24
Epoch [43/50], Iter [155/209] Loss: 2.2200, average_loss: 1.8357
24
24
24
24
24
Epoch [43/50], Iter [160/209] Loss: 2.3952, average_loss: 1.8352
24
24
24
24
24
Epoch [43/50], Iter [165/209] Loss: 1.4877, average_loss: 1.8308
24
24
24
24
24
Epoch [43/50], Iter [170/209] Loss: 2.3884, average_loss: 1.8408
24
24
24
24
24
Epoch [43/50], Iter [175/209] Loss: 1.3724, average_loss: 1.8406
24
24
24
24
24
Epoch [43/50], Iter [180/209] Loss: 1.7226, average_loss: 1.8384
24
24
24
24
24
Epoch [43/50], Iter [185/209] Loss: 1.8035, average_loss: 1.8381
24
-
```

```
Epoch [43/50], Iter [190/209] Loss: 1.7115, average_loss: 1.8427
Epoch [43/50], Iter [195/209] Loss: 1.7136, average_loss: 1.8370
Epoch [43/50], Iter [200/209] Loss: 1.3880, average_loss: 1.8348
Epoch [43/50], Iter [205/209] Loss: 1.9905, average_loss: 1.8348
```

[illegible]



24  
24  
6

Starting epoch 44 / 50

Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24

Epoch [44/50], Iter [5/209] Loss: 1.4728, average\_loss: 2.0181

24  
24  
24  
24  
24

Epoch [44/50], Iter [10/209] Loss: 1.6379, average\_loss: 1.7969

24  
24  
24  
24  
24

Epoch [44/50], Iter [15/209] Loss: 1.6500, average\_loss: 1.8593

24  
24  
24  
24  
24

Epoch [44/50], Iter [20/209] Loss: 2.3626, average\_loss: 1.8966

24  
24  
24  
24  
24

Epoch [44/50], Iter [25/209] Loss: 1.8309, average\_loss: 1.8365

24  
24  
24  
24  
24

Epoch [44/50], Iter [30/209] Loss: 1.7968, average\_loss: 1.8264

24  
24  
24  
24  
24

Epoch [44/50], Iter [35/209] Loss: 2.1729, average\_loss: 1.8367

24  
24  
24  
24  
24

Epoch [44/50], Iter [40/209] Loss: 2.0639, average\_loss: 1.8512

24  
24  
24  
24  
24

Epoch [44/50], Iter [45/209] Loss: 1.3825, average\_loss: 1.8384

24  
24  
24  
24  
24

Epoch [44/50], Iter [50/209] Loss: 2.0201, average\_loss: 1.8468

24  
24  
24  
24  
24

Epoch [44/50], Iter [55/209] Loss: 1.6218, average\_loss: 1.8324

24  
24  
24  
24

24  
Epoch [44/50], Iter [60/209] Loss: 1.6769, average\_loss: 1.8078  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [65/209] Loss: 1.7819, average\_loss: 1.8363  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [70/209] Loss: 1.8262, average\_loss: 1.8257  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [75/209] Loss: 1.6782, average\_loss: 1.8211  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [80/209] Loss: 1.2612, average\_loss: 1.8396  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [85/209] Loss: 2.4810, average\_loss: 1.8429  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [90/209] Loss: 1.7945, average\_loss: 1.8424  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [95/209] Loss: 1.7434, average\_loss: 1.8569  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [100/209] Loss: 2.1109, average\_loss: 1.8577  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [105/209] Loss: 1.9753, average\_loss: 1.8504  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [110/209] Loss: 2.2068, average\_loss: 1.8505  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [115/209] Loss: 1.8122, average\_loss: 1.8574  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [120/209] Loss: 2.1156, average\_loss: 1.8855  
24  
24  
24

24  
24  
Epoch [44/50], Iter [125/209] Loss: 1.4089, average\_loss: 1.8884  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [130/209] Loss: 1.6102, average\_loss: 1.8760  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [135/209] Loss: 1.3539, average\_loss: 1.8704  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [140/209] Loss: 1.8579, average\_loss: 1.8727  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [145/209] Loss: 2.0398, average\_loss: 1.8768  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [150/209] Loss: 1.4580, average\_loss: 1.8716  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [155/209] Loss: 2.0438, average\_loss: 1.8723  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [160/209] Loss: 2.0254, average\_loss: 1.8651  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [165/209] Loss: 1.4339, average\_loss: 1.8640  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [170/209] Loss: 1.6461, average\_loss: 1.8656  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [175/209] Loss: 1.6182, average\_loss: 1.8594  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [180/209] Loss: 1.5349, average\_loss: 1.8529  
24  
24  
24  
24  
24  
Epoch [44/50], Iter [185/209] Loss: 1.9535, average\_loss: 1.8487  
24  
24



```

Epoch [44/50], Iter [190/209] Loss: 1.7399, average_loss: 1.8473
Epoch [44/50], Iter [195/209] Loss: 1.5178, average_loss: 1.8461
Epoch [44/50], Iter [200/209] Loss: 2.1949, average_loss: 1.8470
Epoch [44/50], Iter [205/209] Loss: 1.5711, average_loss: 1.8442

```

[illegible]

[illegible]

24  
24  
6

Starting epoch 45 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24

Epoch [45/50], Iter [5/209] Loss: 1.6179, average\_loss: 1.8774

24  
24  
24  
24  
24

Epoch [45/50], Iter [10/209] Loss: 1.3744, average\_loss: 1.8144

24  
24  
24  
24  
24

Epoch [45/50], Iter [15/209] Loss: 1.8607, average\_loss: 1.7643

24  
24  
24  
24  
24

Epoch [45/50], Iter [20/209] Loss: 1.5140, average\_loss: 1.7664

24  
24  
24  
24  
24

Epoch [45/50], Iter [25/209] Loss: 2.0308, average\_loss: 1.8061

24  
24  
24  
24  
24

Epoch [45/50], Iter [30/209] Loss: 2.1800, average\_loss: 1.7987

24  
24  
24  
24  
24

Epoch [45/50], Iter [35/209] Loss: 2.7178, average\_loss: 1.8357

24  
24  
24  
24  
24

Epoch [45/50], Iter [40/209] Loss: 2.8621, average\_loss: 1.8670

24  
24  
24  
24  
24

Epoch [45/50], Iter [45/209] Loss: 1.7709, average\_loss: 1.8664

24  
24  
24  
24  
24

Epoch [45/50], Iter [50/209] Loss: 2.2295, average\_loss: 1.9122

24  
24  
24  
24  
24

Epoch [45/50], Iter [55/209] Loss: 1.6603, average\_loss: 1.8863

24  
24  
24  
24  
24

```
Epoch [45/50], Iter [60/209] Loss: 2.4317, average_loss: 1.8844
24
24
24
24
24
Epoch [45/50], Iter [65/209] Loss: 1.3160, average_loss: 1.8745
24
24
24
24
24
Epoch [45/50], Iter [70/209] Loss: 1.7344, average_loss: 1.8751
24
24
24
24
24
Epoch [45/50], Iter [75/209] Loss: 1.6927, average_loss: 1.8581
24
24
24
24
24
Epoch [45/50], Iter [80/209] Loss: 1.9651, average_loss: 1.8547
24
24
24
24
24
Epoch [45/50], Iter [85/209] Loss: 1.7134, average_loss: 1.8632
24
24
24
24
24
Epoch [45/50], Iter [90/209] Loss: 1.2012, average_loss: 1.8441
24
24
24
24
24
Epoch [45/50], Iter [95/209] Loss: 1.8237, average_loss: 1.8480
24
24
24
24
24
Epoch [45/50], Iter [100/209] Loss: 1.7495, average_loss: 1.8415
24
24
24
24
24
Epoch [45/50], Iter [105/209] Loss: 1.8321, average_loss: 1.8310
24
24
24
24
24
Epoch [45/50], Iter [110/209] Loss: 2.4066, average_loss: 1.8242
24
24
24
24
24
Epoch [45/50], Iter [115/209] Loss: 1.8704, average_loss: 1.8317
24
24
24
24
24
Epoch [45/50], Iter [120/209] Loss: 2.2952, average_loss: 1.8382
24
24
24
24
```

```
24
Epoch [45/50], Iter [125/209] Loss: 1.9482, average_loss: 1.8387
24
24
24
24
Epoch [45/50], Iter [130/209] Loss: 1.3312, average_loss: 1.8320
24
24
24
24
24
Epoch [45/50], Iter [135/209] Loss: 1.8576, average_loss: 1.8269
24
24
24
24
24
Epoch [45/50], Iter [140/209] Loss: 1.5807, average_loss: 1.8165
24
24
24
24
24
Epoch [45/50], Iter [145/209] Loss: 2.5122, average_loss: 1.8246
24
24
24
24
24
Epoch [45/50], Iter [150/209] Loss: 2.2373, average_loss: 1.8269
24
24
24
24
24
Epoch [45/50], Iter [155/209] Loss: 2.6694, average_loss: 1.8272
24
24
24
24
24
Epoch [45/50], Iter [160/209] Loss: 2.2056, average_loss: 1.8353
24
24
24
24
24
Epoch [45/50], Iter [165/209] Loss: 1.6972, average_loss: 1.8328
24
24
24
24
24
Epoch [45/50], Iter [170/209] Loss: 1.6088, average_loss: 1.8246
24
24
24
24
24
Epoch [45/50], Iter [175/209] Loss: 1.4237, average_loss: 1.8264
24
24
24
24
24
Epoch [45/50], Iter [180/209] Loss: 2.1105, average_loss: 1.8226
24
24
24
24
24
Epoch [45/50], Iter [185/209] Loss: 2.4685, average_loss: 1.8310
24
24
24
```

```
Epoch [45/50], Iter [190/209] Loss: 3.0761, average_loss: 1.8394
Epoch [45/50], Iter [195/209] Loss: 1.7379, average_loss: 1.8386
Epoch [45/50], Iter [200/209] Loss: 2.1684, average_loss: 1.8360
Epoch [45/50], Iter [205/209] Loss: 1.7904, average_loss: 1.8410
```





[illegible]

24  
6

Starting epoch 46 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24

Epoch [46/50], Iter [5/209] Loss: 1.4918, average\_loss: 1.8576

24  
24  
24  
24  
24

Epoch [46/50], Iter [10/209] Loss: 1.8850, average\_loss: 1.8267

24  
24  
24  
24  
24

Epoch [46/50], Iter [15/209] Loss: 1.7055, average\_loss: 1.8409

24  
24  
24  
24  
24

Epoch [46/50], Iter [20/209] Loss: 1.4358, average\_loss: 1.8160

24  
24  
24  
24  
24

Epoch [46/50], Iter [25/209] Loss: 1.8789, average\_loss: 1.8628

24  
24  
24  
24  
24

Epoch [46/50], Iter [30/209] Loss: 2.3117, average\_loss: 1.8851

24  
24  
24  
24  
24

Epoch [46/50], Iter [35/209] Loss: 1.5126, average\_loss: 1.8895

24  
24  
24  
24  
24

Epoch [46/50], Iter [40/209] Loss: 2.5369, average\_loss: 1.8997

24  
24  
24  
24  
24

Epoch [46/50], Iter [45/209] Loss: 1.9581, average\_loss: 1.9229

24  
24  
24  
24  
24

Epoch [46/50], Iter [50/209] Loss: 1.9115, average\_loss: 1.9125

24  
24  
24  
24  
24

Epoch [46/50], Iter [55/209] Loss: 2.1971, average\_loss: 1.9312

24  
24  
24  
24  
24

Epoch [46/50], Iter [60/209] Loss: 1.8888, average\_loss: 1.8888

```
Epoch [46/50], Iter [60/209] Loss: 1.2809, average_loss: 1.9023
24
24
24
24
24
Epoch [46/50], Iter [65/209] Loss: 2.2508, average_loss: 1.9071
24
24
24
24
24
Epoch [46/50], Iter [70/209] Loss: 1.4801, average_loss: 1.8932
24
24
24
24
24
Epoch [46/50], Iter [75/209] Loss: 1.9597, average_loss: 1.8920
24
24
24
24
24
Epoch [46/50], Iter [80/209] Loss: 2.0157, average_loss: 1.8776
24
24
24
24
24
Epoch [46/50], Iter [85/209] Loss: 1.7519, average_loss: 1.8734
24
24
24
24
24
Epoch [46/50], Iter [90/209] Loss: 1.4093, average_loss: 1.8574
24
24
24
24
24
Epoch [46/50], Iter [95/209] Loss: 1.7360, average_loss: 1.8472
24
24
24
24
24
Epoch [46/50], Iter [100/209] Loss: 1.6004, average_loss: 1.8419
24
24
24
24
24
Epoch [46/50], Iter [105/209] Loss: 1.7487, average_loss: 1.8402
24
24
24
24
24
Epoch [46/50], Iter [110/209] Loss: 1.8880, average_loss: 1.8422
24
24
24
24
24
Epoch [46/50], Iter [115/209] Loss: 1.4668, average_loss: 1.8453
24
24
24
24
24
Epoch [46/50], Iter [120/209] Loss: 2.0391, average_loss: 1.8500
24
24
24
24
24
```

24  
Epoch [46/50], Iter [125/209] Loss: 1.9727, average\_loss: 1.8541  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [130/209] Loss: 1.8408, average\_loss: 1.8542  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [135/209] Loss: 1.4319, average\_loss: 1.8431  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [140/209] Loss: 1.6496, average\_loss: 1.8537  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [145/209] Loss: 1.2280, average\_loss: 1.8502  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [150/209] Loss: 2.1217, average\_loss: 1.8460  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [155/209] Loss: 1.6995, average\_loss: 1.8397  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [160/209] Loss: 1.9722, average\_loss: 1.8410  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [165/209] Loss: 2.0528, average\_loss: 1.8441  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [170/209] Loss: 2.3393, average\_loss: 1.8470  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [175/209] Loss: 2.3546, average\_loss: 1.8428  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [180/209] Loss: 1.8243, average\_loss: 1.8370  
24  
24  
24  
24  
24  
Epoch [46/50], Iter [185/209] Loss: 1.7005, average\_loss: 1.8356  
24  
24  
24

```
Epoch [46/50], Iter [190/209] Loss: 1.7528, average_loss: 1.8353
Epoch [46/50], Iter [195/209] Loss: 1.8707, average_loss: 1.8307
Epoch [46/50], Iter [200/209] Loss: 1.9023, average_loss: 1.8300
Epoch [46/50], Iter [205/209] Loss: 1.5215, average_loss: 1.8263
```

[illegible]

[illegible]

```
Starting epoch 47 / 50
Learning Rate for this epoch: 1e-05
24
24
24
24
24
Epoch [47/50], Iter [5/209] Loss: 1.0524, average_loss: 1.8135
24
24
24
24
24
Epoch [47/50], Iter [10/209] Loss: 2.1377, average_loss: 1.7396
24
24
24
24
24
Epoch [47/50], Iter [15/209] Loss: 2.1992, average_loss: 1.7730
24
24
24
24
24
Epoch [47/50], Iter [20/209] Loss: 2.0360, average_loss: 1.7362
24
24
24
24
24
Epoch [47/50], Iter [25/209] Loss: 1.9576, average_loss: 1.7734
24
24
24
24
24
Epoch [47/50], Iter [30/209] Loss: 1.7876, average_loss: 1.7799
24
24
24
24
24
Epoch [47/50], Iter [35/209] Loss: 2.7619, average_loss: 1.7970
24
24
24
24
24
Epoch [47/50], Iter [40/209] Loss: 2.1513, average_loss: 1.8017
24
24
24
24
24
Epoch [47/50], Iter [45/209] Loss: 1.6634, average_loss: 1.8184
24
24
24
24
24
Epoch [47/50], Iter [50/209] Loss: 1.5486, average_loss: 1.8064
24
24
24
24
24
Epoch [47/50], Iter [55/209] Loss: 1.6138, average_loss: 1.7967
24
24
24
24
24
Epoch [47/50], Iter [60/209] Loss: 1.4048, average_loss: 1.7847
```



24  
24  
24  
24  
24  
Epoch [47/50], Iter [65/209] Loss: 1.9281, average\_loss: 1.7827  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [70/209] Loss: 1.5104, average\_loss: 1.7922  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [75/209] Loss: 1.4747, average\_loss: 1.7944  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [80/209] Loss: 1.6268, average\_loss: 1.8019  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [85/209] Loss: 1.8357, average\_loss: 1.7978  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [90/209] Loss: 2.0838, average\_loss: 1.8073  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [95/209] Loss: 1.6592, average\_loss: 1.8168  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [100/209] Loss: 2.2395, average\_loss: 1.8347  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [105/209] Loss: 1.5601, average\_loss: 1.8262  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [110/209] Loss: 1.4716, average\_loss: 1.8181  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [115/209] Loss: 1.5613, average\_loss: 1.8283  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [120/209] Loss: 2.5238, average\_loss: 1.8364  
24  
24  
24  
24  
24

Epoch [47/50], Iter [125/209] Loss: 1.7105, average\_loss: 1.8355  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [130/209] Loss: 1.7546, average\_loss: 1.8337  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [135/209] Loss: 1.5213, average\_loss: 1.8272  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [140/209] Loss: 1.4724, average\_loss: 1.8272  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [145/209] Loss: 1.6784, average\_loss: 1.8332  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [150/209] Loss: 1.2075, average\_loss: 1.8251  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [155/209] Loss: 1.9440, average\_loss: 1.8230  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [160/209] Loss: 2.1761, average\_loss: 1.8306  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [165/209] Loss: 1.9990, average\_loss: 1.8383  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [170/209] Loss: 1.7928, average\_loss: 1.8292  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [175/209] Loss: 1.0976, average\_loss: 1.8278  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [180/209] Loss: 1.3626, average\_loss: 1.8283  
24  
24  
24  
24  
24  
Epoch [47/50], Iter [185/209] Loss: 2.5436, average\_loss: 1.8363  
24  
24  
24  
24

```
Epoch [47/50], Iter [190/209] Loss: 1.6189, average_loss: 1.8381
Epoch [47/50], Iter [195/209] Loss: 2.1276, average_loss: 1.8340
Epoch [47/50], Iter [200/209] Loss: 2.1674, average_loss: 1.8377
Epoch [47/50], Iter [205/209] Loss: 2.1071, average_loss: 1.8352
```

[illegible]

[illegible]

~  
Updating best test loss: 2.70958

Starting epoch 48 / 50  
Learning Rate for this epoch: 1e-05

24  
24  
24  
24  
24  
Epoch [48/50], Iter [5/209] Loss: 1.9153, average\_loss: 1.7341  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [10/209] Loss: 2.1126, average\_loss: 1.8564  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [15/209] Loss: 2.2470, average\_loss: 1.8421  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [20/209] Loss: 2.5411, average\_loss: 1.8464  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [25/209] Loss: 1.9534, average\_loss: 1.7876  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [30/209] Loss: 1.8391, average\_loss: 1.7742  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [35/209] Loss: 1.9529, average\_loss: 1.7959  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [40/209] Loss: 1.7193, average\_loss: 1.7916  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [45/209] Loss: 1.8390, average\_loss: 1.7759  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [50/209] Loss: 1.7176, average\_loss: 1.7633  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [55/209] Loss: 1.8234, average\_loss: 1.7892  
24  
24  
24  
24  
24  
Epoch [48/50], Iter [60/209] Loss: 2.0279, average\_loss: 1.7937

```
Epoch [48/50], Iter [60/209] Loss: 2.0279, average_loss: 1.7937
24
24
24
24
24
Epoch [48/50], Iter [65/209] Loss: 2.3499, average_loss: 1.8063
24
24
24
24
24
Epoch [48/50], Iter [70/209] Loss: 2.4463, average_loss: 1.8155
24
24
24
24
24
Epoch [48/50], Iter [75/209] Loss: 2.3866, average_loss: 1.8349
24
24
24
24
24
Epoch [48/50], Iter [80/209] Loss: 1.6672, average_loss: 1.8218
24
24
24
24
24
Epoch [48/50], Iter [85/209] Loss: 2.0848, average_loss: 1.8432
24
24
24
24
24
Epoch [48/50], Iter [90/209] Loss: 1.7964, average_loss: 1.8397
24
24
24
24
24
Epoch [48/50], Iter [95/209] Loss: 1.5282, average_loss: 1.8416
24
24
24
24
24
Epoch [48/50], Iter [100/209] Loss: 1.2653, average_loss: 1.8296
24
24
24
24
24
Epoch [48/50], Iter [105/209] Loss: 1.5985, average_loss: 1.8362
24
24
24
24
24
Epoch [48/50], Iter [110/209] Loss: 1.3858, average_loss: 1.8169
24
24
24
24
24
Epoch [48/50], Iter [115/209] Loss: 1.8558, average_loss: 1.8218
24
24
24
24
24
Epoch [48/50], Iter [120/209] Loss: 1.4362, average_loss: 1.8130
24
24
24
24
24
```

```
Epoch [48/50], Iter [125/209] Loss: 2.4367, average_loss: 1.8147
24
24
24
24
24
Epoch [48/50], Iter [130/209] Loss: 2.0994, average_loss: 1.8167
24
24
24
24
24
Epoch [48/50], Iter [135/209] Loss: 2.2453, average_loss: 1.8153
24
24
24
24
24
Epoch [48/50], Iter [140/209] Loss: 2.1838, average_loss: 1.8198
24
24
24
24
24
Epoch [48/50], Iter [145/209] Loss: 1.8587, average_loss: 1.8219
24
24
24
24
24
Epoch [48/50], Iter [150/209] Loss: 1.3258, average_loss: 1.8216
24
24
24
24
24
Epoch [48/50], Iter [155/209] Loss: 1.3230, average_loss: 1.8205
24
24
24
24
24
Epoch [48/50], Iter [160/209] Loss: 1.4777, average_loss: 1.8141
24
24
24
24
24
Epoch [48/50], Iter [165/209] Loss: 1.3421, average_loss: 1.8069
24
24
24
24
24
Epoch [48/50], Iter [170/209] Loss: 2.3977, average_loss: 1.8193
24
24
24
24
24
Epoch [48/50], Iter [175/209] Loss: 2.3083, average_loss: 1.8159
24
24
24
24
24
Epoch [48/50], Iter [180/209] Loss: 2.6566, average_loss: 1.8172
24
24
24
24
24
Epoch [48/50], Iter [185/209] Loss: 2.0952, average_loss: 1.8230
24
24
24
24
```



```
Epoch [48/50], Iter [190/209] Loss: 3.0341, average_loss: 1.8265
Epoch [48/50], Iter [195/209] Loss: 1.8301, average_loss: 1.8272
Epoch [48/50], Iter [200/209] Loss: 1.8151, average_loss: 1.8311
Epoch [48/50], Iter [205/209] Loss: 1.2908, average_loss: 1.8277
```

$\frac{1}{\sqrt{\pi}}$

[illegible]

```
Starting epoch 49 / 50
Learning Rate for this epoch: 1e-05
24
24
24
24
24
Epoch [49/50], Iter [5/209] Loss: 1.8233, average_loss: 2.1231
24
24
24
24
24
Epoch [49/50], Iter [10/209] Loss: 1.5137, average_loss: 2.0641
24
24
24
24
24
Epoch [49/50], Iter [15/209] Loss: 2.4552, average_loss: 2.0711
24
24
24
24
24
Epoch [49/50], Iter [20/209] Loss: 1.4273, average_loss: 2.0303
24
24
24
24
24
Epoch [49/50], Iter [25/209] Loss: 1.9332, average_loss: 1.9595
24
24
24
24
24
Epoch [49/50], Iter [30/209] Loss: 1.4310, average_loss: 1.9513
24
24
24
24
24
Epoch [49/50], Iter [35/209] Loss: 1.7475, average_loss: 1.9334
24
24
24
24
24
Epoch [49/50], Iter [40/209] Loss: 1.8816, average_loss: 1.9538
24
24
24
24
24
Epoch [49/50], Iter [45/209] Loss: 1.8863, average_loss: 1.9158
24
24
24
24
24
Epoch [49/50], Iter [50/209] Loss: 1.6966, average_loss: 1.9018
24
24
24
24
24
Epoch [49/50], Iter [55/209] Loss: 1.1121, average_loss: 1.8897
24
24
24
24
24
Epoch [49/50], Iter [60/209] Loss: 1.4891, average_loss: 1.8573
...
```

24  
24  
24  
24  
24  
Epoch [49/50], Iter [65/209] Loss: 1.2580, average\_loss: 1.8430  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [70/209] Loss: 1.6596, average\_loss: 1.8401  
24  
24  
24  
24  
Epoch [49/50], Iter [75/209] Loss: 1.5799, average\_loss: 1.8312  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [80/209] Loss: 1.7040, average\_loss: 1.8314  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [85/209] Loss: 1.6547, average\_loss: 1.8402  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [90/209] Loss: 1.4948, average\_loss: 1.8247  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [95/209] Loss: 1.6313, average\_loss: 1.8247  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [100/209] Loss: 1.5623, average\_loss: 1.8168  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [105/209] Loss: 2.2458, average\_loss: 1.8287  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [110/209] Loss: 1.2836, average\_loss: 1.8129  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [115/209] Loss: 1.7438, average\_loss: 1.8098  
24  
24  
24  
24  
24  
Epoch [49/50], Iter [120/209] Loss: 1.6800, average\_loss: 1.8097  
24  
24  
24  
24  
24

```
Epoch [49/50], Iter [125/209] Loss: 1.5928, average_loss: 1.7979
24
24
24
24
24
Epoch [49/50], Iter [130/209] Loss: 1.7611, average_loss: 1.8015
24
24
24
24
24
Epoch [49/50], Iter [135/209] Loss: 1.7879, average_loss: 1.8093
24
24
24
24
24
Epoch [49/50], Iter [140/209] Loss: 2.7688, average_loss: 1.8161
24
24
24
24
24
Epoch [49/50], Iter [145/209] Loss: 1.6453, average_loss: 1.8121
24
24
24
24
24
Epoch [49/50], Iter [150/209] Loss: 1.4479, average_loss: 1.8091
24
24
24
24
24
Epoch [49/50], Iter [155/209] Loss: 2.4296, average_loss: 1.8089
24
24
24
24
24
Epoch [49/50], Iter [160/209] Loss: 1.5588, average_loss: 1.8095
24
24
24
24
24
Epoch [49/50], Iter [165/209] Loss: 2.6296, average_loss: 1.8162
24
24
24
24
24
Epoch [49/50], Iter [170/209] Loss: 1.8487, average_loss: 1.8198
24
24
24
24
24
Epoch [49/50], Iter [175/209] Loss: 1.4293, average_loss: 1.8162
24
24
24
24
24
Epoch [49/50], Iter [180/209] Loss: 1.5696, average_loss: 1.8139
24
24
24
24
24
Epoch [49/50], Iter [185/209] Loss: 2.6946, average_loss: 1.8229
24
24
24
24
```

```
Epoch [49/50], Iter [190/209] Loss: 1.5779, average_loss: 1.8159
Epoch [49/50], Iter [195/209] Loss: 1.4125, average_loss: 1.8240
Epoch [49/50], Iter [200/209] Loss: 1.6792, average_loss: 1.8208
Epoch [49/50], Iter [205/209] Loss: 2.6428, average_loss: 1.8265
```

[illegible]



[illegible]

```
Starting epoch 50 / 50
Learning Rate for this epoch: 1e-05
24
24
24
24
24
Epoch [50/50], Iter [5/209] Loss: 1.4049, average_loss: 1.4330
24
24
24
24
24
Epoch [50/50], Iter [10/209] Loss: 1.7458, average_loss: 1.5867
24
24
24
24
24
Epoch [50/50], Iter [15/209] Loss: 1.6147, average_loss: 1.5070
24
24
24
24
24
Epoch [50/50], Iter [20/209] Loss: 1.6834, average_loss: 1.6297
24
24
24
24
24
Epoch [50/50], Iter [25/209] Loss: 1.1740, average_loss: 1.6214
24
24
24
24
24
Epoch [50/50], Iter [30/209] Loss: 1.9958, average_loss: 1.6707
24
24
24
24
24
Epoch [50/50], Iter [35/209] Loss: 1.8703, average_loss: 1.7066
24
24
24
24
24
Epoch [50/50], Iter [40/209] Loss: 2.0792, average_loss: 1.7490
24
24
24
24
24
Epoch [50/50], Iter [45/209] Loss: 1.6857, average_loss: 1.7446
24
24
24
24
24
Epoch [50/50], Iter [50/209] Loss: 1.2600, average_loss: 1.7419
24
24
24
24
24
Epoch [50/50], Iter [55/209] Loss: 1.9485, average_loss: 1.7682
24
24
24
24
24
Epoch [50/50], Iter [60/209] Loss: 1.6429, average_loss: 1.7792
24
```

24  
24  
24  
24  
Epoch [50/50], Iter [65/209] Loss: 1.4751, average\_loss: 1.7748  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [70/209] Loss: 2.3979, average\_loss: 1.7805  
24  
24  
24  
24  
Epoch [50/50], Iter [75/209] Loss: 1.6650, average\_loss: 1.7812  
24  
24  
24  
24  
Epoch [50/50], Iter [80/209] Loss: 1.7112, average\_loss: 1.7766  
24  
24  
24  
24  
Epoch [50/50], Iter [85/209] Loss: 1.4420, average\_loss: 1.7665  
24  
24  
24  
24  
Epoch [50/50], Iter [90/209] Loss: 1.5130, average\_loss: 1.7759  
24  
24  
24  
24  
Epoch [50/50], Iter [95/209] Loss: 2.0085, average\_loss: 1.7901  
24  
24  
24  
24  
Epoch [50/50], Iter [100/209] Loss: 1.6862, average\_loss: 1.7947  
24  
24  
24  
24  
Epoch [50/50], Iter [105/209] Loss: 2.2076, average\_loss: 1.7877  
24  
24  
24  
24  
Epoch [50/50], Iter [110/209] Loss: 1.6909, average\_loss: 1.7883  
24  
24  
24  
24  
Epoch [50/50], Iter [115/209] Loss: 1.5821, average\_loss: 1.7809  
24  
24  
24  
24  
Epoch [50/50], Iter [120/209] Loss: 1.6053, average\_loss: 1.7954  
24  
24  
24  
24  
Epoch [50/50], Iter [125/209] Loss: 1.2810, average loss: 1.7929

24  
24  
24  
24  
24  
Epoch [50/50], Iter [130/209] Loss: 1.8424, average\_loss: 1.7968  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [135/209] Loss: 1.2345, average\_loss: 1.7937  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [140/209] Loss: 1.6936, average\_loss: 1.8005  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [145/209] Loss: 2.3774, average\_loss: 1.8145  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [150/209] Loss: 2.0572, average\_loss: 1.8168  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [155/209] Loss: 2.1406, average\_loss: 1.8132  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [160/209] Loss: 1.8711, average\_loss: 1.8120  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [165/209] Loss: 1.7523, average\_loss: 1.8121  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [170/209] Loss: 2.4181, average\_loss: 1.8184  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [175/209] Loss: 1.7037, average\_loss: 1.8235  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [180/209] Loss: 2.1370, average\_loss: 1.8251  
24  
24  
24  
24  
24  
Epoch [50/50], Iter [185/209] Loss: 2.4530, average\_loss: 1.8277  
24  
24  
24  
24  
24

[illegible]

[illegible]



## View example predictions

In [10]:

```
net.eval()

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

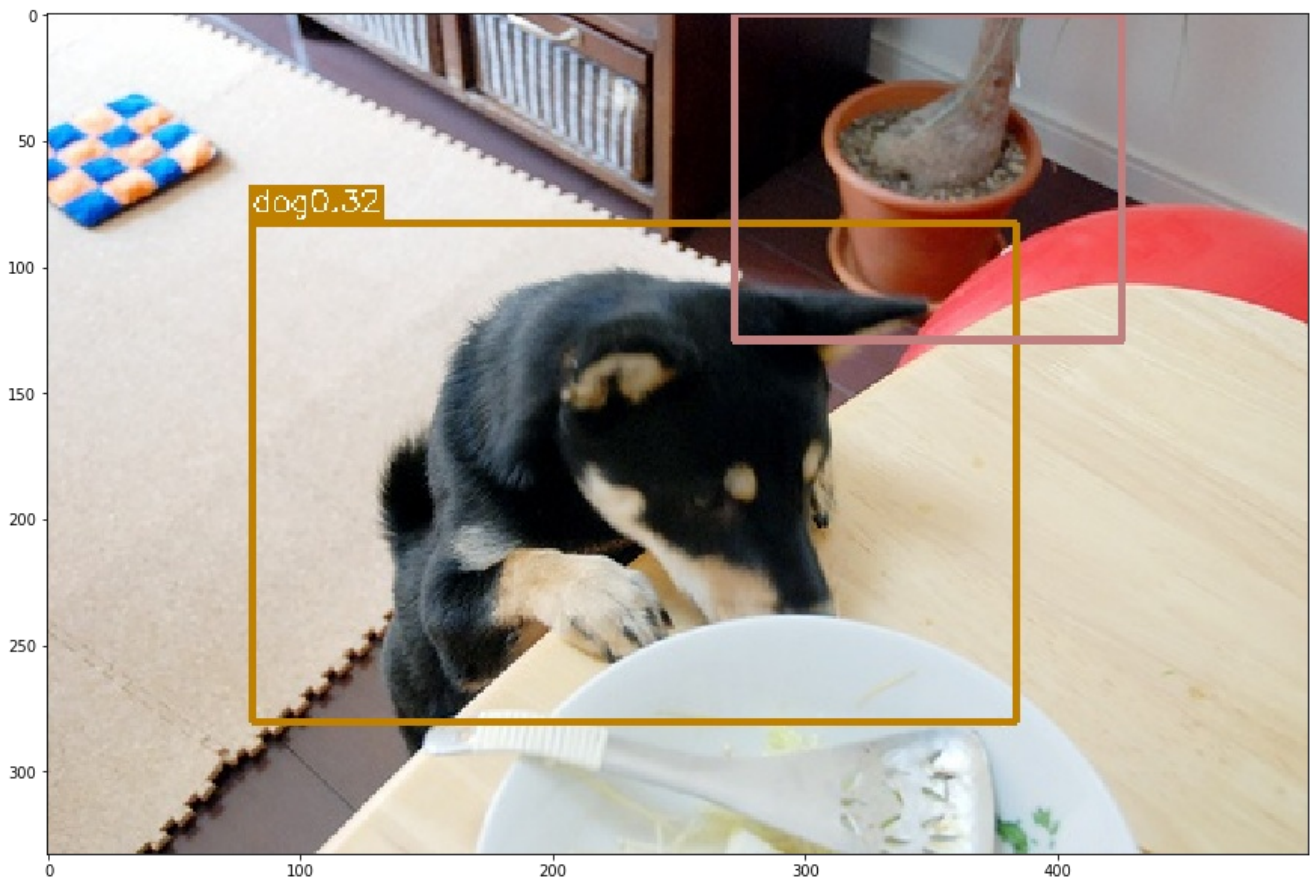
print('predicting...')
result = predict_image(net, image_name, root_img_directory=file_root_test)
for left_up, right_bottom, class_name, _, prob in result:
    color = COLORS[VOC_CLASSES.index(class_name)]
    cv2.rectangle(image, left_up, right_bottom, color, 2)
    label = class_name + str(round(prob, 2))
    text_size, baseline = cv2.getTextSize(label, cv2.FONT_HERSHEY_SIMPLEX, 0.4, 1)
    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[10]:

<matplotlib.image.AxesImage at 0x7f0cdffd6b70>



## Evaluate on Test

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



In [11]:

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

---Evaluate model on test samples---

100%|██████████| 4950/4950 [02:45<00:00, 29.87it/s]

```
---class aeroplane ap 0.4420003998064223---  
---class bicycle ap 0.6231354748842958---  
---class bird ap 0.4651185302992713---  
---class boat ap 0.23673331033736233---  
---class bottle ap 0.1805988712973443---  
---class bus ap 0.6110381390529425---  
---class car ap 0.6463257180956071---  
---class cat ap 0.714802442891229---  
---class chair ap 0.32208158090331185---  
---class cow ap 0.5153431394557259---  
---class diningtable ap 0.3099033590060779---  
---class dog ap 0.6153448026283326---  
---class horse ap 0.6898247815961045---  
---class motorbike ap 0.553187791431618---  
---class person ap 0.5143320132499212---  
---class pottedplant ap 0.18182869887225622---  
---class sheep ap 0.4577297495379029---  
---class sofa ap 0.4927174743359368---  
---class train ap 0.631403452332796---  
---class tvmonitor ap 0.4871978560939636---  
---map 0.48453237930542115---
```

In [12]:

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